Theoretical Economics 15 (2020), 1135–1173 1555-7561/20201135

Optimal incentive contract with endogenous monitoring technology

Anqi Li
Department of Economics, Washington University in St. Louis

Ming Yang
Fuqua School of Business, Duke University

Recent technology advances have enabled firms to flexibly process and analyze sophisticated employee performance data at a reduced and yet significant cost. We develop a theory of optimal incentive contracting where the monitoring technology that governs the above procedure is part of the designer's strategic planning. In otherwise standard principal–agent models with moral hazard, we allow the principal to partition agents' performance data into any finite categories, and to pay for the amount of information the output signal carries. Through analysis of the trade-off between giving incentives to agents and saving the monitoring cost, we obtain characterizations of optimal monitoring technologies such as information aggregation, strict monotone likelihood ratio property, likelihood ratio–convex performance classification, group evaluation in response to rising monitoring costs, and assessing multiple task performances according to agents' endogenous tendencies to shirk. We examine implications of these results for workforce management and firms' internal organizations.

Keywords. Incentive contract, endogenous monitoring technology.

JEL classification. D86, M5, M15.

1. Introduction

Recent technology advances have enabled firms to flexibly process and analyze sophisticated employee performance data at a reduced and yet significant cost. Speech analytics software, natural language processing tools, and cloud-based systems are increasingly used to convert hard-to-process contents into succinct and meaningful performance ratings such as "satisfactory" and "unsatisfactory" (Murff et al. 2011, Singer 2013, Kaplan 2015). This paper develops a theory of optimal incentive contracting where the *monitoring technology* that governs the above procedure is part of the designer's strategic planning.

Anqi Li: anqili@wustl.edu
Ming Yang: ming.yang@duke.edu

We thank four anonymous referees, Nick Bloom, George Mailath, Ilya Segal, Chris Shannon, Joel Sobel, Jacob Steinhardt, Bob Wilson, and seminar participants at Caltech, Duke–UNC, Johns Hopkins Carey, Northwestern, Decentralization Conference 2017, Stanford, UCSB, UCSD, University of Washington, and University of Chicago for feedback and suggestions. Lin Hu and Jessie Li provided generous assistances for the numerical analysis. All errors are our own.

Available at https://econtheory.org. https://doi.org/10.3982/TE3130

Our research agenda is motivated by the case of call center performance management reported by Singer (2013). It has long been recognized that the conversations between call center agents and customers contain useful performance indicators such as customer sentiment, voice quality and tone, etc. Recently, the advent of speech analytics software has finally enabled the processing and analysis of these contents, as well as their conversions into succinct and meaningful performance ratings such as "satisfactory" and "unsatisfactory." On the one hand, running speech analytics software consumes server space and power, and this procedure has been increasingly outsourced to third parties in order to take advantage of the latest development in cloud computing. On the other hand, managers now have considerable freedom to decide which facets of the customer conversation to utilize, thanks to the increased availability of products whose specialties range from emotion detection to word spotting.

We formalize the flexibility and cost associated with the design and implementation of the monitoring technology in otherwise standard principal–agent models with moral hazard. Specifically, we allow the monitoring technology to partition agents' performance data into any finite categories, at a cost that increases with the amount of information the output signal carries (hereafter *monitoring cost*). An incentive contract pairs the monitoring technology with a wage scheme that maps realizations of the output signal to different wages. An optimal contract minimizes the sum of expected wage and monitoring cost, subject to agents' incentive constraints.

Our main result gives characterizations of optimal monitoring technologies in general environments, showing that the assignment of Lagrange multiplier-weighted likelihood ratios to performance categories is positive assortative in the direction of agent utilities. Geometrically, this means that optimal monitoring technologies comprise convex cells in the space of likelihood ratios or their transformations. This result provides practitioners with the needed formula for sorting employees' performance data into performance categories, and exploiting its geometry yields insights into workforce management and firms' internal organizations.

Our proof strategy works directly with the principal's Lagrangian. It handles general situations featuring multiple agents and multiple tasks, in which the direction of sorting vector-valued likelihood ratios into performance categories is non-obvious a priori. It also overcomes the technical challenge whereby perturbations of the sorting algorithm affect wages endogenously through the Lagrange multipliers of agents' incentive constraints, generating new effects that are difficult to assess using standard methods.

We give three applications of our result. In the single-agent model considered by Hölmstrom (1979), we show that the assignment of likelihood ratios to wage categories is positive assortative and follows a simple cutoff rule. The monitoring technology aggregates potentially high-dimensional performance data into rank-ordered performance ratings, and the output signal satisfies the strict monotone likelihood ratio property with respect to the order induced by likelihood ratios. Solving the cutoff likelihood ratios yields consistent findings with recent developments in manufacturing, retail, and healthcare sectors, where decreases in the data processing cost have been shown to increase the fineness of the performance grids (Bloom and Reenen 2006, Murff et al. 2011, Ewenstein et al. 2016).

In the multi-agent model considered by Hölmstrom (1982), the optimal monitoring technology partitions vectors of individual agents' likelihood ratios into convex polygons. Based on this result, we then compare individual and group performance evaluations from the angle of monitoring cost, showing that firms should switch from individual evaluation to group evaluation in response to rising monitoring costs. This result formalizes the theses of Alchian and Demsetz (1972) and Lazear and Rosen (1981) that either team or tournament should be the dominant incentive system when individual performance evaluation is too costly to conduct. It is consistent with the finding of Bloom and Reenen (2006), namely the lack of access to information technologies (IT) increases the use of group performance evaluation among otherwise similar firms.

In the presence of multiple tasks as in Hölmstrom and Milgrom (1991), the resources spent on the assessment of a task performance should increase with the agent's endogenous tendency to shirk the corresponding task. Using simulation, we apply this result to the study of, e.g., how improved precision of some task measurements (caused by, e.g., the availability of high-quality scanner data measuring the skillfulness in scanning items) would affect the resources spent on the assessments of other task performances (e.g., projecting warmth to customers).

1.1 *Related literature*

Earlier studies on contracting with costly experiments (in the sense of Blackwell 1953) include, but are not limited to, Baiman and Demski (1980) and Dye (1986), in which case the principal can pay an external auditor to draw a signal from an exogenous distribution; as well as Hölmstrom (1979), Grossman and Hart (1983), and Kim (1995), in which signal distributions are ranked based on the incentive costs they incur. In these studies, the principal can change the probability space generated by the agent's hidden effort and, in the first two studies, through paying stylized costs. In contrast, we focus on the conversion of the agent's performance data into performance ratings while taking the former's probability space as given. Also, while our assumption that the monitoring cost increases with the amount of information carried by the output signal is suitable for studying data processing and analysis, it could be ill-suited for modeling the cost of running experiments in general.

The current work differs from the existing studies on rational inattention (hereafter RI) in three aspects. First, early developments in RI pioneered by Sims (1998), Sims (2003), Maćkowiak and Wiederholt (2009), and Woodford (2009) sought to explain the stickiness of macroeconomic variables by information processing costs, whereas we examine the implication of costly yet flexible monitoring for principal–agent relationships.[1] Second, we focus mainly on partitional monitoring technologies, because in reality, adding non-performance-related factors into employee ratings could have dire

[1]Yang (2020) studies a security design problem in which a rationally inattentive buyer can obtain any signal about an uncertain state at a cost that is proportional to entropy reduction. Other recent efforts to introduce RI into strategic environments include, but are not limited to, Matějka and McKay (2012), Martin (2017), and Ravid (2017).

consequences such as appeals, lawsuits, and excessive turnover.[2] Finally, our monitoring cost function nests entropy as a special case.

Recent works by Crémer et al. (2007), Jäger et al. (2011), Sobel (2015), and Dilmé (2018) examine the optimal language used between organization members who share a common interest but face communication costs. The absence of conflicting interests, hence incentive constraints, distinguishes these works from ours.

The remainder of this paper is organized as follows. Section 2 introduces the baseline model, Section 3 presents main results, Sections 4 and 5 investigate extensions of the baseline model, and Section 6 concludes. See Appendices A and B for omitted proofs and additional materials.

2. Baseline model

2.1 *Setup*

Primitives. A risk-neutral principal faces a risk-averse agent, who earns a utility $u(w)$ from spending a nonnegative wage $w \geq 0$ and incurs a cost $c(a)$ from privately exerting high effort or low effort $a \in \{0, 1\}$. The function $u : \mathbb{R}_+ \to \mathbb{R}$ satisfies $u(0) = 0$, $u' > 0$, and $u'' < 0$, whereas the function $c : \{0, 1\} \to \mathbb{R}_+$ satisfies $c(1) = c > c(0) = 0$.

Each effort choice $a \in \{0, 1\}$ generates a probability space (Ω, Σ, P_a), where Ω is a finite-dimensional Euclidean space that comprises the agent's performance data, Σ is the Borel sigma-algebra on Ω, and P_a is the probability measure on (Ω, Σ) conditional on the agent's effort being a. The P_as are assumed to be mutually absolutely continuous, and the probability density function p_as they induce are well defined and everywhere positive.

Incentive contract. An incentive contract $\langle \mathcal{P}, w(\cdot) \rangle$ is a pair of *monitoring technology* $\mathcal{P}$ and *wage scheme* $w : \mathcal{P} \to \mathbb{R}_+$. The former represents a human- or machine-operated system that governs the processing and analysis of the agent's performance data, whereas the latter maps outputs of the first-step procedure to different levels of wages. In the main body of this paper, $\mathcal{P}$ can be any partition of Ω with at most K cells that are all of positive measures,[3] and $w : \mathcal{P} \to \mathbb{R}_+$ maps each cell A of $\mathcal{P}$ to a nonnegative wage $w(A) \geq 0$.[4] The upper bound K for the *rating scale* $|\mathcal{P}|$ can be any integer greater than 1, and it is taken as given throughout the analysis.[5]

For any data point $\omega \in \Omega$, let $A(\omega)$ denote the unique *performance category* that contains ω, and let $w(A(\omega))$ denote the wage associated with $A(\omega)$. Time evolves as follows.

[2]See standard human resource textbooks for this subject matter. Saint-Paul (2017) demonstrates the validity of entropy as an information cost in decision problems in which the decision rule is a deterministic function of an exogenous state variable.

[3]In Appendix B.2, we allow the monitoring technology to be any mapping from Ω to lotteries on finite performance categories. If the lottery is degenerate, then the monitoring technology is partitional.

[4]Appendix B.1 examines the case where the agent faces an individual rationality constraint.

[5]The upper bound K, while stylized, guarantees the existence of optimal incentive contract(s). Judging from the simulation exercises we have so far conducted, the optimal rating scale is typically smaller than K even when μ is small (see, e.g., Figure 1).

Stage 1. The principal commits to $\langle \mathcal{P}, w(\cdot) \rangle$.

Stage 2. The agent privately chooses $a \in \{0, 1\}$.

Stage 3. Nature draws ω from Ω according to P_a.

Stage 4. The monitoring technology outputs $A(\omega)$.

Stage 5. The principal pays $w(A(\omega))$ to the agent.

Implementation cost. A monitoring technology $\mathcal{P} = \{A_1, \ldots, A_N\}$ outputs a signal $X : \Omega \to \mathcal{P}$, whose probability distribution given the agent's effort choice a is represented by a vector $\boldsymbol{\pi}(\mathcal{P}, a) = (P_a(A_1), \ldots, P_a(A_N), 0, \ldots, 0)$ in the K-dimensional simplex. While X is often taken as given in the existing principal–agent literature, here it is chosen by the principal as part of the incentive contract.

Given the agent's effort choice a, the principal incurs the following cost from implementing an incentive contract $\langle \mathcal{P}, w(\cdot) \rangle$:

$$\sum_{A \in \mathcal{P}} P_a(A) w(A) + \mu \cdot H(\mathcal{P}, a).$$

This total implementation cost has two parts. The first part $\sum_{A \in \mathcal{P}} P_a(A) w(A)$, i.e., the *incentive cost*, has been the central focus of the existing principal–agent literature. The second part $\mu \cdot H(\mathcal{P}, a)$, hereafter termed the *monitoring cost*, represents the cost associated with the processing and analysis of the agent's performance data. In particular, $\mu > 0$ is an exogenous parameter which we will further discuss in Section 3.4. Meanwhile, $H(\mathcal{P}, a)$ captures the amount of information carried by the output signal and is assumed to satisfy the following properties.

Assumption 1. *There exists a function $h : \Delta^K \to \mathbb{R}_+$ such that $H(\mathcal{P}, a) = h(\boldsymbol{\pi}(\mathcal{P}, a))$ for any pair $(\mathcal{P}, a)$ of monitoring technology and agent effort. Furthermore, h satisfies the following properties.*

(a) For any probability vector $(\pi_1, \ldots, \pi_K) \in \Delta^K$ and any permutation σ on $\{1, \ldots, K\}$, we have $h(\pi_1, \ldots, \pi_K) = h(\pi_{\sigma(1)}, \ldots, \pi_{\sigma(K)})$.

(b) For any two probability vectors $(0, \pi_2, \ldots)$ and $(\pi_1', \pi_2', \ldots) \in \Delta^K$ that differ only in the first two elements and satisfy $\pi_2, \pi_1', \pi_2' > 0$ and $\pi_2 = \pi_1' + \pi_2'$, we have $h(0, \pi_2, \ldots) < h(\pi_1', \pi_2'', \ldots)$

Inspired by the basic principles of information theory, Assumption 1 stipulates that the amount of information carried by the output signal should depend only on the latter's probability distribution and must increase with the fineness of the monitoring technology. Apart from probabilities, nothing else matters, not even the naming or the contents of the performance categories. Assumption 1 is satisfied by, e.g., the entropy $-\sum_{A \in \mathcal{P}} P_a(A) \log_2 P_a(A)$ of the output signal, the bits of information $\log_2 |\mathcal{P}|$ carried by the output signal, etc.[6] In Section 2.2, we motivate the use of this assumption in the example of call center performance management.

The principal's problem. Consider the problem of inducing high effort from the agent.[7] Define a random variable $Z : \Omega \to \mathbb{R}$ as

$$Z(\omega) = 1 - \frac{p_0(\omega)}{p_1(\omega)} \quad \forall \omega,$$

where $p_0(\omega)/p_1(\omega)$ is the *likelihood ratio* associated with data point ω. Note that $\mathbf{E}[Z \mid a = 1] = 0$, and the range of Z is a subset of $(-\infty, -1)$. For any set $A \in \Sigma$ of positive measure, define the *z-value* of A as

$$z(A) = \mathbf{E}[Z \mid A; a = 1].$$

In words, $z(A)$ represents the average value of Z given that the true performance data point belongs to A.

A contract $\langle \mathcal{P}, w(\cdot) \rangle$ is incentive compatible for the agent if

$$\sum_{A \in \mathcal{P}} P_1(A) u\big(w(A)\big) - c \geq \sum_{A \in \mathcal{P}} P_0(A) u\big(w(A)\big)$$

or, equivalently,

$$\sum_{A \in \mathcal{P}} P_1(A) u\big(w(A)\big) z(A) \geq c, \tag{IC}$$

and it satisfies the agent's limited liability constraint if

$$w(A) \geq 0 \quad \forall A \in \mathcal{P}. \tag{LL}$$

An optimal incentive contract that induces high effort from the agent (optimal incentive contract for short) minimizes the total implementation cost given high effort, subject to the agent's incentive compatibility constraint and limited liability constraint:

$$\min_{\langle \mathcal{P}, w(\cdot) \rangle} \sum_{A \in \mathcal{P}} P_1(A) w(A) + \mu \cdot H(\mathcal{P}, 1) \quad \text{subject to (IC) and (LL).}$$

In what follows, we will denote any solution to this problem by $\langle \mathcal{P}^*, w^*(\cdot) \rangle$.

2.2 *Monitoring cost*

In this section, we first illustrate Assumption 1 in the context of call center performance management:

Example 1. In the example described in Section 1, a piece of performance data comprises the major characteristics of a call history (e.g., customer sentiment and voice quality) encoded in binary digits, and the monitoring technology represents the speech

[6] The bit is a basic unit of information in information theory, computing, and digital communications. In information theory, one bit is defined as the maximum information entropy of a binary random variable.

[7] The problem of inducing low effort from the agent is standard.

analytics program that categorizes these binary digits into performance ratings. To formalize the design flexibility, we allow the monitoring technology to partition the agent's performance data into any $N \leq K$ categories, where K can be any integer greater than 1. The cost of running the monitoring technology is assumed to increase with the amount of the processed information, whose definition varies from case to case. For example, if the monitoring technology runs many times among many identical agents, then the optimal design should minimize the average steps it takes to find the performance category that contains the true performance data point. By now, it is well known that this quantity equals approximately the entropy of the output signal. In contrast, if the monitoring technology runs only a few times for a few number of agents, then the worst-case (or unamortized) amount of the processed information is best captured by the bits of information carried by the output signal (see, e.g., Cover and Thomas 2006). In both cases, the quantity of our interest depends only on the probability distribution of the output signal and nothing else. $\diamondsuit$

We next introduce the concept of *setup cost* and distinguish it from our notion of monitoring cost.

Example 1 (Continued). As its name suggests, *setup cost* refers the cost incurred to set up the infrastructure that facilitates data processing and analysis, e.g., fast Fourier transformation (FFT) chips (which transform sound waves into their major characteristics coded in binary digits), recording devices, etc.

The major role of setup cost is to change the probability space (Ω, Σ, P_a)s. For example, design improvements in FFT chips enable more frequent sampling of sound waves and thus cause (Ω, Σ, P_a)s to change. In what follows, we will ignore the setup cost and take the probability spaces of the agent's performance data as given. That said, one can certainly embed our analysis into a two-stage setting in which the principal first incurs the setup cost and then the monitoring cost. Results below do carry over to this new setting. $\diamondsuit$

3. Analysis

3.1 *Preview*

Example 2. Suppose $u(w) = \sqrt{w}$, Z is uniformly distributed over $[-1/2, 1/2]$ given $a = 1$, and $H(\mathcal{P}, a) = f(|\mathcal{P}|)$ for some strictly increasing function $f : \{2, \ldots, K\} \to \mathbb{R}_+$. Below we walk through the key steps in solving the optimal incentive contract, give closed-form solutions, and discuss their practical implications.

Optimal wage scheme. We first solve for the optimal wage scheme for any given monitoring technology $\mathcal{P}$ as in Hölmstrom (1979). Specifically, label the performance categories as $A_1, \ldots, A_N$, and write $\pi_n = P_1(A_n)$ and $z_n = z(A_n)$ for $n = 1, \ldots, N$. Assume $z_j \neq z_k$ for some $j, k \in \{1, \ldots, N\}$ to make the analysis interesting. The principal's problem is then

$$\min_{\{w_n\}_{n=1}^N} \sum_{n=1}^N \pi_n w_n$$

$$\text{subject to } \sum_{n=1}^N \pi_n \sqrt{w_n} z_n \geq c \tag{IC}$$

$$\text{and } w_n \geq 0, n = 1, \ldots, N. \tag{LL}$$

Straightforward algebra yields the expression for the minimal incentive cost:

$$c^2 \left[\sum_{n=1}^N \pi_n \underbrace{\max\{0, z_n\}^2}_{w_n} \right]^{-1}.$$

A careful inspection of this expression reveals Hölmstrom's (1979) *sufficient statistics principle*, namely z-value is the only part of the performance data that provides the agent with incentives.

Optimal monitoring technology. We next solve for the optimal monitoring technology. First, note that the principal should partition the performance data based only on their z-values, and that different performance categories must attain different z-values and hence different wages. Indeed, if the contrary were true, then it follows from the sufficient statistic principle and Assumption 1(b) that merging performance categories of the same z-value saves the monitoring cost without affecting the incentive cost and, thus, constitutes an improvement to the original monitoring technology.

A more interesting question concerns how we should assign the various data points, identified by their z-values, to different performance categories. In the baseline model featuring a single agent and binary efforts, the answer to this question is relatively straightforward: assign high (resp. low) z-values to high-wage (resp. low-wage) categories. Here is a quick proof of this result: since the left-hand side of the (IC) constraint is supermodular in wages and z-values, if our conjecture were false, then reshuffling data points as above while holding the probabilities of the performance categories constant reduces the incentive cost while leaving the monitoring cost unaffected.

When extending the above intuition to general settings featuring multiple agents or multiple actions, we face two challenges. First, in the case where z-values and wages are vectors, the direction of sorting these objects is not obvious a priori. Second, changes in the sorting algorithm affect wages endogenously through the Lagrange multipliers of the agent's incentive constraints, and the resulting effects could be new and difficult to assess using standard methods.

The proof strategy presented in Section 3.3 overcomes these challenges, showing that the assignment of Lagrange multiplier-weighted z-values to performance categories must be positive assortative in the direction of agent utilities. Geometrically, this means that any optimal monitoring technology must comprise convex cells in the space of z-values or their transformations. Theorems 1, 3, and 5 formalize these statements.

Implications. An important feature of the optimal monitoring technology is *information aggregation*—a term used by human resource practitioners to refer to the aggregation of potentially high-dimensional performance data into rank-ordered performance ratings such as "satisfactory" and "unsatisfactory."

The geometry of optimal monitoring technology yields insights into the practical issues covered in Sections 3.4, 4.3, and 5.1. Consider, for example, optimal performance grids. In the current example, it can be shown that the optimal N-partitional monitoring technology divides the space $[-1/2, 1/2]$ of z-values into N disjoint intervals $[\widehat{z}_{n-1}, \widehat{z}_n)$, $n = 1, \dots, N$, where $\widehat{z}_0 = -1/2$ and $\widehat{z}_N = 1/2$. The optimal cut points $\{\widehat{z}_n\}_{n=1}^{N-1}$ can be solved as

$$\min_{\{\widehat{z}_n\}_{n=1}^{N-1}} c^2 \left[\sum_{n=1}^{N} \pi_n \max\{0, z_n\}^2\right]^{-1} - \mu \cdot f(N),$$

where

$$\pi_n = \int_{\widehat{z}_{n-1}}^{\widehat{z}_n} dZ = \widehat{z}_n - \widehat{z}_{n-1}$$

and

$$z_n = \frac{1}{\pi_n} \int_{\widehat{z}_{n-1}}^{\widehat{z}_n} Z\, dZ = \frac{1}{2}[\widehat{z}_n + \widehat{z}_{n-1}].$$

Straightforward algebra yields

$$\widehat{z}_n = \frac{2n-1}{4N-2}, \quad n = 1, \dots, N-1.$$

Based on this result, as well as the functional form of f, we can then solve for the optimal rating scale N and, hence, the optimal incentive contract completely. ♢

3.2 *Main results*

This section analyzes optimal incentive contracts. Results below hold true except perhaps on a measure zero set of data points. The same disclaimer applies to the remainder of this paper.

We first define Z *convexity*.

Definition 1. A set $A \in \Sigma$ is Z-*convex* if

$$\left\{\omega \in \Omega : Z(\omega) = (1-s) \cdot Z\big(\omega'\big) + s \cdot Z\big(\omega''\big) \text{ for some } s \in (0,1)\right\} \subset A$$

holds for any $\omega', \omega'' \in A$ such that $Z(\omega') \neq Z(\omega'')$.

In words, a set $A \in \Sigma$ is Z-convex if whenever it contains two data points of different z-values, it must also contain all data points of intermediate z-values. Let $Z(A)$ denote the image of any set $A \in \Sigma$ under the mapping Z. In the case where $Z(\Omega)$ is a connected set in $\mathbb{R}$, the above definition is equivalent to the convexity of $Z(A)$ in $\mathbb{R}$.

A few assumptions before we go into detail. The next assumption says that the distribution of Z has no atom or hole.

Assumption 2. *The random variable Z is distributed atomlessly on a connected set $Z(\Omega)$ in $\mathbb{R}$ given high effort $a = 1$.*

Assumption 3. *The set $Z(\Omega)$ is compact in $\mathbb{R}$.*

The next assumption imposes regularities on the monitoring cost function: part (a) holds for the bits of information carried by the output signal, and part (b) holds for the entropy of the output signal.

Assumption 4. *The function $h : \Delta^K \to \mathbb{R}_+$ satisfies one of the following conditions.*

(a) There exists a strictly increasing function $f : \{1, \dots, K\} \to \mathbb{R}_+$ such that $h(\boldsymbol{\pi}(\mathcal{P}, a)) = f(|\mathcal{P}|)$.

(b) The function h is continuous.

We now state our main results. The next theorem shows that any optimal incentive contract assigns data points of high (resp. low) z-values to high-wage (resp. low-wage) categories. Under Assumption 2, this can be achieved by first dividing z-values into disjoint intervals and then backing out the partition of the original data space accordingly. The result is an aggregation of potentially high-dimensional performance data into rank-ordered performance ratings, as well as a wage scheme that is strictly increasing in these performance ratings.

Theorem 1. *Assume Assumption 1, and let $\langle \mathcal{P}^*, w^*(\cdot)\rangle$ be any optimal incentive contract that induces high effort from the agent. Then $\mathcal{P}^*$ comprises Z-convex cells labeled as $A_1, \dots, A_N$, where $0 = w^*(A_1) < \cdots < w^*(A_N)$. Assume, in addition, Assumption 2. Then there exists $\inf Z(\Omega) := \widehat{z}_0 < \widehat{z}_1 < \cdots < \widehat{z}_N := \sup Z(\Omega)$ such that $A_n = \{\omega : Z(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ for $n = 1, \dots, N$.*

The next theorem proves the existence of an optimal incentive contract.

Theorem 2. *An optimal incentive contract that induces high effort from the agent exists under Assumptions 1–4.*

Throughout this paper, proofs omitted from the main text are provided in Appendix A.

3.3 *Proof sketch for Theorem 1*

The proof of Theorem 1 consists of three steps. The intuitions of Steps 1 and 2 have already been discussed in Example 2. Step 3 is new.

Step 1. We first take any monitoring technology $\mathcal{P}$ as given and solve for the optimal wage scheme as in Hölmstrom (1979):

$$\min_{w:\mathcal{P}\to\mathbb{R}_+} \sum_{A\in\mathcal{P}} P_1(A)w(A) \quad \text{subject to (IC) and (LL)}. \tag{1}$$

The next lemma restates Hölmstrom's (1979) *sufficient statistic principle*.

Lemma 1. *Let $w^*(\cdot;\mathcal{P})$ be any solution to problem (1). Then there exists $\lambda > 0$ such that $u'(w^*(A;\mathcal{P})) = 1/(\lambda z(A))$ for any $A \in \mathcal{P}$ such that $w^*(A;\mathcal{P}) > 0$.*

Step 2. We next demonstrate that different performance categories must attain different z-values and different wages.

Lemma 2. *Assume Assumption 1. Let $\langle \mathcal{P}^*, w^*(\cdot)\rangle$ be any optimal incentive contract that induces high effort from the agent, and label the cells of $\mathcal{P}^*$ as $A_1, \dots, A_N$ such that $z(A_1) \leq \cdots \leq z(A_N)$. Then $z(A_1) < 0 < \cdots < z(A_N)$ and $0 = w^*(A_1) < \cdots < w^*(A_N)$.*

Step 3. We finally demonstrate that the assignment of z-values to wage categories is positive assortative. In Example 2, we sketched a proof based on supermodularity, and pointed out the difficulties of extending that argument to multidimensional environments. The argument below overcomes these difficulties.

Take any optimal incentive contract, and let A_j and A_k be any two distinct performance categories, where $z(A_j) \neq z(A_k)$ by Lemma 2. Fix any $\epsilon > 0$, and take any $A'_\epsilon \subset A_j$ and $A''_\epsilon \subset A_k$ such that $P_1(A'_\epsilon) = P_1(A''_\epsilon) = \epsilon$ and $z(A'_\epsilon) = z' \neq z(A''_\epsilon) = z''$. In words, A'_ϵ and A''_ϵ have the same probability ϵ given $a = 1$ but different z-values that are independent of ϵ. Lemma 3 in Appendix A.1.1 proves the existence of A'_ϵ and A''_ϵ when ϵ is small.

Consider a perturbation to the monitoring technology that "swaps" A'_ϵ and A''_ϵ. After this perturbation, the new performance categories, denoted by $A_n(\epsilon)$s, become $A_j(\epsilon) = (A_j \backslash A'_\epsilon) \cup A''_\epsilon$, $A_k(\epsilon) = (A_k \backslash A''_\epsilon) \cup A'_\epsilon$, and $A_n(\epsilon) = A_n$ for $n \neq j, k$. Since this perturbation has no effect on the probabilities of the performance categories given $a = 1$, it does not affect the monitoring cost by Assumption 1(a). Meanwhile, it changes the principal's Lagrangian to (ignore the (LL) constraint)

$$\mathcal{L}(\epsilon) = \sum_n \pi_n\big[w_n(\epsilon) - \lambda(\epsilon) u\big(w_n(\epsilon)\big) z_n\big] + \lambda(\epsilon) c,$$

where π_n denotes the probability of A_n (equivalently $A_n(\epsilon)$) given $a = 1$, $w_n(\epsilon)$ denotes the optimal wage at $A_n(\epsilon)$, and $\lambda(\epsilon)$ denotes the Lagrange multiplier associated with the (IC) constraint. A careful inspection of this Lagrangian leads to the following conjecture: in order to minimize $\mathcal{L}(\epsilon)$, the assignment of Lagrange multiplier-weighted z-values to performance categories must be positive assortative in the direction of agent utilities.

To develop intuition, we assume differentiability and obtain

$$\begin{aligned}
\mathcal{L}'(0) &= \sum_n \pi_n w'_n(0) - \lambda'(0) \underbrace{\left[\sum_n \pi_n u\big(w_n(0)\big) z_n(0) - c\right]}_{\text{(i)} \quad =0} \\
&\quad - \lambda(0)\left[\sum_n \pi_n \cdot \underbrace{u'\big(w_n(0)\big) z_n(0)}_{\text{(ii)} \quad =1/\lambda(0)} \cdot w'_n(0) + \sum_n \pi_n u\big(w_n(0)\big) z'_n(0)\right] \\
&= -\lambda(0) \sum_n \pi_n u\big(w_n(0)\big) z'_n(0) \\
&= \lambda(0)\big(z'' - z'\big)\big[u\big(w_k(0)\big) - u\big(w_j(0)\big)\big].
\end{aligned}$$

In the above expression, (i) $= 0$ because the (IC) constraint binds under the original contract, and (ii) $= 1/\lambda(0)$ by Lemma 1. These findings resolve our concerns raised in Section 3.1, showing that our perturbation has negligible effects on the Lagrange multipliers and wages.

To complete the proof, note that $\mathcal{L}'(0) \geq 0$ by optimality, and that $\mathcal{L}'(0) \neq 0$ because $\lambda(0) > 0$, $z'' \neq z'$, and $w_j(0) \neq w_k(0)$ (Lemma 2). Combining these observations yields $\mathcal{L}'(0) > 0$, so our conjecture is indeed true. The Z-convexity of optimal performance categories is now immediate: if an optimal performance category contains extreme but not intermediate z-values, then the assignment of z-values goes in the wrong direction, and an improvement can be constructed.

The above proof strategy yields the endogenous direction of sorting raw performance data into performance categories, which is relatively straightforward in the baseline model but is less so in later extensions. The proof in Appendix A.1 does *not* assume differentiability and handles the limited liability constraint, too.

3.4 *Implications*

Strict monotone likelihood ratio property. Theorem 1 implies that the signal generated by any optimal monitoring technology must satisfy the strict monotone likelihood ratio property (hereafter *strict MLRP*) with respect to the order induced by z-values.

Definition 2. For any $A, A' \in \Sigma$ of positive measures, write $A \overset{z}{\preceq} A'$ if $z(A) \leq z(A')$.

Corollary 1. *The signal $X : \Omega \to \mathcal{P}^*$ generated by any optimal monitoring technology $\mathcal{P}^*$ satisfies strict MLRP with respect to $\overset{z}{\preceq}$, i.e., any $A, A' \in \mathcal{P}^*$ satisfy $A \overset{z}{\preceq} A'$ if and only if $z(A) < z(A')$.*

While the signal generated by any monitoring technology trivially satisfies the weak MLRP with respect to $\overset{z}{\preceq}$ (i.e., replace $<$ with $\leq$ in Corollary 1), it violates the strict MLRP if there are multiple performance categories that attain the same z-value. By contrast, the signal generated by any optimal monitoring technology must satisfy the strict MLRP with respect to $\overset{z}{\preceq}$, because merging performance categories of the same z-value saves the monitoring cost without affecting the incentive cost.

Comparative statics. The parameter μ captures factors that affect the (opportunity) cost of data processing and analysis. Factors that reduce μ include, but are not limited to, the advent of IT-based human resource management systems in the 1990s, advancements in speech analytics, and increases in computing power.

To facilitate comparative statics analysis, we write any choice of optimal incentive contract as $\langle \mathcal{P}^*(\mu), w^*(\cdot;\mu)\rangle$ to make its dependence on μ explicit.

Proposition 1. *Fix any $0 < \mu < \mu'$. For any choices of $\langle \mathcal{P}^*(\mu), w^*(\cdot;\mu)\rangle$ and $\langle \mathcal{P}^*(\mu'), w^*(\cdot;\mu')\rangle$,*

(i) $\sum_{A \in \mathcal{P}(\mu)} P_1(A) w^*(A;\mu) \leq \sum_{A \in \mathcal{P}(\mu')} P_1(A) w^*(A;\mu')$.

(ii) $H(\mathcal{P}^*(\mu), 1) \geq H(\mathcal{P}^*(\mu'), 1)$.

(iii) $|\mathcal{P}^*(\mu)| \geq |\mathcal{P}^*(\mu')|$ *under Assumption 4(a).*

Part (i) follows from the optimalities of $\mathcal{P}^*(\mu)$ and $\mathcal{P}^*(\mu')$. Parts (ii) and (ii) are immediate.

Proposition 1 shows that as data processing and analysis become cheaper, the principal pays less wage on average, and the information carried by the output signal becomes finer. In the case where the monitoring cost is an increasing function of the rating scale (see, e.g., Hook et al. 2011), the optimal rating scale is non-increasing in μ. For other monitoring cost functions such as entropy, we can first compute the cutoff z-values and then the optimal rating scale as in Example 2.[8] Figure 1 plots the numerical solutions obtained in a special case.

The above findings are consistent with several strands of empirical facts. Among others, access to IT has proven to increase the fineness of the performance grids among manufacturing companies, holding other things constant (Bloom and Reenen 2010; Bloom et al. 2012).[9] Crowd-sourcing the processing and analysis of real-time data has enabled the "exact individual diagnosis" that separates distinctive and mediocre performers in companies like General Electric and Zalando (Ewenstein et al. 2016).

4. Extension: Multiple agents

4.1 *Setup*

Each of the two agents $i = 1, 2$ earns a payoff $u_i(w_i) - c_i(a_i)$ from spending a nonnegative wage $w_i \geq 0$ and exerting high effort or low effort $a_i \in \{0, 1\}$. The function $u_i : \mathbb{R}_+ \to \mathbb{R}$ satisfies $u_i(0) = 0$, $u_i' > 0$ and $u_i'' < 0$, whereas the function $c_i : \{0, 1\} \to \mathbb{R}_+$ satisfies $c_i(1) = c_i > c_i(0) = 0$.

Each effort profile $\mathbf{a} = a_1 a_2$ generates a probability space $(\Omega, \Sigma, P_{\mathbf{a}})$, where Ω is a finite-dimensional Euclidean space that comprises both agents' performance data, Σ is the Borel sigma-algebra on Ω, and $P_{\mathbf{a}}$ is the probability measure on (Ω, Σ) conditional on the effort profile being $\mathbf{a}$. The $P_{\mathbf{a}}$s are assumed to be mutually absolutely continuous, and the probability density function $p_{\mathbf{a}}$s they induce are well defined and everywhere positive.

In this new setting, a monitoring technology $\mathcal{P}$ can be any partition of Ω with at most K cells that are all of positive measures, and a wage scheme $\mathbf{w} : \mathcal{P} \to \mathbb{R}_+^2$ maps each cell A of $\mathcal{P}$ to a vector $\mathbf{w}(A) = (w_1(A), w_2(A))^\top$ of nonnegative wages. For any data point ω, let $A(\omega)$ denote the unique performance category that contains ω, and let $\mathbf{w}(A(\omega))$ be the wage vector associated with $A(\omega)$. Time evolves as follows.

[8] In general, this is not an easy task because perturbations of cutoff z-values (which differ from the perturbation considered in Section 3.3) affect wages endogenously through the Lagrange multipliers of the incentive constraints.

[9] See the appendices of Bloom and Reenen (2006) for survey questions regarding the fineness of the performance grids, e.g., "[e]ach employee is given a red light (not performing), an amber light (doing well and meeting targets), a green light (consistently meeting targets, very high performer) and a blue light (high performer capable of promotion of up to two levels)," versus "reward is based on an individual's commitment to the company measured by seniority."

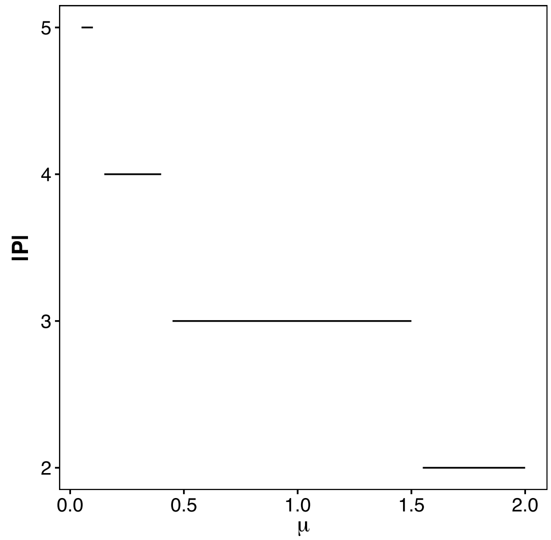

FIGURE 1. Plot of the optimal rating scale against μ: entropy cost, $u(w) = \sqrt{w}$, $Z \sim U[-1/2, 1/2]$, $c = 1$, $K = 100$.

Stage 1. The principal commits to $\langle \mathcal{P}, \mathbf{w}(\cdot) \rangle$.

Stage 2. Agent i privately chooses $a_i \in \{0, 1\}$, $i = 1, 2$.

Stage 3. Nature draws ω from Ω according to $P_{\mathbf{a}}$.

Stage 4. The monitoring technology outputs $A(\omega)$.

Stage 5. The principal pays $w_i(A(\omega))$ to agent $i = 1, 2$.

Consider the problem of inducing both agents to exert high effort. Write $\mathbf{1}$ for $(1, 1)^\top$, and define a vector-valued random variable $\mathbf{Z} = (Z_1, Z_2)^\top$ as

$$Z_i(\omega) = 1 - \frac{p_{a_i=0, a_{-i}=1}(\omega)}{p_{\mathbf{1}}(\omega)} \quad \forall \omega \in \Omega, i = 1, 2.$$

Define the $\mathbf{z}$-value of any set $A \in \Sigma$ of positive measure as $(z_1(A), z_2(A))^\top$, where

$$z_i(A) = \mathbf{E}[Z_i \mid A; \mathbf{a} = \mathbf{1}] \quad \forall i = 1, 2.$$

A contract is incentive compatible for agent i if

$$\sum_{A \in \mathcal{P}} P_{\mathbf{1}}(A) u_i\big(w_i(A)\big) z_i(A) \geq c_i, \tag{IC$_i$}$$

and it satisfies agent i's limited liability constraint if

$$w_i(A) \geq 0 \quad \forall A \in \mathcal{P}. \tag{LL$_i$}$$

An optimal contract minimizes the total implementation cost given the high effort profile, subject to both agents' incentive compatibility constraints and limited liability constraints:

$$\min_{\langle \mathcal{P}, \mathbf{w}(\cdot)\rangle} \sum_{A\in\mathcal{P}} P_{\mathbf{a}}(A) \sum_{i=1}^{2} w_i(A) + \mu \cdot H(\mathcal{P}, \mathbf{1}) \quad \text{subject to } (\mathrm{IC}_i) \text{ and } (\mathrm{LL}_i),\ i = 1, 2.$$

4.2 *Analysis*

The next definition generalizes Z-convexity.

Definition 3. A set $A \in \Sigma$ is $\mathbf{Z}$-convex if

$$\{\omega \in \Omega : \mathbf{Z}(\omega) = (1-s)\cdot \mathbf{Z}(\omega') + s\cdot \mathbf{Z}(\omega'') \text{ for some } s \in (0,1)\} \subset A$$

holds for any $\omega', \omega'' \in A$ such that $\mathbf{Z}(\omega') \neq \mathbf{Z}(\omega'')$.

The next two assumptions impose regularities on the principal's problem analogously to Assumptions 2 and 3.

Assumption 5. *The random variable $\mathbf{Z}$ is distributed atomelessly on a connect set $\mathbf{Z}(\Omega)$ in $\mathbb{R}^2$ given $\mathbf{a} = \mathbf{1}$.*

Assumption 6. *The set $\mathbf{Z}(\Omega)$ is compact in $\mathbb{R}^2$ and has $\dim \mathbf{Z}(\Omega) = 2$.*

The next theorems extend Theorems 1 and 2 to encompass multiple agents.

Theorem 3. *Assume Assumptions 1, 5, and 6. Then any optimal monitoring technology $\mathcal{P}^*$ comprises $\mathbf{Z}$-convex cells that constitute convex polygons in $\mathbb{R}^2$.*

Theorem 4. *An optimal incentive contract that induces high effort from both agents exists under Assumptions 1, 4, 5, and 6.*

Proof sketch. The proof strategy developed in Section 3.3 is useful for handling vector-valued z-values and wages. As before, we fix any $\epsilon > 0$, and take any subsets A'_ϵ and A''_ϵ of two distinct performance categories A_j and A_k, respectively, such that $P_{\mathbf{1}}(A'_\epsilon) = P_{\mathbf{1}}(A''_\epsilon) = \epsilon$ and $\mathbf{z}(A'_\epsilon) := \mathbf{z}' \neq \mathbf{z}(A''_\epsilon) := \mathbf{z}''$ (Lemma 5 in Appendix A.2.1 proves the existence of sets that satisfy weaker properties). After perturbing the monitoring technology as in Section 3.3, the principal's Lagrangian becomes (again ignore the (LL) constraints)

$$\mathcal{L}(\epsilon) = \sum_n \pi_n \left[\sum_i w_{i,n}(\epsilon) - \lambda_i(\epsilon) u_i\big(w_{i,n}(\epsilon)\big) z_{i,n}(\epsilon) - c_i\right],$$

where π_n denotes the probability of A_n (equivalently, $A_n(\epsilon)$) given $\mathbf{a} = \mathbf{1}$, $w_{i,n}(\epsilon)$ denotes agent i's optimal wage at $A_n(\epsilon)$, and $\lambda_i(\epsilon)$ denotes the Lagrange multiplier associated

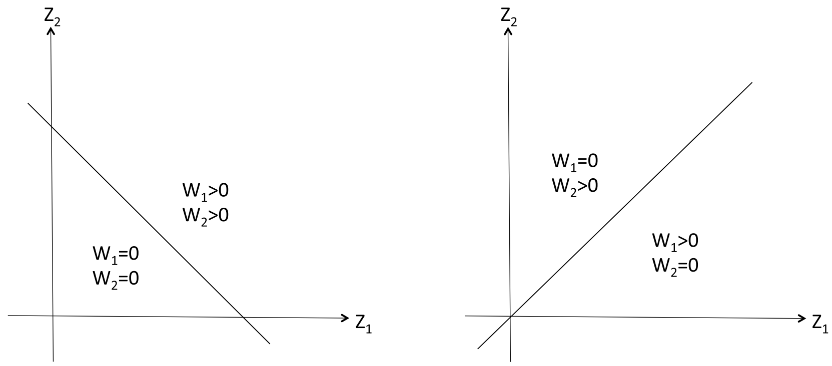

FIGURE 2. Bi-partitional contracts: team and tournament.

with the (IC$_i$) constraint. Assuming differentiability, we obtain

$$\mathcal{L}'(0) = -\sum_{n=1}^{N} \pi_n \cdot \mathbf{u}_n^\top \begin{pmatrix} \lambda_1(0) & 0 \\ 0 & \lambda_2(0) \end{pmatrix} \frac{d}{d\epsilon}\mathbf{z}_n(\epsilon)\bigg|_{\epsilon=0} = (\mathbf{u}_k - \mathbf{u}_j)^\top \big(\widehat{\mathbf{z}}'' - \widehat{\mathbf{z}}'\big),$$

where

$$\mathbf{u}_n := \big(u_1\big(w_{i,n}(0)\big), u_2\big(w_{i,n}(0)\big)\big)^\top \quad \text{for } n = 1, \dots, N$$

and

$$\widehat{\mathbf{z}} := \begin{pmatrix} \lambda_1(0) & 0 \\ 0 & \lambda_2(0) \end{pmatrix} \mathbf{z} \quad \text{for } \mathbf{z} = \mathbf{z}', \mathbf{z}''.$$

Now, since $\mathcal{L}'(0) \geq 0$ by optimality, the assignment of the Lagrange multiplier-weighted $\mathbf{z}$-values to performance categories must be "positive assortative," where the direction of sorting is given by the vector of agents' utilities. The $\mathbf{Z}$-convexity of optimal performance categories then follows from the same reason as in Section 3.3. □

Implications. Solving the optimal convex polygons is computationally hard. However, since the boundaries of convex polygons consist of straight line segments in $\mathbf{Z}(\Omega)$, the following observations are immediate under Assumption 5.

- Any bi-partitional contract takes the form of either a team or a tournament, and it is fully captured by the intercept and slope of the straight line as depicted in Figure 2.
- Contracts that evaluate and reward agents on an individual basis are fully determined by the individual performance cutoffs as depicted in Figure 3.

4.3 *Application: Individual evaluation versus group evaluation*

This section examines the difference between individual and group performance evaluations from the angle of monitoring cost. To obtain the sharpest insights, we assume that agents are *technologically independent*.

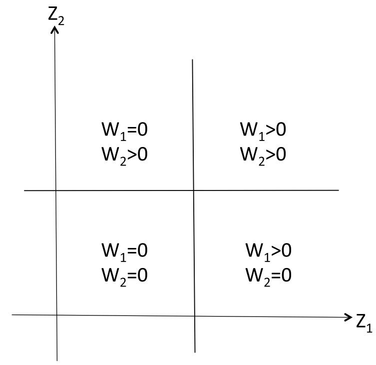

Figure 3. An individual incentive contract.

Assumption 7. *There exist probability spaces* $\{(\Omega_i, \Sigma_i, P_{i,a_i})\}_{i,a_i}$ *as in Section 2 such that* $(\Omega, \Sigma, P_{\mathbf{a}}) = (\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, P_{1,a_1} \times P_{2,a_2})$ *for any* $\mathbf{a} \in \{0, 1\}^2$.

In the language of contract theory, Assumption 7 rules out any *technology linkage* (i.e., ω_i depends on a_{-i}) or *common productivity shock* (i.e., ω_1, ω_2 are correlated given $\mathbf{a}$) between agents.

The next definition is standard.

Definition 4. (i) A partition $\mathcal{P}$ of Ω is an *individual monitoring technology* if for any $A \in \mathcal{P}$, there exist $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$ such that $A = A_1 \times A_2$; otherwise $\mathcal{P}$ is a *group monitoring technology.*

(ii) Let $\mathcal{P}$ be any individual monitoring technology. Then a mapping $\mathbf{w} : \mathcal{P} \to \mathbb{R}^2_+$ is an *individual wage scheme* if $w_i(A_i \times A'_{-i}; \mathcal{P}) = w_i(A_i \times A''_{-i}; \mathcal{P})$ for any $i = 1, 2$ and any $A_i \times A'_{-i}, A_i \times A''_{-i} \in \mathcal{P}$; otherwise $\mathbf{w} : \mathcal{P} \to \mathbb{R}^2_+$ is a *group wage scheme.*

(iii) A contract $\langle \mathcal{P}, \mathbf{w} : \mathcal{P} \to \mathbb{R}^2_+ \rangle$ is an *individual incentive contract* if $\mathcal{P}$ is an individual monitoring technology and $\mathbf{w} : \mathcal{P} \to \mathbb{R}^2_+$ is an individual wage scheme; otherwise it is a *group incentive contract.*

By definition, a group incentive contract either conducts group performance evaluations or pairs individual performance evaluations with group incentive pays. Under Assumption 7, the second option is suboptimal by the sufficient statistics principle or Hölmstrom (1982), thus reducing the comparison between individual and group incentive contracts to that of individual and group performance evaluations.

Define I as the ratio between the minimal cost of implementing bi-partitional incentive contracts and that of implementing individual incentive contracts (the latter, by definition, have at least four performance categories). Note that $I < 1$ is a definitive indicator that group evaluation is optimal whereas individual evaluation is not. The next corollary is immediate.

Corollary 2. *Under Assumptions 1, 4(a), 5, 6, and 7,* $I < 1$ *when* μ *is large.*

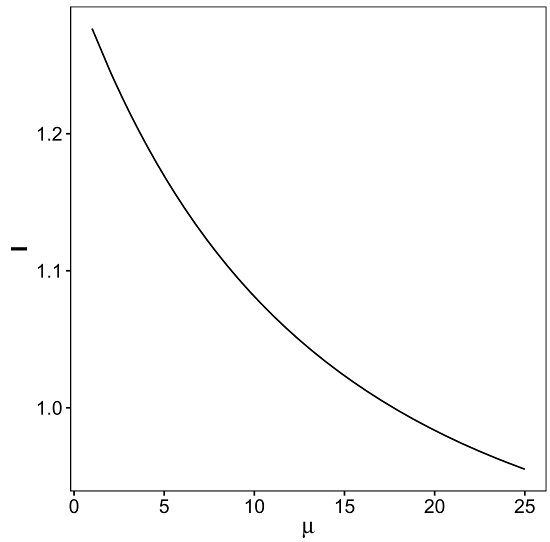

FIGURE 4. Plot of I against μ: entropy cost, $u_i(w) = \sqrt{w}$, $Z_i \sim U[-1/2, 1/2]$, and $c_i = 1$ for $i = 1, 2$.

Beyond the case considered in Corollary 2, we can compute I numerically based on the prior discussion about how to parameterize bi-partitional and individual incentive contracts. Figure 4 plots the solutions obtained in a special case.

Our result formalizes the theses of Alchian and Demsetz (1972) and Lazear and Rosen (1981), namely either team or tournament should be the dominant incentive system when individual performance evaluation is too costly to conduct. It enriches the insights of Hölmstrom (1982), Green and Stokey (1983), and Mookherjee (1984), which attribute the use of group incentive contracts to the technological dependence between agents while abstracting away from the issue of data processing and analysis. Recently, these views are reconciled by Bloom and Reenen (2006), who find—just as our theory predicts—that companies make different choices between individual and group performance evaluations despite being technologically similar, and that group performance evaluation is most prevalent when the capacity to sift out individual-level information is limited by, e.g., the lack of IT access.[10] In the future, it will be interesting to nail down the role of IT in Bloom and Reenen (2006), and to replicate that study for recent advancements in data technologies.

5. Extension: Multiple actions

In this section, the agent's action space $\mathcal{A}$ is a finite set, and taking an action $a \in \mathcal{A}$ incurs a cost $c(a)$ to the agent and generates a probability space (Ω, Σ, P_a) as in Section 2. The

[10]See the survey questions in Bloom and Reenen (2006) regarding the choices between individual and group evaluations, e.g., "employees are rewarded based on their individual contributions to the company" and "compensation is based on shift/plant-level outcomes." The former is regarded as an advanced but expensive managerial practice and is more prevalent among companies with better IT access, other things being equal.

principal wishes to induce the most costly action a^*, defined as the unique solution to $\max_{a\in\mathcal{A}} c(a)$. For any deviation from a^* to $a \in \mathcal{D} := \mathcal{A} - \{a^*\}$, define a random variable $Z_a : \Omega \to \mathbb{R}$ as

$$Z_a(\omega) = 1 - \frac{p_a(\omega)}{p_{a^*}(\omega)} \quad \forall \omega \in \Omega.$$

For any $a \in \mathcal{D}$ and set $A \in \Sigma$ of positive measure, define

$$z_a(A) = \mathbf{E}\big[Z_a \mid A; a^*\big].$$

A contract is incentive compatible if for any $a \in \mathcal{D}$,

$$\sum_{A\in\mathcal{P}} P_{a^*}(A)u\big(w(A)\big)z_a(A) \geq c\big(a^*\big) - c(a). \tag{IC$_a$}$$

An optimal incentive contract $\langle \mathcal{P}^*, w^*(\cdot)\rangle$ that induces a^* solves

$$\min_{\langle \mathcal{P}, w(\cdot)\rangle} \sum_{A\in\mathcal{P}} P_{a^*}(A)w(A) + \mu \cdot H\big(\mathcal{P}, a^*\big) \quad \text{subject to (IC}_a\text{) } \forall a \in \mathcal{D} \text{ and (LL)}.$$

Write $\mathbf{Z}$ for $(Z_a)^\top_{a\in\mathcal{D}}$. For any $|\mathcal{D}|$-vector $\boldsymbol{\lambda} = (\lambda_a)^\top_{a\in\mathcal{D}}$ in $\mathbb{R}^{|\mathcal{D}|}_+$, define a random variable $Z_{\boldsymbol{\lambda}} : \Omega \to \mathbb{R}$ as

$$Z_{\boldsymbol{\lambda}}(\omega) = \boldsymbol{\lambda}^\top \mathbf{Z}(\omega) \quad \forall \omega \in \Omega.$$

The next definition generalizes Z-convexity.

Definition 5. A set $A \in \Sigma$ is $Z_{\boldsymbol{\lambda}}$-convex if

$$\big\{\omega : Z_{\boldsymbol{\lambda}}(\omega) = (1-s)\cdot Z_{\boldsymbol{\lambda}}\big(\omega'\big) + s \cdot Z_{\boldsymbol{\lambda}}\big(\omega''\big) \text{ for some } s \in (0,1)\big\} \subset A$$

holds for any $\omega', \omega'' \in A$ such that $Z_{\boldsymbol{\lambda}}(\omega') \neq Z_{\boldsymbol{\lambda}}(\omega'')$.

The next theorems extend Theorems 1 and 2 to encompass multiple actions.

Theorem 5. *Assume Assumption 1 and Assumption 3 for any $a \in \mathcal{D}$. Then for any optimal incentive contract $\langle \mathcal{P}^*, w^*(\cdot)\rangle$ that induces a^*, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^{|\mathcal{D}|}_+$ with $\max_{a\in\mathcal{D}} \lambda^*_a > 0$ such that all cells of $\mathcal{P}^*$ are $Z_{\boldsymbol{\lambda}^*}$-convex and can be labeled as $A_1, \dots, A_N$, whereby $0 = w^*(A_1) < \cdots < w^*(A_N)$. Assume, in addition, Assumption 2 for any $a \in \mathcal{D}$. Then there exists $-\infty \leq \widehat{z}_0 < \cdots < \widehat{z}_N < +\infty$ such that $A_n = \{\omega : Z_{\boldsymbol{\lambda}^*}(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ for $n = 1, \dots, N$.*

Theorem 6. *Assume Assumptions 1 and 4, as well as Assumptions 2 and 3 for any $a \in \mathcal{D}$. Then an optimal incentive contract that induces a^* exists.*

In the presence of multiple actions, each performance data point is associated with finitely many z-values, each corresponding to a deviation from a^* that the agent can potentially commit. By establishing that the assignment of Lagrange multiplier-weighted z-values to wage categories is positive assortative, Theorem 5 connects the focus of data processing and analysis to the agent's endogenous tendencies to commit deviations. Intuitively, when λ_a^* is large and, hence, the agent is tempted to commit deviation a, the principal should focus on the processing and analysis of Z_a in order to detect deviation a, and the final performance rating should depend significantly on the assessment of Z_a. The next section gives an application of this result.

5.1 *Application: Multiple tasks*

A single agent can exert either high effort or low effort $a_i \in \{0, 1\}$ in each of the two tasks $i = 1, 2$, and each a_i independently generates a probability space $(\Omega_i, \Sigma_i, P_{i,a_i})$ as in Section 2. The goal of a risk-neutral principal is to induce high effort in both tasks.

Write $\mathbf{a} = a_1 a_2$, $\boldsymbol{\omega} = \omega_1 \omega_2$, $\mathcal{A} = \{11, 01, 10, 00\}$, $\mathbf{a}^* = 11$, and $\mathcal{D} = \{01, 10, 00\}$. For any $i = 1, 2$ and $\omega_i \in \Omega_i$, define

$$Z_i(\omega_i) = 1 - \frac{p_{i,a_i=0}(\omega_i)}{p_{i,a_i=1}(\omega_i)},$$

where p_{i,a_i} is the probability density function induced by P_{i,a_i}. For any $\boldsymbol{\omega} \in \Omega_1 \times \Omega_2$ and $\boldsymbol{\lambda} = (\lambda_{01}, \lambda_{10}, \lambda_{00})^\top \in \mathbb{R}^3_+$, define

$$Z_{\boldsymbol{\lambda}}(\boldsymbol{\omega}) = (\lambda_{01} + \lambda_{00}) \cdot Z_1(\omega_1) + (\lambda_{10} + \lambda_{00}) \cdot Z_2(\omega_2) - \lambda_{00} \cdot Z_1(\omega_1) Z_2(\omega_2).$$

The next corollary is immediate from Theorem 5.

Corollary 3. *Assume Assumption 1 and Assumption 3 for any $a \in \mathcal{D}$. Then for any optimal incentive contract $\langle \mathcal{P}^*, w^*(\cdot) \rangle$ that induces high effort in both tasks, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^3_+$ with $\lambda_{01}^* + \lambda_{00}^*$ and $\lambda_{10}^* + \lambda_{00}^* > 0$ such that all cells of $\mathcal{P}^*$ are $Z_{\boldsymbol{\lambda}^*}$-convex and can be labeled as $A_1, \ldots, A_N$, whereby $0 = w^*(A_1) < \cdots < w^*(A_N)$. Assume, in addition, Assumption 2 for any $a \in \mathcal{D}$. Then there exists $-\infty \le \widehat{z}_0 < \cdots < \widehat{z}_N < +\infty$ such that $A_n = \{\boldsymbol{\omega} : Z_{\boldsymbol{\lambda}^*}(\boldsymbol{\omega}) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ for $n = 1, \ldots, N$.*

In their seminal paper, Hölmstrom and Milgrom (1991) show that when the agent faces multiple tasks, overincentivizing tasks that generate precise performance data may prevent the completion of tasks that generate noisy performance data. That analysis abstracts away from monitoring costs and focuses on the power of linear compensation schemes.

Corollary 3 delivers a different message: if the principal's main problem is to allocate limited resources across the assessments of multiple task performances, then the optimal resource allocation should reflect the agent's endogenous tendency to shirk each task. The usefulness of this result is illustrated by the next example.

Example 3. A cashier faces two tasks: to scan items and to project warmth to customers. A piece of performance data consists of the scanner data recorded by the point of sale (POS) system, as well as the feedback gathered from customers. By Corollary 3, the ratio

$$R = \frac{\lambda_{01}^* + \lambda_{00}^*}{\lambda_{10}^* + \lambda_{00}^*}$$

captures how the principal should allocate limited resources across the assessments of the skillfulness in scanning items and warmth. Intuitively, a small R arises when the cashier is reluctant to project warmth to customers, in which case resources should be concentrated on the assessment of warmth, and the final performance rating should depend significantly on such assessment.

We examine how the optimal resource allocation depends on the precision of raw performance data. We make the following assumptions as in Hölmstrom and Milgrom (1991).

- First, $\omega_i = a_i + \xi_i$ for $i = 1, 2$, where ξ_is are independent normal random variables with mean zero and variances σ_i^2s.
- Second, the cashier's utility of consumption has a constant absolute risk aversion (CARA), i.e., $u(w) = 1 - \exp(-\gamma w)$.

Unlike Hölmstrom and Milgrom (1991), we do not confine ourselves to linear wage schemes.

In the case where the monitoring cost is an increasing function of the rating scale, we compute R for different values of σ_1^2, holding $\sigma_2^2 = 1$ and $|\mathcal{P}| = 2$ fixed. Our findings are reported in Figure 5. Assuming that our parameter choices are reasonable, we arrive at the following conclusion: as skillfulness becomes easier to measure—thanks to the availability of high quality scanner data—the cashier becomes more afraid to shirk the scanning task and less so about projecting coldness to customers; to correct the agent's incentive, resources should be shifted toward the processing and analysis of customer feedback and away from that of scanner data. In the future, one can test this prediction by running field experiments such as that of Bloom et al. (2013). For example, one can randomize the quality of scanner data among otherwise similar stores and examine its effect on the resource allocation between scanner data and customer feedback. $\Diamond$

6. Conclusion

We conclude by posing a few open questions. First, our work is broadly related to the burgeoning literature on information design (see, e.g., Bergemann and Morris 2019 for a survey), and we hope it inspires new research questions such as how to conduct costly yet flexible monitoring in long-term employment relationships. Second, our theory may guide investigations into empirical issues such as how advancements in big data technologies have affected the design and implementation of monitoring technologies, and whether they can partially explain the heterogeneity in the internal organizations of otherwise similar firms. We hope someone, maybe us, will carry out these research agendas in the future.

Appendix A: Omitted proofs

A.1 *Proofs for Section 3*

In this appendix, we identify any N-partitional contract $\langle \mathcal{P}, w(\cdot)\rangle$ with its corresponding tuple $\langle A_n, \pi_n, z_n, w_n\rangle_{n=1}^N$, where A_n is a generic cell of $\mathcal{P}$, $\pi_n = P_1(A_n)$, $z_n = z(A_n)$, and

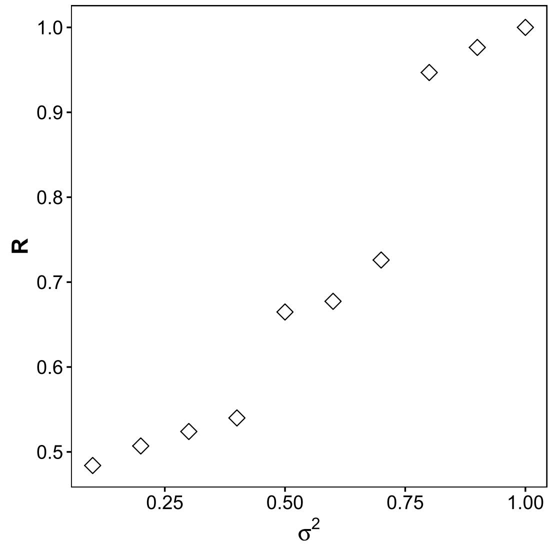

FIGURE 5. Plot of R against σ_1^2: $H(\mathcal{P}, a) = f(|\mathcal{P}|)$, $|\mathcal{P}| = 2$; $u(w) = 1 - \exp(-0.5w)$; $c(00) = 0$, $c(01) = 0.3$, $c(10) = 0.2$ and $c(11) = 0.5$; ξ_1 and ξ_2 are normally distributed with mean zero and $\sigma_2^2 = 1$.

$w_n = w(A_n)$. In addition, we assume without loss of generality (w.l.o.g.) that $z_1 \leq \cdots \leq z_N$.

A.1.1 *Useful lemmas*

PROOF OF LEMMA 1. The wage-minimization problem for any given monitoring technology as in Lemma 1 is

$$\min_{\langle \tilde{w}_n \rangle} \sum_n \pi_n \tilde{w}_n - \lambda \left[\sum_n \pi_n u(\tilde{w}_n) z_n - c \right] - \sum_n \eta_n \tilde{w}_n,$$

where λ and η_n denote the Lagrange multipliers associated with the (IC) constraint and the (LL) constraint at $\tilde{w}_n$, respectively. Differentiating the objective function with respect to $\tilde{w}_n$ and setting the result equal to 0, we obtain $\lambda z_n u'(w_n) = 1 - \eta_n/\pi_n$, and so $u'(w_n) = 1/(\lambda z_n)$ if and only if $w_n > 0$. □

PROOF OF LEMMA 2. Fix any optimal incentive contract that induces high effort from the agent, and let $\langle A_n, \pi_n, z_n, w_n \rangle_{n=1}^N$ be its corresponding tuple. By Assumption 1(b), if $w_j = w_k$ for some $j \neq k$, then merging A_j and A_k has no effect on the incentive cost but strictly reduces the monitoring cost, which contradicts the optimality of the original contract. Then from the assumption that $z_1 \leq \cdots \leq z_N$, it follows that $0 \leq w_1 < \cdots < w_N$ and, by Lemma 1, $z_1 < \cdots < z_N$. This last observation, coupled with $\sum_n \pi_n z_n = 0$, implies $z_1 < 0$ and hence $w_1 = 0$, because otherwise replacing w_1 with 0 reduces the expected wage and relaxes the (IC) constraint while keeping the (LL) constraint satisfied. Finally, combining $w_n > 0$ for $n \geq 2$ and Lemma 1 yields $z_n > 0$ for $n \geq 2$. □

Lemma 3. *For any $A \in \Sigma$ such that $P_1(A) > 0$ and any $\epsilon \in (0, P_1(A)]$, there exists $A_\epsilon \subset A$ such that $P_1(A_\epsilon) = \epsilon$ and $z(A_\epsilon) = z(A)$.*

Proof. Let A be as above. Since P_1 admits a density, it follows that for any $t \in (0, P_1(A)]$, there exists $B_t \subset A$ such that $P_1(B_t) = t$, and $Z(\omega') \leq Z(\omega)$ for any $\omega \in B_t$ and $\omega' \in A \setminus B_t$. Likewise, there exists $C_t \subset A$ such that $P_1(C_t) = t$, and $Z(\omega') \geq Z(\omega)$ for any $\omega \in C_t$ and $\omega' \in A \setminus C_t$. For $t = 0$, define $B_0 = C_0 = \varnothing$.

Let ϵ be as above, and notice two things about the set $B_t \cup C_{\epsilon-t}$, where $t \in [0, \epsilon]$. First, $P_1(B_t \cup C_{\epsilon-t}) = \epsilon$ by construction. Second, $z(B_t \cup C_{\epsilon-t})$ is continuous in t, because P_1 admits a density. The second observation, coupled with $z(B_t) \geq z(A)$ and $z(C_{\epsilon-t}) \leq z(A)$ for any $t \in (0, \epsilon)$, implies the existence of $t \in [0, \epsilon]$ such that $z(B_t \cup C_{\epsilon-t}) = z(A)$. Thus, let $A_\epsilon = B_t \cup C_{\epsilon-t}$, and we are done. $\square$

Proof of Theorem 1. Take any optimal incentive contract that induces high effort from the agent, and let $\langle A_n, \pi_n, z_n, w_n \rangle_{n=1}^N$ be its corresponding tuple. Suppose, to the contrary, that some A_j is not Z-convex. By Definition 1, there exist $A', A'' \subset A_j$ and $\tilde{A} \subset A_k$ for some $k \neq j$, such that (i) $P_1(A'_\epsilon), P_1(A''_\epsilon), P_1(\tilde{A}_\epsilon) > 0$, and (ii) $z' \neq z''$ and $\tilde{z} = (1-s)z' + sz''$ for some $s \in (0,1)$, where $z' := z(A')$, $z'' := z(A'')$, and $\tilde{z} := z(\tilde{A})$. Then by Lemma 3, for any $\epsilon \in (0, \min\{P_1(A'), P_1(A''), P_1(\tilde{A})\})$, there exist $A'_\epsilon \subset A'$, $A''_\epsilon \subset A''$, and $\tilde{A}_\epsilon \subset \tilde{A}$ such that (i) $P_1(A'_\epsilon) = P_1(A''_\epsilon) = P_1(\tilde{A}_\epsilon) = \epsilon$, and (ii) $z(A'_\epsilon) = z'$, $z(A''_\epsilon) = z''$, and $z(\tilde{A}_\epsilon) = \tilde{z}$.

Consider two perturbations to the monitoring technology: (a) move A'_ϵ from A_j to A_k and $\tilde{A}_\epsilon$ from A_k to A_j; (b) move $\tilde{A}_\epsilon$ from A_k to A_j and A''_ϵ from A_k to A_k. By construction, neither perturbation affects the probability distribution of the output signal given high effort and hence the monitoring cost. Below we demonstrate that one of them strictly reduces the incentive cost compared to the original (optimal) contract, thus reaching a contradiction. The conclusion is that all A_ns are Z-convex.

Perturbation (a). Let $\langle A_n(\epsilon), \pi_n, z_n(\epsilon) \rangle_{n=1}^N$ be the corresponding tuple to the monitoring technology after perturbation (a), where $A_j(\epsilon) = (A_j \cup \tilde{A}_\epsilon) \setminus A'_\epsilon$, $A_k(\epsilon) = (A_k \cup A'_\epsilon) \setminus \tilde{A}_\epsilon$, and $A_n(\epsilon) = A_n$ for any $n \neq j, k$. Straightforward algebra shows that

$$\begin{cases} z_j(\epsilon) = z_j + \dfrac{s(z''-z')}{\pi_j}\epsilon \\ z_k(\epsilon) = z_k - \dfrac{s(z''-z')}{\pi_k}\epsilon \\ z_n(\epsilon) = z_n \quad \forall n \neq j, k. \end{cases} \tag{2}$$

Take any wage profile $\langle w_n(\epsilon) \rangle_{n=1}^N$ with $w_1(\epsilon) = 0$ such that the (IC) constraint remains binding after this perturbation, i.e.,

$$\sum_{n=1}^N \pi_n u(w_n(\epsilon)) z_n(\epsilon) = \sum_{n=1}^N \pi_n u(w_n) z_n = c. \tag{3}$$

A careful inspection of (2) and (3) reveals the existence of $M > 0$ independent of ϵ such that when ϵ is small, we can construct a wage profile as above that satisfies, in addition, $|w_n(\epsilon) - w_n| < M\epsilon$ for any $n = 1, \cdots, N$ and, hence, the (LL) constraint by Lemma 2.[11]

With a slight abuse of notation, write $\dot{w}_n(\epsilon) = (w_n(\epsilon) - w_n)/\epsilon$ and $\dot{z}_n(\epsilon) = (z_n(\epsilon) - z_n)/\epsilon$, and note that $\dot{w}_1(\epsilon) = 0$.[12] When ϵ is small, expanding (3) using the twice-differentiability of $u(\cdot)$ and $|w_n(\epsilon) - w_n| \sim \mathcal{O}(\epsilon)$ yields

$$\sum_{n=1}^{N} \pi_n u(w_n) z_n = \sum_{n=1}^{N} \pi_n \big(u(w_n) + u'(w_n) \cdot \dot{w}_n(\epsilon) \cdot \epsilon + \mathcal{O}\big(\epsilon^2\big)\big)\big(z_n + \dot{z}_n(\epsilon) \cdot \epsilon\big).$$

Multiplying both sides of this equation by the Lagrange multiplier $\lambda > 0$ associated with the (IC) constraint prior to the perturbation and rearranging, we obtain

$$\sum_{n=1}^{N} \pi_n \cdot u'(w_n) \cdot \lambda z_n \cdot \dot{w}_n(\epsilon) = -\lambda \sum_{n=1}^{N} u(w_n) \cdot \pi_n \dot{z}_n(\epsilon) + \mathcal{O}(\epsilon),$$

Simplifying using $\dot{w}_1(\epsilon) = 0$, $u'(w_n) = 1/(\lambda z_n)$ for $n \geq 2$ (Lemmas 1 and 2), and (2) yields

$$\sum_{n=1}^{N} \pi_n \dot{w}_n(\epsilon) = s\big[u(w_k) - u(w_j)\big]\big(\lambda z'' - \lambda z'\big) + \mathcal{O}(\epsilon). \tag{4}$$

Perturbation (b). Repeating the above argument for perturbation (b) yields

$$\sum_{n=1}^{N} \pi_n \dot{w}_n(\epsilon) = -\lambda(1-s)\big[u(w_k) - u(w_j)\big]\big(z'' - z'\big) + \mathcal{O}(\epsilon). \tag{5}$$

Then from $u(w_j) \neq u(w_k)$ (Lemma 2), $z' \neq z''$ (by assumption), and $\lambda > 0$, it follows that the right-hand side of either (4) or (5) is strictly negative when ϵ is small. Thus, for either perturbation (a) or (b), we can construct a wage profile that incurs a lower incentive cost than the original optimal contract, and this leads to a contradiction. □

Proof of Theorem 2. By Theorem 1, any optimal monitoring technology with at most $N \in \{2, \dots, K\}$ cells is fully characterized by $N - 1$ cut points $\widehat{z}_1, \dots, \widehat{z}_{N-1}$, where $\min Z(\Omega) \leq \widehat{z}_1 \leq \cdots \leq \widehat{z}_{N-1} \leq \max Z(\Omega)$. Fix any $N \in \{2, \dots, K\}$, and write $\widehat{\mathbf{z}}$ for a generic vector $(\widehat{z}_1, \dots, \widehat{z}_{N-1})^\top$ of cut points. Equip the set of cut point vectors

$$\mathcal{Z}_N = \big\{\widehat{\mathbf{z}} : \min Z(\Omega) \leq \widehat{z}_1 \leq \cdots \leq \widehat{z}_{N-1} \leq \max Z(\Omega)\big\}$$

[11] To be precise, recall that $u(w_n)$, $z_n > 0$ for $n \geq 2$ by Lemma 2, so $z_n(\epsilon) > 0$ for $n \geq 2$ when ϵ is small. For such ϵ and $\langle z_n(\epsilon)\rangle_{n=2}^{N}$, solving $\sum_{n=2}^{N} \pi_n u(x_n) z_n(\epsilon) = \sum_{n=2}^{N} \pi_n u(w_n) z_n$ yields the desired wage profile $\langle w_n(\epsilon)\rangle_{n=2}^{N}$ as above when ϵ is small.

[12] We do *not* assume that $w_n(\epsilon)$ and $z_n(\epsilon)$ are differentiable with respect to ϵ here or in the remainder of this paper.

with the sup norm $\|\cdot\|$,[13] and note that $\mathcal{Z}_N$ is compact by Assumption 3. Let $W(\widehat{\mathbf{z}})$ denote the minimal incentive cost for inducing high effort when the monitoring technology is given by $\widehat{\mathbf{z}} \in \mathcal{Z}_N$. Note that $W(\widehat{\mathbf{z}})$ is finite if and only if $\min Z(\Omega) < \widehat{z}_n < \max Z(\Omega)$ for some n, since the last condition implies that $z(A) \not\equiv 0$ across the performance category As generated by $\widehat{\mathbf{z}}$, and so $W(\widehat{\mathbf{z}})$ can be obtained by solving the wage minimization problem considered in Lemma 1.

We proceed in two steps.

Step 1. Show that $W(\widehat{\mathbf{z}})$ is continuous in $\widehat{\mathbf{z}}$ for any given $N \in \{2, \dots, K\}$. Fix any $\widehat{\mathbf{z}} \in \mathcal{Z}_N$ such that $W(\widehat{\mathbf{z}})$ is finite, and consider w.l.o.g. the case where $\widehat{z}_n$s are all distinct. For a generic performance category $A_n = \{\omega : Z(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ generated by $\widehat{\mathbf{z}}$, we use π_n and z_n to denote its probability given $a = 1$ and z-value, respectively, and use w_n to denote the optimal wage at A_n. Then, for a sufficiently small $\delta > 0$, we take any δ-perturbation $\widehat{\mathbf{z}}^\delta$ to $\widehat{\mathbf{z}}$ such that $\|\widehat{\mathbf{z}}^\delta - \widehat{\mathbf{z}}\| < \delta$. For a generic performance category $A_n^\delta = \{\omega : Z(\omega) \in [\widehat{z}_{n-1}^\delta, \widehat{z}_n^\delta)\}$ generated by $\widehat{\mathbf{z}}^\delta$, we use π_n^δ to denote its probability given $a = 1$, and z_n^δ to denote its z-value.

Now, fix any $\epsilon > 0$, and consider a wage profile that pays, at A_n^δ, $w_n + \epsilon$ if $z_n^\delta > 0$ and w_n otherwise. This wage profile satisfies the (LL) constraint by construction. Under Assumptions 2 and 3, it satisfies the (IC) constraint when δ is sufficiently small, because

$$\lim_{\delta \to 0} \sum_n \pi_n^\delta u(w_n + 1_{z_n^\delta > 0} \cdot \epsilon) z_n^\delta = \sum_n \pi_n u(w_n + 1_{z_n > 0} \cdot \epsilon) z_n > c,$$

where the inequality uses the fact that $z_n > 0$ for some n, obtained from combining $\sum_n \pi_n z_n = 0$ and $z_n \not\equiv 0$. In addition, since

$$\lim_{\delta \to 0} \sum_n \pi_n^\delta (w_n + 1_{z_n^\delta > 0} \cdot \epsilon) = \sum_n \pi_n (w_n + 1_{z_n > 0} \cdot \epsilon),$$

we obtain, when δ is sufficiently small,

$$W(\widehat{\mathbf{z}}^\delta) - W(\widehat{\mathbf{z}}) \le \sum_n \pi_n^\delta (w_n + 1_{z_n^\delta > 0} \cdot \epsilon) - \sum_n \pi_n w_n < \epsilon,$$

where the first inequality uses the fact that the above constructed wage profile is not necessarily optimal when the monitoring technology is given by $\widehat{\mathbf{z}}^\delta$. Finally, interchanging the roles between $\widehat{\mathbf{z}}$ and $\widehat{\mathbf{z}}^\delta$ in the above argument yields $W(\widehat{\mathbf{z}}) - W(\widehat{\mathbf{z}}^\delta) < \epsilon$, thus proving that $|W(\widehat{\mathbf{z}}^\delta) - W(\widehat{\mathbf{z}})| < \epsilon$ when δ is sufficiently small.

Step 2. Under Assumption 4(a), the existence and finiteness of

$$W_N := \min_{\widehat{\mathbf{z}} \in \mathcal{Z}_N} W(\widehat{\mathbf{z}})$$

for any $N \in \{2, \dots, K\}$ follow from Step 1 and the compactness of $\mathcal{Z}_N$. Let m_N denote the minimal rating scale attained by the solution(s) to the above minimization problem. Solving

$$\min_{2 \le N \le K} W_N + \mu \cdot f(m_N)$$

yields the solution(s) to the principal's problem.

[13] We use $\|\cdot\|$ to denote the sup norm in the remainder of this paper.

Under Assumption 4(b), we can write the principal's problem as

$$\min_{\widehat{\mathbf{z}} \in \mathcal{Z}_K} W(\widehat{\mathbf{z}}) + \mu \cdot h\big(\boldsymbol{\pi}(\widehat{\mathbf{z}})\big),$$

where $\boldsymbol{\pi}(\widehat{\mathbf{z}})$ compiles the probabilities of the output signal generated by $a = 1$ and $\widehat{\mathbf{z}}$, and it is clearly continuous in $\widehat{\mathbf{z}}$. The existence of solution(s) to this problem then follows from Step 1 and the compactness of $\mathcal{Z}_K$. □

A.2 *Proofs for Section 4*

In this appendix, we identify any N-partitional contract $\langle \mathcal{P}, \mathbf{w}(\cdot)\rangle$ with its corresponding tuple $\langle A_n, \pi_n, \mathbf{z}_n, \mathbf{w}_n\rangle_{n=1}^N$, where A_n is a generic cell of $\mathcal{P}$, $\pi_n = P_{\mathbf{1}}(A_n)$, $\mathbf{z}_n = (z_{1,n}, z_{2,n})^\top = (z_1(A_n), z_2(A_n))^\top$, and $\mathbf{w}_n = (w_{1,n}, w_{2,n})^\top = (w_1(A_n), w_2(A_n))^\top$.

A.2.1 *Useful lemmas* The next lemma generalizes Lemmas 1 and 2 to encompass multiple agents.

Lemma 4. *Assume Assumption 1. Then for any optimal incentive contract, (i) there exist $\lambda_1, \lambda_2 > 0$ such that $u_i'(w_{i,n}) = 1/(\lambda_i z_{i,n})$ if and only if $w_{i,n} > 0$; (ii) $\mathbf{w}_j \neq \mathbf{w}_k$ for any $j \neq k$.*

Proof. The wage minimization problem for any given monitoring technology is

$$\min_{\langle \tilde{w}_{i,n}\rangle} \sum_{i,n} \pi_n \tilde{w}_{i,n} - \sum_i \lambda_i \left[\sum_n \pi_n u_i(\tilde{w}_{i,n}) z_{i,n} - c_i\right] - \sum_{i,n} \eta_{i,n} \tilde{w}_{i,n},$$

where λ_i and $\eta_{i,n}$ denote the Lagrange multipliers associated with the (IC_i) constraint and the (LL_i) constraint at $\tilde{w}_{i,n}$, respectively. Differentiating the objective function with respect to $\tilde{w}_{i,n}$ yields the first-order condition in part (i). The proof of part (ii) closely parallels that of Lemma 2 and is therefore omitted here. □

The next lemma plays an analogous role to that of Lemma 3.

Lemma 5. *Fix any $\delta > 0$ and any $A \in \Sigma$ such that $P_{\mathbf{1}}(A) > 0$, and assume Assumption 6. Then for any $\epsilon \in (0, P_{\mathbf{1}}(A)]$, there exists $A_\epsilon \subset A$ such that $P_{\mathbf{1}}(A_\epsilon) = \epsilon$ and $\|\mathbf{z}(A_\epsilon) - \mathbf{z}(A)\| < \delta$.*

Proof. With a slight abuse of notation, let $\mathcal{P}$ be any finite partition of Ω such that every $B \in \mathcal{P}$ is measurable, and $\|\mathbf{Z}(\omega) - \mathbf{Z}(\omega')\| < \delta$ for any $\omega, \omega' \in B$. Such $\mathcal{P}$ exists because $P_{\mathbf{1}}$ admits a density, and $\mathbf{Z}(\Omega)$ is a compact set in $\mathbb{R}^2$ (Assumption 6). Define $\mathcal{P}^+ = \{B \in \mathcal{P} : P_{\mathbf{1}}(A \cap B) > 0\}$ and $\mathcal{P}^0 = \{B \in \mathcal{P} : P_{\mathbf{1}}(A \cap B) = 0\}$, which are both finite. Note that $\sum_{B \in \mathcal{P}^0} P_{\mathbf{1}}(A \cap B) = 0$, $\sum_{B \in \mathcal{P}^+} P_{\mathbf{1}}(A \cap B) = P_{\mathbf{1}}(A)$, and $\mathbf{z}(A) = \sum_{B \in \mathcal{P}^+} P_{\mathbf{1}}(A \cap B)\mathbf{z}(A \cap B)$.

Since $P_{\mathbf{1}}$ admits a density, it follows that for any $B \in \mathcal{P}^+$, there exists $C_B \subset A \cap B$ such that $P_1(C_B) = P_{\mathbf{1}}(A \cap B)\epsilon / P_{\mathbf{1}}(A)$. Also, $\|\mathbf{z}(C_B) - \mathbf{z}(A \cap B)\| < \delta$ by construction. Thus

$A_\epsilon = \cup_{B \in \mathcal{P}^+} C_B$ is the desired set, because $P_{\mathbf{1}}(A_\epsilon) = \sum_{B \in \mathcal{P}^+} P_{\mathbf{1}}(A \cap B)\epsilon / P_{\mathbf{1}}(A) = \epsilon$, and

$$\begin{aligned}\|\mathbf{z}(A_\epsilon) - \mathbf{z}(A)\| &= \left\| \sum_{B \in \mathcal{P}^+} \frac{P_{\mathbf{1}}(A \cap B)}{P_{\mathbf{1}}(A)} \big(\mathbf{z}(C_B) - \mathbf{z}(A \cap B)\big) \right\| \\ &\le \sum_{B \in \mathcal{P}^+} \frac{P_{\mathbf{1}}(A \cap B)}{P_{\mathbf{1}}(A)} \|\mathbf{z}(C_B) - \mathbf{z}(A \cap B)\| < \delta. \end{aligned}$$

□

Proof of Theorem 3. Take any optimal incentive contract that induces high effort from both agents, and let $\langle A_n, \pi_n, \mathbf{z}_n, \mathbf{w}_n \rangle_{n=1}^N$ be its corresponding tuple. Suppose, to the contrary, that some A_j is not $\mathbf{Z}$-convex. By definition, there exist $A', A'' \subset A_j$ and $\tilde{A} \in A_k$ for some $k \neq j$, such that (i) $P_{\mathbf{1}}(A''_\epsilon), P_{\mathbf{1}}(A''_\epsilon), P_{\mathbf{1}}(\tilde{A}_\epsilon) > 0$, and (ii) $\mathbf{z}' \neq \mathbf{z}''$ and $\tilde{\mathbf{z}} = (1-s)\mathbf{z}' + s\mathbf{z}''$ for some $s \in (0,1)$, where $\mathbf{z}' := \mathbf{z}(A')$, $\mathbf{z}'' := \mathbf{z}(A'')$, and $\tilde{\mathbf{z}} := \mathbf{z}(\tilde{A})$. Then by Lemma 5, for any $\delta > 0$ and any $\epsilon \in (0, \min\{P_{\mathbf{1}}(A'), P_{\mathbf{1}}(A''), P_{\mathbf{1}}(\tilde{A}\}))$, there exist $A'_\epsilon \subset A'$, $A''_\epsilon \subset A''$, and $\tilde{A}_\epsilon \subset \tilde{A}$ such that (i) $P_{\mathbf{1}}(A''_\epsilon) = P_{\mathbf{1}}(A''_\epsilon) = P_{\mathbf{1}}(\tilde{A}_\epsilon) = \epsilon$, and (ii) $\|\mathbf{z}(A'_\epsilon) - \mathbf{z}'\|$, $\|\mathbf{z}(A''_\epsilon) - \mathbf{z}''\|$, $\|\mathbf{z}(\tilde{A}_\epsilon) - \tilde{\mathbf{z}}\| < \delta$.

Consider two perturbations to the monitoring technology: (a) move A'_ϵ from $A_j$to A_k and $\tilde{A}_\epsilon$ from A_k to A_j; (b) move $\tilde{A}_\epsilon$ from A_k to A_j and A''_ϵ from A_j to A_k. By Assumption 1, neither perturbation affects the probability distribution of the output signal given the high effort profile and hence the monitoring cost. Below we demonstrate that one of them strictly reduces the incentive cost compared to the original optimal contract.

Perturbation (a). Let $\langle A_n(\epsilon), \pi_n, \mathbf{z}_n(\epsilon) \rangle_{n=1}^N$ denote the corresponding tuple to the monitoring technology after perturbation (a), where $A_j(\epsilon) = (A_j \cup \tilde{A}_\epsilon) \setminus A'_\epsilon$, $A_k(\epsilon) = (A_k \cup A'_\epsilon) \setminus \tilde{A}_\epsilon$, and $A_n(\epsilon) = A_n$ for any $n \neq j, k$. Straightforward algebra shows that

$$\begin{cases} \mathbf{z}_j(\epsilon) = \mathbf{z}_j + \dfrac{\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}\big(A'_\epsilon\big)}{\pi_j}\epsilon \\ \mathbf{z}_k(\epsilon) = \mathbf{z}_k - \dfrac{\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}\big(A'_\epsilon\big)}{\pi_k}\epsilon \\ \mathbf{z}_n(\epsilon) = \mathbf{z}_n \quad \forall n \neq j, k, \end{cases} \tag{6}$$

where

$$\big\|\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}\big(A'_\epsilon\big) - \big(\tilde{\mathbf{z}} - \mathbf{z}'\big)\big\| \le \big\|\mathbf{z}(\tilde{A}_\epsilon) - \tilde{\mathbf{z}}\big\| + \big\|\mathbf{z}\big(A'_\epsilon\big) - \mathbf{z}'\big\| < \min\Big\{2\delta, 4\max_{\omega \in \Omega}\big\|\mathbf{Z}(\omega)\big\|\Big\}. \tag{7}$$

Define $\mathcal{B}_i = \{n : w_{i,n} = 0\}$ for $i = 1, 2$. Let $\langle \mathbf{w}_n(\epsilon) \rangle_{n=1}^N$ be any wage profile such that for both $i = 1, 2$: (i) $w_{i,n}(\epsilon) = w_{i,n} = 0$ if $n \in \mathcal{B}_i$; (ii) agent i's incentive compatibility constraint remains binding after perturbation (a), i.e.,

$$\sum_{n=1}^N \pi_n u_i\big(w_{i,n}(\epsilon)\big) z_{i,n}(\epsilon) = \sum_{n=1}^N \pi_n u_i(w_{i,n}) z_{i,n} = c_i. \tag{8}$$

A careful inspection of (6), (7), and (8) reveals the existence of $M > 0$ independent of ϵ and δ such that when ϵ is sufficiently small, we can construct a wage profile as above that satisfies, in addition, $\|\mathbf{w}_n(\epsilon) - \mathbf{w}_n\| < M\epsilon$ for all n and, hence, both (LL_i) constraints.

With a slight abuse of notation, write $\dot{\mathbf{w}}_n(\epsilon) = (\mathbf{w}_n(\epsilon) - \mathbf{w}_n)/\epsilon$ and $\dot{\mathbf{z}}_n(\epsilon) = (\mathbf{z}_n(\epsilon) - \mathbf{z}_n)/\epsilon$, and note that $\dot{w}_{i,n}(\epsilon) = 0$ if $n \in \mathcal{B}_i$ for both $i = 1, 2$. Multiplying both sides of (8) by the Lagrange multiplier $\lambda_i > 0$ associated with the (IC_i) constraint prior to the perturbation and expanding, we obtain, when ϵ is small,

$$\sum_{n=1}^{N} \pi_n \cdot u_i'(w_{i,n}) \cdot \lambda_i z_{i,n} \cdot \dot{w}_{i,n}(\epsilon) = -\lambda_i \sum_{n=1}^{N} u_i(w_{i,n}) \cdot \pi_n \dot{z}_{i,n}(\epsilon) + \mathcal{O}(\epsilon).$$

Simplifying using $\dot{w}_{i,n}(\epsilon) = 0$ if $n \in \mathcal{B}_i$, $u'(w_{i,n}) = 1/(\lambda_i z_{i,n})$ if $n \notin \mathcal{B}_i$ (Lemma 4), and (6) yields

$$\sum_{i,n} \pi_n \dot{w}_{i,n} = (\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda\big(\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}(A'_\epsilon)\big) + \mathcal{O}(\epsilon),$$

where $\mathbf{u}_n = (u_1(w_{1,n}), u_2(w_{2,n}))^\top$ for $n = k, j$, and $\Lambda = \left(\begin{smallmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{smallmatrix}\right)$. Further simplifying using (7) and $\tilde{\mathbf{z}} = (1-s)\mathbf{z}' + s\mathbf{z}''$ yields, when δ is small,

$$\begin{aligned} \sum_{i,n} \pi_n \dot{w}_{i,n} &= (\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda\big(\tilde{\mathbf{z}} - \mathbf{z}'\big) + \mathcal{O}(\epsilon) + (\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda\big(\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}(A'_\epsilon) - (\tilde{\mathbf{z}} - \mathbf{z}')\big) \\ &= s(\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda\big(\mathbf{z}'' - \mathbf{z}'\big) + \mathcal{O}(\epsilon) + \mathcal{O}(\delta). \end{aligned} \tag{9}$$

Perturbation (b). Repeating the above argument for perturbation (b) yields

$$\sum_{i,n} \pi_n \dot{w}_{i,n} = -(1-s)(\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda\big(\mathbf{z}'' - \mathbf{z}'\big) + \mathcal{O}(\epsilon) + \mathcal{O}(\delta). \tag{10}$$

Consider two cases.

Case 1: $(\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda(\mathbf{z}'' - \mathbf{z}') \neq 0$. In this case, the right-hand sides of (9) and (10) have the opposite signs when ϵ and δ are sufficiently small. The remainder of the proof closely parallels that of Theorem 1 and is therefore omitted here.

Case 2: $(\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda(\mathbf{z}'' - \mathbf{z}') = 0$. In this case, note that $(\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda \neq \mathbf{0}^\top$ by Lemma 4, where $\mathbf{0}$ denotes the 2-vector of zeros. Then from Assumption 5 (the distribution of $\mathbf{Z}$ is atomless), there exist $B' \subset A'$, $B'' \subset A''$ and $\tilde{B} \subset \tilde{A}$ such that $P_{\mathbf{1}}(B')$, $P_{\mathbf{1}}(B'')$, $P_{\mathbf{1}}(\tilde{B}) > 0$, $\mathbf{z}(\tilde{B}) = (1-s')\mathbf{z}(B') + s'\mathbf{z}(B'')$ for some $s' \in (0,1)$, and $(\mathbf{u}_k - \mathbf{u}_j)^\top \Lambda(\mathbf{z}(B'') - \mathbf{z}(B')) \neq 0$. Replacing A', A'', and $\tilde{A}$ with B', B'', and $\tilde{B}$, respectively, in the above argument gives the desired result. □

Proof of Theorem 4. By Theorem 3, any optimal monitoring technology with at most $N \in \{2, \ldots, K\}$ cells is fully characterized by (i) a finite number q_N of vertices $\mathbf{z}_1, \ldots, \mathbf{z}_{q_N}$ in $\mathbf{Z}(\Omega)$, and (ii) a $q_N \times q_N$ adjacency matrix $\mathbf{M}$, whose lmth entry equals 1 if $\mathbf{z}_l$ and $\mathbf{z}_m$ are connected by a line segment and equals 0 otherwise. By definition, $\mathbf{M}$ is symmetric and is therefore determined by its upper triangle entries, which can be either 0 or 1. Thus $\mathbf{M}$ belongs to $\mathcal{M}_N := \{0,1\}^{q_N \times (q_N - 1)/2}$, which is a finite set.

Fix any $N \in \{2, \ldots, K\}$, and write $\vec{\mathbf{z}}$ for $(\mathbf{z}_1, \ldots, \mathbf{z}_{q_N})^\top$. For any adjacency matrix $\mathbf{M} \in \mathcal{M}_N$, define

$$\mathcal{Z}_N(\mathbf{M}) = \big\{\vec{\mathbf{z}} : (\vec{\mathbf{z}}, \mathbf{M}) \text{ partitions } \mathbf{Z}(\Omega) \text{ into at most } N \text{ convex polygons}\big\},$$

equip $\mathcal{Z}_N(\mathbf{M})$ with the sup norm $\|\cdot\|$, and note that $\mathcal{Z}_N(\mathbf{M})$ is compact by Assumption 6. Let $W(\vec{\mathbf{z}}, \mathbf{M})$ denote the minimal incentive cost for inducing high effort from both agents when the monitoring technology is given by $(\vec{\mathbf{z}}, \mathbf{M})$. Note that $W(\vec{\mathbf{z}}, \mathbf{M})$ is finite if and only if for both $i = 1, 2$, $z_i(A) \not\equiv 0$ across the performance category As generated by $(\vec{\mathbf{z}}, \mathbf{M})$.

We proceed in two steps.

Step 1. Show that $W(\vec{\mathbf{z}}, \mathbf{M})$ is continuous in its first argument for any given $N \in \{2, \ldots, K\}$ and $\mathbf{M} \in \mathcal{M}_N$. Fix any $\vec{\mathbf{z}} \in \mathcal{Z}_N(\mathbf{M})$ such that $W(\vec{\mathbf{z}}, \mathbf{M})$ is finite, and consider w.l.o.g. the case where $\mathbf{z}_l$s are all distinct. For a generic performance category A_n generated by $(\vec{\mathbf{z}}, \mathbf{M})$, we use π_n and $z_{i,n}$ to denote its probability given $\mathbf{a} = \mathbf{1}$ and z_i-value, respectively, and use $w_{i,n}$ to denote the optimal wage for agent i at A_n. Then, for a sufficiently small $\delta > 0$, we take any δ-perturbation $\vec{\mathbf{z}}^\delta \in \mathcal{Z}_N(\mathbf{M})$ to $\vec{\mathbf{z}}$ such that $\|\vec{\mathbf{z}}^\delta - \vec{\mathbf{z}}\| < \delta$. Label the performance categories generated by $(\vec{\mathbf{z}}^\delta, \mathbf{M})$ as A_n^δs, such that for each $n = 1, 2, \ldots$, $\mathbf{z}_l$ is a vertex of $\mathrm{cl}(\mathbf{Z}(A_n))$ if and only if $\mathbf{z}_l^\delta$ is a vertex of $\mathrm{cl}(\mathbf{Z}(A_n^\delta))$. For a generic performance category A_n^δ, we use π_n^δ and $z_{i,n}^\delta$ to denote its probability given $\mathbf{a} = \mathbf{1}$ and z_i-value, respectively.

Now, fix any $\epsilon > 0$, and consider the wage profile that pays, at A_n^δ, $w_{i,n} + \epsilon/2$ to agent i if $z_{i,n}^\delta > 0$ and $w_{i,n}$ otherwise. This wage profile satisfies the (LL_i) constraint by construction. Under Assumptions 5 and 6, the (IC_i) constraint is satisfied when δ is sufficiently small, because

$$\lim_{\delta \to 0} \sum_n \pi_{i,n}^\delta u(w_{i,n} + 1_{z_{i,n}^\delta > 0} \cdot \epsilon/2) z_n^\delta = \sum_n \pi_n u(w_{i,n} + 1_{z_{i,n} > 0} \cdot \epsilon/2) z_{i,n} > c_i,$$

where the inequality uses the fact that $z_{i,n} > 0$ for some n, obtained from combining $\sum_n \pi_n z_{i,n} = 0$ and $z_{i,n} \not\equiv 0$. In addition, since

$$\lim_{\delta \to 0} \sum_{i,n} \pi_n^\delta (w_{i,n} + 1_{z_{i,n}^\delta > 0} \cdot \epsilon/2) = \sum_{i,n} \pi_n (w_{i,n} + 1_{z_{i,n} > 0} \cdot \epsilon/2),$$

we obtain, when δ is sufficiently small,

$$W\big(\vec{\mathbf{z}}^\delta, \mathbf{M}\big) - W(\vec{\mathbf{z}}, \mathbf{M}) \leq \sum_{i,n} \pi_n^\delta (w_{i,n} + 1_{z_{i,n}^\delta > 0} \cdot \epsilon/2) - \sum_{i,n} \pi_n w_{i,n} < \epsilon,$$

where the first inequality uses the fact that the above constructed wage profile is not necessarily optimal when the monitoring technology is given by $(\vec{\mathbf{z}}^\delta, \mathbf{M})$. Finally, interchanging the roles between $\vec{\mathbf{z}}^\delta$ and $\vec{\mathbf{z}}$ in the above argument yields $W(\vec{\mathbf{z}}^\delta, \mathbf{M}) - W(\vec{\mathbf{z}}^\delta, \mathbf{M}) < \epsilon$, thus proving that $|W(\vec{\mathbf{z}}^\delta, \mathbf{M}) - W(\vec{\mathbf{z}}^\delta, \mathbf{M})| < \epsilon$ when δ is sufficiently small.

Step 2. Under Assumption 4(a), the existence and finiteness of

$$W_N := \min_{\mathbf{M} \in \mathcal{M}_N, \vec{\mathbf{z}} \in \mathcal{Z}_N(\mathbf{M})} W(\vec{\mathbf{z}}, \mathbf{M})$$

for any $N \in \{2, \ldots, K\}$ follows from Step 1, the compactness of $\mathcal{Z}_N(\mathbf{M})$, and the finiteness of $\mathcal{M}_N$. Under Assumption 4(b), we can write the principal's problem as

$$\min_{\mathbf{M} \in \mathcal{M}_K, \vec{\mathbf{z}} \in \mathcal{Z}_K(\mathbf{M})} W(\vec{\mathbf{z}}, \mathbf{M}) + \mu \cdot h\big(\boldsymbol{\pi}(\vec{\mathbf{z}}, \mathbf{M})\big),$$

where $\boldsymbol{\pi}(\vec{\mathbf{z}}, \mathbf{M})$ compiles the probabilities of the output signal generated by $\mathbf{a} = \mathbf{1}$ and $(\vec{\mathbf{z}}, \mathbf{M})$, and it is clearly continuous in $\vec{\mathbf{z}}$. The remainder of the proof is identical to that of Theorem 2 and is therefore omitted here. □

A.3 *Proofs for Section 5*

In this appendix, we write $\mathbf{z}(A) = (z_a(A))_{a\in\mathcal{D}}^\top$ for any set $A \in \Sigma$ of positive measure, and identify any N-partitional contract $\langle \mathcal{P}, w(\cdot)\rangle$ with its corresponding tuple $\langle A_n, \pi_n, \mathbf{z}_n, w_n\rangle_{n=1}^N$, where A_n is a generic cell of $\mathcal{P}$, $\pi_n = P_{a^*}(A_n)$, $\mathbf{z}_n = \mathbf{z}(A_n)$, and $w_n = w(A_n)$. In addition, we assume w.l.o.g. that $w_1 \leq \cdots \leq w_N$.

A.3.1 *Useful lemma* The next lemma generalizes Lemmas 1 and 2 to encompass multiple tasks.

Lemma 6. *Assume Assumption 1. Then for any optimal incentive contract that induces a^*, there exists $\boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{D}|}$ with $\|\boldsymbol{\lambda}\| > 0$ such that (i) $u'(w_n) = 1/(\boldsymbol{\lambda}^\top \mathbf{z}_n)$ if and only if $w_n > 0$, and (ii) $\boldsymbol{\lambda}^\top \mathbf{z}_1 < 0 < \boldsymbol{\lambda}^\top \mathbf{z}_2 < \cdots$. In addition, $0 = w_1 < w_2 < \cdots$.*

Proof. The wage-minimization problem for any given monitoring technology is

$$\min_{\langle \tilde{w}_n\rangle} \sum_n \pi_n \tilde{w}_n - \sum_n \pi_n u(\tilde{w}_n) \cdot \boldsymbol{\lambda}^\top \mathbf{z}_n - \sum_n \eta_n \tilde{w}_n,$$

where $\boldsymbol{\lambda}$ denotes the profile of the Lagrange multipliers associated with (IC_a) constraints, and η_n denotes the Lagrange multiplier associated with the (LL) constraint at w_n. Note that $\|\boldsymbol{\lambda}\| > 0$, because otherwise all (IC_a) constraints are slack, so replacing every $w_n > 0$ with $w_n - \epsilon$ for some small but positive ϵ reduces the incentive cost while keeping all (IC_a) constraints and the (LL) constraint satisfied. Given this, we can then differentiate the objective function with respect to w_n and obtain the first-order condition in part (i). The remainder of the proof is identical to that of Lemma 2 and is therefore omitted here. □

Proof of Theorem 5. Take any optimal incentive contract that induces a^*. Let $\langle A_n, \pi_n, \mathbf{z}_n, w_n\rangle_{n=1}^N$ be its corresponding tuple, and let $\boldsymbol{\lambda}$ be the profile of the Lagrange multipliers associated with (IC_a) constraints. Suppose, to the contrary, that some A_j is not $Z_{\boldsymbol{\lambda}}$-convex. Then there exist $A', A'' \subset A_j$ and $\tilde{A} \subset A_k$ for some $k \neq j$, such that (i) $P_{a^*}(A'_\epsilon), P_{a^*}(A''_\epsilon), P_{a^*}(\tilde{A}_\epsilon) > 0$, and (ii) $\boldsymbol{\lambda}^\top \mathbf{z}' \neq \boldsymbol{\lambda}^\top \mathbf{z}''$ and $\boldsymbol{\lambda}^\top \tilde{\mathbf{z}} = (1-s)\boldsymbol{\lambda}^\top \mathbf{z}' + s\boldsymbol{\lambda}^\top \mathbf{z}''$ for some $s \in (0, 1)$, where $\mathbf{z}' := \mathbf{z}(A')$, $\mathbf{z}'' := \mathbf{z}(A'')$, and $\tilde{\mathbf{z}} := \mathbf{z}(\tilde{A})$. Then by Lemma 3, for any $\epsilon \in (0, \min\{P_{a^*}(A'), P_{a^*}(A''), P_{a^*}(\tilde{A})\})$, there exist $A'_\epsilon \subset A'$, $A''_\epsilon \subset A''$, and $\tilde{A}_\epsilon \subset \tilde{A}$ such that (i) $P_{a^*}(A'_\epsilon) = P_{a^*}(A''_\epsilon) = P_{a^*}(\tilde{A}_\epsilon) = \epsilon$, and (ii) $\boldsymbol{\lambda}^\top \mathbf{z}(A'_\epsilon) = \boldsymbol{\lambda}^\top \mathbf{z}'$, $\boldsymbol{\lambda}^\top \mathbf{z}(A''_\epsilon) = \boldsymbol{\lambda}^\top \mathbf{z}'$, and $\boldsymbol{\lambda}^\top \mathbf{z}(\tilde{A}_\epsilon) = \boldsymbol{\lambda}^\top \tilde{\mathbf{z}}$.

Consider two perturbations to the monitoring technology: (a) move A'_ϵ from A_j to A_k and $\tilde{A}_\epsilon$ from A_k to A_j, and (b) move $\tilde{A}_\epsilon$ from A_k to A_j and A''_ϵ from A_j to A_k. By Assumption 1, neither perturbation affects the probability distribution of the output signal given $a = a^*$ and hence the monitoring cost. Below we demonstrate that one of them strictly reduces the incentive cost compared to the original (optimal) contract.

Perturbation (a). Let $\langle A_n(\epsilon), \pi_n, \mathbf{z}_n(\epsilon)\rangle_{n=1}^N$ be the corresponding tuple to the monitoring technology after perturbation (a), where $A_j(\epsilon) = (A_j \cup \tilde{A}_\epsilon) \setminus A'_\epsilon$, $A_k(\epsilon) = (A_k \cup A'_\epsilon) \setminus \tilde{A}_\epsilon$, and $A_n(\epsilon) = A_n$ for any $n \neq j, k$. Straightforward algebra shows that

$$\begin{cases} \mathbf{z}_j(\epsilon) = \mathbf{z}_j + \dfrac{\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}(A'_\epsilon)}{\pi_j}\epsilon \\ \mathbf{z}_k(\epsilon) = \mathbf{z}_k - \dfrac{\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}(A'_\epsilon)}{\pi_k}\epsilon \\ \mathbf{z}_n(\epsilon) = \mathbf{z}_n \quad \forall n \neq j, k, \end{cases} \tag{11}$$

where

$$\|\mathbf{z}(\tilde{A}_\epsilon) - \mathbf{z}(A'_\epsilon)\| \leq \|\mathbf{z}(\tilde{A}_\epsilon)\| + \|\mathbf{z}(A'_\epsilon)\| \leq 2\max_{\omega\in\Omega}\|\mathbf{Z}(\omega)\|. \tag{12}$$

A careful inspection of (11) and (12) reveals the existence of $M > 0$ such that when ϵ is sufficiently small, we can construct a wage profile $\langle w_n(\epsilon)\rangle_{n=1}^N$ with $w_1(\epsilon) = 0$ that satisfies (i) $|w_n(\epsilon) - w_n| < M\epsilon$ for all n and, hence, the (LL) constraint, as well as (ii)

$$0 \leq \sum_{n=1}^N \pi_n u(w_n(\epsilon)) z_{a,n}(\epsilon) - \sum_{n=1}^N \pi_n u(w_n) z_{a,n} \sim \mathcal{O}(\epsilon) \quad \forall a \in \mathcal{D} \tag{13}$$

and, hence, all (IC_a) constraints.[14]

Write $\dot{w}_n = (w_n(\epsilon) - w_n)/\epsilon$ and $\dot{\mathbf{z}}_n(\epsilon) = (\mathbf{z}_n(\epsilon) - \mathbf{z}_n)/\epsilon$. Expanding (13) and multiplying the result by $\boldsymbol{\lambda}$, we obtain, when ϵ is small,

$$\sum_{n=1}^N \pi_n \cdot u'(w_n) \cdot \boldsymbol{\lambda}^\top \mathbf{z}_n \cdot \dot{w}_n(\epsilon) = -\sum_{n=1}^N u(w_n) \cdot \pi_n \cdot \boldsymbol{\lambda}^\top \dot{\mathbf{z}}_n(\epsilon) + \mathcal{O}(\epsilon).$$

Simplifying using $\dot{w}_1(\epsilon) = 0$, $u'(w_n) = 1/(\boldsymbol{\lambda}^\top \mathbf{z}_n)$ for $n \geq 2$ (Lemma 6), and (11) yields

$$\sum_{n=1}^N \pi_n \dot{w}_n(\epsilon) = s[u(w_k) - u(w_j)](\boldsymbol{\lambda}^\top \mathbf{z}'' - \boldsymbol{\lambda}^\top \mathbf{z}') + \mathcal{O}(\epsilon). \tag{14}$$

Perturbation (b). Repeating the above argument for perturbation (b) yields

$$\sum_{n=1}^N \pi_n \dot{w}_n(\epsilon) = -(1-s)[u(w_k) - u(w_j)](\boldsymbol{\lambda}^\top \mathbf{z}'' - \boldsymbol{\lambda}^\top \mathbf{z}') + \mathcal{O}(\epsilon). \tag{15}$$

Now, since $u(w_k) \neq u(w_j)$ by Lemma 6 and $\boldsymbol{\lambda}^\top \mathbf{z}'' \neq \boldsymbol{\lambda}^\top \mathbf{z}'$ by assumption, the right-hand sides of (14) and (15) must have the opposite signs when ϵ is small. The remainder of the proof is identical to that of Theorem 1 and is therefore omitted here. □

[14]To see why, define $\kappa_a = \sum_{n=2}^N \pi_n u(w_n) z_{a,n}$ and $\mathcal{S}_a = \{\langle x_n\rangle_{n=2}^N \in \mathbb{R}^{N-1} : \sum_{n=2}^N x_n z_{a,n} \geq \kappa_a\}$ for each $a \in \mathcal{D}$, and note that $\langle \pi_n u(w_n)\rangle_{n=2}^N \in \cap_{a\in\mathcal{D}} \mathcal{S}_a$. If, to the contrary, we cannot construct a wage profile as above, then there must exist $a', a'' \in \mathcal{D}$ such that $\cap_{a=a',a''}\{\langle x_n\rangle_{n=2}^N \in \mathbb{R}^{N-1} : \sum_{n=2}^N x_n z_{a,n} \geq \kappa_a\} = \{\langle x_n\rangle_{n=2}^N \in \mathbb{R}^{N-1} : \sum_{n=2}^N x_n z_{a',n} = \kappa_{a'}\}$ and, hence, $z_{a'',n} = -z_{a',n}$ for $n = 2, \ldots, N$ and $\kappa_{a''} = -\kappa_{a'}$. In the meantime, we have $\kappa_a \geq c(a^*) - c(a) > 0$ for all $a \in \mathcal{D}$, and so cannot have $\kappa_{a''} = -\kappa_{a'}$, a contradiction.

Proof of Theorem 6. Define

$$\Lambda = \{\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{D}|} \text{ and } \|\boldsymbol{\lambda}\|_{|\mathcal{D}|} = 1\},$$

where $\|\cdot\|_{|\mathcal{D}|}$ denotes the $|\mathcal{D}|$-dimensional Euclidean norm. By Theorem 5, any optimal monitoring technology with at most $N \in \{2, \dots, K\}$ performance categories is fully captured by some $\boldsymbol{\lambda} \in \Lambda$ and $N-1$ cut points $\widehat{z}_1, \dots, \widehat{z}_{N-1}$ such that $\min_{\omega\in\Omega} \boldsymbol{\lambda}^\top \mathbf{Z}(\omega) \le \widehat{z}_1 \le \cdots \le \widehat{z}_{N-1} \le \max_{\omega\in\Omega} \boldsymbol{\lambda}^\top \mathbf{Z}(\omega)$. Fix any $N \in \{2, \dots, K\}$, and write $\widehat{\mathbf{z}}$ for $(\widehat{z}_1, \dots, \widehat{z}_{N-1})$. For each $\boldsymbol{\lambda} \in \Lambda$, define

$$\mathcal{Z}_N(\boldsymbol{\lambda}) = \Big\{\widehat{\mathbf{z}} : \min_{\omega\in\Omega} \boldsymbol{\lambda}^\top \mathbf{Z}(\omega) \le \widehat{z}_1 \le \cdots \le \widehat{z}_{N-1} \le \max_{\omega\in\Omega} \boldsymbol{\lambda}^\top \mathbf{Z}(\omega)\Big\},$$

equip $\mathcal{Z}_N(\boldsymbol{\lambda})$ with the sup norm $\|\cdot\|$, and note that $\mathcal{Z}_N(\boldsymbol{\lambda})$ is compact by Assumption 3. For any given pair $(\boldsymbol{\lambda}, \mathbf{z})$, write the minimal incentive cost for inducing a^* as $W(\boldsymbol{\lambda}, \mathbf{z})$, and note that $W(\boldsymbol{\lambda}, \mathbf{z})$ is finite if and only if two conditions hold: $\lambda_a > 0$ for all $a \in \mathcal{D}$, and $\min_{\omega\in\Omega} \boldsymbol{\lambda}^\top \mathbf{Z}(\omega) < \widehat{z}_n < \max_{\omega\in\Omega} \boldsymbol{\lambda}^\top \mathbf{Z}(\omega)$ for some n. To see the necessity of the first condition, note that if it fails, then there exists $a \in \mathcal{D}$ such that $z_a(A) \equiv 0$ across all the performance category As generated by $(\boldsymbol{\lambda}, \mathbf{z})$. But then the ($\text{IC}_a$) constraint is violated, and hence $W(\boldsymbol{\lambda}, \mathbf{z})$ is infinite, a contradiction.

We proceed in two steps.

Step 1. Show that $W(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$ is continuous in $(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$ for any given $N \in \{2, \dots, K\}$. Fix any $\boldsymbol{\lambda} \in \Lambda$ and any $\widehat{\mathbf{z}} \in \mathcal{Z}_N(\boldsymbol{\lambda})$ such that $W(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$ is finite. W.l.o.g. consider the case where $\widehat{z}_n$s are all distinct. For a generic performance category $A_n = \{\omega : \boldsymbol{\lambda}^\top \mathbf{Z}(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ generated by $(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$, we use π_n to denote its probability given $a = a^*$, $\widehat{\mathbf{z}}$ to denote its ($|\mathcal{D}|$-vector) z-values, and w_n to denote the optimal wage at A_n. Then, for a sufficiently small $\delta > 0$, we take any δ-perturbations $\boldsymbol{\lambda}^\delta \in \Lambda$ and $\widehat{\mathbf{z}}^\delta \in \mathcal{Z}_N(\boldsymbol{\lambda}^\delta)$ to $\boldsymbol{\lambda}$ and $\widehat{\mathbf{z}}$, respectively, such that $\|\boldsymbol{\lambda}^\delta - \boldsymbol{\lambda}\|_{|\mathcal{D}|}$, $\|\widehat{\mathbf{z}}^\delta - \widehat{\mathbf{z}}\| < \delta$. For a generic performance category $A_n^\delta = \{\omega : \boldsymbol{\lambda}^{\delta\top} \mathbf{Z}(\omega) \in [\widehat{z}_{n-1}^\delta, \widehat{z}_n^\delta)\}$) generated by $(\boldsymbol{\lambda}^\delta, \widehat{\mathbf{z}}^\delta)$, we use π_n^δ to denote its probability given $a = a^*$, and $\mathbf{z}_n^\delta$ to denote its ($|\mathcal{D}|$-vector) z-values.

Now, fix any $\epsilon > 0$, and consider the wage profile that pays, at A_n^δ, $w_n + \epsilon$ if $z_{a,n}^\delta > 0$ for all $a \in \mathcal{D}$ and w_n otherwise. This wage profile satisfies the (LL) constraint by construction. Under Assumptions 2 and 3, it satisfies every (IC_a) constraint when δ is small, because

$$\begin{aligned}
&\lim_{\delta\to 0} \sum_n u\Big(w_n + \prod_{a'\in\mathcal{D}} 1_{z_{a',n}^\delta > 0} \cdot \epsilon\Big) \pi_n^\delta z_{a,n}^\delta \\
&= \sum_n u\Big(w_n + \prod_{a'\in\mathcal{D}} 1_{z_{a',n} > 0} \cdot \epsilon\Big) \pi_n z_{a,n} \\
&> \sum_n u(w_n) \pi_n z_{a,n},
\end{aligned}$$

where the last line uses the existence of n such that $\prod_{a'\in\mathcal{D}} 1_{z_{a',n}>0} = 1$, obtained from combining $\sum_n \pi_n z_{a',n} = 0$ and $z_{a',n}$ being strictly increasing in n for all $a' \in \mathcal{D}$. To complete the proof, note that

$$\lim_{\delta\to 0}\sum_n \pi_n^\delta\Big(w_n + \prod_{a\in\mathcal{D}} 1_{z_{a,n}^\delta>0}\cdot\epsilon\Big) = \sum_n \pi_n\Big(w_n + \prod_{a\in\mathcal{D}} 1_{z_{a,n}^\delta>0}\cdot\epsilon\Big),$$

and so

$$W\big(\boldsymbol{\lambda}^\delta,\widehat{\mathbf{z}}^\delta\big) - W(\boldsymbol{\lambda},\widehat{\mathbf{z}}) \le \sum_n \pi_n^\delta\Big(w_n + \prod_{a\in\mathcal{D}} 1_{z_{a,n}^\delta>0}\cdot\epsilon\Big) - \sum_n \pi_n w_n < \epsilon$$

holds when δ is sufficiently small. Finally, interchanging the roles between $(\boldsymbol{\lambda},\mathbf{z})$ and $(\boldsymbol{\lambda}^\delta,\mathbf{z}^\delta)$ in the above argument yields $W(\boldsymbol{\lambda},\widehat{\mathbf{z}}) - W(\boldsymbol{\lambda}^\delta,\widehat{\mathbf{z}}^\delta) < \epsilon$, thus proving that $|W(\boldsymbol{\lambda}^\delta,\widehat{\mathbf{z}}^\delta) - W(\boldsymbol{\lambda},\widehat{\mathbf{z}})| < \epsilon$ when δ is sufficiently small.

Step 2. Under Assumption 4(a), the existence and finiteness of

$$W_N := \min_{\boldsymbol{\lambda}\in\Lambda,\widehat{\mathbf{z}}\in\mathcal{Z}_N(\boldsymbol{\lambda})} W(\boldsymbol{\lambda},\widehat{\mathbf{z}})$$

for any $N \in \{2,\dots,K\}$ follows from Step 1 and the compactness of Λ and $\mathcal{Z}_N(\boldsymbol{\lambda})$. Under Assumption 4(b), we can write the principal's problem as

$$\min_{\boldsymbol{\lambda}\in\Lambda,\widehat{\mathbf{z}}\in\mathcal{Z}_K(\boldsymbol{\lambda})} W(\boldsymbol{\lambda},\widehat{\mathbf{z}}) + \mu\cdot h\big(\boldsymbol{\pi}(\boldsymbol{\lambda},\widehat{\mathbf{z}})\big),$$

where $\boldsymbol{\pi}(\boldsymbol{\lambda},\widehat{\mathbf{z}})$ compiles the probabilities of the output signal generated by $(\boldsymbol{\lambda},\widehat{\mathbf{z}})$, and it is clearly continuous in its arguments. The remainder of the proof is the identical to that of Theorem 2 and is therefore omitted here. □

Appendix B: Other extensions

B.1 *Individual rationality*

In this appendix, let everything be as in the baseline model, except that the agent faces an individual rationality constraint (rather than a limited liability constraint):

$$\sum_{A\in\mathcal{P}} P_1(A)u\big(w(A)\big) \ge c + \underline{u}. \tag{IR}$$

Under this alternative assumption, an optimal incentive contract that induces high effort from the agent (optimal incentive contract for short) minimizes the total implementation cost, subject to the agent's (IC) and (IR) constraints.

Corollary 4. *Under Assumption 1, any optimal monitoring technology comprises Z-convex cells.*

Proof. Take any optimal incentive contract with $\langle A_n, \pi_n, z_n, w_n\rangle_{n=1}^N$ be its corresponding tuple. Assume w.l.o.g. that $z_1 \le \cdots \le z_N$.

Step 1. Show that $z_1 < \cdots < z_N$ and $w_1 < \cdots < w_N$. The wage-minimization problem given $\langle A_n, \pi_n, z_n\rangle_{n=1}^N$ is

$$\min_{\langle \tilde{w}_n\rangle_{n=1}^N} \sum_{n=1}^N \pi_n \tilde{w}_n - \lambda\left[\sum_{n=1}^N \pi_n u(\tilde{w}_n) z_n - c\right] - \gamma\left[\sum_{n=1}^N \pi_n u(\tilde{w}_n) - (c + \underline{u})\right],$$

where λ and γ denote the Lagrange multipliers associated with the (IC) constraint and the (IR) constraint, respectively. Standard arguments show that both λ and γ are strictly positive. Given this, we can then differentiate the objective function with $\tilde{w}_n$ and obtain

$$u'(w_n) = \frac{1}{\lambda z_n + \gamma}.$$

Thus if, to the contrary, $z_j = z_k$ for some $j \neq k$, then we must have $w_j = w_k$. As a result, the principal can strictly benefit from merging A_j and A_k (which reduces the monitoring cost without affecting the incentive cost), which contradicts the optimality of the current contract.

Step 2. Show Z-convexity. Suppose, to the contrary, that some A_j is not Z-convex. Consider first perturbation (a) in the proof of Theorem 1. Take any wage profile $\langle w_n(\epsilon)\rangle_{n=1}^N$ that makes both the (IC) and (IR) constraints binding after this perturbation, i.e.,

$$\sum_{n=1}^N \pi_n u\big(w_n(\epsilon)\big) z_n(\epsilon) = \sum_{n=1}^N \pi_n u(w_n) z_n \tag{16}$$

and

$$\sum_{n=1}^N \pi_n u\big(w_n(\epsilon)\big) = \sum_{n=1}^N \pi_n u(w_n). \tag{17}$$

A careful inspection of (2), (16), and (17) reveals the existence of $M > 0$ such that when ϵ is sufficiently small, we can construct a wage profile as above that satisfies, in addition, $|w_n(\epsilon) - w_n| < M\epsilon$ for any n.[15]

Write $\dot{w}_n(\epsilon) = (w_n(\epsilon) - w_n)/\epsilon$ and $\dot{z}_n(\epsilon) = (z_n(\epsilon) - z_n)/\epsilon$. Expanding λ (16) $+ \gamma$ (17) yields, when ϵ is small,

$$\sum_{n=1}^N \pi_n \cdot u'(w_n) \cdot (\lambda z_n + \gamma) \cdot \dot{w}_n(\epsilon) = -\lambda \sum_{n=1}^N u(w_n) \cdot \pi_n \dot{z}_n(\epsilon) + \mathcal{O}(\epsilon),$$

and simplifying using $u'(w_n) = 1/(\lambda z_n + \gamma)$ and (2) yields

$$\sum_{n=1}^N \pi_n \dot{w}_n(\epsilon) = s\big[u(w_k) - u(w_j)\big]\big(\lambda z'' - \lambda z'\big). \tag{18}$$

[15]To see why, define $\kappa_1 = \sum_{n=1}^N \pi_n u(w_n) z_n$, $\kappa_2 = \sum_{n=1}^N \pi_n u(w_n)$, $\mathcal{S}_1 = \{\langle x_n\rangle_{n=1}^N \in \mathbb{R}^N : \sum_{n=1}^N x_n z_n \geq \kappa_1\}$, and $\mathcal{S}_2 = \{\langle x_n\rangle_{n=1}^N \in \mathbb{R}^N : \sum_{n=1}^N x_n \geq \kappa_2\}$, and note that $\langle \pi_n u(w_n)\rangle_{n=1}^N \in \mathcal{S}_1 \cap \mathcal{S}_2$. Then from $z_1 < \cdots < z_N$ (shown in Step 1), it follows that $\dim \mathcal{S}_1 \cap \mathcal{S}_2 = N$, which coupled with (2) gives the desired result.

Next consider perturbation (b) in the proof of Theorem 1. Similar algebraic manipulations as above show that

$$\sum_{n=1}^{N} \pi_n \dot{w}_n(\epsilon) = -(1-s)\big[u(w_k) - u(w_j)\big]\big(\lambda z'' - \lambda z'\big). \tag{19}$$

Then from $u(w_j) \neq u(w_k)$ and $z'' \neq z'$, it follows that sgn (18) $\neq$ sgn (19). The remainder of the proof is the same as that of Theorem 1 and is therefore omitted here. □

B.2 *Random monitoring technology*

In this appendix, a monitoring technology $\mathbf{q} : \Omega \to \Delta^K$ maps each performance data point to a lottery on K performance ratings, and a wage scheme $w : \{1, \cdots, K\} \to \mathbb{R}_+$ specifies a wage payment for each performance rating. Time evolves as follows:

Stage 1. The principal commits to $\langle \mathbf{q}, w\rangle$.

Stage 2. The agent privately chooses $a \in \{0, 1\}$.

Stage 3. Nature draws $\omega \in \Omega$ according to P_a.

Stage 4. The monitoring technology outputs $n \in \{1, \ldots, K\}$ with probability $q_n(\omega)$.

Stage 5. The principal pays w_n to the agent.

Under any given contract $\langle \mathbf{q}, w\rangle$, the agent receives performance rating n with probability

$$\pi_n = \int q_n(\omega)\, dP_1(\omega)$$

if he exerts high effort. Define $\mathcal{N} = \{n : \pi_n > 0\}$ as the set of performance ratings that is realized with a strictly positive probability given high effort. For each $n \in \mathcal{N}$, define

$$z_n = \int Z(\omega) q_n(\omega)\, dP_1(\omega)/\pi_n$$

as the z-value of performance rating n. For each $n \notin \mathcal{N}$, define $w_n = 0$. Then $\langle \mathbf{q}, w\rangle$ is incentive compatible if

$$\sum_{n\in\mathcal{N}} \pi_n u(w_n) z_n \geq c, \tag{IC}$$

in which case the monitoring cost it incurs is proportional to the mutual information of the agent's performance data and output signal given high effort:

$$H(\mathbf{q}, 1) = \sum_{n\in\mathcal{N}} \int q_n(\omega) \log \frac{q_n(\omega)}{\int q_n(\omega)\, dP_1(\omega)}\, dP_1(\omega).$$

An optimal incentive contract $\langle \mathbf{q}^*, w^* \rangle$ that induces high effort from the agent (optimal incentive contract for short) solves

$$\min_{\langle \mathbf{q}, w \rangle} \sum_{n=1}^{K} \pi_n w_n + \mu \cdot H(\mathbf{q}, 1) \quad \text{subject to (IC) and (LL).}$$

The next theorem gives characterizations of optimal incentive contracts.

Theorem 7. *For any optimal incentive contract $\langle \mathbf{q}^*, w^* \rangle$ that induces high effort from the agent, we have (i) $\mathbf{q}^* : Z(\Omega) \to \Delta^K$, (ii) $\min\{w_n^* : n \in \mathcal{N}^*\} = 0$, and (iii) for any $j, k \in \mathcal{N}^*$, $w_j^* \neq w_k^*$, and $q_k^*(z)/q_j^*(z)$ is strictly increasing in z if and only if $w_j^* < w_k^*$.*

Proof. Since the incentive cost is linear in $\mathbf{q}(\omega)$ whereas the monitoring cost is convex in $\mathbf{q}(\omega)$, it follows that $\mathbf{q}^* : Z(\Omega) \to \Delta^K$ and that $w_j^* \neq w_k^*$ for any $j, k \in \mathcal{N}^*$. Write $\mathcal{N}^* = \{1, \dots, N\}$, and assume w.l.o.g. that $w_1^* < \cdots < w_N^*$. Then $w_1^* = 0$ for the same reason as in the proof of Lemma 2. Differentiating the principal's objective function with respect to $\mathbf{q}(z)$ and setting the result equal to 0 yields

$$-w_n^* + \lambda u(w_n^*) z = \mu \left[\log \frac{q_n^*(z)}{q_1^*(z)} - \log \frac{\pi_n^*}{\pi_1^*} \right] \quad \forall n = 2, \dots, N, \tag{20}$$

where λ denotes the Lagrange multiplier associated with the (IC) constraint. Part (iii) of this theorem then follows from $\lambda > 0$ and, hence, the left-hand side of (20) being strictly increasing in z. □

The next theorem proves the existence of an optimal incentive contract.

Theorem 8. *Assume Assumptions 2 and 3. Then an optimal incentive contract that induces high effort from the agent exists.*

Proof. For any given $\mathbf{q}$, the wage-minimization problem admits solution(s) if and only if $z_j \neq z_k$ for some $j, k \in \mathcal{N}$, in which case we denote the minimal incentive cost by $W(\mathbf{q})$. If z_ns are constant across all $n \in \mathcal{N}$, then set $W(\mathbf{q}) = +\infty$. The principal's problem is then

$$\min_{\mathbf{q}} W(\mathbf{q}) + \mu \cdot H(\mathbf{q}, 1). \tag{21}$$

By (20) and Assumptions 2 and 3, any solution to (21) must be continuously differentiable on $Z(\Omega)$ (taking the usual care of derivatives at end points). Write $C^1(Z(\Omega), \Delta^K)$ as the set of mappings from $Z(\Omega)$ to Δ^K that is continuously differentiable in its argument, and equip $C^1(Z(\Omega), \Delta^K)$ with the sup norm $\|\cdot\|$, i.e., $\|\mathbf{q}' - \mathbf{q}\| = \sup_{z,n} |q_n'(z) - q_n(z)|$ for any $\mathbf{q}, \mathbf{q}' \in C^1(Z(\Omega), \Delta^K)$. Rewrite the principal's problem as

$$\min_{\mathbf{q} \in C^1(Z(\Omega), \Delta^K)} W(\mathbf{q}) + \mu \cdot H(\mathbf{q}, 1), \tag{22}$$

and note that the objective function is continuous in $\mathbf{q}$.

To prove that (22) admits solution(s), note that

$$\inf_{\mathbf{q}\in C^1(Z(\Omega),\Delta^K)} W(\mathbf{q}) + \mu \cdot H(\mathbf{q}, 1)$$

is a finite number, hereafter denoted by x. Let $\{\mathbf{q}^k\}$ be any sequence in $C^1(Z(\Omega), \Delta^K)$ such that $\lim_{k\to\infty} W(\mathbf{q}^k) + \mu \cdot H(\mathbf{q}^k, 1) = x$. Clearly, $\mathbf{q}^k$ is uniformly bounded across ks, and the family $\{\mathbf{q}^k\}$ is equicontinuous by Assumption 3 and the definition of $C^1(Z(\Omega), \Delta^K)$. Thus, a subsequence of $\{\mathbf{q}^k\}$ converges uniformly to some $\mathbf{q}^\infty$ by Helly's selection theorem, and $W(\mathbf{q}^\infty) + \mu \cdot H(\mathbf{q}^\infty, 1) = x$ by the continuity of the objective function. □

References

Alchian, Armen A. and Harold Demsetz (1972), "Production, information costs, and economic organization." *American Economic Review*, 62, 777–795. [1137, 1152]

Baiman, Stanley and Joel S. Demski (1980), "Economically optimal performance evaluation and control systems." *Journal of Accounting Research*, 18, 184–220. [1137]

Bergemann, Dirk and Stephen Morris (2019), "Information design: A unifying perspective." *Journal of Economic Literature*, 57, 44–95. [1155]

Blackwell, David (1953), "Equivalent comparisons of experiments." *Annals of Mathematical Statistics*, 24, 265–272. [1137]

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2013), "Does management matter? Evidence from India." *Quarterly Journal of Economics*, 128, 1–51. [1155]

Bloom, Nicholas and John Van Reenen (2006), "Measuring and explaining management practices across firms and countries." Centre for Economic Performance Discussion Paper 716. [1136, 1137, 1147, 1152]

Bloom, Nicholas and John Van Reenen (2010), "Why do management practices differ across countries?" *Journal of Economic Perspectives*, 24, 203–224. [1147]

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen (2012), "Americans do IT better: U.S. multinationals and the productivity miracle." *American Economic Review*, 102, 167–201. [1147]

Cover, Thomas M. and Joy A. Thomas (2006), *Elements of Information Theory*, second edition. Wiley-Interscience, Hoboken, New Jersey. [1141]

Crémer, Jacques, Luis Garicano, and Andrea Prat (2007), "Language and the theory of the firm." *Quarterly Journal of Economics*, 122, 373–407. [1138]

Dilmé, Francesc (2018), "Optimal languages." Unpublished paper, SSRN 3331077. [1138]

Dye, Ronald A. (1986), "Optimal monitoring policies in agencies." *Rand Journal of Economics*, 17, 339–350. [1137]

Ewenstein, Boris, Bryan Hancock, and Asmus Komm (2016), "Ahead of the curve: The future of performance management." McKinsey Quarterly. [1136, 1147]

Green, Jerry R. and Nancy L. Stokey (1983), "A comparison of tournaments and contracts." *Journal of Political Economy*, 91, 349–364. [1152]

Grossman, Sanford J. and Oliver D. Hart (1983), "An analysis of the principal–agent problem." *Econometrica*, 51, 7–45. [1137]

Hölmstrom, Bengt (1979), "Moral hazard and observability." *Bell Journal of Economics*, 10, 74–91. [1136, 1137, 1141, 1142, 1144]

Hölmstrom, Bengt (1982), "Moral hazard in teams." *Bell Journal of Economics*, 13, 324–340. [1137, 1151, 1152]

Hölmstrom, Bengt and Paul Milgrom (1991), "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, and Organization*, 7, 24–52. [1137, 1154, 1155]

Hook, Caroline, Andrew Jenkins, and Margaret Foot (2011), *Introducing Human Resource Management*, sixth edition. Prentice Hall. [1147]

Jäger, Gerhard, Lars P. Metzger, and Frank Riedel (2011), "Voronoi languages: Equilibria in cheap-talk games with high-dimensional signals and few signals." *Games and Economic Behavior*, 73, 517–537. [1138]

Kaplan, Esther (2015), "The spy who fired me: The human costs of workplace monitoring." Harper's Magazine. [1135]

Kim, Son Ku (1995), "Efficiency of an information system in an agency model." *Econometrica*, 63, 89–102. [1137]

Lazear, Edward P. and Sherwin Rosen (1981), "Rank-order tournaments as optimum labor contracts." *Journal of Political Economy*, 89, 841–864. [1137, 1152]

Maćkowiak, Bartosz and Mirko Wiederholt (2009), "Optimal sticky prices under rational inattention." *American Economic Review*, 99, 769–803. [1137]

Martin, Daniel (2017), "Strategic pricing with rational inattention to quality." *Games and Economic Behavior*, 104, 131–145. [1137]

Matějka, Filip and Alisdair McKay (2012), "Simple market equilibria with rationally inattentive consumers." *American Economic Review: Papers and Proceedings*, 102, 24–29. [1137]

Mookherjee, Dilip (1984), "Optimal incentive schemes with many agents." *Review of Economic Studies*, 51, 433–446. [1152]

Murff, Harvey J., Fern FitzHenry, Michael E. Matheny, Nancy Gentry, Kristen L. Kotter, Kimberly Crimin, Robert S. Dittus, Amy K. Rosen, Peter L. Elkin, Steven H. Brown, and Theodore Speroff (2011), "Automated identification of postoperative complications within an electronic medical record using natural language processing." *Journal of American Medical Association*, 306, 848–855. [1135, 1136]

Ravid, Doron (2017), "Bargaining with rational inattention." Unpublished paper, SSRN 2957890. [1137]

Saint-Paul, Gilles (2017), "A "quantized" approach to rational inattention." *European Economic Review*, 100, 50–71. [1138]

Sims, Christopher A. (1998), "Stickiness." *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356. [1137]

Sims, Christopher A. (2003), "Implications of rational inattention." *Journal of Monetary Economics*, 50, 665–690. [1137]

Singer, Natasha (2013), "In a mood? Call center agents can tell." New York Times, 12. [1135, 1136]

Sobel, Joel (2015), "Broad terms and organizational codes." Unpublished paper, Department of Economics, University of California, San Diego. [1138]

Woodford, Michael D. (2009), "Information-constrained state-dependent pricing." *Journal of Monetary Economics*, 56, S100–S124. [1137]

Yang, Ming (2020), "Optimality of debt under flexible information acquisition." *Review of Economic Studies*, 87, 487–536. [1137]

Co-editor Thomas Mariotti handled this manuscript.

Manuscript received 2 January, 2018; final version accepted 12 November, 2019; available online 15 November, 2019.