Optimal Incentive Contract with Endogenous Monitoring Technology

Anqi Li[*] Ming Yang[†]

This Version: September 2017

Abstract

Recent technology advances give firms more flexibility to utilize employee performance data at a reduced and yet significant cost. This paper develops a theory of optimal incentive contracting where the monitoring technology that governs the above described procedure is part of the contract designer's strategic planning. In otherwise standard principal-agent models with moral hazard, we assume that the principal can partition the agent's raw performance data into any finite categories and pays for the quantity of information that the output signal carries. Through analysis of the trade-off between providing the agent with incentives and saving the cost of data utilization, we obtain characterizations of the optimal monitoring technology such as information aggregation, strict MLRP, likelihood ratio-convex classification, group monitoring as a best response to high monitoring cost, and matching the intensities of monitoring different tasks to the agent's endogenous tendencies to shirk. We examine the implications of these results for workforce management and firms' internal organizations.

Key words: incentive contract; endogenous monitoring technology.

JEL codes: D86, M15, M5.

[*]Department of Economics, Washington University in St. Louis. anqili@wustl.edu.

[†]Fuqua School of Business, Duke University. ming.yang@duke.edu. We thank Nick Bloom, George Mailath, Ilya Segal, Chris Shannon, Joel Sobel, Jacob Steinhardt, Bob Wilson and seminar audience at Caltech, Duke-UNC, Johns Hopkins Carey, Northwestern, Decentralization Conference 2017, Stanford, UCSB, UCSD and U of Washington for helpful comments and suggestions. Jessie Li and Lin Hu provided generous assistance for the numerical analysis. All errors are our own.

1 Introduction

Technology advances have improved workforce monitoring through the utilization of sophisticated data. Recently, speech analytics software has enabled extraction of key performance indicators (e.g., voice quality and tone, customer sentiment) from the conversations between call center agents and customers (Singer (2013)); cloud-based systems are increasingly used to convert extensive employee tracking records (e.g., keyboard stroke, media interaction) into succinct ratings such as "satisfactory" and "unsatisfactory" (Kaplan (2015)); and natural language processing tools have proven effective at assessing physician care quality through analysis of the narrative reports in electronic medical records (Murff et al. (2011)).

This paper develops a theory of optimal incentive contracting where the *monitoring technology* — which, throughout this paper, refers to the human- or machine-operated system that governs the processing and analysis of employee performance data — is part of the contract designer's strategic planning. Our research agenda is motivated by two considerations. First, improvements in technologies facilitate *design flexibility*, as managers can now be very specific about which key performance indicators they want to factor into pay decisions, which aspect of the data should be given special attention to, and whether the goal is to obtain individual- or group-level information, etc. (Bloom and Van Reenen (2006, 2007), Singer (2013), Kaplan (2015))).[1] Second, the *cost* associated with data utilization has remained significant despite rapid growth in related areas,[2] which suggests that the choice of the monitoring technology be made based on meticulous cost-benefit analyses.

The current paper formalizes the flexibility and the cost that recent technology advances bring to the design and implementation of monitoring technologies in otherwise standard principal-agent models with moral hazard. The main results are characterizations of the optimal monitoring technology through the trade-off between providing the agent with incentives and saving the cost of data utilization. We ex-

[1] These questions come up frequently when companies seek customized solutions from developers such as Cornerstone OnDemand, Hubstaff and Kronos. See these companies' websites for detailed accounts.

[2] Kaplan (2015) estimates that the industry on workforce performance analytics is worth 11 billion dollars. Recent articles by *CB insights*, *Harvard Business Review* and *Towers Watson* rank workforce analytics as one of the top areas of HR spending whose lucrative prospect has spurred the growth and M&A of HR startups in the Bay area and across the globe ("Empowering the Organization," 2014, "The Data-ification of HR," 2015, "The Big Data Opportunity for HR and Finance," 2016).

amine the implications of these results for workforce management and firms' internal organizations.

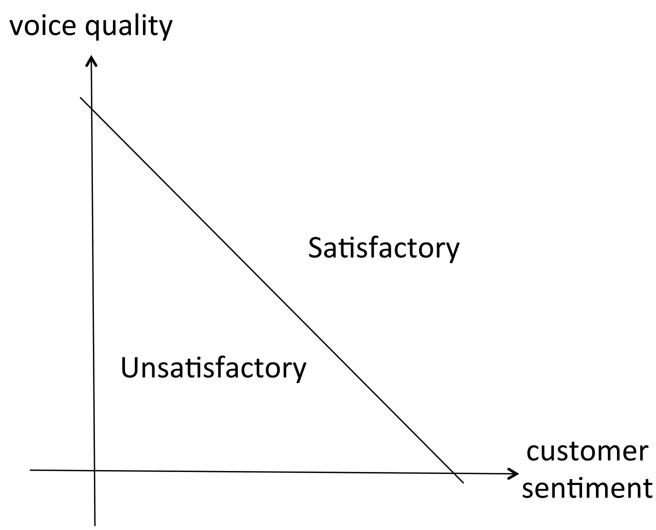

Figure 1: Call center performance management.

To illustrate, consider the example of call center performance management (Singer (2013)), where a conversation between an agent and a customer contains numerous performance indicators such as customer sentiment, voice quality and tone, etc.. In a hypothetical world where data processing and analysis is costless, factoring detailed information along these indicators into the pay decision provides the agent with the strongest incentive to work. But in reality, this solution is difficult to implement, as it requires that call centers pay for the use of sophisticated speech analytics software that consumes significant server space and power.[3] For the sake of cost efficiency, most speech analytics software developed by major vendors classifies the agent's performance into only a few categories such as "satisfactory" and "unsatisfactory." Meanwhile, improved technologies enable call centers to choose between a range of software products with varying emphasis on the above described performance indicators.[4] In the example depicted in Figure 1, this means that the principal can freely adjust the boundary between "satisfactory" and "unsatisfactory," which, by information theory, affects the cost of data processing and analysis only through

[3]With the advent of cloud computing, more companies are paying for software usage rather than buying the software itself.

[4]For example, Beyond Verbal specializes in emotion detection and CallMiner in word spotting.

the probabilities that the agent is rated as such. The optimal boundary balances the trade-off between providing the agent with incentives and saving the cost of data processing and analysis. The current paper is devoted to understanding this trade-off.

Our general framework builds upon otherwise standard agency models with moral hazard, where the raw data on agents' hidden efforts can be high-dimensional objects. An incentive contract consists of a monitoring technology that compresses the raw data into a finite-valued signal, as well as a wage scheme that maps different realizations of this signal to different wages. Motivated by reality, we allow the monitoring technology to be any finite partition of the raw data set, and assume that the cost associated with running the monitoring technology increases with the quantity of information carried by the output signal. An optimal incentive contract that induces any given level of effort minimizes the sum of expected wage and monitoring cost (hereafter, *implementation cost*), subject to agents' incentive constraints.

Our main result shows that the optimal monitoring technology comprises convex cells in the space of likelihood ratios or their transformations. The result that it suffices to consider likelihood ratios or their transformations follows from Holmstrom's (1979) *sufficient statistic principle*, as well as the general property of the monitoring cost function. The result on convexity is more difficult to establish, because perturbing the monitoring technology affects the expected wage *indirectly* through agents' incentive constraints. Our proof strategy works directly with the principal's Lagrangian. By exploiting the general property of the monitoring cost function, the envelope theorem and the first-order condition studied by Holmstrom (1979), we show that the optimal assignment of data points to performance categories is "positive assortative," where the degree of assortativeness is given by the inner product between the likelihood ratio and agents' utilities. We then argue that if a performance category contains data points of extreme but not intermediate likelihood ratios, then the assignment of data points goes wrong against the direction of agents' utilities and an improvement can be constructed.

We give three applications of this result. We first revisit the single-agent model of Holmstrom (1979), where the optimal monitoring technology divides likelihood ratios of the raw performance data into non-overlapping intervals. The resulting performance classifier compresses fine-grained and potentially high-dimensional data into coarse and rank-ordered ratings such as "satisfactory" and "unsatisfactory" (hereafter, *information aggregation*), and the output signal satisfies the *strict monotone* *likelihood ratio property* with respect to the order induced by likelihood ratios. Solving for the intervals of likelihood ratios enables comparative statics analysis of how the optimal incentive contract depends on factors that affect the monitoring cost (e.g., advent of IT-based human resource management systems, advances in natural language processing and voice analysis, improvements in software and computing power, etc.). Results include (i) robust predictions that conform with recent developments in manufacturing, retail and healthcare sectors (Bloom and Van Reenen (2006, 2007); Ewenstein et al. (2016); Murff et al. (2011)), e.g., rating schemes become more fine-grained as data processing and analysis becomes cheaper; and (ii) subtle predictions such as wage variance can be non-monotone in the monitoring cost because of the endogenous interplay between the monitoring technology and the wage scheme.

We next turn to the multi-agent model of Holmstrom (1982), where the optimal monitoring technology partitions vectors of individual agents' likelihood ratios into convex polygons. This theorem enables comparison of individual and group monitoring from the angle of monitoring cost, which in turn predicts that group monitoring is more effective at saving the monitoring cost than individual monitoring is regardless of whether agents are technologically independent or not. This result formalizes the intuitions of Alchian and Demsetz (1972) and Lazear and Rosen (1981) that either team or tournament should be the dominant incentive system when individual performance metrics are too costly to obtain. It is consistent with Bloom and Van Reenen (2006, 2007)'s finding that lack of IT access leads manufacturing companies to choose group monitoring over individual monitoring, and it lends support to the increasing adoption of tech-enabled individual performance management in call centers and retail stores (Singer (2013), Kaplan (2015)).

When the agent can take multiple deviant actions, our analysis suggests that we match the resources spent on the detection of each deviation to the Lagrange multiplier of the corresponding incentive compatibility constraint. Applying this result to the multi-tasking model of Holmstrom and Milgrom (1991) leads to an important lesson, that of matching the intensities of monitoring different tasks to the agent's endogenous tendencies to shirk. Using simulation, we examine the implication of this result for, e.g., how tenure and promotion committees should trade off reading research papers and interpreting teaching evaluations in order to induce faculty members to work hard on both research and teaching.

1.1 Related Literature

Contracting with costly signals Existing studies on contracting with costly signal acquisition limit the kinds of signals that can be used to monitor the agent's performance. For example, in the costly verification model developed by Banker and Datar (1980) and Dye (1986), the principal is restricted to drawing a signal from an exogenous probability distribution. In the linear-contracting model that serves as the workhorse of applied works, the usual assumption is that the principal pays for the reduction of the variance of a Gaussian signal. By contrast, we give the principal considerable freedom to utilize the potentially high-dimensional performance data, and let her pay the cost associated with data utilization. Our analysis jointly predicts information aggregation, strict MLRP, likelihood-ratio convex classification, group monitoring as a best response to high monitoring cost, and matching the intensities of monitoring different tasks to the agent's endogenous tendencies to shirk. The first three results distinguish our theory from earlier works, whereas the last two results are, to our knowledge, new to the theoretical principal-agent literature.

Ranking signals in agency models A number of early works tries to rank signals based on their implications for principal-agent relationships. For example, Holmstrom (1979) shows that adding another signal does not change the wage scheme if and only if the original signal is a sufficient static for the agent's effort. Grossman and Hart (1983) and Kim (1995) demonstrate that performing mean-preserving spreads to the likelihood ratios helps save the incentive cost. The current paper focuses on the design of the monitoring technology. In the spirit of the burgeoning literature on information design (see, e.g., Bergemann and Morris (2016) and the references therein), we allow the principal to optimize over a rich set of signal structures subject to agents' incentive constraints, with the aim of characterizing the optimal signal structure through the tension between the incentive cost and the monitoring cost.

Rational inattention Both the current work and the literature on rational inattention (hereafter, RI) allow decision makers to acquire the most payoff-relevant information at non-trivial costs. The difference is threefold. First, initial developments in RI attribute the stickiness of macroeconomic variables to the cost of information processing (see Sims (1998) and Sims (2003) among others), whereas the current work explores the consequence of flexible information processing in strategic environments.[5] Second, we focus mainly on partitional information structures because in reality, giving random evaluations based on non-performance factors is viewed as bias and is warned against by human resource managers based on concerns such as appeals and turnover (see Bracken et al. (2001), Roberts and Pregitzer (2007) and other standard textbooks on HR management).[6] Finally, we allow for general monitoring cost functions that nest entropy as a special case.

Institution design with limited communication The current paper adds to the literature on institution design with limited communication. Recent developments include Crémer et al. (2007), Jäger et al. (2011), Sobel (2015) and Dilmé (2017), which characterize the optimal language in organizations where players share common interest but communication is costly; and Blumrosen and Feldman (2006) and Blumrosen et al. (2007), which examine mechanism design problems where agents are limited to communicating only a few bits of information. All these studies find it optimal to describe similar state contingencies with the same message, but for different reasons than that behind likelihood-ratio convexity. Specifically, the presence of incentive constraints makes our problem more difficult to analyze than those considered by Crémer et al. (2007), Jäger et al. (2011), Sobel (2015) and Dilmé (2017), whereas the result of Blumrosen and Feldman (2006) and Blumrosen et al. (2007) follows from the fact that the mechanism is dominant strategy incentive compatible and agents' utilities exhibit strict increasing differences.

2 Baseline Model

2.1 Setup

Players A risk-averse agent earns a payoff $u(w)-c(a)$ from spending a non-negative wage $w \geq 0$ and exerting either high effort ($a = 1$) or low effort ($a = 0$), where $u(0) = 0$, $u' > 0$, $u'' < 0$ and $c(1) = c > c(0) = 0$. Each level a of effort generates a probability space (Ω, Σ, P_a), where Ω consists of the agent's raw performance data

[5] Yang (2015) examines a security design problem where a rationally inattentive buyer can flexibly acquire information about the uncertain fundamental. Other recent efforts to introduce RI into the study of microeconomic problems include, but are not limited to Matéjka and McKay (2012) and Martin (2017).

[6] Partitional information structure has been long used by economists to model knowledge and information and is recently considered by Saint-Paul (2011) in the context of RI decision problems.

(e.g., call history), and P_a is the probability measure on (Ω, Σ) conditional on the effort being of a. A risk-neutral principal cannot observe effort, and her goal is to induce high effort. This last assumption is inconsequential, as the problem of inducing low effort is easy to solve.

Incentive contract An incentive contract $\langle \mathcal{P}, w(\cdot)\rangle$ consists of a *monitoring technology* $\mathcal{P}$ and a *wage scheme* $w(\cdot)$. The former represents a human- or machine-operated system that governs the processing and analysis of the raw performance data (e.g., speech analytics tools), whereas the latter maps outputs of the first-step procedure to different levels of wages. Throughout the analysis, we allow $\mathcal{P}$ to be *any* finite partition of Ω whose cells (hereafter, *performance categories*) belong to Σ, and use $w : \mathcal{P} \to \mathbb{R}_+$ to map each cell A of $\mathcal{P}$ to a non-negative wage $w(A) \geq 0$.[7] For each raw data point $\omega \in \Omega$, denote by $A(\omega)$ the unique performance category that contains ω and by $w(A(\omega))$ the wage associated with $A(\omega)$. Time evolves as follows:

1. The principal commits to $\langle \mathcal{P}, w(\cdot)\rangle$;
2. The agent privately chooses $a \in \{0, 1\}$;
3. Nature draws ω from Ω according to P_a;
4. The monitoring technology outputs $A(\omega)$;
5. The principal pays $w(A(\omega))$ to the agent.

Output signal A monitoring technology $\mathcal{P} = \{A_1, \cdots, A_N\}$ outputs a signal $X : \Omega \to \mathcal{P}$ whose probability mass function is defined by $P_X(X = A_n \mid a) = P_a(A_n)$, $n = 1, \cdots, N$. Collecting these probabilities gives rise to a probability vector $\boldsymbol{\pi}(\mathcal{P}, a) = (P_a(A_1), \cdots, P_a(A_N))$. X is taken as exogenously given by the conventional principal-agent literature; here it is determined by the principal's endogenous choice of the monitoring technology.

[7] In general, a monitoring technology can be any mapping from the raw data set to lotteries over finite performance categories. For reasons discussed in Section 1.1, we focus on degenerate lotteries, or equivalently partitional monitoring technologies, in the main body of this paper, and defer the analysis of random monitoring technologies to the online appendix.

Implementation cost For any given level a of effort, *which could be either high or low*, the total cost of implementing an incentive contract is equal to

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

This cost has two parts. The first part $\sum_{A \in \mathcal{P}} P_a(A) w(A)$, the *incentive cost*, has been the central focus of the existing principal-agent literature. The second part $\mu \cdot H(\mathcal{P}, a)$, hereafter the *monitoring cost*, represents the new cost associated with the processing and analysis of the raw performance data (e.g., running speech analytics software on vendor's cloud server). In particular, $\mu > 0$ is an exogenous parameter which we will further discuss in Section 3.1, whereas $H(\mathcal{P}, a) \geq 0$ represents the quantity of the information carried by the output signal and is assumed to satisfy the following properties:

Assumption 1. *There exists a function h such that $H(\mathcal{P}, a) = h(\boldsymbol{\pi}(\mathcal{P}, a))$ for all $(\mathcal{P}, a)$. And for all $N \in \mathbb{N}$ and $(\pi_1, \cdots, \pi_N) \in \Delta^N$,*

(a) $h(\pi_1, \cdots, \pi_N) = h\left(\pi_{\sigma(1)}, \cdots, \pi_{\sigma(N)}\right)$ for all permutation σ on $\{1, \cdots, N\}$;

(b) $h(\pi_1, \cdots, \pi_N) < h(\pi_1', \pi_1'', \cdots, \pi_N)$ for all $\pi_1', \pi_1'' > 0$ where $\pi_1' + \pi_1'' = \pi_1$.

Assumption 1 is inspired by the basic principle of Shannon's (1948) information theory: *the amount of information that a channel processes (here, the monitoring technology) depends only on the frequencies of the messages that it transmits.*[8] In the spirit of Shannon (1948), we assume that the cost associated with running a monitoring technology depends only on the probability distribution of the output signal, that this cost is invariant to the contents or the naming of performance categories, and that it increases as the monitoring technology becomes more fine-grained. These assumptions hold true for bits $\log_2 |\mathcal{P}|$, entropy $-\sum_A P_a(A) \log_2 P_a(A)$, as well as other common measures of the quantity of information carried by the output signal (throughout, $|\mathcal{P}|$ denotes the size of $\mathcal{P}$). The first measure is of practical significance, as $|\mathcal{P}|$ is referred to as the *rating scale* and represents the informativeness of the rating scheme in human resource management (see, e.g., Hook et al. (2011)). Additional

[8]Shannon (1948) insists that "the semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design." See also Cover and Thomas (2006) for a textbook treatment of this subject.

assumptions are needed to ensure that the monitoring cost is indeed the entropy of the output signal. We do *not* make these assumptions for generality's sake and do *not* use them to prove our main results.

Incentive constraints A contract that induces high effort satisfies the agent's limited liability constraint:

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

as well as his incentive compatibility constraint.[9] To formulate the concept of incentive compatibility, we equip each P_a with a probability density function p_a, and assume as in the standard agency literature that the *likelihood ratio* p_0/p_1 is well-defined and its probability distribution is commonly knowledge. Under these assumptions, define a random variable $Z : \Omega \to \mathbb{R}$ by

$$Z = 1 - \frac{p_0}{p_1},$$

and let $Z(A)$ denote the image of any $A \in \Sigma$ under Z.[10] Define the *z-value* of any $A \in \Sigma$ by

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

and write the agent's incentive compatibility constraint as follows:

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

The principal's problem As stated previously, the principal's goal is to induce the agent to exert high effort. Thus an optimal incentive contract minimizes the total implementation cost under high effort, subject to the agent's incentive compatibility and limited liability constraints, i.e.,

$$\min_{\langle \mathcal{P}, w(\cdot) \rangle} \sum_{A \in \mathcal{P}} P_1(A) w(A) + \mu \cdot H(\mathcal{P}, 1) \text{ s.t. (IC) and (LL).} \tag{2.1}$$

In Appendix A.4 we will show that the principal's problem admits solutions under mild regularity conditions. We will denote such solutions by $\langle \mathcal{P}^*, w^*(\cdot) \rangle$.

[9]The online appendix examines the case where the agent faces an individual rationality constraint.
[10]Notice that $Z(\Omega) \subset (-\infty, 1]$ and $\mathbf{E}[Z \mid a = 1] = 0$.

2.2 Examples

2.2.1 Call center performance management: a revisit

In the example delineated in Section 1, a piece of raw performance data, i.e., a conversation between a call center agent and a customer, contains numerous performance indicators such as customer sentiment, voice quality and tone, frequency of alarming words such as "ridiculous," etc.. In a hypothetical world where data processing and analysis is costless, factoring detailed information along these indicators into the pay decision provides the agent with the strongest incentive to work. But in reality, this solution is costly to implement, as it requires, e.g., call centers pay vendors for the use of sophisticated speech analytics software that consumes significant server space and power. For the sake of cost efficiency, most speech analytics software classifies the agent's performance into only a few categories such as "satisfactory" and "unsatisfactory" (see Figure 1 for a graphical illustration).

Meanwhile, enabled by improved technologies, call centers can now choose between a wide range of products with varying emphases on the above described indicators (e.g., Beyond Verbal specializes in emotion detection and CallMiner in word spotting). In the current model, this flexibility manifests itself through the assumption that the principal can adopt any partitional monitoring technology, such as the one depicted in Figure 1, where trading off different performance indicators amounts to adjusting the straight-line boundary between "satisfactory" and "unsatisfactory." By information theory, such adjustment affects the cost of data processing and analysis only through the probabilities that the agent is rated as "satisfactory" and "unsatisfactory." In the opposite situation where *any* other consideration (e.g., naming or contents of performance categories) enters the principal's calculus, too, one is essentially limiting the principal's choice for reasons beyond data processing and analysis.

2.2.2 Setup cost

It is useful to distinguish our monitoring cost from the *setup cost*, namely the cost incurred to set up the infrastructure for data utilization. Examples include the cost of establishing and maintaining electronic medical records (EMR) in the case of pay-for-performance for physicians (Roland (2004), Groves et al. (2013)), as well as the cost of rater training associated with conducting 360-degree performance appraisals among employees (Bracken et al. (2001)).[11][12]

The major role of setup cost is to change the probability space (Ω, Σ, P_a). In the above examples, data collection and maintenance expand (Ω, Σ, P_a), whereas attempts to correct rater biases or to reduce rating errors affect P_a while keeping (Ω, Σ) fixed. Throughout the analysis, we take the probability space and hence the setup cost as given in order to focus on the issue of data utilization. That said, we can embed the current analysis into a richer setting where the principal first incurs the setup cost and then the monitoring cost. Results below carry over to this new setting.

3 Main Results

This section analyzes the optimal incentive contract. Unless otherwise specified, results below hold true except perhaps on a measure-zero set of data points.

We begin with the definition of Z-convexity:

Definition 1. *A set $A \in \Sigma$ is Z-convex if for all $\omega', \omega'' \in A$ where $Z(\omega') \neq Z(\omega'')$,*

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

In words, a set A is Z-convexity if whenever it contains two data points of different z-values, it also contains all data points of in-between z-values. In the case $Z(\Omega)$ is connected in $\mathbb{R}$, this definition reduces to the convexity of $Z(A)$ in $\mathbb{R}$.

The next assumption says that the distribution of Z has no atom or hole; it is made to ensure some regularity in the solution to the principal's problem:

Assumption 2. *$Z(\Omega)$ is connected and the cumulative density function of Z under $a = 1$ is continuous on $Z(\Omega)$.*

We now state the main result of this paper:

[11] 360-degree performance appraisal, as its name suggests, refers to the use of increasingly automated systems to aggregate the feedback received from multiple sources (e.g., supervisors, peers, subordinates) into ratings such as "satisfactory" and "unsatisfactory." It is viewed as one of the most prominent managerial innovations driven by IT and has been adopted by more than one third of the U.S. companies and 90 percent of Fortune 500 companies.

[12] The magnitude of setup cost varies, ranging from virtually zero in the example of call center performance management (customer conversations have long been recorded by call centers but have only recently been utilized for evaluation and compensation purposes) to something significant in the case of pay-for-performance for physicians.

Theorem 1. *Assume Assumption 1. Then any* $\mathcal{P}^*$ *consists of finite* Z*-convex cells. Assume, in addition, Assumption 2. Then there exist* $\inf Z(\Omega) = \widehat{z}_0 < \widehat{z}_1 < \cdots < \widehat{z}_N = \sup Z(\Omega)$ *such that* $\mathcal{P}^* = \{A_1, \cdots, A_N\}$*, where* $A_n = \{\omega : Z(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ *for* $n = 1, \cdots, N$.

Theorem 1 says that the optimal monitoring technology, just like most real-world performance classification schemes, features *information aggregation*, in that it classifies fine-grained and potentially high-dimensional performance data into coarse and rank-ordered grades such as "satisfactory" and "unsatisfactory." Moreover, this is achieved by sorting data points of high (resp. low) z-values into high (resp. low) performance categories, and in some cases by first dividing z-values into non-overlapping intervals and then backing out the partition of the raw data set accordingly.

Proof sketch Theorem 1 can be understood in two steps. The first step involves solving for the optimal wage scheme for any given monitoring technology $\mathcal{P}$, i.e.,

$$\min_{w:\mathcal{P}\to\mathbb{R}_+} \sum_{A\in\mathcal{P}} P_1(A)w(A), \text{ s.t. (IC) and (LL).} \tag{3.1}$$

Denote the solution to this problem (if exists) by $w^*(\cdot;\mathcal{P})$. Define

$$\underline{z} = \frac{1}{u'(0)}.$$

The next lemma restates Holmstrom's (1979) *sufficient statistic principle*:

Lemma 1. *If* $w^*(\cdot;\mathcal{P})$ *exists, then there exists* $\lambda > 0$ *such that* $u'(w^*(A;\mathcal{P})) = 1/\max\{\lambda z(A), \underline{z}\}$ *for all* $A \in \mathcal{P}$.

Lemma 1 reiterates the fact that z-value is a sufficient statistic for the agent's performance when the monitoring technology (and hence the Lagrange multiplier λ of the (IC) constraint) is exogenously given. From this, it follows that when monitoring is flexible and yet costly, the principal should focus on the processing and analysis of z-values and discard the part of the data that is orthogonal to the z-value. This gives rise to the result on information aggregation.

The second step is to solve for the optimal monitoring technology. Specifically, let $W(\mathcal{P})$ denote the incentive cost incurred by $w^*(\cdot;\mathcal{P})$, and rewrite the principal's problem as follows:

$$\min_{\mathcal{P}} W(\mathcal{P}) + H(\mathcal{P}, 1). \tag{3.2}$$

To gain intuition into the result on Z-convexity, we perturb the optimal monitoring technology by reassigning raw data points from one performance category to another, which in turn affects the implementation cost directly through the probability measures and z-values of the performance categories, and indirectly through the Lagrange multipliers of the incentive constraints. These effects are absent from Holmstrom (1979) and are difficult to assess in general.

To make progress, consider a particular kind of perturbation that swaps ϵ-measured data points of z-values z' and z'' across performance categories j and k. By Assumption 1, this perturbation has *no* effect on the probability distribution of the output signal and hence the monitoring cost. In the meantime, it changes the principal's Lagrangian to the following (ignore the (LL) constraint for illustrative purpose):

$$\mathcal{L}(\epsilon) = \sum_n \pi_n w_n^*(\epsilon) - \lambda(\epsilon) \left[\sum_n \pi_n u(w_n^*(\epsilon)) z_n(\epsilon) - c \right], \tag{3.3}$$

where π_n and $w_n^*(\epsilon)$ denote the probability measure and the optimal wage associated with A_n, respectively, and $\lambda(\epsilon)$ denotes the Lagrange multiplier associated with the (IC) constraint. Differentiating both sides of Equation (3.3) with respect to ϵ yields

$$\begin{aligned}
\mathcal{L}'(0) = & \sum_n \pi_n w_n^{*\prime}(0) - \lambda'(0) \underbrace{\left[\sum_n \pi_n u(w_n^*(0)) z_n(0) - c \right]}_{(1) \quad = 0} \\
& - \lambda(0) \left[\sum_n \pi_n \cdot \underbrace{u'(w_n^*(0)) z_n(0)}_{(2) \quad = 1/\lambda(0)} \cdot w_n^{*\prime}(0) + \sum_n \pi_n u(w_n^*(0)) z_n'(0) \right] \\
= & -\lambda(0) \sum_n \pi_n u(w_n^*(0)) z_n'(0) \\
= & \lambda(0)[u(w_k^*) - u(w_j^*)](z'' - z'),
\end{aligned}$$

where (1) $= 0$ because the (IC) constraint binds and (2) $= 1/\lambda(0)$ by Lemma 1. Since $\mathcal{L}'(0) \geq 0$, it follows that the optimal assignment of z-values to performance categories is positive assortative, meaning that if a performance category contains extreme but not intermediate z-values, then the assignment goes wrong direction and hence is not optimal. This completes the proof of the result on Z-convexity.

3.1 Implications

Strict MLRP Theorem 1 suggests that the output signal induced by the optimal monitoring technology satisfy the strict monotone likelihood ratio property with respect to the order induced by z-values. Formally,

Definition 2. *For any totally ordered set* $(\mathcal{P}, \preceq)$, $X : \Omega \to \mathcal{P}$ *satisfies*

(i) the monotone likelihood ratio property (MLRP) *w.r.t.* $\preceq$ *if for all* $A, A' \in \mathcal{P}$*, we have* $z(A) \leq z(A')$ *if and only if* $A \preceq A'$*;*

(ii) the strict monotone likelihood ratio property (strict MLRP) *w.r.t.* $\preceq$ *if for all* $A, A' \in \mathcal{P}$*, we have* $z(A) < z(A')$ *if and only if* $A \preceq A'$*.*

Definition 3. $A \overset{z}{\preceq} A'$ *if* $z(A) \leq z(A')$*.*

While the signal induced by any arbitrary monitoring technology satisfies the MLRP with respect to $\overset{z}{\preceq}$, it violates the strict MLRP with respect to $\overset{z}{\preceq}$ in the case a single z-value can be associated with multiple performance categories. By contrast, the signal induced by the optimal monitoring technology always satisfies the strict MLRP with respect to $\overset{z}{\preceq}$, because in the above described situation, merging performance categories of the same z-value has no effect on the incentive cost but saves the monitoring cost.

Cutoff z-values Theorem 1 suggests that we solve the principal's problem by first applying the optimal cutoffs to z-values and then backing out the partition of the raw data set accordingly. The implication of this observation is twofold:

- First, we can prove existence of the optimal incentive contract under mild regularity conditions.
- Second, solving for the optimal cutoff z-values leads to a full characterization of the optimal incentive contract and enables comparative statics analysis.

The first exercise is straightforward and is relegated to Appendix A.4. The second exercise is more challenging, because perturbing cutoff z-values — which affects the probability measures of the performance categories in Equation (3.3), too — entails subtle effects that depend directly the Lagrange multiplier of the incentive compatibility constraint. Below we explicitly solve for the optimal cutoff z-values in an example:

Example 1. Suppose that $u(w) = \sqrt{w}$, that Z is uniformly distributed over $[-1, 1]$ and that $H(\mathcal{P}, a) = f(|\mathcal{P}|)$. Fix $|\mathcal{P}| = N$, and let $\widehat{z}_0 = -1$ and $\widehat{z}_N = 1$. Tedious but straightforward algebra leads to the following expression for the total implementation cost:

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

where

$$\pi_n = \int_{Z(\omega)\in[\widehat{z}_{n-1},\widehat{z}_n)} dP_1(\omega) = \frac{1}{2}\left[\widehat{z}_n - \widehat{z}_{n-1}\right]$$

and

$$z_n = \int_{Z(\omega)\in[\widehat{z}_{n-1},\widehat{z}_n)} Z(\omega)dP_1(\omega)/\pi_n = \frac{1}{2}\left[\widehat{z}_n + \widehat{z}_{n-1}\right]$$

for $n = 1, \cdots, N-1$. Differentiating the above formula with respect to $\widehat{z}_1, \cdots, \widehat{z}_{N-1}$ and setting the result equal to zero gives us

$$2\widehat{z}_n = \max\{0, z_{n+1}\} + \max\{0, z_n\}$$

for $n = 1, \cdots, N-1$. Solving this system of equations yields $\widehat{z}_n = \frac{2n-1}{2N-1}$, $n = 1, \cdots, N-1$.

Comparative statics The parameter μ captures factors that affect the (opportunity) cost of data processing and analysis, e.g., advent of IT-based HR management systems since the late 90's, advances in natural language processing and voice analysis, improvements in software and computing power, etc..[13] To facilitate comparative statics analysis, we denote optimal contracts by $\langle \mathcal{P}^*(\mu), w^*(\cdot;\mu)\rangle$ in order to make

[13] Demand-side factors that increase the adoption of IT-based HR management systems include tough product market competition, loose labor market regulation, etc. (Bloom and Van Reenen (2006, 2007, 2010)).

their dependence on μ explicit.

Proposition 1. *Fix $0 < \mu < \mu'$. Then for any choices of $\langle \mathcal{P}^*(\mu), w^*(\cdot;\mu)\rangle$ and $\langle \mathcal{P}^*(\mu'), w^*(\cdot;\mu')\rangle$, (i) $W(\mathcal{P}^*(\mu)) < W(\mathcal{P}^*(\mu'))$ and $H(\mathcal{P}^*(\mu),1) > H(\mathcal{P}^*(\mu'),1)$, and (ii) $|\mathcal{P}^*(\mu)| > |\mathcal{P}^*(\mu')|$ in the case where $H(\mathcal{P},a) = f(|\mathcal{P}|)$ and f is increasing.*

Proposition 1 says that as μ decreases, the principal can pay less wage on average without hurting the agent's incentive, and the information carried by the output signal becomes more fine-grained. In the case where the monitoring cost is an increasing function of the rating scale, it follows from Part (ii) that the optimal rating scale is decreasing in μ. For other monitoring cost functions, we numerically assess the relationship between the optimal rating scale and μ and report results in Figure 2.[14]

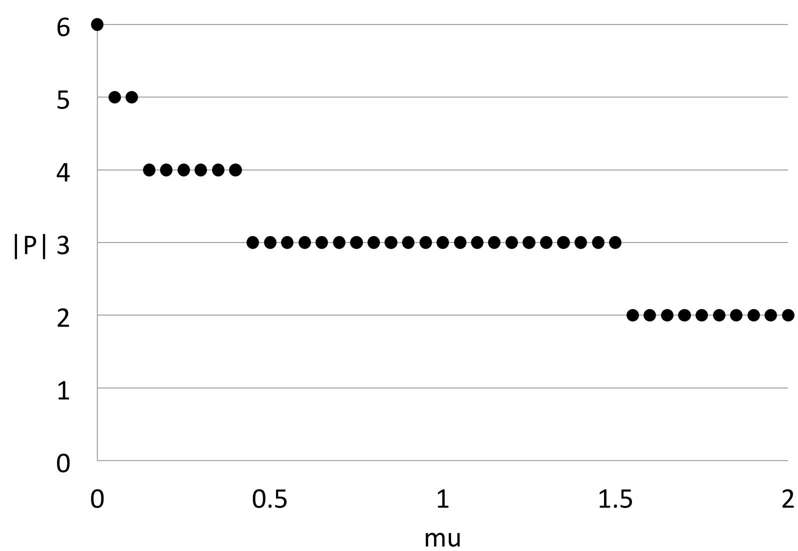

Figure 2: $|\mathcal{P}^*|$ at varying μ's: entropy cost, $u(w) = \sqrt{w}$, $Z \sim U[-1,1]$, $c = 1$.

Our predictions are consistent with several strands of empirical facts. For example, access to IT has proven to increase the adoption of fine-grained performance classification schemes among manufacturing companies, other things being equal (Bloom and Van Reenen (2006, 2007, 2010), Bloom et al. (2012)).[15] Crowd-sourcing real-time

[14]The ability to perform comparative statics analysis with respect to μ distinguishes our theory from alternative explanations for wage compression such as subjective performance evaluation, e.g., Levin (2003).

[15]See appendices of Bloom and Van Reenen (2006, 2007) for survey questions regarding the fine-grainedness of the performance classification scheme, e.g., "each employee is given a red light (not performing), amber light (doing well and meeting targets), a green light (consistently meeting targets

data from meetings, problem sessions, etc., has enabled "exact individual analysis" separates distinctive and mediocre performers in companies like GE and Zalando (Ewenstein et al. (2016)). Finally, natural language processing tools have improved the precision of physician care monitoring through analysis of the narrative reports in EMR (Murff et al. (2011)).

Figure 3: Wage distributions under $\mu = 0.05$ and $\mu = 1.4$: entropy cost, $u(w) = \sqrt{w}$, $Z \sim U[-1, 1]$, $c = 1$.

The impact of μ on the overall wage distribution is more subtle because of, once

very high performer) and a blue light (high performer capable of promotion of up to two levels)," and "rewards is based on an individual's commitment to the company measured by seniority."

again, the endogenous interplay between the monitoring technology and the wage scheme. To formally assess this statement, we plot the wage distribution under different μ's in Figure 3. A careful inspection of this figure reveals, for example, that wage variance can actually decrease with μ. This observation, albeit somewhat surprising at first sight, is not difficult to understand after all: while a small μ makes it easy to extract informative and potentially highly varying signals from the raw data, it also leads to the concentration of rewards among only a few exceptional performers (see, e.g., the upper panel of Figure 3). The overall impact on wage variance is ambiguous and depends on which effect dominates the other.

Minor extensions The online appendix investigates several minor extensions, including cases where (i) the agent faces an individual rationality constraint, (ii) the monitoring cost represents a communication cost shared between players, (iii) the principal observes a signal for free (e.g., plant-level output) before deciding how to utilize the raw performance data, and (iv) the principal can adopt random monitoring technologies.

4 Extensions

4.1 Multiple Agents

Setup Each of the two agents $i = 1, 2$ earns a payoff $u_i(w_i) - c_i(a_i)$ from spending a non-negative wage $w_i \geq 0$ and exerting either high effort ($a_i = 1$) or low effort ($a_i = 0$), where . Each effort profile $\mathbf{a} = a_1 a_2$ generates a probability space $(\Omega, \Sigma, P_{\mathbf{a}})$, where Ω consists of agents' raw performance data and $P_{\mathbf{a}}$ is the probability measure on (Ω, Σ) conditional on the effort profile being of $\mathbf{a}$. A risk-neutral principal cannot observe efforts, and her goal is to induce both agents to exert high effort. This last restriction is inconsequential, as the problem of eliciting any other effort profile can be solved analogously.

Incentive contract An incentive contract $\langle \mathcal{P}, \mathbf{w}(\cdot) \rangle$ consists of a monitoring technology $\mathcal{P}$ and a wage scheme $\mathbf{w}(\cdot)$. As before, we allow $\mathcal{P}$ to be any finite partition of Ω whose cells belong to Σ, and use $\mathbf{w} : \mathcal{P} \to \mathbb{R}^2_+$ to map each cell A of $\mathcal{P}$ to a vector $\mathbf{w}(A) = (w_1(A), w_2(A))$ of non-negative wages. For each $\boldsymbol{\omega} \in \Omega$, denote by

$A(\boldsymbol{\omega})$ the unique performance category that contains $\boldsymbol{\omega}$, and by $\mathbf{w}(A(\boldsymbol{\omega}))$ the wage vector associated with $A(\boldsymbol{\omega})$. Time evolves as follows:

1. The principal commits to $\langle \mathcal{P}, \mathbf{w}(\cdot) \rangle$;
2. Agents independently choose $a_i \in \{0, 1\}$, $i = 1, 2$;
3. Nature draws $\boldsymbol{\omega}$ from Ω according to $P_{\mathbf{a}}$;
4. The monitoring technology outputs $A(\boldsymbol{\omega})$;
5. The principal pays $w_i(A(\boldsymbol{\omega}))$ to agent $i = 1, 2$.

Equip each $P_{\mathbf{a}}$ with a probability density function $p_{\mathbf{a}}$ and write $\mathbf{1} = (1, 1)$. Define $\mathbf{Z} = (Z_1, Z_2)$, where

$$Z_i = 1 - \frac{p_{a_i=0, a_{-i}=1}}{p_{\mathbf{1}}}$$

for $i = 1, 2$. For each $i = 1, 2$ and $A \in \Sigma$, define the z_i-value of A by

$$z_i(A) = \mathbf{E}\left[Z_i \mid A; \mathbf{a} = \mathbf{1}\right].$$

A contract is incentive compatible for agent i if

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

and it satisfies agent i's limited liability constraint if

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

An optimal contract minimizes the implementation cost under the high effort profile, subject to agents' incentive compatibility and limited liability constraints, i.e.,

$$\min_{\langle \mathcal{P}, \mathbf{w}(\cdot) \rangle} \sum_{A \in \mathcal{P}} P_{\mathbf{a}}(A) \sum_{i=1}^{2} w_i(A) + \mu \cdot H(\mathcal{P}, \mathbf{1}) \text{ s.t. (IC}_i\text{) and (LL}_i\text{)}, \ i = 1, 2. \tag{4.1}$$

Optimal multi-agent contract The next definition generalizes the concept of Z-convexity:

Definition 4. *A set $A \in \Sigma$ is* $\mathbf{Z}$*-convex if for all $\boldsymbol{\omega}', \boldsymbol{\omega}'' \in A$ where $\mathbf{Z}(\boldsymbol{\omega}') \neq \mathbf{Z}(\boldsymbol{\omega}'')$,*

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

The next assumption ensures some regularity in the optimal incentive contract:

Assumption 3. $\mathbf{Z}(\Omega)$ *is connected in $\mathbb{R}^2$ and the cumulative density function of* $\mathbf{Z}$ *under $\mathbf{a} = \mathbf{1}$ is continuous on $\mathbf{Z}(\Omega)$.*

The next result extends Theorem 1 to encompass multiple agents:

Theorem 2. *Assume Assumption 1. Then any $\mathcal{P}^*$ consists of finite* $\mathbf{Z}$*-convex cells. Assume, in addition, Assumption 3. Then each $\mathbf{Z}(A)$ is a convex polygon in $\mathbf{Z}(\Omega)$.*

Theorem 2 says that any optimal monitoring technology comprises finite $\mathbf{Z}$-convex cells, and can in some cases be obtained from first dividing the space of $\mathbf{z}$-values into convex polygons and then backing out the partition of the raw data set accordingly. Proving this result requires a non-trivial extension of earlier arguments. As before, we swap ϵ-measured data points of values $\mathbf{z}'$ and $\mathbf{z}''$ across performance categories j and k. The resulting impact on the principal's Lagrangian is

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

where $\widehat{\mathbf{u}}_n$ is a profile of augmented agent utilities at performance category $n = j, k$. Since $\mathcal{L}'(0) \geq 0$, the optimal assignment of data points to performance categories is "positive assortative," where the degree of assortativeness is given by the inner product between the $\mathbf{z}$-value and the augmented utility profile. Geometrically, this means that a data point is more likely to be assigned to a performance category as its $\mathbf{z}$-value moves in the direction of the augmented utility profile. The result on $\mathbf{Z}$-convexity follows, because if a performance category contains data points of extreme but not intermediate $\mathbf{z}$-values, then the assignment of data points goes wrong against the direction of agent utilities and an improvement can be constructed.

Solving for the optimal monitoring technology is computationally hard. Nevertheless, the fact that the boundaries of each $\mathbf{Z}(A)$ consist of straight line segments in $\mathbf{Z}(\Omega)$ implies the following:

- First, we can prove existence of the optimal incentive contract under mild regularity conditions. See Appendix A.4 for further details;

- Second, it is easy to compute the optimal incentive contract within certain classes. For example, any bi-partitional contract takes the form of either a team or a tournament and is parameterized by the intercept and slope of the straight line depicted in Figure 4. Meanwhile, any contract that monitors and rewards agents on an individual basis is fully determined by the independent performance cutoffs of individual agents (see Figure 5 for a graphical illustration). These observations provide a useful starting point for the upcoming analysis.

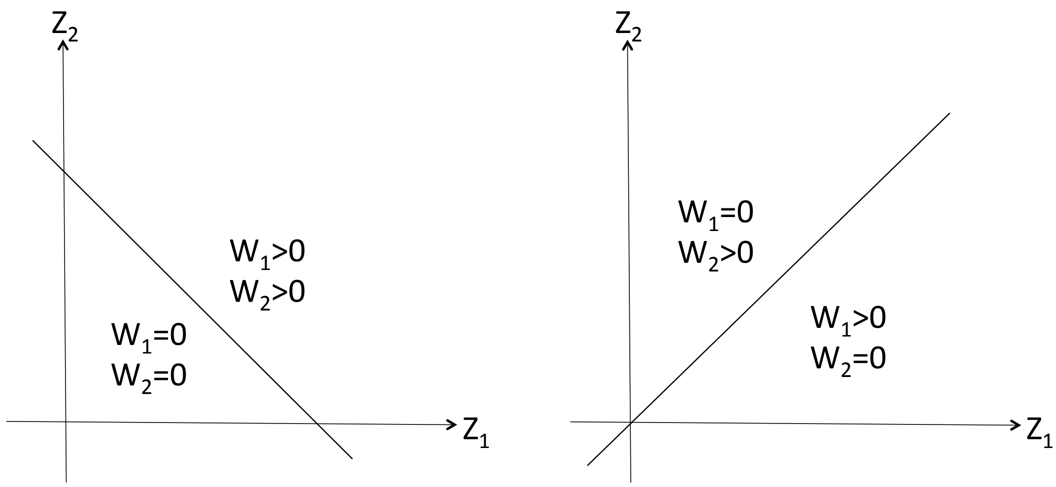

Figure 4: Bi-partitional contract: team and tournament.

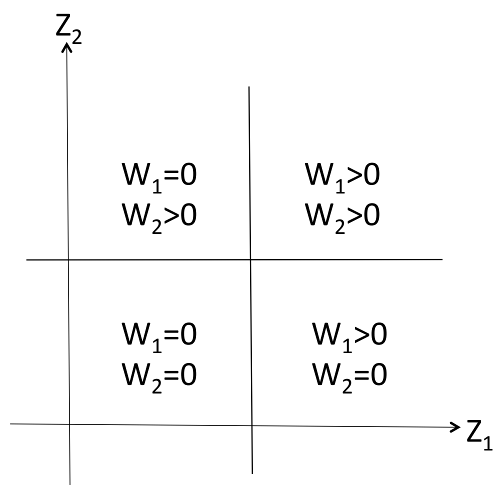

Figure 5: Quad-partitional individual contract.

4.1.1 Application: individual vs. group monitoring

Theorem 2 suggests that the principal use group monitoring as a best response to high monitoring cost. To obtain the sharpest formulation, we assume that agents are *technologically independent* in the remainder of the analysis:

Assumption 4. *For each* $\mathbf{a} \in \{0,1\}^2$*,*

$$(\Omega, \Sigma, P_{\mathbf{a}}) = (\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, P_{1,a_1} \times P_{2,a_2}),$$

where each $(\Omega_i, \Sigma_i, P_{i,a_i})$ *is a probability space and* P_{i,a_i} *depends only on* i *and* a_i*.*

In the language of contract theory, Assumption 4 rules out any kind of *technology linkage* (i.e., ω_i depends on a_{-i}) or *common productivity shock* (i.e., ω_i, ω_j are correlated given $\mathbf{a}$) between agents.

Definition 5. $\langle \mathcal{P}, \mathbf{w}(\cdot;\mathcal{P})\rangle$ *is an* individual incentive contract *if* $\mathcal{P}$ *and* $\mathbf{w}(\cdot;\mathcal{P})$ *are both* individual*; otherwise it is a* group incentive contract*. Specifically,*

(i) $\mathcal{P}$ *is an* individual monitoring technology *if for every* $A \in \mathcal{P}$*, there exist* $A_i \in \Sigma_i$ $i = 1,2$ *such that* $A = A_1 \times A_2$*; otherwise it is a* group monitoring technology*;*

(ii) A wage scheme $\mathbf{w}(\cdot;\mathcal{P})$ *measurable with respect to an individual monitoring technology* $\mathcal{P}$ *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 all* $i = 1,2$ *and* $A_i \times A'_{-i}, A_i \times A''_{-i} \in \mathcal{P}$*; otherwise it is a* group wage scheme.

In words, a group incentive contract either conducts group monitoring or combines individual monitoring with group incentive pays. The next lemma, which follows immediately from the sufficient statistic principle or Holmstrom (1982), shows that the second option is sub-optimal if agents are technologically independent:

Lemma 2. *Assume Assumption 4. Then any* $\langle \mathcal{P}, w^*(\cdot;\mathcal{P})\rangle$ *is a group incentive contract if and only if* $\mathcal{P}$ *is a group monitoring technology.*

Proof. The "only if" direction holds true by definition. To prove the "if" direction, take any individual monitoring $\mathcal{P}$ and let λ_i denote the Lagrange multiplier of the (IC_i) constraint. A straightforward extension of Lemma 1 shows that

$$w_i(A) = (u')^{-1}\left(\frac{1}{\max\{\lambda_i z_i(A_i), \underline{z}\}}\right)$$

for all $A \in \mathcal{P}$. $\square$

To assess the optimality of group monitoring, we divide the cost associated with implementing the optimal bi-partitional contract by that of the optimal individual incentive contract. Denote the result by I, where $I < 1$ is a definitive indicator that group monitoring is optimal. Straightforward algebra shows that $I < 1$ when μ is large in the case $H(\mathcal{P}, a) = f(|\mathcal{P}|)$ and f is increasing. For other monitoring cost functions such as entropy, we compute I based on the prior discussion of how to parameterize bi-partitional and individual incentive contracts (see the end of the previous section) and report results in Figure 6.

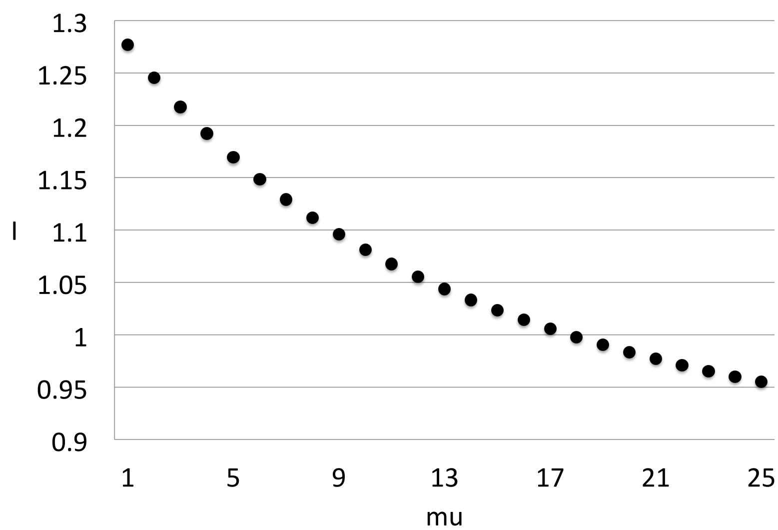

Figure 6: I at varying μ's: entropy cost, $u(w) = \sqrt{w}$, $Z \sim U[-1, 1]$, $c = 1$.

Results so far suggest that the principal use group monitoring as a best response to high monitoring cost. This finding formalizes the intuitions of Alchian and Demsetz (1972) and Lazear and Rosen (1981) that either team or tournament should be the dominant incentive system when individual performance metrics are too costly to obtain.[16] It enriches the analyses of Holmstrom (1982), Green and Stokey (1983)

[16] Alchian and Demsetz (1972) notes that "metering costs must be incurred to monitor each other ... If there is a net increase in productivity available by team production, net of the metering cost associated with disciplining the team, then team production will be relied upon ..." Lazear and Rosen (1981) observes that: "in a modern, complex business organization ... the costs of measurement for each conceivable candidate are prohibitively expensive. Instead, it might be said that those in the running are "tested" by assessments of performance at lower positions ... It is in these situations that the conditions seem ripe for tournaments to be the dominant incentive contract institution."

and Mookherjee (1984), which limit the use of group incentive contract among technologically interdependent agents while ignoring the potential cost arising from monitoring. Recently, these views are reconciled by Bloom and Van Reenen (2006, 2007), which find — just as the current theory predicts — that companies make significantly different choices between individual and that group monitoring despite being technologically similar,[17] and that group monitoring is mostly used when the company's capacity to conduct individual monitoring is limited by, e.g., the lack of IT access.

Our prediction is supported by the increasing adoption of tech-enabled individual performance management in call centers and retail stores. Prior to the advent of speech analytics software, call centers relied mostly on group performance indicators (e.g., queuing delay) for structuring feedbacks and incentive pays, because distilling individual-level information from call records was very costly, if not impossible. Empowered by cloud-based point of sales (POS) systems, retail and fast food stores are making more promotion and probation decisions based on the individual-level information extracted from employees' real-time locations and transactions (Kaplan (2015)).

4.2 Multiple Actions

Setup In the baseline model, suppose instead that the agent can take one of the finite actions in $\mathcal{A}$ whereas the principal wants to induce an arbitrary action $a^* \in \mathcal{A}$. For each deviant action $a \in \mathcal{D} = \mathcal{A} - \{a^*\}$, define a random variable $Z_a : \Omega \to \mathbb{R}$ by

$$Z_a = 1 - \frac{p_a}{p_{a^*}},$$

and define the z_a-value of $A \in \Sigma$ by

$$z_a(A) = 1 - \frac{P_a(A)}{P_{a^*}(A)}.$$

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

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

[17] See the survey questions of Bloom and Van Reenen (2006, 2007) regarding the choice between monitoring individual-level performance or shift/plant-level performance. The former is regarded as an advanced but expensive managerial practice in the manufacturing sector.

An optimal contract solves

$$\min_{\langle \mathcal{P}, w(\cdot)\rangle} \sum_{A \in \mathcal{P}} P_{a^*}(A) w(A) + \mu \cdot H(\mathcal{P}, a^*) \text{ s.t. } (\text{IC}_a) \text{ for all } a \in \mathcal{D} \text{ and (LL).} \tag{4.2}$$

Optimal incentive contract Take any profile of non-negative real numbers $\boldsymbol{\lambda} = (\lambda_a)_{a \in \mathcal{D}} \in \mathbb{R}_+^{|\mathcal{D}|}$, and define the random variable $Z_{\boldsymbol{\lambda}} : \Omega \to \mathbb{R}$ by

$$Z_{\boldsymbol{\lambda}} = \sum_{a \in \mathcal{D}} \lambda_a \cdot Z_a.$$

The next definition generalizes the concept of Z-convexity:

Definition 6. *A set $A \in \Sigma$ is $Z_{\boldsymbol{\lambda}}$-convex if for all $\omega', \omega'' \in A$ where $Z_{\boldsymbol{\lambda}}(\omega') \neq Z_{\boldsymbol{\lambda}}(\omega'')$,*

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

The next result extends Theorem 1 to encompass multiple deviant actions:

Theorem 3. *Assume Assumption 1. Then for any $\mathcal{P}^*$, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}_+^{|\mathcal{D}|}$ such that each $A \in \mathcal{P}^*$ is $Z_{\boldsymbol{\lambda}^*}$-convex. Assume, in addition, Assumption 2 for all $a \in \mathcal{D}$. Then there exist $-\infty \leq \widehat{z}_0 < \cdots < \widehat{z}_N < +\infty$ such that $\mathcal{P}^* = \{A_1, \cdots, A_N\}$, where $A_n = \{\omega : Z_{\boldsymbol{\lambda}^*}(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ for $n = 1, \cdots, N$.*

Theorem 3 says that in the presence of multiple deviant actions, the resources spent on the detection of each deviation should exactly match the Lagrange multiplier of the corresponding incentive compatibility constraint. The proof of this result requires a careful handling of the complementary slackness constraint, but the intuition is straightforward: when λ_a^* is large and hence the agent is tempted to commit deviation a, the focus of the performance analysis should be on whether the agent has committed deviation a or not, and the overall grade should significantly reflect the agent's performance Z_a in resisting this deviation.

4.2.1 Application: multiple tasks

A single agent can exert either high effort ($a_i = 1$) or low effort ($a_i = 0$) in each of the two tasks $i = 1, 2$. Each a_i *independently* generates a probability space $(\Omega_i, \Sigma_i, P_{i,a_i})$, where Ω_i consists of the raw data on the agent's performance in task i, and P_{i,a_i} is the probability measure on (Ω_i, Σ_i) conditional on the effort in task i being of a_i. The principal does not observe efforts, and her goal is to induce high effort in both tasks.

Write $\mathbf{a} = a_1a_2$ and $\boldsymbol{\omega} = \omega_1\omega_2$. Let $\mathcal{A} = \{11, 01, 10, 00\}$, $\mathbf{a}^* = 11$ and $\mathcal{D} = \{01, 10, 00\}$, and define $Z_{\mathbf{a}}$ for each $\mathbf{a} \in \mathcal{D}$. By the assumption of independence, we have

$$Z_{01}(\boldsymbol{\omega}) = Z_{01}(\omega_1),$$

$$Z_{10}(\boldsymbol{\omega}) = Z_{10}(\omega_2),$$

and

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

For any given profile $\boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{D}|}$ of non-negative real numbers, define

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

A straightforward extension of Theorem 3 leads to the following:

Corollary 1. *Assume Assumptions 1 and 2 for all* $\mathbf{a} \in \mathcal{D}$. *Then for any* $\mathcal{P}^*$, *there exist*

(i) $\boldsymbol{\lambda}^*$ *where* $\lambda^*_{\mathbf{a}} \geq 0$ *for all* $\mathbf{a} \in \mathcal{D}$ *and* $\lambda^*_{01} + \lambda^*_{00}, \lambda^*_{10} + \lambda^*_{00} > 0$,

(ii) $-\infty \leq \widehat{z}_0 < \cdots < \widehat{z}_N < +\infty$,

such that $\mathcal{P}^* = \{A_1, \cdots, A_N\}$, *where* $A_n = \{\boldsymbol{\omega} : Z_{\boldsymbol{\lambda}^*}(\boldsymbol{\omega}) \in [\widehat{z}_{n-1}, \widehat{z}_n)\}$ *for* $n = 1, \cdots, N$.

In a seminar paper, Holmstrom and Milgrom (1991) argues that when the agent faces multiple tasks of different measurabilities (see below for the formal definition), over-incentivizing the easy-to-measure task prevents the completion of the difficult-to-measure task. Thus, the primary focus in managing multi-task agency relationships should be on the power of the compensation scheme, and this is particularly true when the monitoring technology can only be taken as exogenously given.

Corollary 1 suggests that we focus instead on the determination of monitoring intensities which, by Corollary 1, should exactly match the agent's endogenous tendencies to shirk. The usefulness of this result is illustrated by the next example:

Example 2. Suppose that task one is teaching and task two is research, and let ω_1 represent teaching evaluation and ω_2 research output. By Corollary 1, the following ratio:

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

captures the sensitivity of the overall performance grade with respect to changes in the teaching evaluation and that of research output, respectively. Intuitively, a big R arises when the agent is tempted to shirk teaching obligations, and the principal best responds by focusing on the interpretation of teaching evaluations and treating the result as a serious component of the overall grade. While R is difficult to solve analytically because of, once again, endogeneity issues, we can nevertheless assess its relationship between model primitives through numerical analysis.

As in Holmstrom and Milgrom (1991), we investigate the relationship between R and effort measurability. To keep consistency, we assume that

- $\omega_i = a_i + \xi_i$, $i = 1, 2$, where each ξ_i is normally distributed with mean zero and variance σ_i^2. Under these assumptions, a decrease in σ_i^2 increases the *measurability* of a_i in the sense of Blackwell (1953);

- The agent has a CARA utility of consumption $u(w) = 1 - \exp(-\gamma w)$;

By contrast, we do *not* limit attention to linear wage schemes.

The numerical exercise examines the case of $|\mathcal{P}| = 2$, where we fix $\sigma_2^2 = 1$ and plot R against σ_1^2. Assuming that our parameter choices are reasonable, the result depicted in Figure 7 delivers the following message: when making important decisions such as tenure and promotion, university committees should focus on reading research papers and recommendation letters rather than parsing the increasing amount of details contained in teaching evaluations (e.g., enlarging rating scale, written feedback, midterm evaluation, private conversation with director of undergraduate studies), and this is true even if the goal is to induce faculty members to work hard on both research and teaching.

5 Conclusion

We develop a theory of optimal incentive contracting where the monitoring technology that governs the processing and analysis of the agent's performance data is part of the principal's strategic planning. We formalize the flexibility and cost that recent

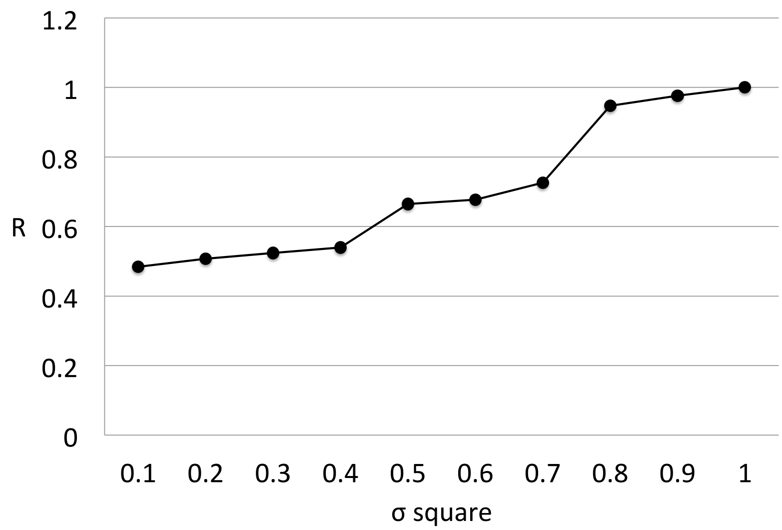

Figure 7: R at various σ_1^2's: $H(\mathcal{P}, a) = f(|\mathcal{P}|)$, $|\mathcal{P}| = 2$; $u(w) = 1 - \exp(-.5w)$; $c(00) = 0$, $c(01) = 0.3$, $c(10) = 0.2$ and $c(11) = 0.5$; ξ_1 and ξ_2 are normally distributed with mean zero and $\sigma_2^2 = 1$.

technology advances introduce into the design and implementation of monitoring technologies, and characterize the optimal monitoring technology through the tension between the incentive cost and the monitoring cost. In various classical settings, we derive testable predictions for the optimal incentive contract and examine their implications for workforce management and firms' internal organizations.

We conclude by suggesting potential avenues for future research. First, as mentioned in Section 1.1, our work is broadly related to the burgeoning literature on information design, as our principal is optimizing over a rich set of information structures subject to the agent's incentive constraints. We hope that the above developed techniques are useful for the study of related issues, such as how review sites like Yelp should disclose information to customers with limited attention spans, and how managers should structrue monitoring activities in long-term agency relationships. Second, it will be desirable to complement our theoretical research with quantifications of the effect of technology advances on the design and implementation of monitoring technologies. We hope that someone, maybe ourselves, will carry out this research agenda in the future.

A Omitted Proofs

A.1 Proof of Theorem 1

Throughout this section, let $\langle A_n, \pi_n, z_n, w_n \rangle_{n=1}^N$ be the tuple associated with an arbitrary incentive contract $\langle \mathcal{P}, w(\cdot) \rangle$, where $\pi_n = P_1(A_n)$, $z_n = z(A_n)$ and $w_n = w(A_n)$ for $n = 1, \cdots, N$. In the case $w(\cdot) = w^*(\cdot; \mathcal{P})$, write w_n^* rather than w_n, and drop the dependence on $\mathcal{P}$ for convenience. Assume w.l.o.g. that $z_1 \leq \cdots \leq z_N$.

We begin by proving Lemma 1:

Proof. Fix any $\mathcal{P}$. Let λ and η_n denote the Lagrange multiplier associated with the (IC) constraint and the (LL) constraint at w_n, respectively, and rewrite the principal's problem as follows:

$$\min_{\{w_n\}} \sum_{n=1}^N \pi_n w_n - \lambda \left[\sum_{n=1}^N \pi_n u(w_n) z_n - c \right] - \sum_{n=1}^N \eta_n w_n.$$

Taking derivative respect to w_n and setting the result equal to zero, we obtain:

$$u'(w_n^*) = \frac{1 - \eta_n / \pi_n}{\lambda z_n}.$$

From this, it follows that $w_n^* > 0 \iff u'(w_n^*) = 1/\lambda z_n \iff \lambda z_n > \underline{z}$, or equivalently $u'(w_n^*) = 1/\max\{\lambda z_n, \underline{z}\}$. $\square$

We next introduce a useful lemma:

Lemma 3. *Assume Assumption 1. Then any optimal incentive contract satisfies* $z_1 < \cdots < z_N$ *and* $0 = w_1^* < w_2^* < \cdots < w_N^*$.

Proof. Notice two things. First, if $w_j^* = w_k^*$ for some $j \neq k$, then merging A_j and A_k saves the monitoring cost without affecting the incentive cost, a contradiction. Second, the very fact that $\sum_{n=1}^N \pi_n z_n = 0$ implies that some z_n's are negative. Putting together these observations with Lemma 1, we see that under any optimal incentive contract, each performance category attains a distinct z-value and only one performance category attains a negative z-value. This leads to the conclusion that $z_1 < \cdots < z_N$ and that $0 = w_1^* < \cdots < w_N^*$. $\square$

We now finalize the proof of Theorem 1:

Proof. Take any $\langle \mathcal{P}^*, w^*(\cdot)\rangle$ with $\langle A_n, \pi_n, z_n, w_n^*\rangle_{n=1}^N$ being the associated tuple. In case some $A_j \in \mathcal{P}^*$ is not Z-convex, there exist $A', A'', \tilde{A} \in \Sigma$ such that

(i) $P_1(A') = P_1(A'') = P_1(\tilde{A}) = \epsilon$ for some small $\epsilon > 0$;

(ii) $A', A'' \subset A_j$ and $\tilde{A} \subset A_k \in \mathcal{P}^*$ for some $k \neq j$;

(iii) $\tilde{z} = (1-s)z' + sz''$ for some $s \in (0,1)$, where $\tilde{z} = z(\tilde{A})$, $z' = z(A')$ and $z'' = z(A'')$.

Consider the following perturbations to $\mathcal{P}^*$:

(a) Assign A' to A_k and $\tilde{A}$ to A_j (hereafter "swapping" A' and $\tilde{A}$);

(b) Assign $\tilde{A}$ to A_j and A'' to A_k (hereafter "swapping" $\tilde{A}$ and A'').

By construction, these perturbations have no effect on the probability distribution of the output signal and hence the monitoring cost. We now argue that one of them helps save the incentive cost.

Consider first perturbation (a). Specifically, let $\langle A_n(\epsilon), \pi_n, z_n(\epsilon)\rangle_{n=1}^N$ be tuple associated with the monitoring technology after perturbation (a). Straightforward algebra shows the following:

$$\begin{cases} z_j'(0) = \dfrac{s\,(z''-z')}{\pi_j}, \\ z_k'(0) = -\dfrac{s\,(z''-z')}{\pi_k}, \\ z_n'(0) = 0 \ \forall n \neq j,k. \end{cases}$$

Let $\langle w_n(\epsilon)\rangle_{n=1}^N$ be any wage profile (which clearly exists) such that $w_1(\epsilon) = w_1(0) = 0$ and the agent's (IC) constraint binds after the perturbation, i.e.,

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

Differentiating both sides of the above equation with respect to ϵ and multiplying the result by λ (the Lagrange multiplier of the (IC) constraint prior to the perturbation), we obtain:

$$\sum_{n=1}^{N} \pi_n \cdot u'\left(w_n^*\right) \cdot \lambda z_n \cdot w_n'(0) = -\lambda \left[u\left(w_j^*\right) \cdot \pi_j z_j'(0) + u\left(w_k^*\right) \cdot \pi_k z_k'(0)\right]$$
$$= s\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(\lambda z'' - \lambda z'\right). \tag{A.1}$$

Meanwhile, since $u'\left(w_n^*\right) = \frac{1}{\lambda z_n}$ for all $n \geq 2$ and $w_1'(0) = 0$, it follows that

$$u'\left(w_n^*\right) \cdot \lambda z_n \cdot w_n'(0) = w_n'(0) \text{ for all } n = 1, \cdots, N.$$

Plugging in this result into (A.1) yields

$$\sum_{n=1}^{N} \pi_n w_n'(0) = s\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(\lambda z'' - \lambda z'\right), \tag{A.2}$$

where the left-hand side represents the marginal effect of our perturbation on the incentive cost.

Now consider perturbation (b). Similar algebraic manipulation yields

$$\sum_{n=1}^{N} \pi_n w_n'(0) = -(1-s)\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(\lambda z'' - \lambda z'\right). \tag{A.3}$$

Since $u\left(w_j^*\right) \neq u\left(w_k^*\right)$ by Lemma 3, it follows that (A.2) and (A.3) have the opposite signs. Thus for one of the above described perturbations, we can construct a wage profile that incurs a lower implementation cost than the optimal incentive contract does. This leads to a contradiction. □

A.2 Proof of Theorem 2

Throughout this section, let $\langle A_n, \pi_n, \mathbf{z}_n, \mathbf{w}_n \rangle_{n=1}^N$ denote the tuple associated with any arbitrary incentive contract $\langle \mathcal{P}, \mathbf{w}(\cdot) \rangle$, where $\pi_n = P_{\mathbf{1}}(A_n)$, $\mathbf{z}_n = \mathbf{z}(A_n)$ and $\mathbf{w}_n = \mathbf{w}(A_n)$ for $n = 1, \cdots, N$. In the case $\mathbf{w}(\cdot) = \mathbf{w}^*(\cdot; \mathcal{P})$, write $\mathbf{w}_n^*$ rather than $\mathbf{w}_n$ and drop the dependence on $\mathcal{P}$ for convenience.

We first introduce a useful lemma:

Lemma 4. *For any optimal incentive contract, (i) there exist* $\lambda_1, \lambda_2 > 0$ *such that*

$$u_i'\left(w_{i,n}^*\right) = \frac{1}{\max\left\{\lambda_i z_{i,n}, \underline{z}\right\}}$$

for $i = 1, 2$ *and* $n = 1, \cdots, N$*; (ii)* $\mathbf{w}_j^* \neq \mathbf{w}_k^*$ *for all* $j \neq k$*.*

Proof. Take $\mathcal{P}^*$ as given. Let λ_i and $\eta_{i,n}$ denote the Lagrange multiplier associated with the (IC_i) constraint and the (LL_i) constraint at $w_{i,n}$, respectively, and reduce the principal's problem to the following:

$$\min_{\{w_{i,n}\}} \sum_{i=1}^{2}\sum_{n=1}^{N} \pi_n w_{i,n} - \sum_{i=1}^{2} \lambda_i \left[\sum_{n=1}^{N} \pi_n u_i(w_{i,n}) z_{i,n} - c_i\right] - \sum_{i=1}^{2}\sum_{n=1}^{N} \eta_{i,n} w_{i,n}.$$

Taking derivative with respect to $w_{i,n}$ and setting the result equal to zero yields the first-order condition described in Part (i). The remainder of the proof follows that of Lemma 3. □

Proof of Theorem 2:

Proof. Take any $\langle \mathcal{P}^*, \mathbf{w}^*(\cdot)\rangle$ with $\langle A_n, \pi_n, \mathbf{z}_n, \mathbf{w}_n^*\rangle_{n=1}^N$ being the associated tuple. In case some $A_j \in \mathcal{P}^*$ is not $\mathbf{Z}$-convex, there exist $A', A'', \tilde{A} \in \Sigma$ such that

(i) $P_{\mathbf{1}}(A') = P_{\mathbf{1}}(A'') = P_{\mathbf{1}}(\tilde{A}) = \epsilon$ for some small $\epsilon > 0$;

(ii) $A', A'' \subset A_j$ and $\tilde{A} \subset A_k$ for some $k \neq j$;

(iii) $\tilde{\mathbf{z}} = (1-s)\cdot \mathbf{z}' + s \cdot \mathbf{z}''$ for some $s \in (0,1)$, where $\mathbf{z}' = \mathbf{z}(A')$, $\mathbf{z}'' = \mathbf{z}(A'')$ and $\tilde{\mathbf{z}} = \mathbf{z}(\tilde{A})$.

Consider first the perturbation that swaps A' and $\tilde{A}$. Let $\langle A_n(\epsilon), \pi_n, \mathbf{z}_n(\epsilon)\rangle_{n=1}^N$ be the tuple associated with the monitoring technology after the perturbation. Straightforward algebra shows that

$$\begin{cases} \mathbf{z}_j'(0) = \dfrac{s\cdot(\mathbf{z}''-\mathbf{z}')}{\pi_j}, \\ \mathbf{z}_k'(0) = -\dfrac{s\cdot(\mathbf{z}''-\mathbf{z}')}{\pi_k}, \\ \mathbf{z}_n'(0) = \mathbf{0} \ \forall n \neq j, k. \end{cases}$$

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 (which clearly exists) such that (i) $w_{i,n}(\epsilon) = 0$ for all $i = 1, 2$ and $n \in \mathcal{B}_i^c$, and (ii) both agents' incentive compatibility constraints bind after the perturbation, i.e., for $i = 1, 2$,

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

Differentiating both sides of the above equation with respect to ϵ and multiplying the result by λ_i (the Lagrange multiplier associated with the (IC$_i$) constraint prior to the perturbation), we obtain:

$$\begin{aligned} &\sum_{n=1}^N \pi_n \cdot u_i'\left(w^*_{i,n}\right) \cdot \lambda_i z_{i,j} \cdot w'_{i,n}(0) \\ =& -\lambda_i \left[u_i\left(w^*_{i,j}\right) \cdot \pi_j z'_{i,j}(0) + u_i\left(w^*_{i,k}\right) \cdot \pi_k z'_{i,k}(0)\right] \;\; i = 1, 2. \end{aligned} \tag{A.4}$$

Meanwhile, since $u_i'\left(w^*_{i,n}\right) = \frac{1}{\lambda_i z_{i,n}}$ for all $n \in \mathcal{B}_i$ and $w'_{i,n}(0) = 0$ for all $n \in \mathcal{B}_i^c$, it follows that

$$u_i'\left(w^*_{i,n}\right) \cdot \lambda_i z_{i,n}(0) \cdot w'_{i,n}(0) = w'_{i,n}(0) \;\forall n, i.$$

Plugging this result into (A.4) and summing up the results over i yields

$$\sum_{i=1}^2 \sum_{n=1}^N \pi_n w'_{i,n}(0) = s \cdot \left(\mathbf{u}^*_k - \mathbf{u}^*_j\right)^\top \cdot \Lambda \cdot \left(\mathbf{z}'' - \mathbf{z}'\right), \tag{A.5}$$

where $\mathbf{u}^*_n = \left(u_1\left(w^*_{1,n}\right), u_2\left(w^*_{2,n}\right)\right)^\top$ for $n = k, j$ and $\Lambda = \left(\begin{smallmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{smallmatrix}\right)$.

Now consider the perturbation that swaps A'' and $\tilde{A}$. Similar algebraic manipulation yields

$$\sum_{i=1}^2 \sum_{n=1}^N \pi_n w'_{i,n}(0) = -(1-s) \cdot \left(\mathbf{u}^*_k - \mathbf{u}^*_j\right)^\top \cdot \Lambda \cdot \left(\mathbf{z}'' - \mathbf{z}'\right). \tag{A.6}$$

Since $\mathbf{u}^*_k - \mathbf{u}^*_j \neq \mathbf{0}$ by Lemma 4, there are two cases to consider:

(i) If $\left(\mathbf{u}^*_k - \mathbf{u}^*_j\right)^\top \cdot \Lambda \cdot \left(\mathbf{z}'' - \mathbf{z}'\right) \neq 0$, then sgn (A.5) $\neq$ sgn (A.6), and the remainder of the proof follows that of Theorem 1.

(ii) If $\left(\mathbf{u}^*_k - \mathbf{u}^*_j\right)^\top \cdot \Lambda \cdot \left(\mathbf{z}'' - \mathbf{z}'\right) = 0$, then there exist $B' \subset A'$, $B'' \subset A''$ and $\tilde{B} \subset \tilde{A}$ where $P_{\mathbf{1}}(B') = P_{\mathbf{1}}(B'') = P_{\mathbf{1}}(\tilde{B}) < \epsilon$, $\mathbf{z}(\tilde{B}) = (1-s') \cdot \mathbf{z}(B') + s' \cdot \mathbf{z}(B'')$ for some $s' \in (0,1)$ and $\left(\mathbf{u}_k^* - \mathbf{u}_j^*\right)^\top \cdot \Lambda \cdot (\mathbf{z}(B'') - \mathbf{z}(B')) \neq 0$. Replacing A', A'' and $\tilde{A}$ with B', B'' and $\tilde{B}$ in the above argument leads to the desired result.

□

A.3 Proof of Theorem 3

Throughout this section, let $\langle A_n, \pi_n, (z_{a,n})_{a\in\mathcal{D}}, w_n\rangle_{n=1}^N$ be the tuple associated with an arbitrary incentive contract $\langle \mathcal{P}, w(\cdot)\rangle$, where $\pi_n = P_{a^*}(A_n)$, $z_{a,n} = z_a(A_n)$ and $w_n = w(A_n)$ for $n = 1, \cdots, N$. Take any profile of non-negative real numbers $\boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{D}|}$ and define $z_{\boldsymbol{\lambda},n} = \sum_{a\in\mathcal{D}} \lambda_a \cdot z_a(A_n)$ for $n = 1, \cdots, N$. In the case $w(\cdot) = w^*(\cdot;\mathcal{P})$, write w_n^* rather than w_n and drop the dependence on $\mathcal{P}$ for convenience.

We begin by introducing a useful lemma:

Lemma 5. *For any optimal incentive contract, (i) there exists $\boldsymbol{\lambda}^* \in \mathbb{R}_+^{|\mathcal{D}|}$ such that*

$$u'(w_n^*) = \frac{1}{\max\{z_{\boldsymbol{\lambda}^*,n}, \underline{z}\}}$$

for $n = 1, \cdots, N$ and $z_{\boldsymbol{\lambda}^*,1} < \cdots < z_{\boldsymbol{\lambda}^*,N}$; (ii) $0 = w_1^* < w_2^* < \cdots < w_N^*$.

Proof. Take $\mathcal{P}^*$ as given. Let λ_a^* and η_n denote the Lagrange multiplier associated with the (IC_a) constraint and the (LL) constraint at w_n, respectively, rewrite the principal's problem as follows:

$$\min_{\{w_n\}} \sum_{n=1}^N \pi_n w_n - \sum_{a\in\mathcal{D}} \lambda_a^* \left[\sum_{n=1}^N \pi_n u(w_n) z_{a,n} - c(a^*) + c(a)\right] - \sum_{n=1}^N \eta_n w_n.$$

Taking derivative with respect to w_n and the remainder of the proof follows that of Lemmas 1 and 3. □

We now finalize the proof of Theorem 3:

Proof. Take any $\langle \mathcal{P}^*, w^*(\cdot)\rangle$ with $\langle A_n, \pi_n, (z_{a,n}), w_n^*\rangle_{n=1}^N$ being the associated tuple and $\boldsymbol{\lambda}^*$ the Lagrange multipliers of the incentive compatibility constraints. In the case some $A_j \in \mathcal{P}^*$ is not $Z_{\boldsymbol{\lambda}^*}$-convex, there exist $A', A'', \tilde{A} \in \Sigma$ such that

(i) $P_{a^*}(A') = P_{a^*}(A'') = P_{a^*}(\tilde{A}) = \epsilon$ for some small $\epsilon > 0$;

(ii) $A', A'' \subset A_j$ and $\tilde{A} \subset A_k \in \mathcal{P}^*$ for some $k \neq j$;

(iii) $\tilde{z}_{\boldsymbol{\lambda}^*} = (1-s) \cdot z'_{\boldsymbol{\lambda}^*} + s \cdot z''_{\boldsymbol{\lambda}^*}$, where $z'_{\boldsymbol{\lambda}^*} = z_{\boldsymbol{\lambda}^*}(A')$, $z''_{\boldsymbol{\lambda}^*} = z_{\boldsymbol{\lambda}^*}(A'')$ and $\tilde{z}_{\boldsymbol{\lambda}^*} = z_{\boldsymbol{\lambda}^*}(\tilde{A})$.

Consider first the perturbation that swaps A' and $\tilde{A}$. Let $\langle A_n(\epsilon), \pi_n, (z_{a,n}(\epsilon)) \rangle_{n=1}^N$ be the tuple associated with the monitoring technology after the perturbation. Straightforward algebra shows that for each $a \in \mathcal{D}$,

$$\begin{cases} z_{a,j}(\epsilon) = \dfrac{s\left(z''_a - z'_a\right)}{\pi_j} \cdot \epsilon + \mathcal{O}(\epsilon^2), \\ z_{a,k}(\epsilon) = -\dfrac{s\left(z''_a - z'_a\right)}{\pi_k} \cdot \epsilon + \mathcal{O}(\epsilon^2), \\ z_{a,n}(\epsilon) = z_{a,n} \ \forall n \neq j, k. \end{cases}$$

Take any wage profile $\langle w_n(\epsilon) \rangle_{n=1}^N$ (which clearly exists) such that $w_1(\epsilon) = w_1(0) = 0$ and

$$\sum_{n=1}^N \pi_n u(w_n(\epsilon)) z_{a,n}(\epsilon) \geq c\left(a^*\right) - c(a)$$

for all $a \in \mathcal{D}$. Multiplying the above inequalities by their respective Lagrange multipliers prior to the perturbation and summing up the results,

$$\sum_{n=1}^N \pi_n u(w_n(\epsilon)) z_{\boldsymbol{\lambda}^*,n}\left(\epsilon\right) \geq \sum_{a \in \mathcal{D}} \lambda_a^* \left(c\left(a^*\right) - c(a)\right).$$

Differentiating both sides of this inequality with respect to ϵ,

$$\begin{aligned} &\epsilon \cdot \left[\sum_{n=1}^N \pi_n \cdot u'\left(w_n^*\right) \cdot z_{\boldsymbol{\lambda}^*,n} \cdot w_n'(0) + s\left[u\left(w_j^*\right) - u\left(w_k^*\right)\right]\left(z''_{\boldsymbol{\lambda}^*} - z'_{\boldsymbol{\lambda}^*}\right)\right] \\ &+ \mathcal{O}\left(\epsilon^2\right) \geq 0. \end{aligned} \tag{A.7}$$

Meanwhile, since $u'\left(w_n^*\right) = \frac{1}{z_{\boldsymbol{\lambda}^*,n}}$ for all $n \geq 2$ and $w_1'(0) = 0$, it follows that

$$u'\left(w_n^*\right) \cdot z_{\boldsymbol{\lambda}^*,n} \cdot w_n'(0) = w_n'(0) \ \forall n \in \mathbb{N}.$$

Plugging this result into the above inequality,

$$\epsilon \cdot \sum_{n=1}^N \pi_n w_n'(0) \geq s\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(z''_{\boldsymbol{\lambda}^*} - z'_{\boldsymbol{\lambda}^*}\right) \cdot \epsilon + \mathcal{O}\left(\epsilon^2\right). \tag{A.8}$$

Now consider the perturbation that swaps A'' and $\tilde{A}$. Similar algebraic manipulation shows that

$$\epsilon \cdot \sum_{n=1}^{N} \pi_n w_n'(0) \geq -(1-s)\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(z''_{\boldsymbol{\lambda}^*} - z'_{\boldsymbol{\lambda}^*}\right) \cdot \epsilon + \mathcal{O}\left(\epsilon^2\right). \tag{A.9}$$

Since $s\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(z''_{\boldsymbol{\lambda}^*} - z'_{\boldsymbol{\lambda}^*}\right)$ and $-(1-s)\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(z''_{\boldsymbol{\lambda}^*} - z'_{\boldsymbol{\lambda}^*}\right)$ have the opposite signs, we conclude that for one of the above described perturbations, we can construct a wage profile that saves the implementation cost. This leads to a contradiction. □

A.4 Existence of Optimal Incentive Contract

Assumption 5. *h takes one of the following forms:*

(a) $h(\boldsymbol{\pi}(\mathcal{P}, a)) = f(|\mathcal{P}|)$ *where* $f : \mathbb{N} \to \mathbb{R}$ *is increasing and is unbounded above;*

(b) h *is continuous on* Δ^K *for some* $K \geq 2$ *and is undefined otherwise.*

A.4.1 Baseline model

Assumption 6. $Z(\Omega)$ *is compact.*

Theorem 4. *Assume Assumption 1, 2, 5 and 6. Then Problem (2.1) admits solutions.*

Proof. By Theorem 1, any $\mathcal{P}^*$ with at most K cells is determined by $K-1$ cutoff z-values $\widehat{z}_1, \cdots, \widehat{z}_{K-1}$ where $\inf Z(\Omega) \leq \widehat{z}_1 \leq \cdots \leq \widehat{z}_{K-1} \leq \sup Z(\Omega)$. Let $\widehat{\mathbf{z}} = (\widehat{z}_1, \cdots, \widehat{z}_{K-1})$. Define

$$\mathbb{Z}_K = \{\widehat{\mathbf{z}} : \inf Z(\Omega) \leq \widehat{z}_1 \leq \cdots \leq \widehat{z}_{K-1} \leq \sup Z(\Omega)\}$$

and equip $\mathbb{Z}_K$ with the sup-norm $\|\cdot\|$. $\mathbb{Z}_K$ is compact by Assumption 6. Denote by $W(\widehat{\mathbf{z}})$ the minimum incentive cost when the cutoff z-values are given by $\widehat{\mathbf{z}}$. Notice that $W(\widehat{\mathbf{z}}) \in (0, +\infty)$ for all $\widehat{\mathbf{z}}$ where $\inf Z(\Omega) < \widehat{z}_n < \sup Z(\Omega)$ for some n.

We proceed in several steps:

Step 1. Show that $W(\widehat{\mathbf{z}})$ is continuous in $\widehat{\mathbf{z}}$. Specifically, take any $\widehat{\mathbf{z}}, \widehat{\mathbf{z}}' \in \mathbb{Z}_K$ where $\|\widehat{\mathbf{z}}' - \widehat{\mathbf{z}}\| < \delta$ for some small $\delta > 0$. Let π_n' and z_n' denote the probability measure and the z-value of $A_n' = \left\{\omega : Z(\omega) \in [\widehat{z}_{n-1}', \widehat{z}_n']\right\}$, respectively, and let w_n^* be the optimal wage at $A_n = \{\omega : Z(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n]\}$ when the cutoff z-values are given by $\widehat{\mathbf{z}}$. Take any $\epsilon > 0$ and consider the wage profile that pays $w_n^* + \epsilon$ at every A_n'. By Assumptions 2 and 6, this wage profile satisfies the (LL) constraint, and it satisfies the (IC) constraint when δ is small:

$$\lim_{\delta \to 0} \sum_{n=1}^{K} u(w_n^* + \epsilon)\widehat{\pi}_n' \widehat{z}_n' = \sum_{n=1}^{K} u(w_n^* + \epsilon)\widehat{\pi}_n \widehat{z}_n > c.$$

In addition, it enables us to bound the difference between $W\left(\widehat{\mathbf{z}}'\right)$ and $W\left(\widehat{\mathbf{z}}\right)$ as follows when δ is small:

$$W\left(\widehat{\mathbf{z}}'\right) - W\left(\widehat{\mathbf{z}}\right) \leq \sum_{n=1}^{K} (w_n^* + \epsilon)\pi_n' - \sum_{n=1}^{K} w_n^* \pi_n \sim \mathcal{O}(\epsilon).$$

Interchanging the roles between $\widehat{\mathbf{z}}$ and $\widehat{\mathbf{z}}'$ in the above derivation yields $W\left(\widehat{\mathbf{z}}\right) - W\left(\widehat{\mathbf{z}}'\right) \sim \mathcal{O}(\epsilon)$ as $\delta \to 0$. Putting these results together, we see that $\left|W\left(\widehat{\mathbf{z}}'\right) - W\left(\widehat{\mathbf{z}}\right)\right| \sim \mathcal{O}(\epsilon)$ as $\delta \to 0$.

Step 2. In Case (a), Problem (2.1) becomes

$$\min_{K \in \mathbb{N}, \widehat{\mathbf{z}} \in \mathbb{Z}_K} W(\widehat{\mathbf{z}}) + \mu \cdot f(K).$$

In Case (b), rewrite Problem (2.1) as follows:

$$\min_{\widehat{\mathbf{z}} \in \mathbb{Z}_K} W(\widehat{\mathbf{z}}) + \mu \cdot h(\widehat{\mathbf{z}}),$$

and notice that $h(\widehat{\mathbf{z}})$ is continuous in $\widehat{\mathbf{z}}$. By Assumptions 5 and 6, the problem admits solutions in both cases.

□

A.4.2 Multiple agents

Assumption 7. $\mathbf{Z}(\Omega)$ *is compact and* $\dim \mathbf{Z}(\Omega) = 2$.

Theorem 5. *Assume Assumptions 1, 3, 5 and 7. Then Problem (4.1) admits solutions.*

Proof. By Theorem 2, any $\mathcal{P}^*$ with at most K cells is determined by at most $N(K)$ line segments $b = (\mathbf{z}, \mathbf{z}')$ with end points $\mathbf{z}, \mathbf{z}' \in \mathbf{Z}(\Omega)$. Let $\mathbf{b} = (b_1, \cdots, b_{N(K)})$. Define

$$\mathcal{B}_{N(K)} = \{\mathbf{b} : \mathbf{b} \text{ partitions } \mathbf{Z}(\Omega) \text{ into at most } K \text{ convex polygons}\}$$

and equip this set with the sup-norm. $\mathcal{B}_{N(K)}$ is clearly compact. Denote by $W_i(\mathbf{b})$ the minimum incentive cost incurred to agent i when the partition of $\mathbf{Z}(\Omega)$ is formed by $\mathbf{b}$. Notice that $W_i(\mathbf{b}) \in (0, +\infty)$ if and only if $z_i(A) \neq z_i(A')$ for some $A, A' \in \mathcal{P}^*$.

We proceed in several steps:

Step 1. Show that $W_i(\mathbf{b})$ is continuous in $\mathbf{b}$. Specifically, take any $\mathbf{b}, \mathbf{b}' \in \mathcal{B}_{N(K)}$ where $\|\mathbf{b}' - \mathbf{b}\| < \delta$ for some small $\delta > 0$, and write the partitions formed by $\mathbf{b}$ and $\mathbf{b}'$ as $\{A_1, A_2, \cdots\}$ and $\{A'_1, A'_2, \cdots\}$, respectively. Use $\mathbf{b}(A)$ to denote the boundaries of a typical cell A of these partitions. Numbering cells in such a way that for each n and $b \in \mathbf{b}(A_n)$, there exists $b' \in \mathbf{b}(A'_n)$ such that $\|b - b'\| < \delta$.

Let w_n^* denote the optimal wage at A_n when the partition is formed by $\mathbf{b}$. Take any $\epsilon > 0$ and consider the wage profile that pays $w_n^* + \epsilon$ at every A'_n. Repeating the argument for Theorem 4 shows that this wage profile satisfies (LL_i) and (IC_i) when δ is small, and it gives us the following bound: $|W_i(\mathbf{b}') - W_i(\mathbf{b})| \sim \mathcal{O}(\epsilon)$, as $\delta \to 0$ under Assumptions 3 and 7.

Step 2. In Case (a), Problem (2.1) becomes

$$\min_{K \in \mathbb{N}, \mathbf{b} \in \mathcal{B}_{N(K)}} W_1(\mathbf{b}) + W_2(\mathbf{b}) + \mu \cdot f(N(K)).$$

In Case (b), rewrite Problem (2.1) as follows:

$$\min_{\mathbf{b} \in \mathcal{B}_{N(K)}} W_1(\mathbf{b}) + W_2(\mathbf{b}) + \mu \cdot h(\mathbf{b}),$$

and notice that $h(\mathbf{b})$ is continuous in $\mathbf{b}$. By Assumptions 5 and 7, the problem admits solutions in both cases.

□

A.4.3 Multiple actions

Theorem 6. *Assume Assumptions 1, 2, 5 and 6 for all $a \in \mathcal{D}$. Then Problem (4.2) admits solutions.*

Proof. Clearly $\mathcal{P}^* = \{\Omega\}$ if $a^* = \arg\min_{a\in\mathcal{A}} c(a)$, so assume $a^* \neq \arg\min_{a\in\mathcal{A}} c(a)$ in the remainder of the analysis. Define

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

where $\|\cdot\|_{|\mathcal{D}|}$ denotes the $|\mathcal{D}|$-dimensional Euclidean norm. For each $\boldsymbol{\lambda} \in \Lambda$, define

$$\mathbb{Z}_{K,\boldsymbol{\lambda}} = \left\{ \widehat{\mathbf{z}} : \sum_{a\in\mathcal{D}} \lambda_a \cdot \inf Z_a(\Omega) \leq \widehat{z}_1 \leq \cdots \leq \widehat{z}_{K-1} \leq \sum_{a\in\mathcal{D}} \lambda_a \cdot \sup Z_a(\Omega) \right\}$$

and equip $\mathbb{Z}_{K,\boldsymbol{\lambda}}$ with the sup-norm $\|\cdot\|$. Let $W(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$ be the minimum incentive cost when A_n's are given by $\{\omega : Z_{\boldsymbol{\lambda}}(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n]\}$, $n = 1, \cdots, K$. Notice that $W(\boldsymbol{\lambda}, \widehat{\mathbf{z}}) \in (0, +\infty)$ for all $(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$ where $\boldsymbol{\lambda} \in \mathbb{R}_{++}^{|\mathcal{D}|}$ and $\sum_{a\in\mathcal{D}} \lambda_a \cdot \inf Z_a(\Omega) < \widehat{z}_n < \sum_{a\in\mathcal{D}} \lambda_a \cdot \sup Z_a(\Omega)$ for some n.

We proceed with the following steps:

Step 1. Show that $W(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$ is continuous in $(\boldsymbol{\lambda}, \widehat{\mathbf{z}})$. Specifically, take any $\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in \Lambda$, $\widehat{\mathbf{z}} \in \mathbb{Z}_{K,\boldsymbol{\lambda}}$ and $\widehat{\mathbf{z}}' \in \mathbb{Z}_{K,\boldsymbol{\lambda}'}$ where $\|\boldsymbol{\lambda}' - \boldsymbol{\lambda}\|_{|\mathcal{D}|}$, $\|\widehat{\mathbf{z}}' - \widehat{\mathbf{z}}\| < \delta$ for some small $\delta > 0$. Let w_n^* denote the optimal wage at A_n when A_n's are given by $\{\omega : Z_{\boldsymbol{\lambda}}(\omega) \in [\widehat{z}_{n-1}, \widehat{z}_n]\}$, $n = 1, \cdots, K$. Take any $\epsilon > 0$ and consider the wage profile that pays $w_n^* + \epsilon$ at every A_n'. Repeating the same argument for Theorem 4 shows that this wage profile satisfies the (LL) constraint and all (IC_a) constraints when δ is small, and it gives us the following bound: $|W(\boldsymbol{\lambda}', \widehat{\mathbf{z}}') - W(\boldsymbol{\lambda}, \widehat{\mathbf{z}})| \sim \mathcal{O}(\epsilon)$, as $\delta \to 0$ if Assumptions 2 and 6 hold true for all $a \in \mathcal{D}$.

Step 2. In Case (a), Problem (4.2) becomes

$$\min_{K\in\mathbb{N}, \boldsymbol{\lambda}\in\Lambda, \widehat{\mathbf{z}}\in\mathbb{Z}_{K,\boldsymbol{\lambda}}} W(\boldsymbol{\lambda}, \widehat{\mathbf{z}}) + \mu \cdot f(K).$$

In Case (b), rewrite Problem (4.2) as

$$\min_{\boldsymbol{\lambda}\in\Lambda,\widehat{\mathbf{z}}\in\mathbb{Z}_{K,\boldsymbol{\lambda}}} W\left(\boldsymbol{\lambda},\widehat{\mathbf{z}}\right)+\mu\cdot h(\boldsymbol{\lambda},\widehat{\mathbf{z}}),$$

where $h(\boldsymbol{\lambda},\widehat{\mathbf{z}})$ is continuous in both arguments. By Assumptions 5 and 6, the problem admits solutions in both cases.

□

B Online Appendix (For Online Publication Only)

B.1 Individual Rationality

In the baseline model, suppose instead that the agent faces an outside option that confers a reservation utility $\underline{u}$ at the contracting stage. In this new setting, the wage scheme is a mapping $w:\mathcal{P}\to\mathbb{R}$, and a contract is individually rational if

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

An optimal incentive contract minimizes the total implementation cost under high effort, subject to (IC) and (IR).

Corollary 2. *Assume Assumption 1. Then any $\mathcal{P}^*$ consists of finite Z-convex cells.*

Proof. To begin with, let us take $\mathcal{P}^*$ as given, and let λ and γ denote the Lagrange multiplier of the (IC) constraint and the (IR) constraint, respectively. Rewrite the principal's problem as follows:

$$\min_{\{w_n\}}\sum_{n=1}^{N}\pi_n w_n-\lambda\left[\sum_{n=1}^{N}\pi_n u(w_n)z_n-c\right]-\gamma\left[\sum_{n=1}^{N}\pi_n u(w_n)-c+\underline{u}\right].$$

Taking derivative with respect to w_n and setting the result equal to zero, we obtain:

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

Plugging this first-order condition into the proof of Lemma 3, we see that $z_1<\cdots<z_N$ and $w_1^*<\cdots<w_N^*$.

Now suppose, to the contrary, that some $A_j \in \mathcal{P}^*$ is not Z-convex. Consider first the perturbation that swaps A' and $\tilde{A}$ in the proof of Theorem 1. Let $\langle z_n(\epsilon)\rangle_{n=1}^N$ denote the z-values of the performance categories after this perturbation. Straightforward algebra shows that

$$\begin{cases} z_j'(0) = \dfrac{s\left(z'' - z'\right)}{\pi_j}, \\ z_k'(0) = -\dfrac{s\left(z'' - z'\right)}{\pi_k}, \\ z_n'(0) = 0, \forall n \neq j, k. \end{cases}$$

Let $\langle w_n(\epsilon)\rangle_{n=1}^N$ be any wage profile that makes both the (IC) constraint and the (IR) constraint bind after the perturbation, i.e.,

$$\lambda \cdot \sum_{n=1}^N \pi_n u(w_n(\epsilon)) z_n(\epsilon) = \lambda \cdot c$$
$$\text{and } \gamma \cdot \sum_{n=1}^N \pi_n u(w_n(\epsilon)) = \gamma \cdot (c + \underline{u}).$$

Differentiating these equations with respect to ϵ and summing up the results,

$$\sum_{n=1}^N \pi_n \cdot u'\left(w_n^*\right) \cdot (\lambda z_n + \gamma) \cdot w_n'(0) = -\lambda\left[u\left(w_j^*\right) \cdot \pi_j z_j'(0) + u\left(w_k^*\right) \cdot \pi_k z_k'(0)\right]. \quad \text{(B.1)}$$

Meanwhile, since $u'\left(w_n^*\right) = \frac{1}{\lambda z_n + \gamma}$ for all $n = 1, \cdots, N$, it follows that

$$u'\left(w_n^*\right) \cdot (\lambda z_n + \gamma) \cdot w_n'(0) = w_n'(0), n = 1, \cdots, N.$$

Plugging this result into (B.1),

$$\sum_{n=1}^N \pi_n w_n'(0) = s\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right]\left(\lambda z'' - \lambda z'\right). \quad \text{(B.2)}$$

Now consider the perturbation that swaps $\tilde{A}$ and A''. Similar algebraic manipulation yields

$$\sum_{n=1}^{N} \pi_n w_n'(0) = -(1-s)\left[u\left(w_k^*\right) - u\left(w_j^*\right)\right](\lambda z'' - \lambda z'). \tag{B.3}$$

Since $u\left(w_j^*\right) \neq u\left(w_k^*\right)$, it follows that sgn (B.2) $\neq$ sgn (B.4), and the remainder of the proof follows that of Theorem 1. □

B.2 Shared Monitoring Cost

In the baseline model, suppose instead that a fraction $\beta \in [0,1]$ of the monitoring cost is borne by the agent.

Corollary 3. *Assume Assumption 1. Then any $\mathcal{P}^*$ consists of finite Z-convex cells $A_1, \cdots, A_N$ if $u\left(w^*\left(A_n\right)\right) + \beta \frac{\partial h}{\partial \pi_n}\left(P_0\left(A_1\right), \cdots, P_0\left(A_N\right)\right)$ differs across $n = 1, \cdots, N$.*

Proof. Rewrite the agent's incentive compatibility constraint as follows:

$$\sum_{n=1}^{N} \pi_n u(w_n) z_n - c \geq \beta\left[h\left(\pi_1, \cdots, \pi_N\right) - h\left(\pi_1 z_1, \cdots, \pi_N z_N\right)\right]. \tag{IC$_\beta$}$$

In case some $A_j \in \mathcal{P}^*$ is not Z-convex, consider first the perturbation that swaps A' and $\tilde{A}$. Let $\langle w_n(\epsilon)\rangle_{n=1}^N$ be any wage profile such that (i) $w_1(\epsilon) = w_1(0) = 0$, and (ii) the (IC$_\beta$) constraint binds after the perturbation, i.e.,

$$\sum_{n=1}^{N} \pi_n u(w_n(\epsilon)) z_n(\epsilon) - c = \beta\left[h\left(\pi_1, \cdots, \pi_N\right) - h\left(\pi_1 z_1(\epsilon), \cdots, \pi_N z_N(\epsilon)\right)\right].$$

Straightforward algebra shows that

$$\sum_{n=1}^{N} \pi_n w_n'(0) = s(v_k - v_j)(\lambda z'' - \lambda z'), \tag{B.4}$$

where

$$v_n = u\left(w_n^*\right) + \beta \frac{\partial h}{\partial \pi_n}(P_0(A_1), \cdots, P_0(A_N)), n = k, j.$$

Consider next the perturbation that swaps $\tilde{A}$ and A''. Similar algebraic manipulation yields

$$\sum_{n=1}^{N} \pi_n w_n'(0) = -(1-s)(v_k - v_j)\left(\lambda z'' - \lambda z'\right). \tag{B.5}$$

By assumption, we have sgn (B.4) $\neq$ sgn (B.5), and the remainder of the proof follows that of Theorem 1. □

B.3 Contingent Monitoring Technology

In the baseline model, suppose instead that players observe the realization $s \in S$ of a signal before the principal implements the monitoring technology that she pre-commits to. In this news setting, an incentive contract is a profile of contingent monitoring technologies and wage schemes $\langle \mathcal{P}(s), w(\cdot; s)\rangle_{s\in S}$, where each $\mathcal{P}(s)$ is a finite partition of Ω, and each $w(\cdot; s)$ is a mapping between $\mathcal{P}(s)$ and non-negative reals. Time evolves as follows:

1. The principal commits to $\langle \mathcal{P}(s), w(\cdot; s)\rangle_{s\in S}$;
2. The agent privately chooses $a \in \{0, 1\}$;
3. Nature draws $\omega \in \Omega$ and $s \in S$;
4. Players observe s and the unique cell $A(\omega; s)$ of $\mathcal{P}(s)$ that contains ω;
5. The principal pays the promised wage $w(A(\omega; s); s)$ to the agent.

For simplicity, suppose that ω and s are independently distributed for any given a, and define $Z : \Omega \to \mathbb{R}$ and the z-value of any $A \in \Sigma$ the same as before. Let S a finite set, and define

$$Z(s) = 1 - \frac{P_0(s)}{P_1(s)}$$

for each $s \in S$.

Assumption 8. *$Z(s) \neq 0$ for all $s \in S$.*

Corollary 4. *Assume Assumptions 1 and 8. Then each $\mathcal{P}^*(s)$ consists of finitely Z-convex cells.*

Proof. For each $s \in S$ and $A_{n,s} \in \mathcal{P}^*(s)$ (the n^{th} cell of partition s), let $z_{n,s}$, $w^*_{n,s}$ and $\pi_{n,s}$ denote the z-value of $A_{n,s}$, the optimal wage at $A_{n,s}$ and the probability measure of $A_{n,s}$ under $a = 1$, respectively. Take $\langle \mathcal{P}^*(s) \rangle_{s \in S}$ as given and rewrite the principal's problem as follows:

$$\min_{\{w_{n,s}\}} \sum_{n,s} P_1(s)\pi_{n,s} w_{n,s} - \lambda \left[\sum_{n,s} P_1(s)\pi_{n,s} u(w_{n,s}) z_{n,s} Z(s) - c \right] - \sum_{n,s} \eta_{n,s} w_{n,s},$$

where λ and $\eta_{n,s}$ denote the Lagrange multiplier of the (IC) constraint and the (LL) constraint at $w_{n,s}$, respectively. Taking derivative with respect to $w_{n,s}$ and setting the result equal to zero,

$$u'\left(w^*_{n,s}\right) = \frac{1}{\max\left\{\lambda z_{n,s} Z(s), \underline{z}\right\}} \;\forall n, s.$$

Plugging this first-order condition into the proof of Lemma 3, we see that $z(A_{1,s}) < z(A_{2,s}) < \cdots$ and $w^*(A_{1,s}) = 0 < w^*(A_{2,s}) < \cdots$ for all $s \in S$.

Now suppose to the contrary that $A_{j,s} \in \mathcal{P}^*(s)$ is not Z-convex for some $j \in \mathbb{N}$ and $s \in S$. In this situation, there exist $A', A'', \tilde{A} \in \Sigma$ where

(i) $P_1(A') = P_1(A'') = P_1(\tilde{A}) = \epsilon$ for some small $\epsilon > 0$;

(ii) $A', A'' \subset A_{j,s}$ and $\tilde{A} \subset A_{k,s} \in \mathcal{P}^*(s)$ for some $k \neq j$;

(iii) $\tilde{z} = (1-t)z' + tz''$ for some $t \in (0,1)$, where $z(A') = z'$, $z(A'') = z''$ and $z(\tilde{A}) = \tilde{z}$.

Consider first the perturbation that swaps A' and $\tilde{A}$ in state s. Let $\langle z_{n,s'}(\epsilon) \rangle_{n \in \mathbb{N}, s' \in S}$ denote the z-values of all performance categories in all states after this perturbation, where straightforward algebra shows that

$$\begin{cases} z'_{j,s}(0) = \dfrac{t\left(z'' - z'\right)}{\pi_{j,s}}, \\ z'_{k,s}(0) = -\dfrac{t\left(z'' - z'\right)}{\pi_{k,s}}, \\ z'_{n,s}(0) = 0 \;\forall n \neq j, k, \\ z'_{n,s'}(0) = 0 \;\forall n \text{ and } s' \neq s. \end{cases}$$

Take any wage profile $\langle w_{n,s'}(\epsilon) \rangle_{n \in \mathbb{N}, s' \in S}$ (which clearly exists) such that $w_{1,s'}(\epsilon) = w_{1,s'}(0) = 0$ for all $s' \in S$ and the incentive compatibility constraint binds after the perturbation, i.e.,

$$\sum_{s',n} \pi_{n,s'} P_1(s') \, u(w_{n,s'}(\epsilon)) \, z_{n,s'}(\epsilon) \, Z(s') = c.$$

Taking derivative with respect to ϵ and multiplying the result by λ (the Lagrange multiplier of the (IC) constraint prior to the perturbation), we obtain:

$$\begin{aligned} &\sum_{s',n} \pi_{n,s'} P_1(s') \cdot u'\left(w^*_{n,s'}\right) \cdot \lambda z_{n,s'} Z(s') \cdot w'_{n,s'}(0) \\ =& -\lambda P_1(s) \, Z(s) \left[u\left(w^*_{j,s}\right) \cdot \pi_{j,s} z'_{j,s}(0) + u\left(w^*_{k,s}\right) \cdot \pi_{k,s} z'_{k,s}(0)\right] \\ =& t P_1(s) \, Z(s) \left[u\left(w^*_{k,s}\right) - u\left(w^*_{j,s}\right)\right] (\lambda z'' - \lambda z'). \end{aligned} \tag{B.6}$$

Meanwhile, since $u'\left(w^*_{n,s'}\right) = \frac{1}{\lambda z_{n,s'} Z(s')}$ for all $n \geq 2$ and $s' \in S$ whereas $w'_{1,s'}(0) = 0$ for all $s' \in S$, it follows that

$$u'\left(w^*_{n,s'}\right) \cdot \lambda z_{n,s'} Z(s') \cdot w'_{n,s'}(0) = w'_{n,s'}(0) \ \ \forall n, s'.$$

Plugging this result into (B.6),

$$\begin{aligned} &\sum_{s',n} \pi_{n,s'} P_1(s') \, w'_{n,s'}(0) \\ =& t P_1(s) \, Z(s) \left[u\left(w^*_{k,s}\right) - u\left(w^*_{j,s}\right)\right] (\lambda z'' - \lambda z'). \end{aligned} \tag{B.7}$$

Consider next the perturbation that swaps A'' and $\tilde{A}$ in state s. Similar algebraic manipulation yields

$$\begin{aligned} &\sum_{s',n} \pi_{n,s'} P_1(s') \, w'_{n,s'}(0) \\ =& -(1-t) P_1(s) \, Z(s) \left[u\left(w^*_{k,s}\right) - u\left(w^*_{j,s}\right)\right] (\lambda z'' - \lambda z'). \end{aligned} \tag{B.8}$$

Since $w^*_{j,s} \neq w^*_{k,s}$ and $Z(s) \neq 0$ (Assumption 8), we have sgn (B.7) $\neq$ sgn (B.8), and the remainder of the proof follows that of Theorem 1. $\square$

B.4 Random Monitoring Technology

Setup An incentive contract $\langle \mathbf{q}, w \rangle$ consists of a monitoring technology $\mathbf{q} : \Omega \to \Delta^N$ and a wage scheme $w : \{1, \cdots, N\} \to \mathbb{R}_+$. The former maps each raw data point ω to a lottery $\mathbf{q}(\omega) = (q_1(\omega), \cdots, q_N(\omega))$ over finite N performance categories and is said to be *partitional* if $\mathbf{q}(\omega)$ is degenerate for all $\omega \in \Omega$, whereas the latter assigns a non-negative wage $w_n \geq 0$ to each performance category n. Time evolves as follows:

1. The principal commits to $\langle \mathbf{q}, w \rangle$;
2. The agent privately chooses $a \in \{0, 1\}$;
3. Nature draws $\omega \in \Omega$ according to P_a;
4. The monitoring technology outputs n with probability $q_n(\omega)$, $n = 1, \cdots, N$;
5. The principal pays the promised wage w_n where $0 = w_1 < \cdots < w_N$.

For each $A \in \Sigma$ and $n = 1, \cdots, N$, define

$$q_n(A) = \int_{\omega \in A} q_n(\omega) dP_1(\omega).$$

For each $n = 1, \cdots, N$, let $\pi_n = q_n(\Omega)$ and define

$$z_n = \int_{\omega \in \Omega} Z(\omega) q_n(\omega) dP_1(\omega) / \pi_n.$$

A contract is incentive compatible if

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

An optimal incentive contract $\langle \mathbf{q}^*, w^* \rangle$ solves

$$\min_{\langle \mathbf{q}, w \rangle} \sum_{n=1}^{N} \pi_n w_n + \mu \cdot H(\mathbf{q}, 1) \text{ s.t. (IC) and (LL)}, \tag{B.9}$$

where $\mu \cdot H(\mathbf{q}, 1)$ represents the monitoring cost under $a = 1$.

Partitional monitoring technology Whether randomization helps save the monitoring cost or not varies from case to case. Below we give an example where focusing on partitional monitoring technologies is w.l.o.g.:

Theorem 7. *If* $H(\mathbf{q}, a) = f(N)$*, then* $\mathbf{q}^*$ *is partitional.*

Proof. Take any $A', A'' \in \Sigma$ where $P_1(A') = P_1(A'') = \epsilon$ and $z(A') = z' < z(A'') = z''$, and write $q_n^*(A') = q_n'$ and $q_n^*(A'') = q_n''$ for $n = 1, \cdots, N$. Suppose to the contrary that $q_1', q_1'' \in (0,1)$. Take any $m \geq 2$ where $q_m' \in (0,1)$, and consider a perturbation where $q_1'(\epsilon) = q_1' + \epsilon$, $q_1''(\epsilon) = q_1'' - \epsilon$, $q_m'(\epsilon) = q_m' - \epsilon$ and $q_m''(\epsilon) = q_m'' + \epsilon$. By construction, this perturbation has no effect on $\pi_1, \cdots, \pi_N$ and hence the monitoring cost.

Let $\{z_n(\epsilon^2)\}_{n=1}^N$ denote the z-values of the performance categories after this perturbation. Straightforward algebra shows that

$$\begin{cases} z_1(\epsilon^2) = z_1(0) - \dfrac{(z'' - z')}{\pi_1} \cdot \epsilon^2 + \mathcal{O}\left(\epsilon^4\right), \\ z_m(\epsilon^2) = z_m(0) + \dfrac{(z'' - z')}{\pi_m} \cdot \epsilon^2 + \mathcal{O}\left(\epsilon^4\right), \\ z_n(\epsilon^2) = z_n(0) \ \forall n \neq 1, m. \end{cases}$$

Plugging this result into the proof sketch of Section 3 leads to the following contradiction:

$$\mathcal{L}(\epsilon^2) = \mathcal{L}(0) - \lambda(0)[u(w_m^*) - u(w_1^*)](z' - z) \cdot \epsilon^2 + \mathcal{O}(\epsilon^4) < \mathcal{L}(0),$$

from which it follows that $q_1(\omega) \in \{0, 1\}$ for all $\omega \in \Omega$. Repeating the argument for $n = 2, \cdots, N-1$ gives the desired result. □

Mutual information We now give an example where randomization helps save the implementation cost. The next assumption is taken from the literature on rational inattention:

Assumption 9. *There exists* $K \geq 2$ *such that*

$$H(\mathbf{q}, a) = \sum_{n=1}^N \int q_n(\omega) \log \frac{q_n(\omega)}{\int q_n(\omega) dP_a(\omega)} dP_a(\omega)$$

if $N \leq K$ *and is undefined otherwise.*

The next theorem shows that in the current setting, the optimal monitoring technology exists and features information aggregation, and the chance that it assigns the agent to high performance categories increases with z:

Theorem 8. *Suppose that u is smooth and that Assumptions 6 and 9 hold. Then $\mathbf{q}^*$ exists, is unique and constitutes a mapping from $Z(\Omega)$ to the interior of some Δ^N, $2 \leq N \leq K$, and $q_n^*(z)/q_m^*(z)$ increases with z for all $n > m$.*

Proof. Fix any $2 \leq N \leq K$ and let $W(\mathbf{q})$ be the minimum incentive cost under $\mathbf{q} : \Omega \to \Delta^N$. Notice that $W(\mathbf{q})$ is linear in $\mathbf{q}(\omega)$ for each ω, because replacing $\mathbf{q}(\omega)$ with a lottery that assigns $\mathbf{q}'$ to ω with probability $1-s$ and $\mathbf{q}''$ to ω with probability s such that $q_n(\omega) = (1-s)q_n' + sq_n''$ for all n has no effect on π_n or z_n, $n = 1, \cdots, N$ and hence the incentive cost. Meanwhile, since $H(\mathbf{q}, 1)$ is convex in $\mathbf{q}(\omega)$ for each ω, it follows that the principal's objective function $W(\mathbf{q}) + \mu \cdot H(\mathbf{q}, 1)$ is convex in $\mathbf{q}(\omega)$ for each ω.

The implication of this result is twofold. First, if there exist $A', A'' \in \Sigma$ where $P_1(A') = P_1(A'') = \epsilon$, $z(A') = z(A'')$ and yet $\mathbf{q}(A') \neq \mathbf{q}(A'')$, then assigning $\frac{1}{2}(\mathbf{q}(A') + \mathbf{q}(A''))$ to all $\omega \in A' \cup A''$ has no effect on the incentive cost but saves the monitoring cost, a contradiction. Thus it is w.l.o.g. to restrict attention to $\mathbf{q} : Z(\Omega) \to \Delta^N$.

Second, the first-order approach is valid, namely the problem can be solved by differentiating the principal's objective function with respect to $\mathbf{q}(z)$ and setting the result equal to zero. The solution denoted by $\mathbf{q}^*$, if exists, is unique and is continuously differentiable, and its range lies in the interior of Δ^N by Assumption 9. In the case $Z(\Omega)$ is compact, it follows that $\mathbf{q}^*$ is equicontinuous in z. Thus the principal's problem can be reformulated as

$$\min_{\substack{\mathbf{q}:Z(\Omega)\to\Delta^N \\ \text{is equicontinuous}}} W(\mathbf{q}) + \mu \cdot H(\mathbf{q}, 1),$$

and the existence of solution follows from Helly's selection theorem.

Finally, tedious algebra yields the following first-order condition with respect to $q_n(z)$:

$$-w_n + \lambda u(w_n) z = \mu \left[\log \frac{q_n(z)}{q_1(z)} - \log \frac{\int q_n(z) dP_1(z)}{\int q_1(z) dP_1(z)} \right], \tag{FOC}$$

where λ denotes the Lagrange multiplier of the (IC) constraint. Rearranging yields

$$\lambda\left[u(w_n) - u(w_m)\right]z - (w_n - w_m) = \mu\left[\log\frac{q_n(z)}{q_m(z)} - \log\frac{\int q_n(z)dP_1(z)}{\int q_m(z)dP_1(z)}\right],$$

from which it follows that $q_n(z)/q_m(z)$ is increasing in z for all $n > m$. □

References

ALCHIAN, A., AND H. DEMSETZ (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62(5), 777-795.

BAIMAN, S., AND J. DEMSKI (1980): "Economically Optimal Performance Evaluation and Control Systems," *Journal of Accounting Research*, 18(S), 184-220.

BANKER, R., AND S. DATAR (1989): "Sensitivity, Precision and Linear Aggregation of Signals for Performance Evaluation," *Journal of Accounting Research*, 27(1), 21-39.

BERGEMANN, D., AND S. MORRIS (2016): "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium," *American Economic Review*, 106(5), 586-591.

BLACKWELL, D. (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265-272.

BLOOM, N., AND J. VAN REENEN (2006): "Measuring and Explaining Management Practices Across Firms and Countries," *Centre for Economic Performance Discussion Paper*, No. 716.

——— (2007): "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics*, 122(4): 1351-1408.

——— (2010): "Why Do Management Practices Differ Across Countries?," *Journal of Economic Perspectives*, 24(1), 203-224.

BLOOM, N., R. SADUN, AND J. VAN REENEN (2012): "Americans Do IT Better: U.S. Multinationals and the Productivity Miracle," *American Economic Review*, 102(1), 167-201.

Blumrosen, L., and M. Feldman (2006): "Implementation with a Bounded Action Space," *Proceedings of the 7th ACM Conference on Electronic Commerce*, 62-71.

Blumrosen, L., N. Nisan, and I. Segal (2007): "Auctions with Severely Bounded Communication," *Journal of Artificial Intelligence Research*, 28, 233-266.

Bracken, D., C. Timmreck, and A. Church (2001): *The Handbook of Multisource Feedback*. San Francisco, CA: Jossey-Bass.

Cover, T. M., and J. A. Thomas (2006): *Elements of Information Theory.* Hoboken, NJ: John Wiley & Sons, 2nd ed.

Crémer J., L. Garicano, and A. Prat (2007): "Language and the Theory of the Firm," *Quarterly Journal of Economics*, 122(1), 373-407.

Dilmé, F. (2017): "Optimal Languages," *Working Paper*, University of Bonn.

Dye, R. (1986): "Optimal Monitoring Policies in Agencies," *The Rand Journal of Economics*, 17(3), 339-350.

Ewenstein, B., B. Hancock, and A. Komm (2016): "Ahead of the Curve: The Future of Performance Management," *McKinsey Quarterly*, May.

Green, J. R. and N. L. Stokey (1983): "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91(3), 349-364.

Grossman, S., and O. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51(1), 7-45.

Groves, P., B. Kayyali, D. Knott and S. Van Kuiken (2013): "The 'Big Data' Revolution in Healthcare," *McKinsey & Company*, January.

Holmstrom, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 10(1), 74-91.

——— (1982): "Moral Hazard in Teams," *The Bell Journal of Economics*, 13(2), 324-340.

Holmstrom, B., and P. Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7(S), 24-52.

Hook, B., A. Jenkins, and M. Foot (2011): *Introducing Human Resource Management*. Pearson, 6th ed.

Jäger, G., L. Metzger, and F. Riedel (2011): "Voronoi Languages: Equilibria in Cheap-Talk Games with High-Dimensional Signals and Few Signals," *Games and Economic Behavior*, 73, 517-537.

Kaplan, E. (2015): "The Spy Who Fired Me: The Human Costs of Workplace Monitoring," *Harper's Magazine*, March.

Kim, S. K. (1995): "Efficiency of an Information System in an Agency Model," *Econometrica*, 63(1), 89-102.

Lazear, E., and S. Rosen (1981): "Rank-Order Tournaments as Optimal Labor Contracts," *Journal of Political Economy*, 89(5), 841-864.

Levin, J. (2003): "Relational Incentive Contracts," *American Economic Review*, 93(3), 835-857.

Martin, D. (2017): "Strategic Pricing with Rational Inattention to Quality," *Games and Economic Behavior*, 104, 131-145.

Matéjka, F., and A. McKay (2012): "Simple Market Equilibria with Rationally Inattentive Consumers," *American Economic Review: Papers and Proceedings*, 102(3), 24-29.

Mookherjee, D. (1984): "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, LI, 433-446.

Murff, H., F. FitzHenry, M. Matheny, K. Kotter, K. Crimin, R. Dittus, A. Rosen, P. Elkin, S. Brown (2011): "Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing," *Journal of American Medical Association*, 306(8), 848-855.

Roberts, G., and M. Pregitzer (2007): "Why Employees Dislike Performance Appraisals," *Regent Global Business Review*, May-June.

Roland, M. (2004): "Linking Physicians' Pay to the Quality of Care: A Major Experiment in the United Kingdom," *New England Journal of Medicine*, 351, 1448-1454.

Saint-Paul, G. (2011): "A "Quantized" Approach to Rational Inattention," *Working Paper*, Toulouse School of Economics.

Shannon, C. (1948): "A Mathematical Theory of Communication," *Bell Labs Technical Journal*, 27(3), 379-423.

Sims, C. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49(1), 317-356.

——— (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665-690.

Singer, N. (2013): "In a Mood? Call Center Agents Can Tell," *New York Times*, October 12.

Sobel, J. (2015): "Broad Terms and Organizational Codes," *Working Paper*, UCSD.

Yang, M. (2015): "Optimality of Debt under Flexible Information Acquisition," *Working Paper*, Fuqua Business School.

"Empowering the Organization: 2014 HR Service Delivery and Technology Survey," *Towers Watson*, June 2014.

"The Big Data Opportunity for HR and Finance," *Harvard Business Review*, April 24, 2016.

"The Data-ification of HR: Startups Bringing Big Data to Recruitment & Talent Management," *CB Insights*, March 2015.