Optimal Incentive Contract with Endogenous Monitoring Technology

Anqi Li∗  Ming Yang†

This Version: December 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 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 performance classification, group evaluation as a best response to high monitoring cost, and dividing resources across the assessments of various tasks according 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, U of Washington and U of Chicago for comments and suggestions. Lin Hu and Jessie Li 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., customer sentiment, voice quality and tone) 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 strokes, media interactions) 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 a human- or machine-operated system that governs the processing and analysis of employee performance data — is part of the 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 results, etc. (Bloom and Van Reenen (2006, 2007), Singer (2013), Kaplan (2015)). Second, the cost associated with data utilization has remained significant despite rapid growth in related areas, suggesting that the choice of the monitoring technology be made based on meticulous cost-benefit analyses.

The current paper formalizes the flexibility and 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 examine the

---

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

implications of these results for workforce management and firms’ internal organizations.

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 too costly to implement, as it requires that call centers pay for the use of sophisticated speech analytics software that consumes significant server space and power.\footnote{With the advent of cloud computing, more companies are paying for software usage rather than buying the software itself.} 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.” In the meantime, improved technologies enable call centers to choose between a range of software products with varying emphasis on the above described performance indicators.\footnote{For example, Beyond Verbal specializes in emotion detection and CallMiner in word spotting.} 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 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 framework builds upon otherwise standard principal-agent 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 raw data into a finite-valued signal, as well as a wage scheme that maps realizations of this signal to different levels of 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 (hereafter, monitoring cost) increases with the quantity of information that the output signal carries. An optimal incentive contract that induces any given level of effort minimizes the sum of expected wage and monitoring 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 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 evaluations from the angle of monitoring cost, which in turn predicts that group evaluation is more effective at saving the monitoring cost than individual evaluation 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 assemble. It is consistent with Bloom and Van Reenen (2006, 2007)’s finding that lack of IT access leads manufacturing companies to choose group evaluation 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 increase the intensity of monitoring each deviation with the Lagrange multiplier of the corresponding incentive constraint. Applying this result to the multi-tasking model of Holmstrom and Milgrom (1991) leads to an important lesson, namely dividing resources across the assessments of various tasks according 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 reward 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 principal can only pay for the reduction of the variance of a Gaussian signal. By contrast, we give the principal considerable freedom to utilize potentially high-dimensional performance data and let her pay the associated cost. Our analysis jointly predicts information aggregation, strict MLRP, likelihood-ratio convex performance classification, group evaluation as a best response to high monitoring cost, and dividing resources across the assessments of various tasks according 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 has tried to rank signals based on their implications for principal-agent relationships. For example, Holmstrom (1979) shows that adding a second signal to the model has no effect on the wage if the likelihood ratio of the first signal is a sufficient statistic for the agent’s effort. Likewise, Grossman and Hart (1983) and Kim (1995) demonstrate that performing mean-preserving spreads to likelihood ratios helps save the incentive cost. The current paper takes the raw data as given and examines a new aspect of employee monitoring: data utilization. Our goal to characterize the optimal monitoring technology subject to agent’s incentive constraints resembles that of the burgeoning literature on information design (see, e.g., Bergemann and Morris (2017) for a thorough survey). However, our focus on moral hazard problems is different. Methodologically, we solve a joint optimization problem through analysis of the endogenous interplay between the monitoring technology and the wage scheme.

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 informa-
tion processing (see Sims (1998), Sims (2003), MacKowiak and Wiederholt (2009) and Woodford (2009) among others), whereas the current work explores the consequence of flexible information processing in strategic environments.

Second, we focus mainly on partitional information structures because in reality, the use of random performance classification schemes, e.g., assigning agents’ performances into different categories based on non-performance factors, is viewed as bias and is warned against by managers based on concerns such as appeals, lawsuits and turnover (see Bracken et al. (2001), Roberts and Pregitzer (2007) and other standard textbooks on HR management).

Finally, our monitoring cost functions 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 have 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 to the mediator. 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 that of Crémer et al. (2007), Jäger et al. (2011), Sobel (2015) and Dilmé (2017). Meanwhile, the results of Blumrosen and Feldman (2006) and Blumrosen et al. (2007) exploit the fact that the mechanism is dominant strategy incentive compatible and agents’ utilities have strict increasing differences in types and allocations.

The remainder of this paper proceeds as follows: Section 2 introduces the baseline model; Section 3 presents the main results; Section 4 investigates several extensions; Section 5 concludes. See Appendix A for omitted proofs and the online appendix for

---

5 Yang (2015) investigates a security design problem where a rationally inattentive buyer can flexibly acquire information about the uncertain fundamental. Other recent efforts to examine the impact of RI for strategic interactions include, but are not limited to Matéjka and McKay (2012), Martin (2017) and Ravid (2017).

6 We defer the analysis of random performance classification schemes to the online appendix.

7 Partition information structure has been long used by economists to model knowledge and information. It is recently considered by Saint-Paul (2011), which justifies the relevance of entropic cost in RI decision problems where the choice variable is a deterministic function of the exogenous state variable.
additional results.

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 \((\Omega, \Sigma, P_a)\), where \( \Omega \) consists of the agent’s raw performance data (e.g., call history), and \( P_a \) is the probability measure on \((\Omega, \Sigma)\) conditional on the effort being of \( a \). For concreteness’ sake, let \( \Omega \) be a finite-dimensional Euclidean space and \( \Sigma \) the Borel sigma-algebra on \( \Omega \). \( P_0 \) and \( P_1 \) are mutually absolutely continuous, and the probability density function \( p_a \) (i.e., \( P_a(A) = \int_A p_a d\omega \) for all \( A \in \Sigma \)) exists and is everywhere positive for each \( a \).

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. In the main body of this paper, we allow \( \mathcal{P} \) to be any finite partition of \( \Omega \) that comprises measurable cells, and use \( w : \mathcal{P} \to \mathbb{R}_+ \) to map each cell \( A \) of \( \mathcal{P} \) to a non-negative wage \( w(A) \geq 0 \). For each raw data point \( \omega \in \Omega \), denote by \( A(\omega) \) the unique performance category that contains \( \omega \) and by \( w(A(\omega)) \) the wage associated with \( A(\omega) \). Time evolves as follows:

1. the principal commits to \( \langle \mathcal{P}, w(\cdot) \rangle \);

\[8^{8}\] In general, a monitoring technology can be any mapping from raw data to lotteries on finite performance categories. For reasons discussed in Section 1.1, we consider degenerate lotteries, or equivalently partitional monitoring technologies, in the main body of this paper. In the online appendix, we demonstrate the robustness of our results to randomization, i.e., assigning the same data point to different performance categories probabilistically.

\[9^{9}\] The online appendix examines the case where the agent faces an individual rationality constraint.
2. the agent privately chooses \( a \in \{0, 1\} \);
3. Nature draws \( \omega \) from \( \Omega \) 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 \( X : \Omega \rightarrow \mathcal{P} \) whose probability mass function under any given level \( a \) of effort is defined by \( P_X(X = A_n | a) = P_a(A_n), n = 1, \cdots, N \). Collecting these probabilities yields \( \pi(\mathcal{P}, a) = (P_a(A_1), \cdots, P_a(A_N)) \). \( X \) 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.

**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 \( H(\mathcal{P}, a) = h(\pi(\mathcal{P}, a)) \) for all \((\mathcal{P}, a)\). Further, the following hold true for all \( N \in \mathbb{N} \) and \((\pi_1, \cdots, \pi_N) \in \Delta^N:\)

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

(b) \( h(\pi_1, \cdots, \pi_N) < h(\pi_1', \pi_1'', \cdots, \pi_N) \) for all \( \pi_1, \pi_1', \pi_1'' > 0 \) such that \( \pi_1 = \pi_1' + \pi_1'' \).
Assumption 1 is inspired by the basic principle of Shannon’s (1948) information theory: the amount of information processed by a channel (here, the monitoring technology) depends only on the frequencies of the messages that it transmits. 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 $\log_2 |P|$, 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, $|P|$ denotes the size of $P$). The first measure is of practical significance, as $|P|$ 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 assumptions are needed to ensure the use of entropic monitoring cost. 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:

$$w(A) \geq 0 \quad \forall A \in P,$$

(LL)

as well as his incentive compatibility constraint. Let $p_0/p_1$ be the likelihood ratio associated with the raw data point. Define a random variable $Z : \Omega \to \mathbb{R}$ by

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

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.

While not being the focus of the current analysis, an alternative interpretation of $H(P,a)$ is the complexity cost associated with the writing and implementation of the incentive contract (see, e.g., Dye (1985), Rubinstein (1998) and Battigalli and Maggi (2002)).

See Cover and Thomas (2006) for a textbook treatment of this subject matter, as well as Saint-Paul (2011) for the relevance of entropic cost in RI decision problems where the choice variable is a deterministic function of the exogenous state variable.
and let $Z(A)$ denote the image of any $A \in \Sigma$ under $Z$. Define the $z$-value of any set $A \in \Sigma$ of positive measure by
\[ z(A) = E[Z \mid A; a = 1], \]
and write the agent’s incentive compatibility constraint as follows:
\[ \sum_{A \in \mathcal{P}: P_1(A) > 0} P_1(A) u(w(A)) z(A) \geq c. \quad (IC) \]

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

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

**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, 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 too 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).

---

\(^{13}\)Notice that $Z(\Omega) \subset (-\infty, 1]$ and that $E[Z \mid a = 1] = 0$. 

---
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 consideration, 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 subsequent 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 prior to conducting 360-degree performance appraisals among employees (Bracken et al. (2001)).

The major role of setup cost is to change the probability space $(\Omega, \Sigma, P_a)$. In the above examples, data collection and maintenance expands $(\Omega, \Sigma, P_a)$, whereas investment in rating skills (e.g., correct biases, reduce errors) affects $P_a$ while keeping $(\Omega, \Sigma)$ 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 our 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.

---

14360-degree performance appraisal, as its name suggests, refers to the use of increasingly automated systems to convert 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.

15The 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 thoroughly utilized for evaluation and compensation purposes) to something significant in the case of pay-for-performance for physicians.
3 Main Results

This section analyzes the optimal incentive contract. Results below hold true except perhaps on a measure-zero set of data points. In particular, we will focus on performance categories of positive measures and ignore those of measure zero (there are at most finitely many such sets). The same disclaimer applies to the remainder of our analysis.

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

The next assumption says that the distribution of $Z$ under high effort has no hole or atom; it is made to ensure some regularity on the principal’s problem:

**Assumption 2.** $Z(\Omega)$ is a connected subset of $\mathbb{R}$ and the distribution of $Z$ under $a = 1$ is atomless.

We now state the main result of this paper:

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

Theorem 1 says that the optimal monitoring technology, just like most real-world performance classification schemes, features *information aggregation*, namely 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.,

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

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

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

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 $u'(w^*(A; \mathcal{P})) = 1/\max\{\lambda z(A), \hat{z}\}$ 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 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:

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

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, 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 $\epsilon$-measure (under $a = 1$) data points with $z$-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

\footnote{The quantifier “if exists” rules out the trivial case where $z(A) = constant$ for all $A \in \mathcal{P}$ and hence high effort cannot be induced.}
the principal’s Lagrangian to the following (ignore the \( LL \) constraint for ease of exposition):

\[
\mathcal{L}(\epsilon) = \sum_n \pi_n w^*_n(\epsilon) - \lambda(\epsilon) \left[ \sum_n \pi_n u(w^*_n(\epsilon))z_n(\epsilon) - c \right],
\]

(3.3)

where \( \pi_n \) is probability measure of \( A_n \) under \( a = 1 \), \( w^*_n(\epsilon) \) the optimal wage associated with \( A_n \) and \( \lambda(\epsilon) \) the Lagrange multiplier associated with the \( IC \) constraint. Notice that the wage and the Lagrange multiplier can depend on \( \epsilon \) whereas the probability measure doesn’t.

Differentiating both sides of Equation (3.3) with respect to \( \epsilon \) and rearranging, we obtain:

\[
\mathcal{L}'(0) = \sum_n \pi_n w^*_n(0) - \lambda'(0) \left[ \sum_n \pi_n u(w^*_n(0))z_n(0) - c \right] \]

\[
\quad - \lambda(0) \left[ \sum_n \pi_n u(w^*_n(0))z_n(0) \right] \]

\[
\quad = - \lambda(0) \sum_n \pi_n u(w^*_n(0))z_n'(0) \]

\[
\quad = \lambda(0)[u(w^*_k(0)) - u(w^*_j(0))](z'' - z'),
\]

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, i.e., 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 stipulates that the performance 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.** \( A \preceq A' \) if \( z(A) \leq z(A') \).
Corollary 1. $X : \Omega \to \mathcal{P}^*$ satisfies strict MLRP w.r.t. $\preceq$, namely for all $A, A' \in \mathcal{P}^*$, $A \preceq A'$ if and only if $z(A) < z(A')$.

While the signal induced by any arbitrary monitoring technology always satisfies the weak MLRP with respect to $\preceq$ (i.e., replacing “$<$” with “$\leq$” in Corollary 1), it violates the strict MLRP with respect to $\preceq$ in the case where a single $z$-value is attained by multiple performance categories. By contrast, the signal induced by the optimal monitoring technology must satisfy the strict MLRP with respect to $\preceq$, 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 4 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 further 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 relegated to Theorem 4 of 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 consequences 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 notice that $\tilde{z}_0 = -1$ and $\tilde{z}_N = 1$. 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

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

16
and
\[ z_n = \int_{\omega \in [\hat{z}_{n-1}, \hat{z}_n]} Z(\omega) dP_1(\omega) / \pi_n = \frac{1}{2} [\hat{z}_n + \hat{z}_{n-1}] \]

for \( n = 1, \cdots, N - 1 \). Differentiating the above formula with respect to \( \hat{z}_1, \cdots, \hat{z}_{N-1} \) and setting the result equal to zero, we obtain:

\[ 2\hat{z}_n = \max\{0, z_{n+1}\} + \max\{0, z_n\} \]

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

**Comparative statics** The parameter \( \mu \) 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.\(^{17}\) To facilitate comparative statics analysis, we denote optimal contracts by \( \langle \mathcal{P}^*(\mu), \mathcal{w}^*(\cdot; \mu) \rangle \) in order to make their dependence on \( \mu \) explicit.

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

Proposition 1 says that as \( \mu \) decreases, the principal pays less wage on average 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, the optimal rating scale is decreasing in \( \mu \). For other monitoring cost functions, we numerically assess the relationship between the optimal rating scale and \( \mu \) and report results in Figure 2.\(^{18}\)

Our predictions are consistent with several strands of empirical facts. For example, access to IT has facilitated the adoption of fine-grained performance classification schemes among manufacturing companies, other things being equal \( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \).\(^{17}\) Demand-side factors that increase the adoption of IT-based HR management systems include tough product market competition, loose labor market regulation, etc. \( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \).\(^{18}\) The ability to perform comparative statics analysis with respect to \( \mu \) distinguishes our theory from alternative explanations for wage compression such as subjective performance evaluation, e.g., Levin (2003).\(^{18}\)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)

\( \text{[Bloom and Van Reenen (2006, 2007, 2010)]} \)
Reenen (2006, 2007, 2010), Bloom et al. (2012). Crowd-sourcing real-time 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 2: $|\mathcal{P}^*|$ at varying $\mu$’s: entropy cost, $u(w) = \sqrt{w}$, $Z \sim U[-1, 1]$, $c = 1$.

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

---

19See 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 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.”
Figure 3: Wage distributions under $\mu = 0.05$ and $\mu = 1.4$: entropy cost, $u(w) = \sqrt{w}$, $Z \sim U[-1, 1]$, $c = 1$.

**Minor extensions** The online appendix investigates several minor extensions, including cases where (i) the agent faces an individual rationality constraint, (ii) the principal observes a signal for free (e.g., plant-level output) before deciding how to utilize the raw performance data, and (iii) the monitoring technology can map raw data to lotteries over finite performance categories.
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 $u_i(0) = 0$, $u'_i > 0$, $u''_i < 0$ and $c_i(1) = c_i > c_i(0) = 0$. Each effort profile $a = a_1a_2$ generates a probability space $(\Omega, \Sigma, P_a)$, where $\Omega$ consist of agents’ raw performance data, and $P_a$ is the probability measure on $(\Omega, \Sigma)$ conditional on the effort profile being of $a$. As before, let $\Omega$ be a finite-dimensional Euclidean space and $\Sigma$ the Borel sigma-algebra on $\Omega$. $P_a$’s are mutually absolutely continuous, and the probability density function $p_a$ is well-defined and is everywhere positive for each $a$.

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 pure effort profile can be solved analogously.

Incentive contract An incentive contract $\langle P, w(\cdot) \rangle$ consists of a monitoring technology $P$ and a wage scheme $w(\cdot)$. As before, $P$ can be any finite partition of $\Omega$ that comprises positive-measure cells, and $w : P \to \mathbb{R}^2_+$ maps each cell $A$ of $P$ to a vector $w(A) = (w_1(A), w_2(A))$ of non-negative wages. For each $\omega \in \Omega$, denote by $A(\omega)$ the unique performance category that contains $\omega$, and by $w(A(\omega))$ the wage vector associated with $A(\omega)$. Time evolves as follows:

1. the principal commits to $\langle P, w(\cdot) \rangle$;
2. agents independently choose $a_i \in \{0, 1\}$, $i = 1, 2$;
3. Nature draws $\omega$ from $\Omega$ according to $P_a$;
4. the monitoring technology outputs $A(\omega)$;
5. the principal pays $w_i(A(\omega))$ to agent $i = 1, 2$.

Write $1 = (1, 1)$. Define $Z = (Z_1, Z_2)$ by

$$Z_i = 1 - \frac{P_{a_1=0,a_2=1}}{p_1} i = 1, 2,$$
and denote by $Z(A)$ the image of any $A \in \Sigma$ under $Z$. For each $i = 1, 2$ and $A \in \Sigma$ of positive measure, define the $z_i$-value of $A$ by

$$z_i(A) = E[Z_i \mid A; a = 1].$$

A contract is incentive compatible for agent $i$ if

$$\sum_{A \in \mathcal{P}} P_1 u_i(w_i(A))(A)z_i(A) \geq c_i, \quad (IC_i)$$

and it satisfies agent $i$’s limited liability constraint if

$$w_i(A) \geq 0, \forall A \in \mathcal{P}. \quad (LL_i)$$

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

$$\min_{\langle \mathcal{P}, w(\cdot) \rangle} \sum_{A \in \mathcal{P}} P_a(A) \sum_{i=1}^2 w_i(A) + \mu \cdot H(\mathcal{P}, 1) \text{ s.t. } (IC_i) \text{ and } (LL_i), i = 1, 2. \quad (4.1)$$

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

**Definition 3.** A set $A \in \Sigma$ is $Z$-convex if for all $\omega', \omega'' \in A$ where $Z(\omega') \neq Z(\omega'')$,

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

The next assumption ensures some regularity on our problem analogously to Assumption 2:

**Assumption 3.** $Z(\Omega)$ is a connected subset of $\mathbb{R}^2$ and the distribution of $Z$ under $a = 1$ is atomless.

The next result extends Theorem 1 to encompass multiple agents:

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

21
Theorem 2 says that any optimal monitoring technology comprises finite $Z$-convex cells, and can in some cases be obtained from first dividing the space of $z$-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 $\epsilon$-measure data points of $z$-values $z'$ and $z''$ across performance categories $j$ and $k$. The resulting impact on the principal’s Lagrangian is

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

where $\hat{u}_n$ is a profile of augmented agent utilities at performance category $n = j, k$. Since $L'(0) \geq 0$, 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 $z$-value and the augmented utility profile. Geometrically, this means that data points are more likely to be assigned to a performance category as their $z$-values move closer in the direction of the augmented utility profile. The result on $Z$-convexity follows, because if a performance category contains data points of extreme but not intermediate $z$-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, from the fact that the boundaries of convex polygons are straight line segments in $Z(\Omega)$, we deduce the following:

- first, we can prove existence of the optimal incentive contract under further regularity conditions (see Theorem 5 of Appendix A.4 for details);
- second, assessing the performances of certain kinds of contracts remains straightforward. For example, any bi-partitional contract represents either a team or a tournament and is parameterized by the intercept and slope of the straight line depicted in Figure 4. Likewise, any contract that evaluates and rewards agents on an individual basis is captured 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.
4.1.1 Application: individual vs. group evaluation

Theorem 2 suggests that the principal use group performance evaluation as a best response to high monitoring cost. To formalize this intuition, we limit attention to the case of technologically independent agents:

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

\[
(\Omega, \Sigma, P_a) = (\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, P_{1,a_1} \times P_{2,a_2}),
\]

where each probability space \((\Omega_i, \Sigma_i, P_{i,a_i})\) satisfies the properties stated at the begin-
ning of Section 3

In the language of contract theory, Assumption 4 rules out any kind of technology linkage (i.e., $\omega_i$ depends on $a_{-i}$) or common productivity shock (i.e., $\omega_1, \omega_2$ are correlated given $a$) between agents.

**Definition 4.** (i) $\mathcal{P}$ is an individual monitoring technology if every $A \in \mathcal{P}$ satisfies $A = A_1 \times A_2$ for some $A_i \in \Sigma_i$, $i = 1, 2$; otherwise it is a group monitoring technology;

(ii) $w : \mathcal{P} \to \mathbb{R}_+^2$ is an individual wage scheme if $\mathcal{P}$ is an individual monitoring technology and $w_i(A_i \times A'_{-i}; \mathcal{P}) = w_i(A_i \times A''_{-i}; \mathcal{P})$ 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;

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

In words, a group incentive contract either conducts group evaluation or combines individual evaluation with group incentive pay. 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 and take any $\mathcal{P}$ for which the optimal wage scheme $w^*(\cdot; \mathcal{P})$ exists. Then $\langle \mathcal{P}, w^*(\cdot; \mathcal{P}) \rangle$ is a group incentive contract if and only if $\mathcal{P}$ is a group monitoring technology.

**Proof.** The “if” direction is by definition. The “only if” direction follows from Holmstrom (1982).

To assess the optimality of group performance evaluation, we divide the cost of 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 evaluation is optimal. Straightforward algebra shows that $I < 1$ when $\mu$ is large in the case where $H(\mathcal{P}, a) = f(|\mathcal{P}|)$ 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 $\mu$'s: entropy cost, $u(w) = \sqrt{w}$, $Z \sim U[-1, 1]$, $c = 1$.

Our 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 assemble.\textsuperscript{20} It enriches the analyses of Holmstrom (1982), Green and Stokey (1983) and Mookherjee (1984), which limit the use of group incentive contract among technologically interdependent agents while ignoring the monitoring cost. 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 group evaluations despite being technologically similar.\textsuperscript{21} Further, group evaluation is most common among companies where the capacity for conducting individual evaluation 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

\textsuperscript{20}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.”

\textsuperscript{21}See the survey questions of Bloom and Van Reenen (2006, 2007) regarding the choice between individual and group evaluations, e.g., “employees are rewarded based on their individual contributions to the company,” and “compensation is based on shift/plant-level outcomes.” The former is regarded as an advanced but expensive managerial practice in the manufacturing sector.
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 customer conversations was too costly, if not impossible. Empowered by cloud-based point of sales (POS) systems, retail and fast food stores are now making more probation decisions based on the individual-level information extracted from employees’ real-time locations and transactions (Kaplan (2015)), e.g., how quickly a cashier fulfills a transaction and how she moves between consecutive sales.

4.2 Multiple Actions

**Setup** In the baseline model, now suppose, instead, that the agent can take one of the finite actions in a set $\mathcal{A}$, and that the principal wants to induce a non-trivial action $a^* \neq \arg \min_{a \in \mathcal{A}} c(a)$. Let $P_a$ be the probability measure over $(\Omega, \Sigma)$ under action $a$, where $P_a$’s are mutually absolutely continuous, and the probability density function $p_a$ is well-defined and is everywhere positive for each $a$.

For each deviant action $a \in \mathcal{D} = \mathcal{A} - \{a^*\}$, define a random variable $Z_a : \Omega \rightarrow \mathbb{R}$ by

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

For each $a \in \mathcal{D}$ and $A \in \Sigma$ of positive measure, define the $z_a$-value of $A$ by

$$z_a(A) = \mathbb{E}[Z_a | A; a^*].$$  

A contract is incentive compatible if the following holds true for all $a \in \mathcal{D}$:

$$\sum_{A \in \mathcal{P}} P_{a^*}(A)u(w(A))z_a(A) \geq c(a^*) - c(a).$$  \hspace{1cm} (IC$_a$)

An optimal contract inducing $a^*$ solves:

$$\min_{(\mathcal{P},w(\cdot))} \sum_{A \in \mathcal{P}} P_{a^*}(A)w(A) + \mu \cdot H(\mathcal{P}, a^*) \text{ s.t. } (IC_a) \text{ for all } a \in \mathcal{D} \text{ and } (LL).$$  \hspace{1cm} (4.2)
**Optimal incentive contract**  Take any profile of non-negative reals \( \lambda = (\lambda_a)_{a \in D} \in \mathbb{R}_{+}^{\lvert D \rvert} \) and define a random variable \( Z_\lambda : \Omega \rightarrow \mathbb{R} \) by

\[
Z_\lambda = \sum_{a \in D} \lambda_a \cdot Z_a.
\]

The next definition generalizes the concept of \( Z \)-convexity:

**Definition 5.** A set \( A \in \Sigma \) is \( Z_\lambda \)-convex if for all \( \omega', \omega'' \in A \) where \( Z_\lambda (\omega') \neq Z_\lambda (\omega'') \),

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

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

**Theorem 3.** Assume Assumption 1. Then for any \( P^* \), there exists \( \lambda^* \in \mathbb{R}_{+}^{\lvert D \rvert} \) with \( \sum_{a \in D} \lambda^*_a > 0 \) such that \( P^* \) comprises \( Z_{\lambda^*} \)-convex cells. Assume, in addition, Assumption 2 for all \( a \in D \). Then there exist \( -\infty \leq \hat{z}_0 < \cdots < \hat{z}_N < +\infty \) such that \( P^* = \{ A_1, \cdots , A_N \} \), where \( A_n = \{ \omega : Z_{\lambda^*} (\omega) \in [\hat{z}_{n-1}, \hat{z}_n) \} \) for \( n = 1, \cdots , N \).

Theorem 3 suggests that we increase the intensity of the monitoring each deviation with the Lagrange multiplier of the corresponding incentive constraint. The proof of this result requires a careful handling of the complementary slackness constraint, but the intuition is straightforward: when \( \lambda^*_a \) is large and hence the agent is tempted to commit deviation \( a \), the principal should focus attention 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 \((\Omega_i, \Sigma_i, P_{i,a_i})\) satisfying the properties described at the beginning of Section 2, where \( \Omega_i \) consists of the agent’s raw performance data on task \( i \), and \( P_{i,a_i} \) is the probability measure on \((\Omega_i, \Sigma_i)\) conditional on the effort spent on task \( i \) being of \( a_i \). The principal does not observe efforts, and her goal is to induce high effort in both tasks.

Write \( a = a_1 a_2 \) and \( \omega = \omega_1 \omega_2 \), and let \( A = \{11, 01, 10, 00\} \), \( a^* = 11 \) and \( D = \)}
For each $i = 1, 2$, define
\[ Z_i = 1 - \frac{p_{i,a_i=0}}{p_{i,a_i=1}}. \]

For each profile $\lambda = (\lambda_{01}, \lambda_{10}, \lambda_{00}) \in \mathbb{R}^{\lvert D \rvert}_+$ of non-negative reals, define
\[
Z_\lambda(\omega) = (\lambda_{01} + \lambda_{00}) \cdot Z_1(\omega_1) + (\lambda_{10} + \lambda_{00}) \cdot Z_2(\omega_2) - \lambda_{00} \cdot Z_1(\omega_1) \cdot Z_2(\omega_2).
\]

A straightforward extension of Theorem 3 shows the following:

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

(i) $\lambda^*$ where $\lambda^*_{01} + \lambda^*_{00}, \lambda^*_{10} + \lambda^*_{00} > 0$,

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

such that $\mathcal{P}^* = \{A_1, \cdots, A_N\}$, where $A_n = \{\omega : Z_\lambda(\omega) \in [\hat{z}_{n-1}, \hat{z}_n)\}$ 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 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 2 suggests a conjugate solution, namely dividing resources across the assessments of various tasks according to 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 $\omega_1$ represent teaching evaluation and $\omega_2$ research output. By Corollary 2, the following ratio:
\[
R = \frac{\lambda^*_{01} + \lambda^*_{00}}{\lambda^*_{10} + \lambda^*_{00}}
\]
captures how sensitive the overall performance grade changes as we vary teaching evaluation rather than research output. 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 significant
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 $\xi_i$ is normally distributed with mean zero and variance $\sigma_i^2$. Under these assumptions, a decrease in $\sigma_i^2$ increases the measurability of $a_i$ in the sense of Blackwell (1953);

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

However, 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 $\sigma_1^2$. 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 technology advances introduce into the design and implementation of monitoring technologies, and characterize the optimal monitoring technology through the tension between incentive cost and 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, and hopefully it provides a starting point for future research such as how review sites like Yelp should disclose information to customers with limited attention spans, and how managers should organize monitoring activities in long-term agency relationships. Second, it will be desirable to enrich our theoretical exercise with empirical quantifications of how technology affects 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 any 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 \( n = 1, \ldots, N \). If \( w(\cdot) \) coincides with the optimal wage scheme under \( \mathcal{P} \), then write \( w_n^* \) rather than \( w_n \) for \( n = 1, \ldots, N \). Assume w.l.o.g. that \( \pi_n > 0 \) for all \( n \) and that \( z_1 \leq \cdots \leq z_N \).

We begin by proving Lemma 1.
Proof. Fix any $\mathcal{P}$ for which the optimal wage scheme exists. Rewrite the principal’s problem as follows:

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

where $\lambda$ denotes the Lagrange multiplier associated with the (IC) constraint and $\eta_n$ the Lagrange multiplier associated with the (LL) constraint at $w_n$. Differentiating the principal’s objective function with 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 first-order condition, it follows that $w_n^* > 0 \iff u'(w_n^*) = 1 / \lambda z_n \iff \lambda z_n > z$, or equivalently $u'(w_n^*) = 1 / \max \{\lambda z_n, z\}$.  

We next introduce a useful lemma:

**Lemma 3.** Assume Assumption 1. Then any optimal incentive contract that induces $a = 1$ 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. Combining these observations with Lemma 1 we see that every performance category should attain a distinct $z$-value, and that only one performance category can attain negative $z$-value. This leads to the conclusion that $z_1 < \cdots < z_N$ and that $0 = w_1^* < \cdots < w_N^*$.  

We now finalize the proof of Theorem 1:

**Proof.** Take any optimal incentive contract $\langle \mathcal{P}^*, w^*(\cdot) \rangle$ 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 $A', A'' \subset A_j$ and $\tilde{A} \subset A_k$ such that

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

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

31
Consider the following perturbations to $P^*$:

(a) Assign $A'$ to $A_k$ and $\tilde{A}$ to $A_j$ (hereafter, “swap” $A'$ and $\tilde{A}$);

(b) Assign $\tilde{A}$ to $A_j$ and $A''$ to $A_k$ (hereafter, “swap” $\tilde{A}$ and $A''$).

By construction, these perturbations do not affect the probability distribution of the output signal under high effort and hence the monitoring cost. We now argue that one of them helps save the incentive cost.

Consider first perturbation (a). Let $\langle A_n(\epsilon), \pi_n, z_n(\epsilon) \rangle_{n=1}^N$ be tuple associated with the monitoring technology after this perturbation. Differentiating $z_n(\epsilon)$ with respect to $\epsilon$ and evaluating the result at $\epsilon = 0$, we obtain:

$$\begin{align*}
    z'_j(0) &= s \left( \frac{z''}{\pi_j} - \frac{z'}{\pi_j} \right), \\
    z'_k(0) &= -s \left( \frac{z''}{\pi_k} - \frac{z'}{\pi_k} \right), \\
    z'_n(0) &= 0 \quad \forall n \neq j, k.
\end{align*}$$

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

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

Such wage profile clearly exists. Differentiating both sides of the above equation with respect to $\epsilon$ and multiplying the result by the Lagrange multiplier $\lambda$ associated with the (IC) constraint prior to the perturbation, we obtain:

$$\sum_{n=1}^N \pi_n \cdot u'(w^*_n) \cdot \lambda z_n \cdot w'_n(0) = -\lambda \left[ u(w^*_j) \cdot \pi_j z'_j(0) + u(w^*_k) \cdot \pi_k z'_k(0) \right]$$

$$= s \left[ u(w^*_k) - u(w^*_j) \right] (\lambda z'' - \lambda z'). \quad (A.1)$$

Meanwhile, since $u'(w^*_n) = \frac{1}{\lambda z_n}$ for all $n \geq 2$ (Lemma [1]) and $w'_1(0) = 0$ ($w_1(\epsilon) = 0 = w^*_1$ by construction), it follows that

$$u'(w^*_n) \cdot \lambda z_n \cdot w'_n(0) = w'_n(0) \quad \text{for all } n = 1, \ldots, N.$$
Plugging in this result into (A.1), we obtain:

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

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

Consider next perturbation (b). Similar algebraic manipulation shows that

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

Since \( u(w_j^*) \neq u(w_k^*) \) by Lemma 3, we conclude 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. \( \square \)

### A.2 Proof of Theorem 2

Throughout this section, let \( \langle A_n, \pi_n, z_n, w_n \rangle_{n=1}^{N} \) denote the tuple associated with any incentive contract \( \langle P, w(\cdot) \rangle \), where \( \pi_n = P_1(A_n) \), \( z_n = z(A_n) \) and \( w_n = w(A_n) \) for \( n = 1, \ldots, N \). Assume w.l.o.g. that \( \pi_n > 0 \) for all \( n \). In case \( w(\cdot) \) coincides with the optimal wage scheme under \( P \), write \( w_n^* \) rather than \( w_n \) for \( n = 1, \ldots, N \).

We first introduce a useful lemma:

**Lemma 4.** For any optimal incentive contract that induces \( a = 1 \), (i) there exist \( \lambda_1, \lambda_2 > 0 \) such that

\[ u_i'(w_{i,n}^*) = \frac{1}{\max \{ \lambda i z_{i,n}, \bar{z} \}} \]

for all \( i = 1, 2 \) and \( n = 1, \ldots, N; \) (ii) \( w_j^* \neq w_k^* \) for all \( j \neq k \).

**Proof.** Take any optimal monitoring technology \( P^* \) as given and rewrite the principal’s problem as follows:

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

33
where $\lambda_i$ denotes the Lagrange multiplier associated with the (IC$_i$) constraint and $\eta_{i,n}$ the Lagrange multiplier associated with the (LL$_i$) constraint at $w_{i,n}$. Differentiating the principal’s objective function with respect to $w_{i,n}$ and setting the result equal to zero, we obtain the first-order condition in Part (i). The remainder of the proof follows that of Lemma 3 and thus is omitted.

Proof of Theorem 2

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

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

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

Consider first the perturbation that swaps $A'$ and $\tilde{A}$. Let $\langle A_n(\epsilon), \pi_n, z_n(\epsilon) \rangle_{n=1}^N$ denote the tuple associated with the monitoring technology after this perturbation. Differentiating $z_n(\epsilon)$ with respect to $\epsilon$ and evaluating the result at $\epsilon = 0$, we obtain:

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

Define $\mathcal{NB}_i = \{n : w^*_{i,n} > 0\}$ as the set of performance categories where agent $i$’s limited liability constraint is non-binding prior to the perturbation. Take any wage profile $\langle w_n(\epsilon) \rangle_{n=1}^N$ such that (i) $w_{i,n}(\epsilon) = 0$ for all $i = 1, 2$ and $n \in \mathcal{NB}^c_i$, and (ii) both agents’ incentive compatibility constraints bind after the perturbation, i.e., for $i = 1, 2$,

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

Such wage profile clearly exists. Differentiating both sides of the above equation with respect to $\epsilon$ and multiplying the result by the Lagrange multiplier $\lambda_i$ associated with
the (IC\(_i\)) constraint prior to the perturbation, we obtain:

\[
\sum_{n=1}^{N} \pi_n \cdot u_i' \left( w_{i,n}^* \right) \cdot \lambda_i z_{i,j} \cdot w_{i,n}'(0) \\
= -\lambda_i \left[ u_i \left( w_{i,j}^* \right) \cdot \pi_j z_{i,j}'(0) + u_i \left( w_{i,k}^* \right) \cdot \pi_k z_{i,k}'(0) \right], \quad i = 1, 2. \tag{A.4}
\]

Meanwhile, since \(u_i' \left( w_{i,n}^* \right) = \frac{1}{\lambda_i z_{i,n}}\) for all \(n \in NB_i\) (Lemma 4) and \(w_{i,n}'(0) = 0\) for all \(n \in NB_i^c\) \((w_{i,n}(\epsilon) = 0 = w_{i,n}^*\) by construction), it follows that

\[
u_i' \left( w_{i,n}^* \right) \cdot \lambda_i z_{i,n}(0) \cdot w_{i,n}'(0) = w_{i,n}'(0) \quad \forall n, i.
\]

Plugging this result into (A.4) and summing up the results over \(i\), we obtain:

\[
\sum_{i=1}^{2} \sum_{n=1}^{N} \pi_n w_{i,n}'(0) = s \cdot \left( u_k^* - u_j^* \right)^\top \cdot \Lambda \cdot \left( z'' - z' \right),\tag{A.5}
\]

where \(u_n^* = \begin{bmatrix} u_1(w_{1,n}^*) \\ u_2(w_{2,n}^*) \end{bmatrix}\) for \(n = k, j\) and \(\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\).

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

\[
\sum_{i=1}^{2} \sum_{n=1}^{N} \pi_n w_{i,n}'(0) = -(1 - s) \cdot \left( u_k^* - u_j^* \right)^\top \cdot \Lambda \cdot \left( z'' - z' \right). \tag{A.6}
\]

Since \(u_k^* - u_j^* \neq 0\) by Lemma \[4\], there are two cases to consider:

(i) If \(\left( u_k^* - u_j^* \right)^\top \cdot \Lambda \cdot \left( z'' - z' \right) \neq 0\), then \(\text{sgn} \ (A.5) \neq \text{sgn} \ (A.6)\), and the remainder of the proof follows that of Theorem \[1\].

(ii) If \(\left( u_k^* - u_j^* \right)^\top \cdot \Lambda \cdot \left( z'' - z' \right) = 0\), then there exist \(B' \subset A', B'' \subset A''\) and \(\tilde{B} \subset \tilde{A}\) such that \(P_1(B') = P_1(B'') = P_1(\tilde{B}) < \epsilon, \ z(\tilde{B}) = (1 - s') \cdot z(B') + s' \cdot z(B'')\) for some \(s' \in (0, 1)\) and \(\left( u_k^* - u_j^* \right)^\top \cdot \Lambda \cdot \left( z(B'') - z(B') \right) \neq 0\). Thus, in the above argument, replacing \(A', A''\) and \(\tilde{A}\) with \(B', B''\) and \(\tilde{B}\) gives the desired result.
A.3 Proof of Theorem 3

Throughout this section, let \( \langle A_n, \pi_n, (z_{a,n})_{a \in D}, w_n \rangle_{n=1}^N \) be the tuple associated with any incentive contract \( \langle P, w(\cdot) \rangle \), where \( \pi_n = \pi_n^*(A_n) \), \( z_{a,n} = z_a(A_n) \) and \( w_n = w(A_n) \) for \( n = 1, \cdots, N \). Assume w.l.o.g. that \( \pi_n > 0 \). If \( w(\cdot) \) coincides with the optimal wage scheme under \( P \), then write \( w^*_n \) rather than \( w_n \) for \( n = 1, \cdots, N \). For any profile of non-negative reals \( \lambda \in \mathbb{R}^{|D|}_+ \) and \( A \in \Sigma \), define \( z_{\lambda}(A) = \sum_{a \in D} \lambda_a \cdot z_a(A) \). Write \( z_{\lambda,n} = z_{\lambda}(A_n) \) for \( n = 1, \cdots, N \).

We begin by introducing a useful lemma:

**Lemma 5.** For any optimal incentive contract that induces \( a^* \), (i) there exists \( \lambda^* \in \mathbb{R}^{|D|}_+ \) with \( \sum_{a \in D} \lambda^*_a > 0 \) such that

\[
\frac{1}{u'(w^*_n)} = \max \left\{ \frac{1}{z_{\lambda^*,n}}, \cdots \right\} \forall n = 1, \cdots, N,
\]

and \( z_{\lambda^*,1} < \cdots < z_{\lambda^*,N} \); (ii) \( 0 = w^*_1 < w^*_2 < \cdots < w^*_N \).

**Proof.** Take any optimal monitoring technology \( P^* \) as given and rewrite the principal’s problem as follows:

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

where \( \lambda^*_a \) denotes the Lagrange multiplier associated with the \( [IC_a] \) constraint and \( \eta_n \) the Lagrange multiplier associated with the \( [LL] \) constraint at \( w_n \). \( \lambda^*_a > 0 \) for some \( a \) because otherwise the agent will deviate to the least costly action. Differentiating the principal’s objective function with respect to \( w_n \) gives the first-order condition in Part (i). The remainder of the proof follows that of Lemmas 1 and 3 and thus is omitted. \( \square \)

We now finalize the proof of Theorem 3.

**Proof.** Take any optimal incentive contract \( \langle P^*, w^*(\cdot) \rangle \) with \( \langle A_n, \pi_n, (z_{a,n})_{a \in D}, w^*_n \rangle_{n=1}^N \) being the associated tuple and \( \lambda^* \) the Lagrange multipliers associated with the incentive compatibility constraints. In case some \( A_j \in P^* \) is not \( Z_{\lambda^*} \)-convex, there exist \( A', A'' \subset A_j \) and \( \tilde{A} \subset A_k \) such that
(i) \( P_a(A') = P_a(A'') = P_a(\bar{A}) = \epsilon \) for some small \( \epsilon > 0 \);

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

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 this perturbation. Differentiating \( z_{a,n}(\epsilon) \) with respect to \( \epsilon \) yields

\[
\begin{align*}
  z_{a,j}(\epsilon) &= \frac{s}{\pi_j} (z''_a - z'_a) \cdot \epsilon + \mathcal{O}(\epsilon^2), \\
  z_{a,k}(\epsilon) &= -\frac{s}{\pi_k} (z''_a - z'_a) \cdot \epsilon + \mathcal{O}(\epsilon^2), \\
  z_{a,n}(\epsilon) &= z_{a,n} \quad \forall \, n \neq j, k,
\end{align*}
\]

where \( z'_a = z_a(A') \) and \( z''_a = z_a(A'') \). Take any wage profile \( \langle w_n(\epsilon) \rangle_{n=1}^N \) such that (i) \( w_1(\epsilon) = w_1^* = 0 \) and (ii) for all \( a \in D \),

\[
\sum_{n=1}^N \pi_n u(w_n(\epsilon)) z_{a,n}(\epsilon) \geq c(a^*) - c(a).
\]

Multiplying the above inequalities by their respective Lagrange multipliers prior to the perturbation and summing up the results, we obtain:

\[
\sum_{n=1}^N \pi_n u(w_n(\epsilon)) z_{a,n}(\epsilon) \geq \sum_{a \in D} \lambda_a^* (c(a^*) - c(a)).
\]

Differentiating both sides of this inequality with respect to \( \epsilon \) yields

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

(A.7)

Meanwhile, since \( u'(w_n^*) = \frac{1}{z_{\lambda^*,n}} \) for all \( n \geq 2 \) (Lemma 5) and \( w_1'(0) = 0 \) (\( w_1(\epsilon) = 0 = w_1^* \) by construction), it follows that

\[
u'(w_n^*) \cdot z_{\lambda^*,n} \cdot w_n'(0) = w_n'(0) \quad \forall \, n.
\]
Plugging this result into the above inequality, we obtain:

\[ \epsilon \cdot \sum_{n=1}^{N} \pi_n w'_n(0) \geq s \left[ u(w^*_k) - u(w^*_j) \right] (z''_{\lambda^*} - z'_{\lambda^*}) \cdot \epsilon + O(\epsilon^2). \quad (A.8) \]

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

\[ \epsilon \cdot \sum_{n=1}^{N} \pi_n w'_n(0) \geq -(1 - s) \left[ u(w^*_k) - u(w^*_j) \right] (z''_{\lambda^*} - z'_{\lambda^*}) \cdot \epsilon + O(\epsilon^2). \quad (A.9) \]

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

\[ \Box \]

### A.4 Existence of Optimal Incentive Contract

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

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

(b) \( h \) is continuous on \( \Delta^K \) for some \( K \geq 2 \).

#### 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 optimal monitoring technology with at most \( K \) cells is associated with \( K - 1 \) cutoffs \( \hat{z}_1, \ldots, \hat{z}_{K-1} \) satisfying \( \min Z(\Omega) \leq \hat{z}_1 \leq \cdots \leq \hat{z}_{K-1} \leq \max Z(\Omega) \). Write \( \hat{Z} = (\hat{z}_1, \ldots, \hat{z}_{K-1}) \). Define

\[ Z_K = \{ \hat{z} : \min Z(\Omega) \leq \hat{z}_1 \leq \cdots \leq \hat{z}_{K-1} \leq \max Z(\Omega) \}, \]

and equip \( Z_K \) with the sup-norm \( \| \cdot \| \). \( Z_K \) is compact by Assumption 6. Denote by \( W(\hat{z}) \) the minimum incentive cost when the cutoffs are given by \( \hat{z} \). Notice that
$W(\hat{z}) \in (0, +\infty)$ as long as $\min Z(\Omega) < \hat{z}_n < \max Z(\Omega)$ for some $n$.

We proceed in several steps:

Step 1. Show that $W(\hat{z})$ is continuous in $\hat{z}$. Take any $\hat{z}, \hat{z}' \in Z_K$ where $\|\hat{z}' - \hat{z}\| < \delta$ for some small $\delta > 0$. Let $\pi_n$ and $z_n$ (resp. $\pi'_n$ and $z'_n$) denote the probability measure and $z$-value of $A_n = \{\omega : Z(\omega) \in [\hat{z}_{n-1}, \hat{z}_n]\}$ (resp. $A'_n = \{\omega : Z(\omega) \in [\hat{z}'_{n-1}, \hat{z}'_n]\}$), and let $w^*_n$ be the optimal wage at $A_n$ when the cutoffs are given by $\hat{z}$.

Take any $\epsilon > 0$ and consider the wage profile that pays $w^*_n + \epsilon$ at every $A'_n$. This wage profile satisfies the (LL) constraint by construction. By Assumptions 2 and 6, it satisfies the (IC) constraint when $\delta$ is small:

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

Finally, it enables us to bound the difference between $W(\hat{z}')$ and $W(\hat{z})$ as follows when $\delta$ is small:

$$W(\hat{z}') - W(\hat{z}) \leq \sum_{n=1}^{K} (w^*_n + \epsilon) \pi'_n - \sum_{n=1}^{K} w^*_n \pi_n \sim O(\epsilon).$$

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

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

$$\min_{K \in \mathbb{N}, \hat{z} \in Z_K} W(\hat{z}) + \mu \cdot f(K).$$

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

$$\min_{\hat{z} \in Z_K} W(\hat{z}) + \mu \cdot h(\hat{z}),$$

and notice that $h(\hat{z})$ is continuous in $\hat{z}$. Then existence of solution follows from Assumption 6.

39
A.4.2 Multiple agents

Assumption 7. \( Z(\Omega) \) is compact and \( \dim Z(\Omega) = 2 \).

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

Proof. By Theorem 2, any optimal monitoring technology with at most \( K \) cells is fully characterized by (1) a finite \( N(K) \) number of vertices in \( z_1, \cdots, z_{N(K)} \in Z(\Omega) \), and (2) an \( N(K) \times N(K) \) adjacency matrix \( M \) whose \( lm \)’th entry equals 1 if \( z_l \) and \( z_m \) are connected by a line segment and 0 otherwise.

Write \( b = (z_1, \cdots, z_{N(K)}) \). For any adjacency matrix \( M \), define

\[
B(M) = \{ b : (b, M) \text{ partitions } Z(\Omega) \text{ into at most } K \text{ convex polygons} \},
\]

and equip \( B(M) \) with the sup-norm \( \| \cdot \| \). \( B(M) \) is clearly compact. Denote by \( W_i(b) \) the minimum wage expenditure incurred to incentivize agent \( i \) when the partition of is given by \( (b, M) \). Notice that \( W_i(b) \in (0, +\infty) \) if and only if \( z_i(A) \neq z_i(A') \) for some cells \( A, A' \) of the partition.

We proceed in several steps:

Step 1. Show that \( W_i(b) \) is continuous in \( b \). Take any \( b, b' \in B(M) \) where \( \| b' - b \| < \delta \) for some small \( \delta > 0 \). Write the partitions formed by \( (b, M) \) and \( (b', M) \) as \( \{ A_1, A_2, \cdots \} \) and \( \{ A'_1, A'_2, \cdots \} \), respectively, where each \( A_n \) is uniquely identified by its boundaries \( Bd(A_n) = \{ lm : M_{lm} = 1 \} \). Numbering cells in such a way that \( Bd(A_n) = Bd(A'_n) \) for all \( n \).

Let \( w^*_{i,n} \) denote the optimal wage at \( A_n \). Take any \( \epsilon > 0 \) and consider the wage profile that pays \( w^*_{i,n} + \epsilon \) at every \( A'_n \). A similar argument as that of Theorem 4 shows that under Assumptions 3 and 7, this wage profile satisfies (LL) and (IC) when \( \delta \) is small. Further, it helps bound \( |W_i(b') - W_i(b)| \sim O(\epsilon) \) as \( \delta \to 0 \).

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

\[
\min_{K \in \mathbb{N}, M, b \in B(M)} W_1(b) + W_2(b) + \mu \cdot f(N(K)).
\]
In Case (b), rewrite Problem (2.1) as follows:

$$\min_{M, b \in B(M)} W_1(b) + W_2(b) + \mu \cdot h(b),$$

and notice that $h(b)$ is continuous in $b$. Then existence of solution follows from Assumption 7.

A.4.3 Multiple actions

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

Proof. Define

$$\Lambda = \left\{ \lambda : \lambda \in \mathbb{R}^{\mid D \mid} \text{ and } \parallel \lambda \parallel_{\mid D \mid} = 1 \right\},$$

where $\parallel \cdot \parallel_{\mid D \mid}$ denotes the $\mid D \mid$-dimensional Euclidean norm. For each $\lambda \in \Lambda$, define

$$Z_K(\lambda) = \left\{ \tilde{z} : \sum_{a \in D} \lambda_a \cdot \min \Omega_a \leq \tilde{z}_1 \leq \cdots \leq \tilde{z}_{K-1} \leq \sum_{a \in D} \lambda_a \cdot \max \Omega_a \right\},$$

and equip $Z_K(\lambda)$ with the sup-norm $\parallel \cdot \parallel$. Denote by $W(\lambda, \tilde{z})$ the minimum incentive cost when the partition is given by $\{\omega : Z_\lambda (\omega) \in [\tilde{z}_{n-1}, \tilde{z}_n]\}$, $n = 1, \cdots, K$. Notice that $W(\lambda, \tilde{z}) \in (0, +\infty)$ as long as $\lambda \in \mathbb{R}^{\mid D \mid}$ and $\sum_{a \in D} \lambda_a \cdot \min \Omega_a < \tilde{z}_n < \sum_{a \in D} \lambda_a \cdot \max \Omega_a$ for some $n$.

We proceed with the following steps:

Step 1. Show that $W(\lambda, \tilde{z})$ is continuous in $(\lambda, \tilde{z})$. Take any $\lambda, \lambda' \in \Lambda$, $\tilde{z}, \tilde{z}' \in Z_K(\lambda')$ where $\parallel \lambda' - \lambda \parallel_{\mid D \mid}, \parallel \tilde{z}' - \tilde{z} \parallel < \delta$ for some small $\delta > 0$. Define $A_n = \{\omega : Z_\lambda (\omega) \in [\tilde{z}_{n-1}, \tilde{z}_n]\}$ and $A'_n = \{\omega : Z_{\lambda'} (\omega) \in [\tilde{z}'_{n-1}, \tilde{z}'_n]\}$ for $n = 1, \cdots, K$. Let $w^*_n$ denote the optimal wage at $A_n$. Take any $\epsilon > 0$ and consider the wage profile that pays $w^*_n + \epsilon$ at every $A'_n$. A similar argument as that of Theorem 4 shows that if Assumptions 2 and 6 hold true for all $a \in D$, then this new wage profile satisfies the (LL) constraint and all (ICa) constraints when $\delta$ is small. Further, it helps bound $|W(\lambda', \tilde{z}') - W(\lambda, \tilde{z})| \sim \mathcal{O}(\epsilon)$ as $\delta \to 0$. 

41
Step 2. In Case (a), Problem (4.2) becomes

\[
\min_{K \in \mathbb{N}, \lambda \in \Lambda, \hat{z} \in \mathbb{Z}(\lambda)} W(\lambda, \hat{z}) + \mu \cdot f(K).
\]

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

\[
\min_{\lambda \in \Lambda, \hat{z} \in \mathbb{Z}(\lambda)} W(\lambda, \hat{z}) + \mu \cdot h(\lambda, \hat{z}),
\]

and notice that \(h(\lambda, \hat{z})\) is continuous in both arguments. Existence of solution follows from the compactness assumption.

\[\square\]

B Online Appendix (For Online Publication Only)

B.1 Individual Rationality

In the baseline model, now suppose, instead, that the agent faces an outside option that confers him a reservation utility of \(u\) at the outset. In this new setting, the wage scheme is a mapping \(w : \mathcal{P} \rightarrow \mathbb{R}\), and it is individually rational if

\[
\sum_{A \in \mathcal{P}} P_1(A)u(w(A)) \geq c + u. \quad \text{(IR)}
\]

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

**Corollary 3.** Assume Assumption [I] Then any \(\mathcal{P}^*\) consists of finite \(Z\)-convex cells.

**Proof.** First, take any optimal monitoring technology \(\mathcal{P}^*\) as given, and let \(\lambda\) and \(\gamma\) denote the Lagrange multipliers of the (IC) and (IR) constraints, respectively. Rewrite the principal’s problem as follows:

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

Differentiating this objective function with respect to \(w_n\) and setting the result equal
to zero, we obtain:

\[ u'(w^*_n) = \frac{1}{\lambda z_n + \gamma}. \]

Since \( z_n \)'s are all different, assume w.l.o.g. that \( z_1 < \cdots < z_N \) and that \( w^*_1 < \cdots < w^*_N \).

Now suppose, to the contrary, that some \( A_j \in P^* \) 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

\[
\begin{cases}
  z'_j(0) = \frac{s (z''_j - z'_j)}{\pi_j}, \\
  z'_k(0) = -\frac{s (z''_k - z'_k)}{\pi_k}, \\
  z'_n(0) = 0, \forall n \neq j, k.
\end{cases}
\]

Take any wage profile \( \langle w_n(\epsilon) \rangle_{n=1}^{N} \) that makes both the (IC) and (IR) constraints binding after the perturbation, i.e.,

\[
\lambda \cdot \sum_{n=1}^{N} \pi_n u(w_n(\epsilon))z_n(\epsilon) = \lambda \cdot c
\]

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

Differentiating these equations with respect to \( \epsilon \) and summing up the results,

\[
\sum_{n=1}^{N} \pi_n \cdot u'(w^*_n) \cdot (\lambda z_n + \gamma) \cdot w'_n(0) = -\lambda \left[ u(w^*_j) \cdot \pi_j z'_j(0) + u(w^*_k) \cdot \pi_k z'_k(0) \right]. \tag{B.1}
\]

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

\[
u'(w^*_n) \cdot (\lambda z_n + \gamma) \cdot w'_n(0) = w'_n(0), \quad n = 1, \cdots, N.
\]

Plugging this result into (B.1), we obtain:

\[
\sum_{n=1}^{N} \pi_n w'_n(0) = s \left[ u(w^*_k) - u(w^*_j) \right] (\lambda z'' - \lambda z'). \tag{B.2}
\]
Now consider the perturbation that swaps $\tilde{A}$ and $A''$. Similar algebraic manipulation yields

$$\sum_{n=1}^{N} \pi_n w'_n(0) = -(1-s) \left[ u(w^*_k) - u(w^*_j) \right] (\lambda z'' - \lambda z'). \quad (B.3)$$

Since $u(w^*_j) \neq u(w^*_k)$, we have $\text{sgn} (B.2) \neq \text{sgn} (B.3)$, and the remainder of the proof follows that of Theorem 1.

**B.2 Contingent Monitoring Technology**

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

1. the principal commits to $(P(s), w(\cdot; s))_{s \in S}$;
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 $P(s)$ that contains $\omega$;
5. the principal pays the promised wage $w(A(\omega; s); s)$ to the agent.

Suppose for simplicity that $\omega$ and the new signal are independently distributed for any given level $a$ of effort. Define $Z : \Omega \to \mathbb{R}$ and the $z$-value of any $A \in \Sigma$ the same as before. Let $Q_a$ denote the p.m.f. of the new signal under effort $a$. Define

$$Z(s) = 1 - \frac{Q_0(s)}{Q_1(s)}$$

for all $s \in S$.

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

**Corollary 4.** Assume Assumptions 1 and 8. Then each $P^*(s)$ comprises finite $Z$-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.

First, take \( \langle \mathcal{P}^*(s) \rangle_{s \in S} \) as given and rewrite the principal’s problem as follows:

\[
\min_{\{w_{n,s}\}} \sum_{n,s} Q_1(s) \pi_{n,s} w_{n,s} - \lambda \left[ \sum_{n,s} Q_1(s) \pi_{n,s} u(w_{n,s}) z_{n,s} Z(s) - c \right] - \sum_{n,s} \eta_{n,s} w_{n,s},
\]

where \( \lambda \) denotes the Lagrange multiplier of the (IC) constraint and \( \eta_{n,s} \) the Lagrange multiplier of the (LL) constraint at \( w_{n,s} \). Differentiating the principal’s objective function with respect to \( w_{n,s} \) and setting the result equal to zero, we obtain:

\[
u'(w_{n,s}^*) = \frac{1}{\max \{ \lambda z_{n,s} Z(s), \check{z} \}} \forall n, s.
\]

Plugging this first-order condition into the proof of Lemma 3, we see that for all \( s \), \( z_{n,s} \)'s are all different and at most one \( w_{n,s}^* \) is equal to zero. Without loss of generality, let \( z_{1,s} < z_{2,s} < \cdots \) and \( w_{1,s}^* = 0 < w_{2,s}^* < \cdots \).

Now suppose, to the contrary, that \( A_{j,s} \in \mathcal{P}^*(s) \) is not \( Z \)-convex for some \( j \in \mathbb{N} \) and \( s \in S \). Then there exist \( A', A'' \subset A_{j,s} \) and \( \tilde{A} \subset A_{k,s} \in \mathcal{P}^*(s) \) for some \( k \neq j \) such that

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

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

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. Differentiating these \( z \)-values with respect to \( \epsilon \) and evaluating the results at \( \epsilon = 0 \), we obtain:

\[
\begin{cases}
    z_{j,s}'(0) = \frac{t(z'' - z')}{\pi_{j,s}'}, \\
    z_{k,s}'(0) = -\frac{t(z'' - z')}{\pi_{k,s}'}, \\
    z_{n,s}'(0) = 0 \forall n \neq j, k, \\
    z_{n,s'}'(0) = 0 \forall n \text{ and } s' \neq s.
\end{cases}
\]
Take any wage profile \( \langle w_{n,s'}(\epsilon) \rangle_{n \in \mathbb{N}, s' \in S} \) such that \( w_{1,s'}(\epsilon) = w_{1,s'}^* = 0 \) for all \( s' \in S \) and the incentive compatibility constraint binds after the perturbation, i.e.,

\[
\sum_{s',n} Q_1(s') \pi_{n,s'} w_{n,s'}(\epsilon) z_{n,s'}(\epsilon) Z(s') = c.
\]

Differentiating the above equation with respect to \( \epsilon \) and multiplying the result by the Lagrange multiplier \( \lambda \) of the (IC) constraint prior to the perturbation, we obtain:

\[
\sum_{s',n} Q_1(s') \pi_{n,s'} \cdot u'(w_{n,s'}^*) \cdot \lambda z_{n,s'} Z(s') \cdot w_{n,s'}'(0)
\]

\[
= -\lambda Q_1(s) Z(s) [u(w_{j,s}^*) \cdot \pi_{j,s} z_{j,s}'(0) + u(w_{k,s}^*) \cdot \pi_{k,s} z_{k,s}'(0)]
\]

\[
= t Q_1(s) Z(s) [u(w_{k,s}^*) - u(w_{j,s}^*)] (\lambda z'' - \lambda z'). \tag{B.4}
\]

Meanwhile, for all \( s' \), we have \( u'(w_{n,s'}^*) = \frac{1}{\lambda z_{n,s'} Z(s')} \) for all \( n \geq 2 \) and \( w_{1,s'}'(0) = 0 \). Thus,

\[
u'(w_{n,s'}^*) \cdot \lambda z_{n,s'} Z(s') \cdot w_{n,s'}'(0) = w_{n,s'}'(0) \forall n, s'.
\]

Plugging this result into (B.4), we obtain:

\[
\sum_{s',n} Q_1(s') \pi_{n,s'} w_{n,s'}'(0)
\]

\[
= t Q_1(s) Z(s) [u(w_{k,s}^*) - u(w_{j,s}^*)] (\lambda z'' - \lambda z'). \tag{B.5}
\]

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

\[
\sum_{s',n} Q_1(s') \pi_{n,s'} w_{n,s'}'(0)
\]

\[
= -(1-t) Q_1(s) Z(s) [u(w_{k,s}^*) - u(w_{j,s}^*)] (\lambda z'' - \lambda z'). \tag{B.6}
\]

Since \( w_{j,s}^* \neq w_{k,s}^* \) and \( Z(s) \neq 0 \) (Assumption [8]), we conclude that \( \text{sgn} \ (B.5) \neq \text{sgn} \ (B.6) \), and the remainder of the proof follows that of Theorem 1.
B.3 Random Monitoring Technology

Setup. In the baseline model, now suppose, instead, that the monitoring technology can be any mapping $q : \Omega \to \Delta^N$, where $N$ is taken as given for now. Time evolves as follows:

1. the principal commits to $(q, w)$;
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

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

as the probability that the state realization is in $A$ and the agent is assigned to category $n$ under high effort. For each $n = 1, \cdots, N$, let $\pi_n = q_n(\Omega)$ be the probability that the agent is assigned to category $n$ under high effort, and define

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

as the $z$-value of category $n$. Then the agent’s incentive compatibility constraint can be written as follows:

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

An optimal incentive contract $(q^*, w^*)$ solves

$$\min_{N, q, \{w_n\}} \sum_{n=1}^{N} \pi_n w_n + \mu \cdot H(q, 1) \text{ s.t. } \{IC\} \text{ and } \{LL\},$$  \hspace{1cm} (B.7)

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

47
Partitional monitoring technology

A monitoring technology is *partitional* if \( q_n(\omega) \in \{0, 1\} \) for all \( \omega \) and \( n \). Below we give an example where focusing on partitional monitoring technologies is w.l.o.g.:

**Theorem 7.** Assume Assumption 2 and \( H(q, a) = f(N) \). Then \( q^* \) is partitional.

*Proof.* Suppose, to the contrary, that \( q^* \) is not partitional. Under Assumption 2 there exists \( A', A'' \in \Sigma \) where (1) \( P_1(A') = P_1(A'') = \epsilon \) for some small \( \epsilon > 0 \), (2) \( z(A') = z' < z(A'') = z'' \) and (3) \( q_j^*(A'), q_m^*(A'), q_j^*(A''), q_m^*(A'') \in (0, 1) \) for some \( j < m \). Consider first the case of \( j = 1 \). Consider the following perturbation to the monitoring technology (parameterized by \( \epsilon \)):

\[
\begin{align*}
q_1(A'; \epsilon) &= q_1^*(A') + \epsilon, \\
q_1(A''; \epsilon) &= q_1^*(A'') - \epsilon, \\
q_m(A'; \epsilon) &= q_m^*(A') - \epsilon, \\
q_m(A''; \epsilon) &= q_m^*(A'') + \epsilon.
\end{align*}
\]

By construction, this perturbation has no effect on \( \pi_n \)'s and hence the monitoring cost. Let \( \{z_n(\epsilon)\}_{n=1}^{N} \) denote the \( z \)-values of the performance categories after this perturbation. Straightforward algebra shows that

\[
\begin{align*}
z_1(\epsilon) &= z_1(0) - \frac{(z'' - z')}{\pi_1} \cdot \epsilon^2 + O(\epsilon^4), \\
z_m(\epsilon) &= z_m(0) + \frac{(z'' - z')}{\pi_m} \cdot \epsilon^2 + O(\epsilon^4), \\
z_n(\epsilon) &= z_n(0) \quad \forall n \neq 1, m.
\end{align*}
\]

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

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

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

**Mutual information**

The next assumption is commonly made by the literature on rational inattention:
**Assumption 9.** There exists $K \geq 2$ such that

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

for all $N \leq K$ and $q : \Omega \to \Delta^N$.

**Theorem 8.** Suppose that $u$ is smooth and that Assumptions 2, 6 and 9 hold. Then $q^*$ exists and is of form $q^* : Z(\Omega) \to \Delta^N$ for some $2 \leq N \leq K$. Further, $q^*_n(z)/q^*_m(z)$ increases with $z$ for all $n > m$.

**Proof.** Fix any $2 \leq N \leq K$ and consider the optimal contract of $N$ categories. Take any $q$ for which the optimal wage scheme exists, and denote by $W(q)$ the minimum incentive cost under $q$. Notice that $W(q)$ is linear in $q(\omega)$ for each $\omega$, because replacing $q(\omega)$ with a lottery $s \circ q(\omega) + (1-s) \circ q''(\omega)$ such that $sq_n(\omega) + (1-s)q''_n(\omega) = q_n(\omega)$ for $n = 1, \ldots, N$ has no effect on $\pi_n$’s or $z_n$’s and hence the incentive cost. Then from the fact that $H(q, 1)$ is convex in $q(\omega)$, it follows that the principal’s objective function $W(q) + \mu \cdot H(q, 1)$ is convex in $q(\omega)$ for each $\omega$.

The implication of this result is twofold. First, it is w.l.o.g. to consider $q : Z(\Omega) \to \Delta^N$, because if the contrary were true, then there exist $A', A'' \in \Sigma$ such that $P_1(A') = P_1(A'') = \epsilon$, $z(A') = z(A'')$ and yet $q(A') \neq q(A'')$. But then applying any lottery $s \circ q(A') + (1-s) \circ q(A'')$ with $s \in (0, 1)$ to all $\omega \in A' \cup A''$ has no effect on the incentive cost but saves the monitoring cost, a contradiction.

Second, the solution to the principal’s problem, if exists, satisfies the following first-order condition obtained from differentiating the objective function with respect to $q(z)$ and setting the result equal to zero ($\lambda$ is the Lagrange multiplier of the (IC) constraint): for all $n > 1$,

$$-w_n + \lambda u(w_n) z = \mu \left[ \log \frac{q_n(z)}{q_1(z)} - \log \frac{\pi_n}{\pi_1} \right].$$

Thus $q^*$, if exists, is continuously differentiable, and hence is equicontinuous by Assumption 6. Thus, rewrite the principal’s problem as follows:

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

and the existence of solution follows from Helly’s selection theorem. \qed
References


