Optimal Incentive Contract with Costly and Flexible Monitoring

ANQI LI, Washington University in St. Louis
MING YANG, Fuqua Business School

1 Introduction

Successful employee monitoring in modern workplaces requires the use of monitoring technologies that transform the growing volume and variety of performance data into information that is meaningful for employee evaluation and compensation. A famous example of monitoring technology is the 360-degree performance appraisal system, which classifies the feedback received from an employee’s supervisors, peers and subordinates into ratings such as “outstanding,” “effective,” “satisfactory” and “unsatisfactory” (Bracken et al. (2001)). Recent advances in IT and data science bring two opportunities and challenges to the design and implementation of monitoring technologies. First, the cost of data processing, storage and communication has been reduced but remains significant. Second, new technologies enable design flexibility, which sparkles debates about how we can classify the various sources of feedback into meaningful performance categories, which rating scale best balances informativeness and interpretability, how we should trade off the monitoring of linked activities, and which mix of individual and group evaluations simultaneously motivates large groups of employees (Kaplan and Norton (1992), Pulakos (2004), Hook et al. (2011), Bryson et al. (2013)). In this paper, we develop a new model that enables the analysis of how the cost and flexibility that recent technological advances bring to the processing, storage and communication of performance data can potentially affect employee incentives and the internal organization of firms.

2 Illustrative Example

Our model builds on an otherwise standard principal-agent model with moral hazard, where we represent all available data about the agent’s hidden effort by a random and potentially high-dimensional performance state. To best represent the cost and flexibility that IT and big data bring to the design and implementation of monitoring technologies, we allow the principal to adopt any monitoring technology that partitions the agent’s performance state space into finite categories, at a cost that increases with the amount of information that the induced performance signal carries. To illustrate, suppose a performance state consists of the ratings given by a senior manager and a junior manager. In a hypothetical world where monitoring is costless, fully revealing the performance state provides the agent with the strongest incentive to work. But in reality, the processing, storage and communication of performance data incur significant costs, giving rise to a common practice supported by information theory (Cover and Thomas (2006)), that of classifying the fine-grained performance data into a limited number of categories (Bracken et al. (2001), Pulakos (2004), Hook et al. (2011)).

1 As of today, IT-based 360-degree performance appraisal is adopted by more than one third of U.S. companies and 90 percent of Fortune 500 companies.

2 On the one hand, Ewen and Edwards (2001) estimates that web-based technologies have reduced the processing cost of multi-source feedback by 80 percent; Soliman (2013) reports that cloud-based technologies are increasingly used for employee performance tracking, processing and storage. On the other hand, a 2014 survey by Towers and Watson ranks data processing and analysis as one of the top areas for HR spending; and a recent CB insight article details how the lucrative prospect of data processing and analysis has spurred the growth and M&A of HR tech startups across the globe (“The Data-ification of HR,” 2015).
Figure 1 depicts two commonly used monitoring technologies, where the one on the left panel labels each performance state as either “satisfactory” or “unsatisfactory” depending on whether the senior manager’s rating exceeds a cut-off or not, while the one on the right panel does so according to whether a weighted score is above or below a threshold. While these monitoring technologies emphasize distinct opinion sources and assign varying contents to the performance categories “satisfactory” and “unsatisfactory,” they can both be adopted by the principal in our model, as long as the performance signals they induce carry about the same amount of information. These assumptions enable us to formalize the cost and flexibility associated with the processing, storage and communication of performance data, and to analyze their impact on employee incentives and the internal organization of firms through the trade-off between the incentive cost and the monitoring cost.

Figure 1: Commonly used monitoring technologies represented by partitions of the performance state space.

3 Baseline Model

In the baseline model, we revisit the classical setup of Holmstrom (1979) where a single agent can influence the distribution of performance states by exerting either high effort or low effort. The main result shows that the optimal monitoring technology compresses those fine-grained and high-dimensional performance states with similar likelihood ratios into the same coarse, single-dimensional performance grade, therefore lending support to the use of multi-source performance appraisal systems aggregate the various kinds of feedback into coarse, rank-ordered ratings — such as “outstanding,” “highly effective,” “satisfactory” and “unsatisfactory” — based on the employee’s achievement measured by an overall score.

This result showcases Holmstrom (1979)’s sufficient statistics principle, which says that the optimal wage scheme for any given monitoring technology depends only on the likelihood ratios of the performance states. This suggests that when monitoring is costly and flexible, the principal should focus on the processing, storage and communication of likelihood ratios, and ignore the part of the performance state that is orthogonal to the likelihood ratio. As a result, the optimal monitoring technology assigns distinct average likelihood ratios to different performance categories, yielding a performance signal that satisfies the strict monotone likelihood ratio property (SMLRP) with respect to the order induced by likelihood ratios. Furthermore, since performance states with
similar likelihood ratios have similar effects on the compensation cost while the monitoring cost is independent of the likelihood ratios, the optimal monitoring technology classifies performance states with similar likelihood ratios into the same performance category and can be obtained from applying a simple cutoff rule to the space of likelihood ratios. Together with the empirical findings of Bloom and Van Reenen [2006, 2007], these results attribute the use of monitoring technologies with different degrees of fine-grainedness to factors that affect the (opportunity) cost of employee monitoring (e.g., access to IT, labor market regulation, production market competition), thereby adding a new explanation to the long-lasting puzzle surveyed by Gibbons and Henderson [2013], that of why the management practices adopted by otherwise similar firms exhibit significant and persistent heterogeneity.

4 Multiple Actions

When the agent can take multiple deviant actions, the optimal incentive contract takes the form of a balanced scorecard (Kaplan and Norton [1992]) which increases the resources spent on the detection of each potential deviation with the Lagrange multiplier of the corresponding incentive compatibility constraint. In the multi-task model considered by Holmstrom and Milgrom [1991], this suggests that the principal should fine-tune the monitoring intensity across tasks according to the agent’s tendency to shirk. In Figure 2, we present a simulation result showing that when the agent simultaneously engages two tasks A and B, the principal should shift focus to the monitoring of task B as the performance state in task A becomes an increasingly precise measure of the agent’s underlying effort. This result has policy implications, e.g., how universities should evaluate the teaching of faculty members who, in addition to teaching, engage in other more-difficult-to-measure tasks such as research and administration.

Figure 2: Horizontal axis represents the informativeness of the performance state in task A. Vertical axis represents the intensity at which task A is monitored relative to task B.

5 Multiple Agents

We finally turn to the multi-agent model considered by Holmstrom [1982], Green and Stokey [1983] and Mookherjee [1984], where the conventional wisdom attributes the use of group incentive contracts (e.g., team, tournament) to the technological interdependence between agents, while limiting
the use of individual incentive contracts among technologically independent agents. Recently, this view has been challenged by Bloom and Van Reenen (2006, 2007), which find that even firms with similar production technologies make significantly different choices between individual and group incentive contracts. We resolve this puzzle from the angle of monitoring cost. Intuitively, group-based monitoring lumps agent assessment together and yields coarser ratings than individual-based monitoring. When monitoring is costly and flexible, the limited monitoring capacity creates an attentional linkage between agents that stipulates the use of group incentive contract even if agents are technologically independent. The main prediction of our result, that employees are less recognized for their individual performance as the monitoring cost increases, other things equal, is supported by the findings of Bloom and Van Reenen (2006, 2007).

References


