THE ACCOUNTING REVIEW
Vol. 82, No. 2
2007
pp. 389–425

The Retention Effect of Withholding Performance Information

Korok Ray
University of Chicago

ABSTRACT: It is a common practice for firms to conduct performance evaluations of their employees and yet to withhold this information from those employees. This paper argues that firms strategically withhold performance information to retain workers. In particular, if the worker enjoys high outside options and is tempted to quit, then the firm chooses not to reveal his performance information in order to keep him on the job. The firm's equilibrium strategy is to fire if performance is sufficiently low, reveal information if performance is sufficiently high, and withhold information otherwise. The pooling equilibrium is robust under a wide variety of settings, such as general cost functions, ability-contingent outside options, nonlinear contracts, nonverifiable output, and multiple stages of production.

Keywords: *performance measurement; performance evaluation; information revelation; disclosure.*

I. INTRODUCTION

A large literature in management and compensation documents the common practice of firms conducting performance evaluations of employees and yet withholding this information from those employees.[1] For example, Murphy and Cleveland (1991) find that managers often lump workers who have different performance levels into a single category, suggesting that the reported distribution of performance is less variable than the true distribution of performance. The standard explanation from the management literature is based on psychology and politics: revealing true output to employees kills morale and creates animosity within the firm; thus, firms tell all workers that they are above average. While these explanations are plausible, this paper argues that there is an economic force driving this phenomenon: firms withhold performance information to retain workers.

[1] According to Saal et al. (1980), leniency, range restriction, and halo error are the effective categories that encompass forms of withholding. See Landy and Farr (1980), Milkovich and Newman (1996), Milkovich and Wigdor (1991), and Mohrman and Lawler (1983) for surveys of the literature.

This paper is based on Chapter 2 of my dissertation submitted to Stanford University. I owe the greatest debt to my thesis advisor, Edward Lazear, for years of guidance. John Roberts, Bill Rogerson, Madhav Rajan, Stefan Reichelstein, Canice Prendergast, Luis Rayo, Andy Skryzpacz, Lars Stole, Milt Harris, two anonymous referees, and seminar participants at Stanford University, the University of Chicago, and the Chicago/Minnesota Theory Conference all provided many helpful comments. Maris Goldmanis and Marco Fiaccadori provided excellent research assistance. Stanford University Graduate School of Business and the University of Chicago Graduate School of Business provided generous financial support.

Editor's note: This paper was accepted by Madhav Rajan.

Submitted September 2005
Accepted July 2006

To show this, I construct an agency model consisting of a risk-neutral principal (the firm) contracting with a risk-neutral agent (the worker) under limited liability. The steering assumption is that the firm observes the worker's job-specific ability, but the worker himself does not. In practice, evaluating performance is a complex activity, and once firms conduct such evaluations they may choose whether and how much of that evaluation to reveal to the worker. Note that this assumption differs from the standard adverse selection models in which the agent holds private information.[2] Here, the principal must decide whether to reveal her private information, possibly for strategic purposes.

For example, consider that a firm hires an employee and observes his job-specific ability. Firms have expertise in the nature of their business and have experience hiring and firing workers, so it is plausible that the firm can better differentiate a high-ability worker from a low-ability worker than the worker himself can. After observing the worker's ability, the firm can choose whether to disclose this information to the worker. Following this disclosure, the worker can choose to quit the firm and work elsewhere, or stay. As is common in most production settings, ability and effort are complements, so more able workers have higher marginal products of effort.

The two key assumptions in the model are complementarity in production and the option to quit at the interim stage. Because of complementarity between effort and ability, a highly able agent works hard because the marginal return from his labor is higher. Thus, the firm wants to reveal output for workers with high ability. On the other hand, the agent has the option to leave the firm after the evaluation and collect his outside option. Limited liability creates a conflict of interest between the principal and the agent, and, hence, there will be times when the firm wishes to retain the agent but the agent wishes to quit. This gives the firm an incentive to withhold information in order to strategically retain workers. Complementarity in production pushes the firm to reveal information, while the retention motive pushes the firm to withhold information.

The trade-off of these two forces creates a unique equilibrium in which the firm reveals information if output is very high or very low, but withholds information if output is in-between. Profit from retaining workers who otherwise would have quit gives the firm incentives to expand the pool. However, complementarity in production makes pooling costly, and this places a bound on the optimal size of the pool. In equilibrium, the firm selects the pool such that the average member of the pool is exactly indifferent between staying at the firm and leaving. This equilibrium of pooling in the middle and separating at the extremes is consistent with performance evaluations in practice (see Milkovich and Wigdor [1991] and surveys such as Murphy and Cleveland [1991]), where workers are either laid-off, promoted, or told that they are average (Milkovich and Newman 1996, 360–370).

The pooling equilibrium is robust under a wide variety of settings. First, there still exists a pooling equilibrium when outside options are contingent on ability. Second, I show the equilibrium holds if the worker's output is nonverifiable, so the firm announces a report of the agent's early-stage output, which the agent may or may not believe. To prevent the firm from arbitrarily biasing its report, it is necessary for the firm to make a transfer payment to the agent contingent on its announcement. Third, I consider nonlinear compensation schemes. While nonlinearity complicates the analysis, making the payoff functions convex and concave in places, I show that, under mild conditions, it is still profitable for the firm to withhold information.

---

[2] See Baiman (1990) and Lambert (2001) for reviews of such models in the accounting literature.

Motivation and Related Literature

Psychology is the most common explanation for the withholding of performance information from employees. Saal and Knight (1988) suggest several widely accepted accounts of withholding, including the desire to be liked, unwillingness to give negative feedback, fear that other managers inflate their ratings, and abnormally high or low criteria.[3] Specifically, Jones et al. (1983) argue that "the individual who receives an average performance assessment may feel punished because of a belief that a higher rating was deserved," which may decrease future productivity. To prevent discouragement among employees, managers refrain from fully disclosing performance information.

Politics is the other common explanation of withholding performance information.[4] For example, Longenecker et al. (1987) conducted in-depth interviews with 60 upper-level executives and found that executives "deliberately distort and manipulate evaluations for political purposes." The underlying justification for political considerations is a concern about "how to best use the evaluation process to motivate and reward subordinates." There is a widespread belief, documented in Milkovich and Newman (1996) that disclosing performance information will create animosity among co-workers and subsequently cause productivity to suffer. Schall (1983) further demonstrates that political considerations often develop into sets of unwritten rules that permit managers to communicate poor performance evaluations to subordinates without formally giving low ratings.[5] According to this literature, political considerations within firms are thus a significant factor of the practice of withholding performance evaluation information.

There are only a handful of papers that explore information disclosure in performance evaluations, and none of them consider the effects of retention. MacLeod (2003) proposes a model of subjective evaluations and finds that the firm will compress the distribution of performance if the principal and agents receive conflicting signals on the agent's performance. Lizzeri et al. (2002) explore the incentive effects of revealing information to employees. They find that under some conditions, the expected cost to the principal of inducing a given level of effort is lower if the agent cannot condition his second-stage effort on early-stage output. Once again, neither paper assumes complementarity across stages or the option to quit at the interim stage.

The model here bears some similarity to the career concerns literature (Gibbons and Murphy 1992; Holmstrom 1999; Meyer and Vickers 1997), though the trade-offs and nature of information asymmetry are different. For example, ability and effort are complements in the model of Holmstrom (1999), but the principal (the market) does not know the manager's ability, causing him to work hard early in his career to influence market perceptions on his ability. In my model, the principal (the firm) knows the agent's ability, so the main trade-off is between disclosure and retention.

There is a rather large body of literature on the voluntary disclosure of private information in asymmetric information games.[6] The classic early results in this literature

---

[3] See Arvey and Jones (1985), who suggest that discipline and punishment have not sufficiently been studied in organizational settings, and where these topics have been studied, evidence shows that performance evaluation is sometimes used as a vehicle for discipline. Kay et al. (1965) show specifically that increasingly threatening assessments yield decreasingly favorable attitudes and declining subsequent constructive improvement in job performance.

[4] See Bernardin and Beatty (1984), who point out that performance evaluation occurs in circumstances that do not always allow for rationality, straightforwardness, or objectivity in evaluation. Cleveland et al. (1986) suggest that managers first determine the overall rating and then go back and fill in the details of the rating.

[5] Schall (1983) deems these rules "tacit understandings (generally unwritten and unspoken) about appropriate ways to interact (communicate) with others in given roles and situations."

[6] For an extensive review of this literature in a capital markets context, see Verrecchia (2001).

(Grossman and Hart 1980; Grossman 1981; Milgrom 1981) showed that private information is fully revealed in equilibrium if disclosure is costless and there is a monotonic ordering of "favorable" versus "unfavorable" news. Later research identified conditions under which this full disclosure or "unraveling" result breaks down and information withholding can take place. These include (1) costly disclosure of information, for example, due to its proprietary nature (Verrecchia 1983); (2) uncertainty as to whether the informed party actually possesses the information (Dye 1985; Jung and Kwon 1988; Shin 1994); and (3) failure of the monotonicity assumption on the "favorableness" of news, since favorable financial market news could increase competition and thus have not only positive, but also negative consequences for the firm (Darrough and Stoughton 1989; Okuno-Fujiwara et al. 1990). While most of the earlier literature focuses on a simple binary choice between disclosure and nondisclosure, a number of the more recent papers also address partial revelation and noisy signals (Newman and Sansing 1993; Okuno-Fujiwara et al. 1990).

My paper addresses a similar question as the disclosure literature: When is it optimal for an informed party not to reveal its private information to an uninformed party? However, my model differs from the existing disclosure literature in at least two ways. First, I consider information revelation in an agency model, whereas most of the disclosure literature focuses on capital markets. Second, and most importantly, the firm in my model can commit to an information partition prior to obtaining private information, whereas in most of the models discussed the decision to reveal is conditional on existing information. This ability to commit is important in obtaining the equilibrium with information withholding, even though there is still monotonicity of the "favorableness" of news. While it would be optimal for the firm *ex post* to fully reveal the higher values from the pooling interval, it is optimal *ex ante* to commit not to do so. The firm will withhold information from both the upper and lower ends of the pool, and the gain from the latter is greater than the loss from the former.[7]

The paper is organized as follows. Section II presents the basic model and solves for the first best outcome. Section III imposes a limited liability constraint and establishes the payoff functions of both parties under the full revelation benchmark—the restricted game in which the principal reveals output. Section IV lays out the partial revelation game in detail, and establishes properties of the principal's and agent's payoff functions. Section V proves the main result, and Section VI shows the robustness of the equilibrium. Section VII concludes.

II. THE MODEL

An employee (an agent) works on a project for a firm. Both parties are risk neutral. The agent has ability θ and exerts effort e at cost $C(e)$, where C, C', and C'' are positive for all $e > 0$. To ease calculations, I assume the marginal cost function is log-concave. This is a weak assumption satisfied by virtually every cost function ever used in applied agency models, such as the class of power functions.[8]

[7] Note that in order to have conflicting incentives for the revelation of the entire pooled interval, we do not need conflicting incentives for revelation of each particular point in the interval, and thus there is no need for multiple audiences as in Darrough and Stoughton (1989) or Newman and Sansing (1993).

[8] See Bagnoli and Bergstrom (2005) for discussion of the assumption of log-concavity. While this assumption simplifies analysis of equilibrium, the main result still holds without it, as shown in Section VI.

The agent produces output:

$$q = \theta e + \varepsilon.$$

The noise term $\varepsilon \in [0,\infty)$ is distributed around a mean of m, with cumulative distribution function (cdf) $G(\cdot)$ and density function $g(\cdot)$. The total value of the project is Vq. Note that q is a function of θe, so ability and effort are complementary inputs to production; a marginal increase in θ increases the marginal productivity of effort. The agent observes his effort, but the firm does not. Ability θ and effort e are noncontractible, but output q is contractible. Noncontractibility of θ represents (1) that contracts are written before knowledge of ability is realized or collected, and (2) that there are exogenous costs of contract complexity. Ability, like effort, is difficult to measure and hence cannot form the basis for contracts.

The ability parameter $\theta \in (0,\infty)$ is distributed randomly with cdf $F(\cdot)$ and density function $f(\cdot)$. The agent does not observe his own ability, but the firm does (as a result of the firm's prior experience with the agent). The firm can choose to reveal or not reveal θ to the agent. The agent will use his knowledge of θ to decide whether to work on the project, and to select his appropriate effort level. Once revealed, ability information is *ex post* verifiable by the agent.[9] So if the principal decides to reveal ability, then she will do it truthfully. This focuses the problem on whether to reveal information at all, and not on whether to distort revealed information.

Finally, both the principal and the agent have outside options. This captures the value from quitting the project and dedicating resources (labor, capital) elsewhere. Let $\bar{u}_1^a$ and $\bar{u}_1^p$ denote the agent's and the principal's outside options before they enter in a contractual relationship with each other (and before the firm observes the agent's ability). Let $\bar{u}_2^a$ and $\bar{u}_2^p$ be the respective outside options right before deciding whether to work on the new project (but after the firm has observed ability). Call $\bar{u}_2 = \bar{u}_2^a + \bar{u}_2^p$ the residual surplus: the total surplus from abandoning the project. I assume that $Vm < \bar{u}_2^a$ and $Vm < \bar{u}_2^p$, so neither party would want to work on the project if no effort is exerted, even if that party received the whole surplus from production. Finally, the principal and agent must satisfy the participation constraint (PC) that their equilibrium payoffs must exceed $\bar{u}_1^p + \bar{u}_2^p$ and $\bar{u}_1^a + \bar{u}_2^a$, respectively.

First Best

This subsection establishes the first best benchmark useful for the rest of the paper: it is efficient for the firm to fully reveal θ, and it is efficient to quit projects when the employee's ability turns out to be low. The first result states that more information is always socially optimal, and so strategic information revelation generates welfare losses. The second result shows that it is efficient to terminate bad projects because firms and workers have outside options.

As mentioned earlier (and proved in later sections), the firm will withhold information in order to retain the worker. Precisely, there are conditions under which a worker prefers to leave the firm but the firm wants him to stay. These conflicts of interest are absent in the social planner's problem. The planner simply terminates workers at the efficient rate, and, hence, there is no reason to withhold ability information. All proofs are in the Appendix.

[9] I relax this assumption later in the paper.

Proposition 1: It is efficient to fully reveal ability.

In fact, withholding information makes the planner (weakly) worse off, since it prevents effort from accurately conditioning on ability. To see this, suppose $\Theta \subset R$ is a nontrivial pooling region. If the planner chooses to pool, then the effort function must be constant over Θ. If the planner reveals information, then the effort function is unconstrained over Θ. So pooling shrinks the planner's choice set of all possible effort functions. This forces the planner to optimize over a smaller set, and, hence, total surplus decreases. A key figure in the planner's optimization problem is the continuation surplus function:

$$S(\theta) = \max_{e}\{EV(\theta e + \varepsilon) - C(e)\} = V(\theta e(\theta) + m) - C(e(\theta)),$$

where E is the expectation with respect to ε. This is the total surplus from continuing for each realization of ability θ. The effort function is increasing because effort and ability are complementary. So the agent works harder if he knows he has more ability because the marginal return to his labor is now higher. The social planner's termination rule takes the form of a cutoff rule. In other words, terminating low ability workers is efficient.

Proposition 2: There exists a cutoff point θ^* such that it is efficient for only workers with $\theta > \theta^*$ to work on the project.

Here, θ^* is the efficient termination rule. The proof of this proposition uses the envelope theorem and the first-order conditions on $e(\cdot)$ to show that the continuation surplus function is increasing. As usual, it is possible to implement the first best with a standard sell-out contract, where the agent pays the firm upfront the equilibrium value of output less equilibrium cost of effort and less the agent's outside option, and the firm is able to give the agent a full share in the output from production.

III. FULL REVELATION

To build intuition and establish preliminary results, I first solve the model for the Full Revelation (FR) case. Assume that the principal can fully commit to reveal θ to the agent. A contract is a tuple $\langle T_p, s_1, s_2, b\rangle$, where $T_p \subseteq R_+$ is the firing rule (fire if $\theta \notin T_p$), s_1 is salary paid for participation in the initial (evaluation) stage,[10] s_2 is salary paid for participation in the project itself, and b is the bonus on final output q. Compensation for the agent is a linear function of final output, so he receives $s_1 + s_2 + bq$ if he ends up working on the project, and s_1 otherwise. This assumption of a sharing rule b and salaries s_1 and s_2 that are independent of θ is reasonable in a setting where bonuses are negotiated before the evaluation stage. Of course, the agent's total compensation still contains risk, since q is still a random variable, realized only at the end of the production stage.[11]

Given this contract, the agent responds by choosing actions $\langle T_a, e(\cdot)\rangle$: a quit rule $T_a \subseteq R_+$ (quit if $\theta \notin T_a$) and an effort function $e(\cdot)$. Let $T = T_a \cap T_p$. Since output must clear

[10] Even though production occurs in a single stage, as described in the setup of the model, production is preceded by an evaluation stage, during which the firm finds out the agent's ability level.

[11] Contracts that are not linear in final output are considered in the "Robustness" section of the paper. While not all results obtained for linear contracts carry over to the general case, I show that the main result is still valid: even with general contracts, there exist conditions under which it is optimal for the firm to withhold information. However, the assumption that the contract cannot be conditioned on ability is important—without this assumption, it would in general be more profitable to retain marginal agents by means of lump-sum payments to these agents rather than by pooling.

both hurdles for the agent to advance, the probability of advancing to the production stage is $P(T) = \Pr(\theta \in T)$. Conditioning on θ, the agent's and principal's continuation utilities are:

$$u(\theta) = E[b(\theta e(\theta) + \varepsilon) + s_2 - C(e(\theta))] = b(\theta e(\theta) + m) + s_2 - C(e(\theta))$$
$$\pi(\theta) = E[(V - b)(\theta e(\theta) + \varepsilon) - s_2] = (V - b)(\theta e(\theta) + m) - s_2.$$

Given a contract $\langle T_p, s, b\rangle$, the agent's problem is:

$$\max_{T_a, e(\cdot)} \int_T u(\theta) f(\theta) d\theta + (1 - P(T))\bar{u}_2^a + s_1$$
subject to (PC)

where T is a function of T_a and (PC) is the participation constraint. In words, if $\theta \in T$, then the agent advances and receives $u(\theta)$. If not, then he receives $\bar{u}_2^a$. He bears $C(e)$ only if he advances. Note that $C(e)$ is embedded in $u(\theta)$ and so does not appear in the above optimization explicitly. If the agent advances, then the principal receives $\pi(\theta)$ for each $\theta \in T$, and if the agent leaves, then the principal receives her outside option $\bar{u}_2^p$. The principal selects a target, bonus, and salary to maximize her expected utility, so she solves:

$$\max_{T_p, b, s_1, s_2} \int_T \pi(\theta) f(\theta) d\theta + (1 - P(T))\bar{u}_2^p - s_1$$
subject to (PC).

Proposition 3: The effort function $e(\theta,b)$ is increasing in θ and b. The continuation payoffs $u(\theta)$ and $\pi(\theta)$ are increasing and convex. The agent and the principal adopt cutoff strategies for separation: $T_a = [\theta_a,\infty)$ and $T_p = [\theta_p,\infty)$.

The first order conditions (F.O.C.) shows that $C'(e(\theta)) = b\theta$, and so the effort function is in fact a function of both the incentives b and the information θ (but is independent of the salaries s_1 and s_2). Because ability and effort are complements, a high θ affects the marginal productivity of effort, so effort will increase for high realizations of ability.

The shape of the effort function $e(\theta)$ determines the shape of the principal's continuation payoff $\pi(\theta)$ and the agent's continuation payoff $u(\theta)$. In particular, if $e(\theta)$ is increasing, then u and π will be convex. Higher ability levels are increasingly profitable for both the principal and the agent. Once again, this is driven by the complementarity between ability and effort. If θ is large, then the agent's marginal productivity of effort increases, and so final output will be large. Monotonicity of the continuation payoff functions implies that both parties will use cutoff strategies for separation. In particular, the agent will continue if and only if $u(\theta) \geq \bar{u}_2^a$, and since u is increasing, this holds if and only if $\theta \geq \theta_a$. Similarly, the principal will continue if and only if $\theta \geq \theta_p$. Thus, the separation rules are fully described by their cutoff values θ_a, θ_p and $\tau = \max\{\theta_a, \theta_p\}$. In addition, $1 - P(T) = F(\tau)$. I will refer to θ_p and θ_a as the principal's and agent's (respectively) targets, hurdles, and termination rules interchangeably.

At the optimum, the agent is indifferent between staying and leaving, and the principal sets θ_p such that $\pi(\theta_p) = \bar{u}_2^p$. The principal sets the lowest possible salary level s_1 to guarantee participation. Given θ, the agent's and principal's expected payoffs are given by:

$$U(\theta) \equiv \begin{cases} u(\theta) & \text{if } \theta > \tau \\ \bar{u}_2^a & \text{otherwise} \end{cases} \qquad \Pi^R(\theta) \equiv \begin{cases} \pi(\theta) & \text{if } \theta > \tau \\ \bar{u}_2^p & \text{otherwise.} \end{cases}$$

In choosing the incentive coefficient b, the principal trades off two separate forces. As she increases b, she induces more effort out of the agent (an incentive effect), which generates more output and hence more profit. But increasing b also reduces the principal's share of final output, since she earns $(V - b)q$. The optimal $b^* \in (0,V)$, so the principal gives the agent some but not full incentives. Because the effort function is increasing in b for each θ, this shows that the second-best effort level is less than the first-best optimum.

Proposition 4: The salary s_1 is positive only if the agent's participation constraint binds in equilibrium. The salary s_2 is positive only if $\theta_a > \theta_p$ in equilibrium.

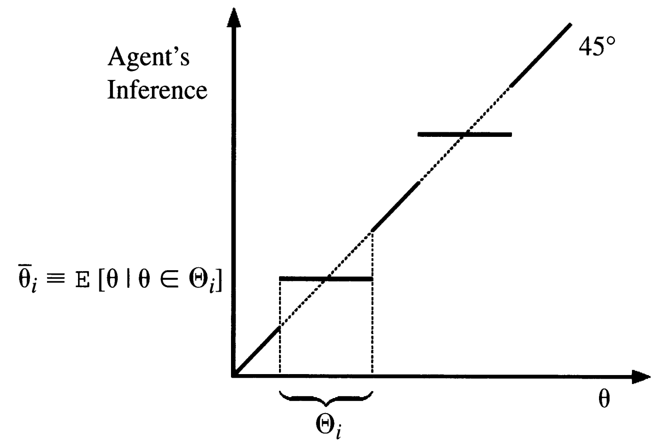
The proposition shows that the salary s_1 is paid only if it is necessary to induce agents to enter the contractual relationship with the firm, since increasing s_1 does not affect the agent's effort level or continuation decision and strictly reduces the principal's payoff from those agents who would have participated anyway. Similarly, s_2 is paid only if it can help the firm retain some agents that would otherwise have preferred to quit after discovering their ability level (i.e., some of those with $\theta_p < \theta < \theta_a$). Furthermore, the proposition shows that the firm will never raise the salary high enough to retain *all* workers who would have quit otherwise: if the salary s_2 is positive in equilibrium, then there are still some people who quit even though the firm wants them to stay (in equilibrium, $\theta_p < \theta < \theta_a$).

IV. PARTIAL REVELATION

Now suppose that the principal has the option of withholding ability information. The firm can now conceal ability over finite disjoint intervals $\Theta_i \subset R$ for $i \in I$ where I is a finite set and Θ_i's are disjoint. Call these Θ_i pooling regions, since the firm pools information from all workers with ability $\theta \in \Theta_i$ together. The intervals Θ_i represent evaluation categories often seen in practice. For example, a worker may only know that his ability is "good," where "good" means a rating between 5 to 7 on a scale of 1 to 10. The firm commits to an information structure $\Theta = \{\Theta_i\}_{i \in I}$ for a finite set I at the time the contract is signed (and before ability is observed), such that if $\theta \in \Theta_k$ for some k after it has been observed by the firm, the firm tells the worker that his output lies within Θ_k. In particular, the firm cannot arbitrarily distort information revelation after the evaluation stage by telling the worker that his ability lies within some other interval Θ_j for $j \neq k$. This focuses analysis on the decision to reveal or not reveal information, but does not address the issue of arbitrary bias in disclosure, which has been explored elsewhere (Prendergast and Topel 1996).

Workers are told in which category their ability lies, so they know that $\theta \in \Theta_i$ for some i. The assumption that the Θ_i's are disjoint guarantees that any worker who is told that his ability lies within Θ_i believes that every other worker with ability within Θ_i is told the same thing. Moreover, because workers are risk-neutral and contracts are linear in θ, if they are told that $\theta \in \Theta_i$ for some i, then they will behave as if they had average productivity within the interval $\bar{\theta}_i \equiv E[\theta|\theta \in \Theta_i]$. Figure 1 illustrates the worker's inference of his ability, given an information structure Θ. Outside of the pooling regions the worker is told his true

FIGURE 1
Agent's Inference



productivity, while within each pooling region the worker assumes that he is average. Formally, this is equivalent to a firm revealing a signal that reveals true output outside of Θ and that maps each Θ_i into $\bar{\theta}_i$. However, allowing firms to select the partition Θ instead of a signaling mechanism is closer to performance categories seen in practice.

The contract is now a tuple $\langle s_1, s_2, b, T_p, \Theta\rangle$ where $\Theta \equiv \{\Theta_i\}_{i \in I}$. I assume the agent is constrained by limited liability (LL), so $s_1, s_2, b \geq 0$. Given this contract, the agent responds by choosing actions $\langle T_a, e(\cdot)\rangle$. The principal now has an additional instrument Θ that she will use to extract rents from the agent. In particular, the principal strategically selects Θ to retain agents who would otherwise prefer to leave. Limited liability restricts the contract space and hence reduces surplus. In particular, limited liability guarantees nonnegative wages to the agent, and thus passes positive rents to the agent. Since the principal holds all of the bargaining power and thus receives all of the surplus in the first best, she now withholds information to take back those rents from the agent on the retention margin.

The agent's effort function $e(\cdot)$ must be constant over each pooling interval Θ_i, since the agent cannot condition on information he does not have. Call this the Measurability Constraint (MC):

$$e(\theta) = \hat{e}_i \quad \forall \theta \in \Theta_i \quad \forall i \in I, \tag{MC}$$

where $\hat{e}_i \in R$ is the level of the function over Θ_i. Recall that the agent is told not just that $\theta \in \Theta$ but that $\theta \in \Theta_i$ for some i. Even though the agent is paid according to θ, he selects his effort with (possibly) imperfect information on θ. This has real productivity consequences for the firm, since the agent's effort choice determines output q. Pooling will affect the agent's effort choice, and, hence, affect the firm's revenue $(V - b)q - s_2 - s_1$.

Figure 2 presents the timeline of the game under partial revelation. At the outset, the principal proposes a contract $\langle T_p, s_1, s_2, b, \Theta\rangle$, which the agent accepts or rejects. If the agent rejects, then the principal and the agent get payoffs $\bar{u}_1^p$ and $\bar{u}_1^a$, respectively. Then, if the agent accepts, then θ is realized. Only the principal observes θ. If $\theta \in \Theta_i$ for some i, then the principal reveals Θ_i. Otherwise, the principal reveals θ. In the actual production stage, the agent decides to stay or leave. If he leaves, then the agent gets $\bar{u}_2^a$ while the

FIGURE 2
Timeline

Evaluation Stage

Principal proposes contract $\langle T_p, s_1, s_2, b, \Theta \rangle$.	Agent accepts/rejects the contract.	θ is realized (only principal observes it).	Principal reveals θ according to Θ.	

Production Stage

Agent stays/leaves.	Agent selects $e(\theta)$ s.t. (MC).	ε is realized. $q = \theta e(\theta) + \varepsilon$.	Agent gets $s_1 + s_2 + bq$.	Principal gets $(V - b)q - s_1 - s_2$.

principal gets $\bar{u}_2^p$. If he stays, then the agent selects effort $e(\theta)$ subject to (MC). So the agent selects $\hat{e}_i$ for all $\theta \in \Theta_i$. Next, ε is realized, and $q = \theta_i \hat{e}_i + \varepsilon$ if $\theta \in \Theta_i$, while $q = \theta e(\theta) + \varepsilon$ otherwise. Finally, the principal receives profit $(V - b)q - s_1 - s_2$ and pays the agent $s_1 + s_2 + bq$.

The principal and agent maximize the same objective functions as before, subject to the additional measurability constraint. Let $\hat{e}(\cdot)$ denote the optimal effort function solving this program. Now we can use the agent's problem to arrive at the shape of the effort function.

Proposition 5: The effort function $\hat{e}(\theta)$ is constant over each Θ_i and increasing elsewhere. In particular $\hat{e}(\theta) = e(E_\theta[\theta | \theta \in \Theta_i])$ for all $\theta \in \Theta_i$ for each $i \in I$.

The shape of the effort function determines the shape of the principal's (and agent's) payoff functions. Let $\hat{u}$ and $\hat{\pi}$ denote the agent's and principal's equilibrium continuation payoffs, respectively, under partial revelation. These payoffs are identical with the full revelation payoffs u and π outside of the pooling regions. But over the pooling regions, these payoffs are a function of e from (MC), so:

$$\hat{u}(\theta) = \begin{cases} b[\theta e(\bar{\theta}_i) + m] - C(e(\bar{\theta}_i)) + s_2 & \text{if } \theta \in \Theta_i \text{ for some } i \\ b[\theta e(\theta) + m] - C(e(\theta)) + s_2 & \text{otherwise} \end{cases}$$

where $\bar{\theta}_i = E_\theta[\theta | \theta \in \Theta_i]$ is the conditional mean of θ over the interval Θ_i. The principal's payoffs depend on the actual value θ and on the effort level, which depends on whether $\theta \in \Theta_i$ for some i:

$$\pi(\theta) = \begin{cases} (V - b)[\theta e(\bar{\theta}_i) + m] - s_2 & \text{if } \theta \in \Theta_i \text{ for some } i \\ (V - b)[\theta e(\theta) + m] - s_2 & \text{otherwise.} \end{cases}$$

Solving the principal's and agent's programs gives the shape of these payoff functions.

Proposition 6: The agent's payoff $\hat{u}(\theta)$ and the principal's payoff $\hat{\pi}(\theta)$ are both upper-semicontinuous, linear over each Θ_i and increasing and convex elsewhere. Both the agent and the principal adopt cutoff strategies for separation.

Convexity for $\hat{u}$ and $\hat{\pi}$ fails because of the discontinuity. Nevertheless, monotonicity is preserved for both functions. The proposition establishes similar properties as the full revelation case, except that now it is necessary to deal with the discontinuities in the functions from the Measurability Constraint. In particular, the separation rule still holds. The agent continues if and only if $\theta > \theta_a$ (or $\bar{\theta}_i > \theta_a$ if $\theta \in \Theta_i$), and the principal continues if and only if $\theta > \theta_p$ (or $\bar{\theta}_i > \theta_p$ if $\theta \in \Theta_i$).

Suppose the principal cannot commit to a revelation scheme. So even if the principal claims to reveal output *ex ante*, she can always reverse this decision *ex post*. Then:

Proposition 7: Without commitment, any pooling region unravels.

This holds because π is increasing. Suppose the principal claims to withhold information by pooling ability over an interval $\Theta_i = [x,y)$. Let $\bar{\theta} = E_\theta[\theta|\theta \in [x,y)]$ be average ability over this region. The agent chooses average effort $e = e(\bar{\theta})$ over this region, and the principal earns average profit $\hat{\pi}(\theta) = (V - b)[\theta\, e(\bar{\theta}) + m] - s_2$. But if the principal observes $\theta > \bar{\theta}$, then she can get $\pi(\theta) > \hat{\pi}(\theta)$ if she reveals it. Since she has no commitment, she will do so. So she, in fact, will only pool over $[x,\bar{\theta})$ and separate over $[\bar{\theta},y)$. Applying this same argument with the candidate pooling interval $[x,\bar{\theta})$ shows that the principal pools below the average of this interval, but separates above it. Repeating this argument *ad infinitum*, the pooling region unravels. See the Appendix for the full formal proof. Therefore the interesting case is when the principal can commit to a revelation strategy. I assume this in what follows.

Intermediate Targets

Recall that the residual surplus $\bar{u}_2 = \bar{u}_2^a + \bar{u}_2^p$ is the sum of the principal's and agent's outside options and represents the value to both parties of abandoning work after the evaluation stage. The logic behind the pooling equilibrium stems from two sequential results. First, the distribution of the residual surplus between the principal and agent determines the ranking of their productivity targets θ_p and θ_a, respectively. How the two parties split the residual surplus $\bar{u}_2$ determines their termination decisions. Second, the ranking of the targets determines whether the principal has an incentive to pool ability. More precisely, if there exists a region in which the principal wants to keep the agent but the agent wants to leave, then the principal will withhold information (pool ability) to retain the agent.

Definition 1: Let $\gamma \equiv \bar{u}_2^a/\bar{u}_2$ denote the agent's share of the residual surplus $\bar{u}_2$.

Because outside options satisfy $\bar{u}_t = \bar{u}_t^a + \bar{u}_t^p$, it is clear that $\gamma \in [0,1]$. High γ means the agent captures most of the residual surplus after a failed project, so γ is one measure of the distribution of residual surplus between the two parties. The distribution of residual surplus determines the ordering of the targets in a clean and intuitive way:

Proposition 8: For each $\bar{u}_2$, there exist $\gamma_1^*(\bar{u}_2), \gamma_2^*(\bar{u}_2) \in (0,1)$ such that if $\gamma > \gamma_1^*(\bar{u}_2)$, then $\theta_a > \theta^* > \theta_p$. If $\gamma < \gamma_2^*(\bar{u}_2)$, then $\theta_a < \theta^* < \theta_p$.

In words, the party who receives most of the residual surplus will set an inefficiently high target, while the other party will set an inefficiently low target. For example, suppose that $\gamma > \gamma_1^*(\bar{u}_2)$, so the agent receives most of the residual surplus. This means his outside option is high (relative to the principal), and, hence, the revealed ability must also be high in order to justify forgoing these attractive outside opportunities. He will tolerate fewer failures, since his alternatives are good, and he therefore sets a high-ability hurdle θ_a. In fact, he quits some projects that are efficient to continue, and so he sets $\theta_a > \theta^*$. Simultaneously the principal receives a small share of the residual surplus and has low outside options relative to the agent. Thus, the agent does not require a high hurdle rate to justify continuation, since her alternatives outside are weak. So she sets a low-ability hurdle, and even continues some projects that are efficient to quit, so $\theta_p < \theta^*$. Similar logic holds if $\gamma < \gamma_2^*$. This result suggests that there will be two classes of equilibria. The distribution of residual surplus will determine whether the equilibrium is separating or pooling.

V. MAIN RESULT

To build intuition behind when the principal will pool and when she will separate, first note that in general pooling is costly. To see this, suppose the principal pools over an interval Θ_i. Let $\bar{\theta}_i = E_\theta[\theta|\theta \in \Theta_i]$ be the average value over the pool. If she reveals θ, then the principal earns $\pi(\theta) = (V - b)(\theta e(\theta) + m) - s_2$, where $e(\theta)$ solves $C'(e(\theta)) = b\theta$. If she pools, she earns $\hat{\pi}(\theta) = (V - b)(\theta e(\bar{\theta}) + m) - s_2$. Since π is convex over Θ_i, Jensen's inequality shows that:

$$E_\theta[\pi(\theta)|\theta \in \Theta_i] > \pi(E_\theta[\theta|\theta \in \Theta_i]) = \pi(\bar{\theta}_i) = \hat{\pi}(\bar{\theta}_i) = E_\theta[\hat{\pi}(\theta)|\theta \in \Theta_i].$$

The last equality holds since $\hat{\pi}$ is linear. Multiply both sides by $1/Pr(\Theta_i)$:

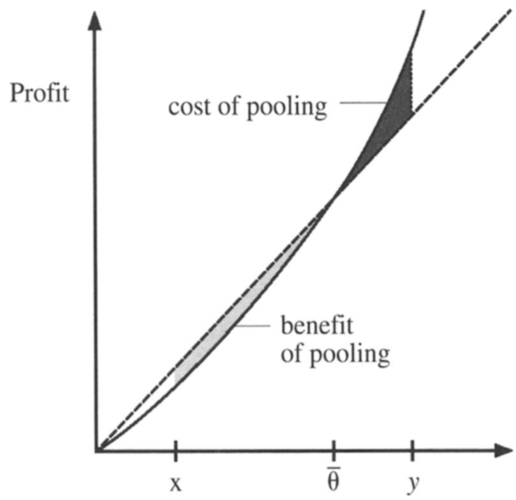
$$\int_{\Theta_i} \pi(\theta)f(\theta)d\theta > \int_{\Theta_i} \hat{\pi}(\theta)f(\theta)d\theta.$$

Hence, revealing earns more profit for the principal than pooling.

Intuitively, the principal reports only the average ability $\bar{\theta}_i = E_\theta[\theta|\theta \in \Theta]$ to everyone in the pool, and the agent chooses effort based on this report. In particular, if an agent produces $\theta > \bar{\theta}_i$, instead of choosing $e(\theta)$, as he would if he knew θ, he chooses $e(\bar{\theta}_i) < e(\theta)$, since the effort function $e(\cdot)$ is increasing. So the stars (those with $\theta > \bar{\theta}_i$) slack off and the slugs ($\theta < \bar{\theta}_i$) work harder, since both think that they are average. Profits are increasing in effort, so the principal loses money on the stars and gains on the slugs. But because the profit function is convex, the loss exceeds the gain. The principal makes so much money off the stars that the cost of telling them that they are average exceeds the benefit of telling the slugs that they are average. It is important to note the reliance on the complementarity of ability and effort. Because of this complementarity, the effort function is increasing and concave in θ, so the principal's profit function is increasing and convex, which permits the use of Jensen's inequality. So complementarity implies that the loss in output from the stars exceeds the gain in output from the slugs. This is illustrated in Figure 3, where the convex and linear dashed lines are the principal's payoffs under separating and pooling, respectively.

Therefore, because pooling is costly, the principal will never pool if she gains nothing from it. Suppose the principal receives most of the residual surplus: $\gamma < \gamma_2^*(\bar{u}_2)$. By Proposition 8, the principal sets a higher target than the agent, so $\theta_p > \theta^* > \theta_a$. If $\theta < \theta_a < \theta_p$, then both parties want the agent to quit, so there is no conflict of interest. The principal

FIGURE 3
Costs and Benefits of Pooling



can fire the agent, or, equivalently, simply reveal θ to the agent and he will quit on his own (since $\theta < \theta_a$). If $\theta > \theta_p$, then the principal wants to retain the agent, and since $\theta_p > \theta_a$, the agent also wants to stay, so again there is no conflict of interest. By the Jensen inequality argument above, pooling is costly and yields no additional benefits to the principal. So the principal will reveal θ to the agent and he will choose to stay. If ability lies at either extreme ($\theta < \theta_a < \theta_p$ or $\theta > \theta_p > \theta_a$), then the interests of both parties are aligned.

Now if $\theta \in (\theta_a,\theta_p)$, then the principal wishes to fire the agent, while the agent wishes to stay. This represents a conflict of interest. In at-will employment contracts, both parties are free to leave at any time, and so ability must clear $\max\{\theta_a, \theta_p\}$ to justify continuation. If $\theta_a < \theta_p$, then $\theta_p = \max\{\theta_a, \theta_p\}$ is the relevant hurdle. Ability fails this hurdle if $\theta_a < \theta < \theta_p$, so the principal can implement her optimal termination decision by firing the agent. Note that simply revealing θ is not sufficient (as it was earlier) because of the conflict of interest; the principal must fire the agent. Thus the principal fires if $\theta < \theta_p$ and fully reveals ability otherwise. Her payoffs are given by $\max\{\pi(\theta),\bar{u}_2^p\}$. In sum, if the principal receives most of the residual surplus, then she sets a higher ability target than the agent, and, hence, can implement her optimal termination rule without resorting to pooling.

Life is different if the tables are turned. Now suppose that the agent receives most of the residual surplus: $\gamma > \gamma_1^*(\bar{u}_2)$. By Proposition 8, the agent sets a higher target than the principal, so $\theta_a > \theta^* > \theta_p$. As before, there is alignment of interest if ability is extreme but conflict of interest otherwise. If ability is very low ($\theta < \theta_p < \theta_a$), then both parties prefer separation. The principal fires the agent, or equivalently reveals θ and the agent leaves on his own. If ability is very high ($\theta > \theta_a > \theta_p$), then both parties prefer continuation. The principal reveals θ and the agent chooses to stay. By Jensen's inequality, pooling is costly and yields no benefits to the principal.

If $\theta \in (\theta_p,\theta_a)$, the agent prefers to quit, but the principal prefers him to stay. If the principal reveals θ, then the agent will quit, leading to a suboptimal outcome for the principal. The key insight is that the principal can retain the worker by withholding information. More precisely, the principal will pool over a large enough region such that the average ability level within the pool satisfies $\bar{\theta} \geq \theta_a$. This ensures that the agent will stay on the job if he is told that he is average. Above this pooling interval, the principal will separate because pooling is costly (by Jensen's inequality). The principal pools as little as possible, only enough to keep workers on the job.

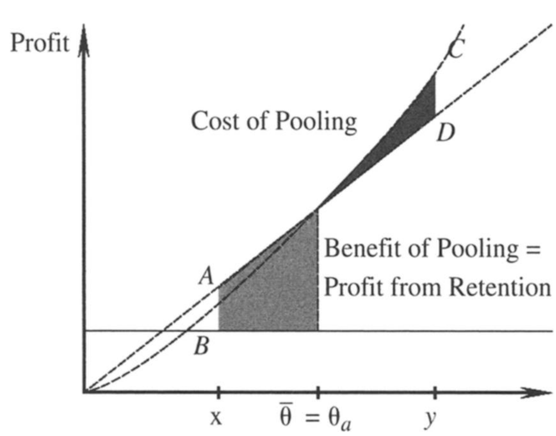
This profit is large enough that the benefits of pooling outweigh the costs. So the value of retention overturns the costs of pooling generated by the convex profit function and Jensen's inequality. In sum, the principal fires the worker if ability is sufficiently low, reveals output if it is sufficiently high, and pools if output is in-between. Collecting these results, we arrive at the main theorem. The full proof and construction of equilibrium lies in the Appendix.

Theorem 1: If $\gamma < \gamma_2^*(\bar{u}_2)$, then the principal fully reveals output. If $\gamma > \gamma_1^*(\bar{u}_2)$, then there exists a unique nontrivial pooling interval $\Theta = \{[x,y)\}$ such that the principal pools over $[x,y)$ and reveals output otherwise.

Figure 4 illustrates the costs and benefits of pooling with profit from retention. The lightly shaded region beneath the curved dashed line represents the principal's profit from retention. This is the net profit from agents who stay on the job because they believe that they are average. Observe that the principal selects the pooling region $[x,y)$ such that the average member of the pool $\bar{\theta}$ is exactly indifferent between staying and leaving ($\bar{\theta} = \theta_a$). The proof shows that the constraint $\bar{\theta} \geq \theta_a$ binds: pooling is costly and the principal does it only to keep agents between x and θ_a on the job.

The non-contractibility of ability is a critical assumption. If θ were contractible, then the principal would not pool in equilibrium. Instead, he would pay the agents who want to leave just enough to induce them to stay. Formally, if the contract now took the form $(s_t(\theta),b(\theta))$, then the principal would select $s_2(\theta)$ for all $\theta \in (\theta_p,\theta_a)$ to make those agents indifferent between staying and leaving. While this incurs the cost of extra fixed payments for the principal, it has two benefits. First, it retains agents that the principal wants to keep. Second, it does this without distorting incentives, as pooling does. Recall that pooling skews incentives since the stars work less thinking that they are average, resulting in a loss for

FIGURE 4
Profit from Retention



the principal. By adjusting salaries instead, the principal achieves her retention objective without bearing the loss in output from the incentive effect on the stars.

To summarize, if the principal receives most of the residual surplus, then she will adopt a firing rule higher than the agent's quit rule, and she will fully reveal ability to the agent. If the agent receives most of the residual surplus, then the agent's target exceeds the principal's, and so the principal will partially reveal ability to the agent. Precisely, the principal will fully reveal ability if it is sufficiently high or low, and withhold information (pool) if it is in-between. The principal withholds information as a retention mechanism.

VI. ROBUSTNESS

This section shows that the pooling equilibrium in Theorem 1 is robust under more general cost functions, ability-contingent outside options, nonverifiable output, and compensation schemes that are not linear in output. I also show that the model can be extended to a full two-stage model (where production occurs and effort choices are made in both stages).

General Cost Functions

Until now the marginal cost function was log-concave, or $C'''/C'' < C''/C'$. While this is a weak assumption satisfied by almost every cost function ever used in agency models, it is not necessary for the main result.[12] In particular, there still exists a pooling equilibrium such that the principal pools ability in order to retain workers.

Proposition 9: If $\theta_p < \theta_a$, then there exists a unique equilibrium such that the principal pools over a non-trivial pooling region Θ.

The set of pooling intervals Θ consists of all regions over which π is concave, combined with the interval $[x,y)$ defined in Theorem 1. As before, the principal will select x and y such that:

$$E_\theta[\theta|\theta \in [x,y)] = \theta_a.$$

So the principal still pools to retain workers who otherwise would have left. In addition, she pools where it is inherently profitable to do so, i.e., where the profit function is concave. Dropping the assumption that $C'''/C'' < C''/C'$ leaves open the possibility that π may be concave, which enlarges the set of pooling regions. But this will not eliminate the original pooling region $[x,y)$ from Theorem 1. So the retention effects of withholding performance information still hold under general cost functions even though the pooling region may now include intervals other than $[x,y)$.

Contingent Outside Options

In the benchmark model I assumed that the agent's outside options $\bar{u}_2^a$ did not depend on ability θ. This section shows that the main result is robust when the agent's outside options are contingent on θ if a single-crossing property holds. To see this, suppose that the agent now has an outside option *function* $\bar{u}_2^a(\theta)$, which is increasing in θ (assume that

[12] Without assuming $C'''/C'' < C''/C'$, the profit function will still be continuous but may be concave in places. By Jensen's inequality, the principal will pool wherever her profit function is concave, since the gain in output from the slugs exceeds the loss in output from the stars. Wherever the profit function is concave, the benefits of pooling outweigh the costs, and so the principal will pool.

the firm's outside options are still independent of θ). This reflects the natural assumption that higher observed ability produces higher outside opportunities for the agent. These outside opportunities can take the form of higher market wages offered to the agent by other firms. Of course, the market's information is only as good as the agent's, so if the firm pools over a region Θ_i, then the outside option will not vary over this region. In other words, the measurability constraint is now:

$$\hat{e}(\theta) = \hat{e}_i \text{ and } \hat{\bar{u}}_2^a(\theta) = \hat{\bar{u}}_1^a \quad \forall \theta \in \Theta_i \quad \forall i \in I \tag{MC$'$}$$

Therefore, the effort function and the outside option function will only vary on the information that the firm reveals to both the agent and the market.

The agent's problem under full revelation is now:

$$\max_{T_a, e(\cdot)} \int_T u(\theta) f(\theta) d\theta + \int_{T^C} \bar{u}_2^a(\theta) f(\theta) d\theta - C(e(\theta)) + s_1$$

subject to (PC), (MC$'$), where $T = T_a \cap T_p$, and T_a and T_p are the agent's and the principal's continuation rules, respectively. So if $\theta \notin T$, the agent now earns an outside option $\bar{u}_2^a(\theta)$. The principal's problem is the same as in the benchmark model.

To guarantee that the main result holds under contingent outside options, we need two relatively weak additional assumptions. The first is a single-crossing property on the agent's outside option function and the agent's equilibrium payoff function, and the second requires that the market's expectations of ability given a pooled estimate be unbiased. If these properties are satisfied, then a form of Theorem 1 still holds.

Proposition 10: Suppose that:

1. there exists a unique θ_0 such that $\bar{u}_2^a(\theta_0) = u(\theta_0)$ and $\bar{u}_2^a(\theta) > u(\theta)$ if and only if $\theta < \theta_0$; and
2. the market's beliefs are unbiased, $\hat{\bar{u}}_i^a = E_\theta[\bar{u}_2^a(\theta) | \theta \in \Theta_i]$, for all $i \in I$.

Then there exists a unique pooling interval whenever $T_p \cap (T_a)^C \neq \emptyset$ and no pooling occurs whenever $T_p \cap (T_a)^C = \emptyset$.

The single-crossing property (the first condition) is necessary to ensure that the agent will adopt a cutoff strategy. Observe that in the benchmark model, the outside option function did not vary with respect to θ and the payoff function was convex, so the single-crossing property automatically held. The result above shows that as long as the outside option functions are relatively well behaved, there still exists a unique pooling equilibrium such that the principal withholds information to retain workers. If the single-crossing property failed, then the continuation region would possibly consist of a disjoint union of intervals, and there is no guarantee that the principal would still pool over certain regions. Finally, observe that if the outside options function were weakly concave, then the single-crossing property would hold automatically, as long as $u(0) < \bar{u}_2^a(0)$, i.e., the agent chooses not to continue if $\theta = 0$. This result holds because of the convexity of the payoffs.

Nonverifiable Ability

Ability or performance within a firm can be difficult to verify. For example, when the firm observes the worker's productivity and announces this productivity to the agent, the worker may have no reason to believe the firm's announcement. In general, this opens a Pandora's Box of problems. Without imposing any constraints, reporting the true value of ability will not form an equilibrium. To see this, suppose that the worker believes the firm's report. Then the firm will not truthfully report ability, since it can simply report an ability level higher than true ability. Since the agent's effort is increasing in (the report of) his ability and the principal's profit is increasing in the agent's effort, this will cause the agent to work harder and the principal to earn more profit. Hence, the principal has an incentive to deviate, and truth-telling is not an equilibrium.[13]

Therefore, some form of commitment will be necessary in order to discipline the principal from arbitrarily distorting her report to the agent. In particular, suppose that the principal can commit to a contract on its announcement to the firm. Let $\hat{\theta}$ denote the firm's announcement of the agent's output, and $t(\hat{\theta})$ be a (nonlinear) transfer to the agent. If this transfer function is increasing, then the principal commits to paying more money to agents with higher reported output. Because the firm can commit to this contract, the agent knows that the firm will not throw money away on inflated reports.

Consider the following equilibrium. Fix a contract $\langle \theta_p, s_1, s_2, b, \Theta \rangle$ where $\Theta = \{\Theta_i\}_{i \in I}$. The principal's report to the agent is:

$$\hat{\theta} = \begin{cases} E_\theta[\theta | \theta \in \Theta_i] & \text{if } \theta \in \Theta_i \text{ for some } i \\ \theta & \text{otherwise.} \end{cases} \tag{1}$$

Suppose the agent believes the principal's report. Then at the interim stage the agent maximizes his continuation utility:

$$u(\hat{\theta}) = b(\hat{\theta}e(\hat{\theta}) + m) - C(e(\hat{\theta})) + s_2 + t(\hat{\theta}).$$

So for each $\theta > \tau = \max\{\theta_a, \theta_p\}$ the agent optimizes his effort function point-wise, yielding the F.O.C. $C'(e) = b\hat{\theta}$. This generates the agent's optimal effort, taking his beliefs as given. The agent chooses the same effort function as under verifiable output. The agent's effort function is $e(\hat{\theta})$, where $\hat{\theta}$ is given by Equation (1).

Let $\pi(\hat{\theta},\theta)$ be the principal's profit function when the true output is θ and the announcement is $\hat{\theta}$:

$$\pi(\hat{\theta},\theta) = (V - b)(\theta e(\hat{\theta}) + m) - s_2 - t(\hat{\theta}).$$

The truth-telling condition for the principal is $\theta \in \text{argmax}_{\hat{\theta}}\ \pi(\hat{\theta},\theta)$. Solving this truth-telling condition shows that the optimal transfer satisfies:

$$t'(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_i \text{ for some } i \\ (V - b)\theta e'(\theta) & \text{otherwise.} \end{cases} \tag{2}$$

[13] Other papers have tackled this issue in different ways. In Prendergast and Topel (1996) the principal disciplines the manager's report with an independent signal of performance, and there will be positive bias in the manager's report in equilibrium. Rajan and Reichelstein (2006) use subjective performance measures to slice up a fixed bonus pool that the firm commits to early in the game.

Observe that the transfer function is independent of the announcement over the pooling region. The next proposition shows that if output is nonverifiable, then it is still possible to sustain the same equilibrium as before with an appropriate transfer function.

Proposition 11: There exists an increasing and convex transfer function such that the principal pools over $\Theta = \{[x,y)\}$ if $\theta_p < \theta_a$.

The convexity of the transfer function comes from the Marginal Condition (2). This convex schedule makes the principal's announcement credible to the agent and sustains the equilibrium. Intuitively, the principal commits to making increasingly large transfers to high-ability workers. This serves as a disciplining force on the principal's announcement; since these transfers are costly, the principal will not arbitrarily distort her announcement. The agent knows this, and hence believes the principal.

Nonlinear Compensation Schemes

So far I have considered only contracts that are linear functions of the final output. While this assumption simplified the analysis, it is still natural to ask how the conclusions would change under more general compensation schemes. Given that it is essentially impossible to obtain precise characterizations of equilibria under completely general compensation schemes, it is unsurprising that there are no straightforward conditions on the primitives of the model that guarantee the existence of a pooling equilibrium in such a general setting. Remarkably, however, it is still possible to show, under fairly weak technical conditions, that the principal will choose to pool ability information to retain workers if, under the optimal contract, there are agents who want to abandon the project even though the principal would like them to stay.

In particular, let the compensation scheme $R:R_+ \rightarrow R_+$ be any continuous and differentiable function of final output.[14] A contract is a pair $\langle T_p, R(\cdot)\rangle$, where T_p is the firing rule (fire if $\theta \notin T_p$) and $R(\cdot)$ is the compensation scheme. Given this contract, the agent responds by choosing actions $\langle T_a, e(\cdot)\rangle$: a quit rule T_a (quit if $\theta \notin T_a$) and an effort function $e(\cdot)$. Let $T = T_a \cap T_p$, so the probability of advancing is $P(T) = \Pr(\theta \in T)$. Conditioning on θ, the agent's and principal's continuation utilities are:

$$u(\theta) = E_\varepsilon R(\theta e(\theta) + \varepsilon) - C(e(\theta))$$
$$\pi(\theta) = V(\theta e(\theta) + m) - E_\varepsilon R(\theta e(\theta) + \varepsilon),$$

where E_ε is the expectation with respect to ε. The agent's and the principal's optimization problems are, respectively:

$$\max_{T_a, e(\cdot)} \int_T u(\theta)f(\theta)d\theta + (1 - P(T))\bar{u}_2^a \text{ and } \max_{T_p, R(\cdot)} \int_T \pi(\theta)f(\theta)d\theta + (1 - P(T))\bar{u}_2^p.$$

Note that this implies that $T_a = \{\theta | u(\theta) \geq \bar{u}_2^a$ and $T_p = \{\theta | \pi(\theta) \geq \bar{u}_2^p\}$.

[14] While I allow compensation schemes that are general functions of final output, it is important that they do not directly depend on ability. With ability-contingent compensation schemes, Condition 2 of Proposition 12 would never be satisfied in an equilibrium: it would be profitable for the principal to deviate by paying agents in some neighborhood of θ_a just enough to induce them to stay. I thank an anonymous referee for suggesting closer inspection of ability-contingent compensation contracts.

Without further assumptions on the reward scheme, little else can be said about the equilibrium choices of the agents. In particular, we can no longer conclude that effort is increasing in ability, that payoff functions are increasing and concave, or that either party will adopt a cutoff strategy for termination.[15] Given all this uncertainty, it is remarkable that fairly weak regularity conditions are sufficient to guarantee that the main result still holds: the principal will withhold ability information in order to retain agents.

Proposition 12: Let T_p and T_a be the equilibrium continuation sets under the constraint of full revelation. Suppose the following conditions hold:

1. There exists $\theta_a > 0$ such that $T_a = [\theta_a, \infty)$.
2. $\pi(\theta_a) > \bar{u}_2^p$ and there exists some nontrivial neighborhood of θ_a, N_{θ_a}, such that $N_{\theta_a} \subset T_p$.
3. The function $y:(N_{\theta_a} \cap (-\infty,\theta_a]) \rightarrow [\theta_a,\infty)$ given by $y(t) = \inf \{x \geq t | E_\theta[u(\theta)|\theta \in [t,x)] \geq \bar{u}_2^a\}$ is well-defined, continuous and differentiable, and $|y'(\theta_a)| < \infty$.

Then, in the unconstrained equilibrium, there exists a nontrivial pooling interval $\Theta = [\alpha,\beta)$.

Condition 1 says that the agent will be using a cutoff strategy in equilibrium. I show in the proof of the proposition that a sufficient (but not necessary) condition for this is that $R' > 0$. As such, the condition is fairly weak.

Condition 2 is a formalization of the statement that the principal would like to retain some of the agents who choose to quit under full revelation. It also requires that the principal *strictly* prefers the marginal-ability agent θ_a to stay. The condition says that there is some neighborhood of θ_a such that the principal would like agents with ability in this neighborhood to stay on the project. A sufficient (though not necessary) condition for this would be that the principal adopts a cutoff strategy with $\theta_p < \theta_a$. In this case, $N_{\theta_a} = (\theta_p,\infty)$.

Condition 3 requires that it is possible to retain agents by pooling around θ_a. Given the continuity of $u(\cdot)$ and the fact that $u(\theta_a) > \bar{u}_2^a$, the first part of the condition would hold for most $u(\cdot)$. The second part of the condition is essentially a regularity condition on the distribution of ε and the cost function, requiring that the right endpoint of the minimal pooling interval does not change at an infinite speed as the interval becomes infinitesimal. The simplest example of a function satisfying Condition 3 is the $y(t)$ function under linear contracts, defined implicitly by $y'(t) = [f(t(t - \theta_a)]/[f(y(t))(y(t) - \theta_a)]$.

Under these conditions, it is profitable for the principal to pool ability information around θ_a, by the same logic as in the linear case. At θ_a, the principal's effective payoff function (i.e., payoffs taking into account the continuation decision) is right-continuous and has a positive jump under full revelation. Therefore, by pooling around θ_a, the principal's possible loss due to lower effort exerted by those to the right of θ_a is only second-order, while the gain due to participation by those to the left of θ_a is first-order, much like in Figure 4.

---

[15] Furthermore, since the expectation operator and the compensation function can no longer be exchanged (due to nonlinearity), previous arguments based on Jensen's inequality no longer hold, and it is also no longer the case that agents who are told their ability is in some Θ_i will choose their effort as if their ability were average within the interval $(\bar{\theta}_i)$.

Two-Stage Production

So far I have assumed that θ is an exogenously given ability parameter. In this subsection I extend the model to a setup where θ is first-stage output that can be influenced by the agent's choice of effort. I show that under certain conditions pooling occurs also in this setting.

Now the agent works on the project over two stages (as opposed to the main model, where the first stage was merely an evaluation period). Effort e_t is exerted in each stage $t = 1, 2$ at cost $C(e_t)$, where C' and C'' are positive. As before, both the agent and the principal are risk-neutral and there is no discounting. In each stage, the agent produces output $q_t = e_t + \varepsilon_t$. The noise terms $\varepsilon_t \in [0,\infty)$ are independently and identically distributed (i.i.d.), and distributed around a mean of m, with cdf $G(\cdot)$ and density function $g(\cdot)$. The agent observes his effort but the firm does not. Output q_t is noncontractible, but final output $q = q_1 q_2$ is contractible. The firm captures value Vq from the project. Each q_t represents stage t output, while Vq measures total project value. Thus, q_t is the project's *internal* output within the firm used for planning and evaluation purposes, while Vq measures the project's *external* value based on market prices. A good example of a production function that fits this framework is a multistage project with an initial research and development component.

After the production stage, output (q_1, q_2) is observable to both parties, but only the firm observes q_1 after the first stage. The firm can choose to reveal q_1 to the agent or not. Once revealed, performance information is *ex post* verifiable by the agent. The principal's and the agents' outside options in stage t are $\bar{u}_t^a$ and $\bar{u}_t^p$, respectively. As before, the residual surplus is defined as $\bar{u}_2 = \bar{u}_2^a + \bar{u}_2^p$ and the principal and agent must satisfy a participation constraint (PC) that their equilibrium payoffs must exceed $\bar{u}_1^p + \bar{u}_2^p$ and $\bar{u}_1^a + \bar{u}_2^a$, respectively.

Similar to the nonlinear extension of the main model, let the compensation scheme $R(q){:}R_+ \rightarrow R_+$ be any continuous and differentiable function of final output. A contract is a pair $\langle T_p, R(\cdot)\rangle$, where T_p is the firing rule (fire if $q_1 \notin T_p$) and $R(\cdot)$ is the compensation scheme. Given this contract, the agent responds by choosing actions $\langle T_a, e_1, e_2(\cdot)\rangle$: a quit rule T_a (quit if $q_1 \notin T_a$), first-stage effort e_1, and second-stage effort function $e_2(\cdot)$. Let $T = T_a \cap T_p$. The probability of advancing is $P(T) = \Pr(q_1 \in T|e_1)$. Conditioning on q_1, the agent's and principal's continuation utilities are:

$$u(q_1) = E_2 R(q_1(e_2(q_1) + \varepsilon_2)) - C(e_2(q_1))$$
$$\pi(q_1) = Vq_1(e_2(q_1) + m) - E_2 R(q_1(e_2(q_1) + \varepsilon)),$$

where E_2 is the expectation with respect to ε_2. The agent's and the principal's optimization problems are, respectively:

$$\max_{T_a, e_1, e_2(\cdot)} \int_T u(q_1) g(q_1 - e_1) dq_1 + (1 - P(T))\bar{u}_2^a + s_1$$

$$\max_{T_p, R(\cdot)} \int_T \pi(q_1) g(q_1 - e_1) dq_1 + (1 - P(T))\bar{u}_2^p - s_1.$$

Note that this implies that $T_a = \{q_1 | u(q_1) \geq \bar{u}_2^a\}$ and $T_p = \{q_1 | \pi(q_1) \geq \bar{u}_2^p\}$.

Proposition 13: Let T_p and T_a be the equilibrium continuation sets under the constraint of full revelation. Suppose the following conditions hold:

1. There exist $q_a > 0$ and $q_p > 0$ such that $T_a = [q_a, \infty)$ and $T_p = [q_p,\infty)$, and $q_p < q_a$. Furthermore, $\pi'(q_1) > 0$, for all $q_1 > 0$.
2. The function $y:(q_p,q_a] \rightarrow [q_a,\infty)$ given by $y(t) = \inf\{x \geq t | E[u(q_1)|q_1 \in [t,x)] \geq \bar{u}_2^a\}$ is well-defined, continuous, and differentiable, and $|y'(q_a)| < \infty$.
3. The function $e_1(t)$ that gives the optimal stage-one effort as a function of the pooling interval $[t,y(t))$ satisfies $e'_1(q_a) \leq 0$.

Then, there will be a non-trivial pooling interval $Q = [t, y(t))$.

The conditions for this proposition are similar to those in Proposition 12 except we now require that *both* agents use cutoff strategies, and we have added an additional condition that deals with the effect of pooling on first-stage effort. In this case, pooling can affect payoffs even off the pooling interval, since pooling can cause the agent to adjust his first-stage effort. This complication was absent from the single-stage model, since ability was given exogenously. The last condition says that pooling on an infinitesimal interval around q_a will not decrease the first-stage effort. Without this, we would not be able to reach the same result, because pooling could then negatively affect first-stage effort and, thus, payoffs both on and off the pooling region. Consequently, the gains from pooling could be annihilated.

Note that the proof of Proposition 13 does not depend on the assumption that outside options are independent of the level of first-stage output. Therefore, the result automatically extends to the case when the principal's and the agent's outside options are increasing functions of first-stage output, $\bar{u}_2^a(q_1)$ and $\bar{u}_2^p(q_1)$, respectively. That is, a more-productive early stage produces higher outside opportunities for the firm and the agent. Since Conditions 1 and 2 of the proposition depend on the forms of the outside options functions, $E[u(q_1)|q_1 \in [t,x)] \geq \bar{u}_2^a$ from Condition 2 becomes $E[u(q_1) - \bar{u}_2^a(q_1)|q_1 \in [t,x)] \geq 0$.

It turns out that in the special case of linear contracts of the form $\langle b, s_1, s_2\rangle$ (and constant outside options), as considered in the central part of this paper, the conditions of the proposition hold whenever $q_a > q_p$. Furthermore, Proposition 8 still holds, so $q_a > q_p$ whenever the agent receives a sufficiently large fraction of the residual surplus.[16]

VII. CONCLUSION

The compensation and management literature has extensively documented that firms withhold performance evaluation information from employees. Most studies cite psychological and political reasons for withholding performance evaluation information. Managers do not wish to depress morale in the workplace, and as Bjerke et al. (1987) found upon interviewing Navy officers, there are in fact psychological incentives for supervisors to withhold performance evaluation information, primarily to secure esteem-building promotions for junior officers. Managers also withhold information to avoid complicated interfirm politics, as noted by Mitchell and O'Reilly (1983). When possible, managers deter tension and animosity between managers and subordinates as well as among subordinates.

This paper advances an economic theory based on strategic retention. The key ingredients in the model here are complementarity in production (between ability and output)

[16] Details and a formal proof are available from the author on request.

and outside options at the interim stage. These assumptions show that in equilibrium, the firm pools workers together if their output is mediocre and reveals performance if their output is at either extreme. The size of the pooling interval trades off the benefits of pooling from retention against the cost of pooling from lost output. As such, this paper offers an economic explanation of a phenomenon that was previously justified on psychological and political grounds.

APPENDIX

Proof of Proposition 1

By pooling, the planner must restrict effort functions to be constant over pooled regions. Therefore by not pooling, the planner maximizes over a larger set (all real-valued functions), and is therefore (weakly) better off. ■

Proof of Proposition 2

Let $X \equiv \{\theta : S(\theta) > \bar{u}_2\}$ be the social planner's continuation set: it is efficient to allow a worker with ability θ to work on the project if and only if $\theta \in X$. Let $P = \Pr(X)$. I will show that X is in fact an interval. Continuation surplus is:

$$S(\theta) = \max_{e(\theta)}\{EV(\theta e + \varepsilon) - C(e)\} = \max_{e(\theta)}\{V(\theta e(\theta) + m) - C(e(\theta))\}.$$

Note that the F.O.C. for the optimal choice of $e(\theta)$ is $V\theta = C'(e(\theta))$. Since $V\theta > V \cdot 0 = 0 = C'(0)$ and C' is increasing, we know that $e(\theta) > 0$. Also note that the F.O.C. implies that $\lim_{\theta\to\infty} e(\theta) = \infty$ and $\lim_{\theta\to 0} e(\theta) = 0$.

Now, by the envelope theorem, $S'(\theta) = Ve(\theta) > 0$, so that $S(\cdot)$ is strictly increasing. Furthermore, $S(\cdot)$ is continuous, and $\lim_{\theta\to\infty} S(\theta) = \infty$ and $\lim_{\theta\to 0} S(\theta) = Vm - C(0) < Vm < \bar{u}_2$. Therefore, the intermediate value theorem implies that there exists θ^* such that $S(\theta^*) = \bar{u}_2$. Since $S(\cdot)$ is strictly increasing, this implies that $S(\theta) \geq \bar{u}_2$ if and only if $\theta > \theta^*$. That is, $X = [\theta^*, \infty)$. ■

Proof of Proposition 3

The agent solves:

$$\max_{T_a, e(\cdot)} \int_T u(\theta) f(\theta) d\theta + (1 - P(T))\bar{u}_2^a + s_1 \text{ subject to (PC)}$$

where $P(T) = \Pr(\theta \in T)$ and $u(\theta) = b(\theta e(\theta) + m) + s_2 - C(e(\theta))$.

For now, let us assume that $0 < b < V$; I will show later in this proof that this will indeed be the case under the optimal contract. Let $\langle e^*(\cdot), T_a^*\rangle$ denote the optimal values. Note that $T_a^* = \{\theta | u(\theta) > \bar{u}_2^a\}$.

The F.O.C. with regard to $e(\cdot)$ gives $C'(e^*(\theta)) = b\theta$ for all $\theta > \tau$. Hence e^* is a function of both θ and b, so write $e^*(\theta,b)$. By the implicit function theorem:

$$\frac{\partial e^*}{\partial \theta}(\theta,b) = \frac{b}{C''(e^*)} > 0 \quad \text{and} \quad \frac{\partial e^*}{\partial b}(\theta,b) = \frac{\theta}{C''(e^*)} > 0.$$

So for any $b > 0$, $e^*(\cdot)$ is increasing in b and θ.

Finally, note that the F.O.C. implies that $e^*(\theta) > 0$: $C'(e(\theta)) = b\theta > 0 = C'(0)$.

Agent's Payoffs

Let $u(\theta,b)$ be the agent's payoff function under the optimal choice of effort. Note that $u(\cdot,\cdot)$ is continuous in both arguments. Furthermore, for any $b > 0$, $u(\cdot,b)$ is an increasing and convex function of θ, since, by the implicit function theorem and the F.O.C.:

$$\frac{\partial u}{\partial \theta}(\theta,b) = be^*(\theta) > 0 \quad \text{and} \quad \frac{\partial^2 u}{(\partial \theta)^2}(\theta,b) = b\frac{\partial e^*}{\partial \theta}(\theta,b) > 0.$$

On the other hand, for any $\theta > 0$, $u(\theta,\cdot)$ is an increasing function of b, since, again by the implicit function theorem and the F.O.C.:

$$\frac{\partial u}{\partial b}(\theta,b) = \theta e^*(\theta,b) + m > 0.$$

Continuity, monotonicity, and convexity of u w.r.t. θ imply the existence of a cutoff strategy, as an application of the intermediate value theorem. In fact, for any $b > 0$, $\lim_{\theta \downarrow 0} u(\theta,b) = bm - C(0) + s_2$ and $\lim_{\theta \uparrow \infty} u(\theta,b) = \infty$.

If $bm - C(0) + s_2 \geq \bar{u}_2^a$, then, since u is increasing in θ, we know that $u(\theta,b) \geq \bar{u}_2^a$ for all θ, so the agent always chooses to work on the project. That is, $T_a = [0,\infty)$. However, Proposition 4 will show that this is not possible in equilibrium.

If $bm - C(0) + s_2 < \bar{u}_2^a$, then the results above imply that there exist two points x, y such that $0 < x < y$ and $u(x,b) < \bar{u}_2^a < u(y,b)$. Therefore, by continuity and monotonicity, there exists a unique $\theta_a(b) > 0$ such that $u(\theta,b) \geq \bar{u}_2^a$ if and only if $\theta \geq \theta_a(b)$. Such a $\theta_a(b;\bar{u}_2^a)$ corresponds to the agent's optimal quit rule and satisfies the marginal condition $u(\theta_a(b;\bar{u}_2^a),b) = \bar{u}_2^a$. Consequently, $T_a = [\theta_a(b;\bar{u}_2^a),\infty)$. Notice that $\theta_a(\cdot;\bar{u}_2^a)$ is decreasing in b, since by the implicit function theorem and the F.O.C.:

$$\frac{\partial \theta_a}{\partial b}(b;\bar{u}_2^a) = -\frac{\theta_a(b;\bar{u}_2^a)e(\theta_a(b;\bar{u}_2^a)) + m}{be(\theta_a(b;\bar{u}_2^a))} < 0.$$

Similar calculations show that $\theta_a(b;\cdot)$ is an increasing and concave function of $\bar{u}_2^a$.

Principal's Payoffs

Let $\pi(\theta,b) \equiv (V - b)(\theta e(\theta) + m) - s_2$. For any $b > 0$, $\pi(\cdot,b)$ is a continuous and measurable function. This function is increasing in θ, since:

$$\frac{\partial \pi}{\partial \theta}(\theta,b) = (V - b)\left(e(\theta,b) + \theta\frac{\partial e}{\partial \theta}(\theta,b)\right) > 0.$$

Furthermore, under our assumptions on $C(\cdot)$, $\pi(\theta,b)$ is convex in θ for all $b > 0$:

$$\frac{\partial^2\pi}{(\partial\theta)^2}(\theta,b) = (V - b)\left(2\,\frac{\partial e}{\partial\theta}(\theta,b) + \theta\,\frac{\partial^2 e}{\partial\theta^2}(\theta,b)\right),$$

$$\frac{\partial^2\pi}{(\partial\theta)^2}(\theta,b) > 0 \Longleftrightarrow 2\,\frac{\partial e}{\partial\theta}(\theta,b) + \theta\,\frac{\partial^2 e}{(\partial\theta)^2}(\theta,b) > 0$$

$$\Longleftrightarrow -\frac{\theta\,\dfrac{\partial^2 e}{(\partial\theta)^2}(\theta,b)}{\dfrac{\partial e}{\partial\theta}(\theta,b)} < 2.$$

By the F.O.C. and the implicit function theorem:

$$-\frac{\theta\dfrac{\partial^2 e}{(\partial\theta)^2}(\theta,b)}{\dfrac{\partial e}{\partial\theta}(\theta,b)} = \frac{C'(e(\theta,b))C'''(e(\theta,b))}{(C''(e(\theta,b)))^2}.$$

Recall that $C'''/C'' < C''/C'$. Thus, the term in the right-hand side is less than 1, and hence, *a fortiori*, less than 2. Therefore, $\frac{\partial^2\pi}{(\partial\theta)^2}(\theta,b) > 0$.

Since π is increasing and convex in θ, we know that $\lim_{\theta\to\infty}\pi(\theta,b) = \infty$. Since $e(0) = 0$, $\lim_{\theta\to 0}\pi(\theta,b) = (V - b)m - s_2 < Vm < \bar{u}_2^p$. Thus, by continuity and the intermediate value theorem, there exists $\theta_p(b;\bar{u}_2^p)$ such that $\pi(\theta_p(b;\bar{u}_2^p),b) = \bar{u}_2^p$. Since π is strictly increasing in θ, this $\theta_p(b;\bar{u}_2^p)$ is unique, and $\pi(\theta,b) \geq \bar{u}_2^p$ if and only if $\theta \geq \theta_p(b;\bar{u}_2^p)$. Thus $T_p = [\theta_p(b;\bar{u}_2^p),\infty)$, so the principal will employ a cutoff strategy. Straight computation yields $\frac{\partial\theta_p}{\partial\bar{u}_2^p}(b;\bar{u}_2^p) > 0$ and $\frac{\partial^2\theta_p}{(\partial\bar{u}_1^p)^2}(b;\bar{u}_2^p) < 0$. Thus $\theta_p(b;\cdot)$ is an increasing and concave function of $\bar{u}_2^p$.

The Equilibrium Choice of b

It was assumed above that $0 < b < V$. I will now show that the principal will never choose $b = 0$ or $b = V$. First note that the results obtained under the assumption $0 < b < V$ showed that for each such value of b there exists an ability level $\tilde{\theta}(b) = \max\{\theta_p(b),\theta_a(b)\}$ such that for all $\theta > \tilde{\theta}$ we have $\pi(\theta) > \bar{u}_p^2$ and $u(\theta) > \bar{u}_a^2$. As long as these values occur with positive probability (a sufficient condition for this would be that $f(\theta) > 0$ for all $\theta > 0$), the expected profit to the principal under *any* $b \in (0,V)$ is greater than $\bar{u}_2^p$.

On the other hand, if the principal chooses $b = V$, then her continuation payoff for any ability level θ is $-s_2 \leq 0 < \bar{u}_p^2$. Thus, the principal always terminates the project and her expected profit is $\bar{u}_2^p$, which is strictly less than the profit she would have obtained by adopting *any* $b \in (0,V)$. Therefore, $b = V$ is never chosen.

Similarly, if the principal chooses $b = 0$, then we know from the first-order condition of the agent that she will choose $e(\theta,0) = 0$ for any θ. As a result, the principal's continuation payoff will be $Vm - s_2 \leq Vm < \bar{u}_2^p$, and she will again choose to terminate. Thus, her expected profit in this case is again only $\bar{u}_2^p$, so that this choice is strictly dominated by any $b \in (0,V)$. Consequently, $b = 0$ is never chosen. ■

Proof of Proposition 4

Note that s_1 does not affect the agent's choice of effort and, since it is paid regardless of the continuation decision, it also does not influence the termination rules. Hence, it has only a direct effect on the principal's and the agent's payoffs, strictly decreasing the former and increasing the latter, given that the agent does accept the contract. Thus, the only case in which the principal would choose to pay $s_1 > 0$ would be if this were necessary to satisfy the agent's participation constraint, i.e., to induce the agent to enter the contractual relationship. By the same reason, it is also clear that the principal will never raise the salary above the level at which the agent's participation constraint just binds.

As for the second part of the proposition, suppose there is an equilibrium in which $s_2 > 0$ and $\theta_p \geq \theta_a$. The principal's objective function is:

$$E\Pi(b,s_1,s_2,\theta_p) = \int_{\theta_p}^{\infty} \pi(\theta)f(\theta)d\theta + F(\theta_p)\bar{u}_2^p - s_1,$$

where $\pi(\theta) = (V - b)(\theta e(\theta) + m) - s_2$. Since we have already observed that s_2 does not affect the effort choice, we know that $\frac{\partial \pi}{\partial s_2} = -1$. Therefore:

$$\frac{\partial}{\partial s_2} E\Pi(b,s_1,s_2,\theta_p) = \int_{\theta_p}^{\infty} \frac{\partial \pi}{\partial s_2}(\theta)f(\theta)d\theta = F(\theta_p) - 1 < 0.$$

Consequently, it is profitable for the principal to decrease s_2 (and this is feasible without violating limited liability, since $s_2 > 0$). This means that the proposed s_2, θ_a, and θ_p cannot be part of an equilibrium. This proves the proposition.

Returning briefly to the agent's problem in selecting θ_a, we can now see that $\theta_a = 0$ will never occur in equilibrium. Recall that $\theta_a = 0$ would occur if and only if $bm - C(0) + s_2 \geq \bar{u}_2^a$. But, since $\theta_p > 0 = \theta_a$, the current proposition implies that $s_2 = 0$. Therefore, $bm - C(0) + s_2 = bm - C(0) < Vm < \bar{u}_2^a$. Contradiction. Consequently, $\theta_a > 0$ and so it must satisfy the condition $u(\theta_a) = \bar{u}_2^a$. ■

Proof of Proposition 5

The agent's continuation payoff is:

$$\hat{u}(\theta) = \begin{cases} b[\theta\hat{e}_i + m] - C(\hat{e}_i) + s_2 & \text{if } \theta \in \Theta_i \text{ for some } i \\ b[\theta e(\theta) + m] - C(e(\theta)) + s_2 & \text{otherwise.} \end{cases}$$

Therefore, under a bonus b, the agent solves $\max_{e(\theta)}\hat{u}(\theta)$.

Outside of the pooling region the agent solves the same problem as in the full revelation case. Thus, we have that the optimal solution is given by the effort function $e(\theta,b)$ from FR, described in Proposition 3.

Within the pooling region, the agent solves a different problem. From (MC) the agent's effort function must be constant over each Θ_i, so he earns $\hat{u}(\theta) = b\theta\hat{e}_i - C(\hat{e}_i) + s_2$ over Θ_i, where $\hat{e}_i$ is the single effort choice over Θ_i. Since the agent's information set is Θ_i, his expected utility is $E_\theta[\hat{u}(\theta)|\Theta_i] = b\bar{\theta}_i\hat{e}_i - C(\hat{e}_i) + s_2$. The F.O.C. shows that for each $i \in I$, $C'(\hat{e}_i) = b\bar{\theta}_i$, defining the optimal effort level on each pooling region. So $C'(\hat{e}_i)$

$= C'(e(\bar{\theta}_i))$, which implies that $\hat{e}_i = e(\bar{\theta}_i)$ where $e(\cdot)$ is the effort function from FR. Combined with Step One, the best response function is defined over the whole domain and is:

$$\hat{e}(\theta,b) = \begin{cases} e(\bar{\theta}_i,b) & \text{if } \theta \in \Theta_i \text{ for some } i \in I \\ e(\theta,b) & \text{otherwise.} \end{cases}$$

This function, although discontinuous at each boundary point of the pooling regions, is a (weakly) increasing function of θ. By Proposition 3, $e(\cdot,b)$ is a strictly increasing function of θ and $\hat{e}(\theta,b)$ is constant over each Θ_i, so it is sufficient to prove that at each discontinuity point the function $\hat{e}(\cdot,b)$ has a positive jump. Formally, by the mean value theorem, for each $i \in I$, $\inf \Theta_i < \bar{\theta}_i < \sup\Theta_i$. Since $e(\cdot,b)$ is continuous and, for each $b > 0$, strictly increasing:

$$\lim_{\theta\uparrow\inf\Theta_i} \hat{e}(\theta,b) = e(\inf \Theta_i,b) < e(\bar{\theta}_i,b) = \lim_{\theta\downarrow\inf\Theta_i} \hat{e}(\theta,b)$$
$$\lim_{\theta\uparrow\sup\Theta_i} \hat{e}(\theta,b) = e(\bar{\theta}_i,b) < e(\sup \Theta_i,b) = \lim_{\theta\downarrow\sup\Theta_i} \hat{e}(\theta,b).$$

Again, monotonicity confirms the idea that the second-stage effort is positively related with the signal. Under the assumption $\Theta_i = [\alpha_i,\beta_i)$ for some $i \in I$, the above computation also shows that $\hat{e}(\cdot,b)$ is an upper-semicontinuous function of θ, since, for all $q \geq 0$, $\limsup_{\theta\to q} \hat{e}(\theta,b) \leq \hat{e}(q,b)$.

By Proposition 3, we also have that the effort level is increasing in b:

$$\frac{\partial \hat{e}}{\partial b}(\theta,b) = \frac{\sigma(\theta)}{C''(\hat{e})} > 0 \quad \text{for } \theta > 0, \text{ where}$$
$$\sigma(\theta) = \begin{cases} E_\theta[\theta|\theta \in \Theta_i] & \text{if } \theta \in \Theta_i \text{ for some } i \\ \theta & \text{otherwise.} \end{cases}$$

■

Proof of Proposition 6

Agent's Payoff

The agent's payoff is given by:

$$\hat{u}(\theta) = \begin{cases} b[\theta e(\bar{\theta}_i) + m] - C(e(\bar{\theta}_i)) + s_2 & \text{if } \theta \in \Theta_i \text{ for some } i \\ b[\theta e(\theta) + m] - C(e(\theta)) + s_2 & \text{otherwise} \end{cases}$$

where $\bar{\theta}_i = E_\theta[\theta|\theta \in \Theta_i]$. The function is clearly linear and increasing in θ over each Θ_i. Off the pooling intervals, the function is equal to $u(\theta)$, which we have shown to be increasing and convex in θ. Finally, since $\hat{e}(\theta,b)$ is an upper-semicontinuous, measurable, and weakly increasing function of θ, $\hat{u}(\theta)$ is also upper continuous, measurable, and *strictly* increasing over its entire domain.

It is also obvious that it is optimal for the agent to use a cutoff strategy based on the same cutoff value θ_a that was defined in the full revelation case. Consider two cases:

1. If θ is revealed to the agent, then he should quit if and only if $\bar{u}_2^a > \hat{u}(\theta) = u(\theta)$, which holds if and only if $\theta < \theta_a$, as shown before.

2. If Θ_i is revealed, the agent should quit if and only if $\bar{u}_2^a > E_\theta[\hat{u}(\theta)|\theta \in \Theta_i] = b(\bar{\theta}_i e(\bar{\theta}_i) + m] - C(e(\bar{\theta}_i)) = u(\bar{\theta}_i)$, which holds if $\bar{\theta}_i < \theta_a$.

Principal's Payoff

The principal's payoff function is:

$$\hat{\pi}(\theta) = \begin{cases} (V - b)[\theta e(\bar{\theta}_i) + m] - s_2 & \text{if } \theta \in \Theta_i \text{ for some } i \\ (V - b)[\theta e(\theta) + m] - s_2 & \text{otherwise.} \end{cases}$$

The function is clearly linear and increasing in θ over each Θ_i. Off the pooling intervals, the function is equal to $\pi(\theta)$, which we have shown to be increasing and convex in θ. Finally, since $\hat{e}(\theta,b)$ is an upper-semicontinuous, measurable, and weakly increasing function of θ, $\hat{\pi}(\theta)$ is also upper continuous, measurable, and *strictly* increasing over its entire domain.

Also, by the same logic as in the agent's case, the principal should terminate the contract if and only if $\sigma(\theta) < \theta_p$, where:

$$\sigma(\theta) = \begin{cases} E_\theta[\theta|\theta \in \Theta_i] & \text{if } \theta \in \Theta_i \text{ for some } i \\ \theta & \text{otherwise.} \end{cases}$$

■

Proof of Proposition 7

Suppose the principal pools over some $\Theta_i = [x,y)$. Construct a sequence of sets F_n such that $F_0 \equiv \Theta_i = [x,y)$, $\bar{\theta}_n = E_\theta[\theta|\theta \in F_n]$, and $F_{n+1} = [x,\bar{\theta}_n)$.

Lemma 1. *If the principal pools over F_n, then without commitment he pools over F_{n+1} and separates over $F_n \backslash F_{n+1}$.*

Proof. (By induction) *Step One.* Her profit over F_0 is the linear function $\hat{\pi}_0(\theta) = (V - b)[\theta e(\bar{\theta}_0) + m] - s_2$ for all $\theta \in F_0$. Since $e(\cdot)$ is increasing, $\pi(\theta) = (V - b)[\theta e(\theta) + m] - s_2 > (V - b)[\theta e(\bar{\theta}_0) + m] - s_2 = \hat{\pi}(\theta)$ for all $\theta > \bar{\theta}_0$. So the principal prefers to reveal θ for all $\theta \in [\bar{\theta}_0,y) = F_0 \backslash F_1$ and pool for all $\theta \in [x,\bar{\theta}_0) = F_1$.

Step n. Her profit over F_n is the linear function $\hat{\pi}_n(\theta) = (V - b)[\theta e(\bar{\theta}_n) + m] - s_2$ for all $\theta \in F_n$. Since $e(\cdot)$ is increasing, $\pi(\theta) = (V - b)[\theta e(\theta) + m] - s_2 > (V - b)[\theta e(\bar{\theta}_n) + m] - s_2 = \hat{\pi}(\theta)$ for all $\theta > \bar{\theta}_n$. So the principal prefers to reveal θ for all $\theta \in [\bar{\theta}_n,y) = F_n \backslash F_{n+1}$ and pool for all $\theta \in [x,\bar{\theta}_n) = F_{n+1}$.

If the principal pools over $F_0 \equiv \Theta_i$, then from the lemma, without commitment the principal pools only over F_{n+1} for each n. But $F_{n+1} \subset F_n$ and $\bar{\theta}_n \rightarrow x$ so $F_n \rightarrow \emptyset$. So the principal does not pool over any interval, and hence not on any finite disjoint union of intervals. So any pooling region Θ unravels. ■

Proof of Proposition 8

Let $\tilde{u}(\theta) = \tilde{b}(\theta e(\theta,\tilde{b}) + m) - C(e(\theta,\tilde{b})) + \tilde{s}_2$ and $\tilde{\pi}(\theta) = (V - \tilde{b})(\theta e(\theta,\tilde{b}) + m) - \tilde{s}_2$ be the payoff functions in fully revealing equilibrium. It was proved above that $\tilde{u}'(\theta) > 0$ and $\tilde{\pi}'(\theta) > 0$ and that $\tilde{u}(\theta_a) = \bar{u}_2^a = \gamma \bar{u}_2$ and $\tilde{\pi}(\theta_p) = \bar{u}_2^p = (1 - \gamma)\bar{u}_2$. Therefore:

$$\theta_a < \theta^* \Longleftrightarrow \gamma \bar{u}_2 = \bar{u}_2^a = \tilde{u}(\theta_a) < \tilde{u}(\theta^*) \Longleftrightarrow \gamma < \hat{\gamma}_1(\bar{u}_2) \equiv \frac{\tilde{u}(\theta^*)}{\bar{u}_2}.$$

$$\theta_p < \theta^* \Longleftrightarrow (1 - \gamma)\bar{u}_2 = \bar{u}_2^p = \tilde{\pi}(\theta_p) < \tilde{\pi}(\theta^*) \Longleftrightarrow \gamma > \tilde{\gamma}_2(\bar{u}_2) \equiv 1 - \frac{\pi(\theta^*)}{\bar{u}_2}.$$

Thus, if $\gamma > \gamma_1^*(\bar{u}_2) \equiv \max\{\hat{\gamma}_1(\bar{u}_2),\hat{\gamma}_2(\bar{u}_2)\}$, then $\theta_a > \theta^* > \theta_p$. And if $\gamma < \gamma_2^*(\bar{u}_2) \equiv \min\{\hat{\gamma}_1(\bar{u}_2),\hat{\gamma}_2(\bar{u}_2)\}$, then $\theta_a < \theta^* < \theta_p$. Finally, note that $\gamma_1^*(\bar{u}_2) \in (0,1)$, because $\tilde{u}(\theta)$, $\tilde{\pi}(\theta) > 0$ and $\tilde{u}(\theta) + \tilde{\pi}(\theta) = V(\theta e(\theta,\tilde{b}) + m) - C(e(\theta,\tilde{b})) < V(\theta e(\theta,V) + m) - C(e(\theta,V)) = S(\theta)$, where the inequality holds because $e(\theta,V)$ is the unique maximizer of $V(\theta e + m) - C(e)$. Therefore, $\tilde{u}(\theta^*)$, $\tilde{\pi}(\theta^*) < S(\theta^*) = \bar{u}_2$. ■

Proof of Theorem 1

Recall that the agent's and the principal's payoffs under full revelation, given participation at θ, are:

$$u(\theta) = b(\theta e(\theta) + m) - C(e(\theta)) + s_2$$

$$\pi(\theta) = (V - b)(\theta e(\theta) + m) - s_2.$$

Let $\Theta_i \subset R$ be a candidate pooling interval. Let $\bar{\theta}_i = E_\theta[\theta|\theta \in \Theta_i]$ be the average ability over Θ_i. As in the proof of Proposition 5, the agent's and principal's payoffs over the pool, given participation, are:

$$\hat{u}(\theta) = b(\theta e(\bar{\theta}_i) + m) - C(e(\bar{\theta}_i)) + s_2$$

$$\hat{\pi}(\theta) = (V - b)(\theta e(\bar{\theta}_i) + m) - s_2.$$

As shown previously, the project will be carried out if and only if $\sigma(\theta) > \max\{\theta_a,\theta_p\}$, where $\sigma(\theta) = \theta$ under full revelation and $\sigma(\theta) = \bar{\theta}_i$ under pooling.

Let the principal's actual payoffs (taking into account the participation decisions) be given by $\Pi^R(\theta)$ under revelation and $\Pi^P(\theta)$ under pooling. Note that pooling does not affect any decisions for $\theta \notin \Theta_i$, so that $\Pi^P(\theta) = \Pi^R(\theta)$ for all $\theta \notin \Theta_i$. For $\theta \in \Theta_i$, we have:

$$\Pi^R(\theta) = \begin{cases} \pi(\theta) & \text{if } \theta \geq \max\{\theta_a,\theta_p\} \\ \bar{u}_2^p & \text{otherwise.} \end{cases} \qquad \Pi^P(\theta) = \begin{cases} \hat{\pi}(\theta) & \text{if } \bar{\theta}_i \geq \max\{\theta_a,\theta_p\} \\ \bar{u}_2^p & \text{otherwise.} \end{cases}$$

Clearly, pooling is profitable if and only if $E_\theta\Pi^P(\theta) > E_\theta\Pi^R(\theta)$. Since $\Pi^P(\theta) = \Pi^R(\theta)$ for all $\theta \notin \Theta_i$, pooling is profitable if and only if $E_\theta[\Pi^P(\theta)|\Theta_i] > E_\theta[\Pi^R(\theta)|\Theta_i]$. In what follows, it will be important to note that, by Jensen's inequality, $E_\theta[\pi(\theta)|\Theta_i] > \pi(E_\theta[\theta|\Theta_i]) = \pi(\bar{\theta}) = \hat{\pi}(\bar{\theta}) = E_\theta[\hat{\pi}(\theta)|\Theta_i]$, where the last equality holds since $\hat{\pi}$ is linear in θ.

Part One

Let $\theta_a < \theta_p$, so $\theta_p = \max\{\theta_a, \theta_p\}$ and $\Pi^R(\theta) = \max\{\pi(\theta),\bar{u}_2^P\}$.
Consider cases:

1. If $\bar{\theta}_i < \theta_p$, then $E_\theta[\Pi^P(\theta)|\Theta_i] = E_\theta[\bar{u}_2^p|\Theta_i] \leq E_\theta[\max\{\bar{u}_2^p,\pi(\theta)\}|\Theta_i] = E_\theta[\Pi^R(\theta)|\Theta_i]$, and the inequality is strict if $\Pr\{\Theta_i \cap (\theta^p,\infty)\} > 0$. Thus, it never pays to pool.
2. If $\bar{\theta}_i > \theta^p$, then $E_\theta[\Pi^P(\theta)|\Theta_i] = E_\theta[\hat{\pi}(\theta)|\Theta_i] < E_\theta[\pi(\theta)|\Theta_i] \leq E_\theta[\Pi^R(\theta)|\Theta_i]$, where the strict inequality is just the conclusion from the Jensen's theorem argument above, and the weak inequality holds because $\Pi^R(\theta) = \max\{\bar{u}_2^p,\pi(\theta)\}$. Consequently, pooling is *strictly* worse than revelation.

Therefore, if $\theta_a < \theta_p$, the principal will not pool over any interval, and hence over any finite union of disjoint intervals: $\Theta = \emptyset$. This proves part one.

Part Two

Now let $\theta_p < \theta_a$, so $\theta_a = \max\{\theta_a, \theta_p\}$. We can immediately make two observations:

1. If $\Theta_i \subset (q_a,\infty)$, then $E_\theta[\Pi^P(\theta)|\Theta_i] = E_\theta[\hat{\pi}(\theta)|\Theta_i] < E_\theta[\pi(\theta)|\Theta_i] = E_\theta[\Pi^R(\theta)|\Theta_i]$, so that there will be no pooling.
2. If $\bar{\theta}_i < \theta_a$, then $\Pi^P(\theta) = \bar{u}_2^p$ and

$$\Pi^R(\theta) = \begin{cases} \bar{u}_2^p & \text{if } \theta < \theta_a \\ \pi(\theta) \geq \pi(\theta_a) > \pi(\theta_p) = \bar{u}_2^p & \text{otherwise.} \end{cases}$$

The inequalities above hold because π is increasing and $\theta_a < \theta_p$. Consequently, $\Pi^R(\theta) \geq \bar{u}_2^p$, and the inequality is strict if $\theta > \theta_a$. It follows that $E_\theta[\Pi^P(\theta)|\Theta_i] = E_\theta[\bar{u}_2^p|\Theta_i] \leq E_\theta[\Pi^R(\theta)|\Theta_i]$, and the inequality is strict if $\Pr\{\Theta_i \cap (\theta^a,\infty)\} > 0$. Thus, it never pays to pool.

Therefore, if $\Theta = \{\Theta_i\}_{i\in I}$ is a profitable pooling region, then it must be the case that, for all $i \in I$, $\Theta_i \cap (-\infty,\theta_a] \neq \emptyset$ and $\bar{\theta}_i \geq \theta_a$. This implies that $\theta_a \in \Theta_i$. Since all of the Θ_i are disjoint intervals, it follows that I is a singleton, i.e., $\Theta = \{[x,y)\}$, for some $y > x > 0$ such that $\bar{\theta} = E_\theta[\theta|\theta \in [x,y)] \geq \theta_a > \theta_p$ and $\theta_a \in [x,y)$. Since Θ is a singleton, I will now change notation slightly and denote $\Theta = [x,y)$.

Now, the principal's payoffs under pooling over $[x,y)$ are:

$$\Pi^P(\theta;x,y) = \begin{cases} \bar{u}_2^p & \text{if } \theta < x \\ \hat{\pi}(\theta) & \text{if } \theta \in [x,y) \\ \pi(\theta) & \text{if } \theta \geq y. \end{cases}$$

The principal's payoffs under revelation are $\Pi^R(\theta) = \Pi^P(\theta;q_a,q_a)$.

Thus, in choosing the optimal $[x,y)$, the principal solves:

$$\max_{x,y} \int_0^x \bar{u}_2^p f(\theta)d\theta + \int_x^y \hat{\pi}(\theta)f(\theta)d\theta + \int_y^\infty \pi(\theta)f(\theta)d\theta - \lambda(\theta_a - \bar{\theta}), \tag{A1}$$

which is the Lagrangian with the constraint $\bar{\theta} \geq \theta_a$. Now $\Pr(\Theta) = \Pr(x < \theta < y) = F(y) - F(x)$, and $\bar{\theta} \equiv E_\theta[\theta|\Theta] = \dfrac{1}{F(y) - F(x)} \displaystyle\int_x^y \theta f(\theta)d\theta$.

Taking derivatives and simplifying yields:

$$\frac{\partial\bar{\theta}}{\partial y} = \frac{f(y)}{\Pr(\Theta)}(y - \bar{\theta}) \quad \text{and} \quad \frac{\partial\bar{\theta}}{\partial x} = \frac{f(x)}{\Pr(\Theta)}(\bar{\theta} - x).$$

Differentiate (A1) with regard to y, x:

$$\hat{\pi}(y)f(y) - \pi(y)f(y) + \lambda\frac{\partial\bar{\theta}}{\partial y} = 0$$

$$\bar{u}_2^p f(x) - \hat{\pi}(x)f(x) + \lambda\frac{\partial\bar{\theta}}{\partial x} = 0.$$

So:

$$\Pr(\Theta)\,\frac{\pi(y) - \hat{\pi}(y)}{y - \bar{\theta}} = \lambda = \Pr(\Theta)\,\frac{\hat{\pi}(x) - \bar{u}_2^p}{\bar{\theta} - x} > 0.$$

Thus, the constraint binds, and $\bar{\theta} = \theta_a$. We can therefore define implicitly $y(x)$:

$$\theta_a = \frac{1}{F(y(x)) - F(x)} \int_x^{y(x)} \theta f(\theta)d\theta$$
$$y'(x) = \frac{f(x)(x - \theta_a)}{f(y)(y - \theta_a)} < 0.$$

Let $\alpha = \lim_{x\uparrow\theta_a} y'(x)$. Noticing that $\lim_{x\uparrow\theta_a} y(x) = \theta_a$, we have, by de l'Hopital:

$$\alpha = \lim_{x\uparrow\theta_a} \frac{f'(x)(x - \theta_a) + f(x)}{(f'(y(x))(y(x) - \theta_a) + f(y(x)))y'(x)} = \frac{f(\theta_a)}{f(\theta_a)\alpha} = \frac{1}{\alpha},$$

as long as $f'(\theta_a) < \infty$. Thus, $\alpha = 1/\alpha$, and therefore $\lim_{x\uparrow\theta_a} y'(x) = -1$.

Since $\Pi^R(\theta) = \Pi^P(\theta;x,y(x))$ outside the pooling region, we know that $E_\theta[\Pi^P(\theta;x,y(x))] > E_\theta[\Pi^R(\theta)]$ if and only if $E_\theta[\Pi^P(\theta;x,y(x))|\Theta] > E_\theta[\Pi^R(\theta)|\Theta]$. Let $S(x) \equiv E_\theta[\Pi^P(\theta;x,y(x))|\Theta] - E_\theta[\Pi^R(\theta)|\Theta]$ be the total surplus gained by pooling over $[x,y)$ (as opposed to revealing fully). Some pooling will occur if and only if $\exists x < \theta_a$ such that $S(x) > 0$.

The surplus $S(x)$ is made up by a gain from those with below-average ability and a loss from those with above-average ability, $S(x) = \mu(x) - \lambda(x)$ where:

$$\mu(x) = \frac{1}{F(y(x)) - F(x)} \int_x^{\theta_a} (\hat{\pi}(\theta) - \bar{u}_2^p)f(\theta)d\theta > 0,$$
$$\lambda(x) = \frac{1}{F(y(x)) - F(x)} \int_{\theta_a}^{y(x)} (\pi(\theta) - \hat{\pi}(\theta))f(\theta)d\theta > 0.$$

The net surplus is given by:

$$\begin{aligned} S(x) &= \mu(x) - \lambda(x) \\ &= \frac{1}{F(y(x)) - F(x)} \left(\int_x^{y(x)} \hat{\pi}(\theta)f(\theta)d\theta - \bar{u}_2^p \int_x^{\theta_a} f(\theta)d\theta - \int_{\theta^a}^{y(x)} \pi(\theta)f(\theta)d\theta \right) \\ &= \frac{1}{F(y(x)) - F(x)} \left(\int_x^{y(x)} \hat{\pi}(\theta)f(\theta)d\theta - \bar{u}_2^p(F(\theta_a) - F(x)) - \int_{\theta_a}^{y(x)} \pi(\theta)f(\theta)d\theta \right). \end{aligned}$$

We want to show that it is profitable to pool, i.e., that $\exists x < \theta_a{:}S(x) > 0$. Notice that this is equivalent to proving:

$$\exists x < \theta_a : \frac{1}{F(y(x)) - F(x)} \int_x^{y(x)} \hat{\pi}(\theta)f(\theta)d\theta > \frac{\bar{u}_2^p(F(\theta_a) - F(x)) + \int_{\theta_a}^{y(x)} \pi(\theta)f(\theta)d\theta}{F(y(x)) - F(x)}.$$

By linearity of $\hat{\pi}$, this is equivalent to:

$$\exists x < \theta_a : \hat{\pi}(\theta_a) > \frac{\bar{u}_2^p(F(\theta_a) - F(x)) + \int_{\theta_a}^{y(x)} \pi(\theta)f(\theta)d\theta}{F(y(x)) - F(x)},$$

where the left-hand side is positive and independent of x. Now, define the function

$$\phi : (0,\theta_a) \rightarrow R$$

$$\phi(t) = \frac{\bar{u}_2^p(F(\theta_a) - F(t)) + \int_{\theta_a}^{y(t)} \pi(\theta)f(\theta)d\theta}{F(y(t)) - F(t)}.$$

We have that ϕ is continuous and, by de l'Hopital:

$$\begin{aligned}
\lim_{t\uparrow\theta_a} \phi(t) &= \lim_{t\uparrow\theta_a} \frac{\bar{u}_2^p(F(\theta_a) - F(t)) + \int_{\theta_a}^{y(t)} \pi(\theta)f(\theta)d\theta}{F(y(t)) - F(t)} \\
&= \lim_{t\uparrow\theta_a} \frac{-\bar{u}_2^p f(t) + y'(t)\pi(y(t))g(y(t))}{y'(t)f(y(t)) - f(t)} \\
&= \lim_{t\uparrow\theta_a} \frac{\pi(y(t))(\theta_a - t) + \bar{u}_2^p(y(t) - \theta_a)}{y(t) - t} \\
&\equiv A - B,
\end{aligned}$$

where the last equality follows by the definition of $y'(x)$. Now:

$$A = \lim_{t\uparrow\theta_a} \frac{\pi(y(t))(\theta_a - t) + \pi(y(t))(y(t) - \theta_a)}{y(t) - t} = \lim_{t\uparrow\theta_a} \pi(y(t)) = \pi(\theta_a) = \hat{\pi}(\theta_a)$$

$$\begin{aligned}
B &= \lim_{t\uparrow\theta_a}\left\{(\pi(y(x)) - \bar{u}_2^p) \cdot \frac{y(t) - \theta_a}{y(t) - t}\right\} \\
&= (\pi(\theta_a) - \bar{u}_2^p) \cdot \lim_{t\uparrow\theta_a} \frac{y(t) - \theta_a}{y(t) - t} \\
&= (\pi(\theta_a) - \bar{u}_2^p) \cdot \lim_{t\uparrow\theta_a} \frac{y'(t)}{y'(t) - 1} \\
&= \frac{1}{2}(\pi(\theta_a) - \bar{u}_2^p) > 0,
\end{aligned}$$

where that last equality follows by $\lim_{t\uparrow\theta_a} y'(t) = -1$. The inequality holds, because $\pi(\theta_a) > \pi(\theta_p) = \bar{u}_2^p$. Thus, $\lim_{t\uparrow\theta_a} \phi(t) = A - B < A = \hat{\pi}(\theta_a)$. By continuity, this means $\exists x < \theta_a$: $\phi(x) < \hat{\pi}(\theta_a)$. Hence, pooling is profitable. ■

Proof of Proposition 9

The equilibrium is by construction. Let Z_j be the intervals over which π is concave, for $j \in J$. If π is convex everywhere, then $J = \emptyset$, and Theorem 1 will hold. Otherwise, since π is concave over Z_j, Jensen's inequality shows that:

$$E_\theta[\pi(\theta)|\theta \in Z_j] < \pi(E_\theta[\theta|\theta \in Z_j]) = \pi(\bar{\theta}_j) = \hat{\pi}(\bar{\theta}_j) = E_\theta[\hat{\pi}(\theta)|\theta \in Z_j]$$

where $\bar{\theta}_j = E_\theta[\theta|\theta \in T_j]$. Multiply both sides by $1/\Pr(Z_j)$ to see that the principal earns more profit by pooling than by revealing output.

The same analysis from the proof of Theorem 1 applies here, so $[x,y)$ will still be a pooling interval where $E_\theta[\theta|\theta \in [x,y)] = \theta_a$. Therefore, let $\Theta = \{Z_j\}_{j\in J} \cup \{[x,y)\}$. It is clear that the principal will pool over Θ and reveal ability outside of Θ. ■

Proof of Proposition 10

Note that by the single-crossing property the agent will adopt a cutoff continuation strategy under full revelation, i.e., $T_a = [\theta_a, \infty)$, where $\theta_a = \theta_0$. This is true because, given θ, it is profitable to continue if and only if $u(\theta) > \bar{u}_2^a(\theta)$. Furthermore, note that the proofs from Proposition 3 that $e'(\theta) > 0$, $u'(\theta) > 0$, $u''(\theta) > 0$, $\pi'(\theta) > 0$, and $\pi''(\theta) > 0$ are still valid in this new setting, since they do not depend on outside options. Furthermore, since π is still increasing and convex, Proposition 3's proof that the principal will also adopt a cutoff strategy remains valid as well. Thus, $T_p = [\theta_p, \infty)$.

The proofs of Propositions 4 and 5 also hold. In particular, effort over pooling regions is given by $\hat{e}(\theta) = e(\bar{\theta}_i)$, which is constant over pooling intervals, strictly increasing elsewhere, and exhibits positive jumps at the endpoints of pooling intervals.

Proposition 6 holds in modified form. In particular, observe that:

1. If θ is revealed to the agent, then he should quit if and only if $\bar{u}_2^a(\theta) > u(\theta)$, which holds if and only if $\theta < \theta_a$, as shown before.
2. If Θ_i is revealed, the agent should quit if and only if $\bar{u}_2^i > E_\theta[\hat{u}(\theta)|\theta \in \Theta_i]$. Given that the market's beliefs are unbiased, $\bar{u}_2^i = E_\theta[\bar{u}_2^a(\theta)|\theta \in \Theta_i]$, so that the condition becomes: quit if and only if $E_\theta[\bar{u}_2^a(\theta) - \hat{u}(\theta)|\theta \in \Theta_i] > 0$.

The principal still adopts a straightforward cutoff strategy based on q_p, as in Proposition 6. Finally, Proposition 7 still clearly holds, but Proposition 8 has no obvious analog.

We are now ready to prove the main result. The proof will be based on the proof of Theorem 1. First, consider the case when $T_p \cap (T_a)^C = \emptyset$. This is equivalent to $\theta_a \leq \theta_p$. We see that the situation is equivalent to part one of Theorem 1, the proof of which still remains valid. Thus there is no pooling.

Next, consider the case when $T_p \cap (T_a)^C \neq \emptyset$, i.e., $\theta_a > \theta_p$. Let Θ_i be a candidate pooling interval. Now, the condition that $\Theta_i \cap (-\infty,\theta_a] \neq \emptyset$ from the proof of Theorem 1 still holds. Next, the condition that $\bar{\theta}_i \geq \theta_a$ from the proof of Theorem 1 now becomes $E_\theta[u(\theta) - \bar{u}_2^a(\theta)|\theta \in \Theta_i] \geq 0$. Together, these conditions imply that $\theta_a \in \Theta_i$ and therefore that there is at most one profitable pooling interval, $\Theta = [x,y)$.

The agent's problem becomes:

$$\max_{x,y} \int_0^x \bar{u}_2^p f(\theta)d\theta + \int_x^y \hat{\pi}(\theta)f(\theta)d\theta + \int_y^\infty \pi(\theta)f(\theta)d\theta - \lambda \int_x^y [u(\theta) - \bar{u}_2^a(\theta)]f(\theta)d\theta, \tag{A2}$$

which is the Lagrangian with the constraint $E[u(\theta) - \bar{u}_2^a(\theta)|\theta \in [x,y)] \geq 0$. Just like in the proof of the original theorem, the first-order conditions show that $\lambda \neq 0$, and therefore:

$$\int_x^y [u(\theta) - \bar{u}_2^a(\theta)] f(\theta) d\theta = 0.$$

Consequently, we can define $y(x)$ implicitly by this relation. By the implicit function theorem:

$$y'(x) = \frac{f(x)(u(x) - \bar{u}_2^a(x))}{f(y)(u(y) - \bar{u}_2^a(y))} < 0.$$

Noting that, by single crossing, $u(\theta_a) = \bar{u}_2^a(\theta_a)$ and $u'(\theta_a) \neq (\bar{u}_2^a)'(\theta_a)$, we can essentially repeat the same argument as in the main theorem to show that $\lim_{x \uparrow \theta_a} y'(x) = -1$.

All that remains to be shown to complete the proof is that pooling over some nontrivial interval of the form $[t, y(t))$ will indeed be profitable. Let $\Pi(t)$ be the expected net profit of the principal under pooling over $[t, y(t))$:

$$\Pi(t) = \int_0^t \bar{u}_2^p f(\theta) d\theta + \int_t^{y(t)} \hat{\pi}(\theta) f(\theta) d\theta + \int_{y(t)}^{\infty} \pi(\theta) f(\theta) d\theta.$$

Observe that, since Π is continuous, the principal will pool about θ_a if $\lim_{x \uparrow \theta_a} \Pi'(x) < 0$.

Now, noting that $\lim_{x \uparrow \theta_a} y'(x) = -1$, $\lim_{x \uparrow \theta_a} \hat{\pi}(x) = \lim_{x \uparrow \theta_a} \pi(x) = \pi(\theta_a) > \bar{u}_2^p$, we get $\lim_{x \uparrow \theta_a} \Pi'(x) = f(\theta_a)(\bar{u}_2^p - \pi(\theta_a)) < 0$. Thus, pooling is profitable. ■

Proof of Proposition 11

Let $\langle \theta_p, s_1, s_2, b, \Theta \rangle$ be a contract where $\Theta = [x, y)$ is the pooling region in Theorem 1. Recall that:

$$t'(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_i \text{ for some } i \\ (V - b)\theta e'(\theta) & \text{otherwise.} \end{cases} \quad \text{(A3)}$$

If there is no pooling, then $t(\theta)$ is fully determined by (A3). Only the intercept must be determined but evidently it is profitable to set it to zero ($t(\theta_a) = 0$) and hence:

$$t^R(\theta) = \int_{q^a}^{\theta} (V - b)\omega e'(\omega) f(\omega) d\omega$$

is the transfer function under revelation. The derivative of t in the interval $[y, \infty)$ satisfies the same condition (A3), but the intercept is zero for all θ in $[x, y)$. So the transfer under pooling is:

$$t^P(\theta) = \begin{cases} 0 & \text{if } \theta \in [x, y) \equiv \Theta \\ \int_y^{\theta} (V - b)\omega e'(\omega) d\omega & \text{otherwise.} \end{cases}$$

Note that on the interval $[y,\infty)$ the function e is the same both for pooling and no pooling. (It solves the same F.O.C.) Evidently, for $\theta > y$ we will have that $t^P(\theta) = t^R(\theta) - t^R(y)$. So:

$$\int_x^\infty \{t^P(\theta) - t^R(\theta)\} f(\theta) d\theta < 0.$$

Hence, pooling over $[x,y)$ will be more profitable than no pooling at all.

From (A3), the transfer function is clearly increasing. Totally differentiating (A3):

$$t''(\theta) = (V - b)(e'(\theta) + \theta e''(\theta)), \text{ so}$$

$$t''(\theta) = (V - b)e'(\theta)\left(1 - C'''(e(\theta))\frac{b\theta}{(C''(e(\theta)))^2}\right)$$

$$= (V - b)e'(\theta)\left(1 - C'''(e(\theta))\frac{C'(e(\theta))}{(C''(e(\theta)))^2}\right) > 0.$$

Since $C'''/C'' < C''/C'$ and $e'(\theta) > 0$. So the transfer function is convex. ■

Proof of Proposition 12

Recall that, under full revelation, the continuation payoffs are:

$$u(\theta) = E_\varepsilon R(\theta e(\theta) + \varepsilon) - C(e(\theta))$$

$$\pi(\theta) = V(\theta e(\theta) + m) - E_\varepsilon R(\theta e(\theta) + \varepsilon).$$

Note that, since $e(\theta)$ is chosen optimally by the agent, the envelope theorem says that $u'(\theta) = E_\varepsilon[R'(\theta e(\theta) + \varepsilon)e(\theta)]$, so that a sufficient condition for condition 1 of the proposition is that $R' > 0$, for this condition would ensure $u'(\theta) > 0$.

Under pooling over some interval $[\alpha,\beta)$, effort has to be constant (say, $\hat{e}(\alpha,\beta)$) over the interval, so, on that interval, payoffs become:

$$\hat{u}(\theta;\alpha,\beta) = E_\theta[E_\varepsilon R(\theta\hat{e}(\alpha,\beta) + \varepsilon)|\theta \in [\alpha,\beta)] - C(\hat{e}(\alpha,\beta))$$

$$\hat{\pi}(\theta;\alpha,\beta) = V(\theta\hat{e}(\alpha,\beta) + m) - E_\varepsilon R(\theta\hat{e}(\alpha,\beta) + \varepsilon).$$

Note that both π and $\hat{\pi}$ are continuous and differentiable in θ.

The optimal effort under pooling is determined by the condition:

$$E_\theta[E_\varepsilon R'(\theta\hat{e}(\alpha,\beta) + \varepsilon)\theta|\theta \in [\alpha,\beta)] = C'(\hat{e}(\alpha,\beta)),$$

which shows that $\hat{e}(\alpha,\beta)$ is continuous in both arguments and therefore that $\hat{\pi}$ is also continuous in α and β. We also see that $\hat{e}(\theta,\theta) = e(\theta)$ and therefore that $\hat{\pi}(\theta,\theta,\theta) = \pi(\theta)$, for all θ.

Now, consider pooling over some interval $[t, y(t))$, where $t \in N_{\theta_a}$ and $y(\cdot)$ is the function from condition 3. Let $\Pi^R(\theta)$ be the principal's effective (participation-adjusted) payoff under full revelation, while $\Pi^P(\theta;t)$ is her effective payoff under pooling over $[t,y(t))$. Note that, for any t, $\Pi^R(\theta) = \Pi^P(\theta;t)$ for all $\theta \notin [t,y(t))$, and that for all $\theta \in [t,y(t))$, we have:

$$\Pi^R(\theta) = \begin{cases} \pi(\theta) & \text{if } \theta \geq \theta_a \\ \bar{u}_2^p & \text{otherwise.} \end{cases} \qquad \Pi^P(\theta;t) = \hat{\pi}(\theta;t,y(t))$$

Let $\Delta(t) \equiv E_\theta \Pi^P(\theta,t,y(t)) - E_\theta \Pi^R(\theta)$. Note that it is profitable to pool over $[t,y(t))$ if and only if $\Delta(t) > 0$. Now:

$$\Delta(t) = \int_t^{\theta_a} \{\hat{\pi}(\theta;t,y(t)) - \bar{u}_2^p\} f(\theta)d\theta - \int_{\theta_a}^{y(t)} \{\pi(\theta) - \hat{\pi}(\theta;t,y(t))\} f(\theta)d\theta.$$

Therefore, $\Delta(\theta_a) = 0$ and:

$$\begin{aligned}\Delta'(\theta_a) &= -(\hat{\pi}(\theta_a,\theta_a,\theta_a) - \bar{u}_2^p)f(\theta_a) - (\pi(\theta_a) - \hat{\pi}(\theta_a,\theta_a,\theta_a))f(\theta_a)y'(\theta_a) \\ &= (\bar{u}_2^p - \pi(\theta_a))f(\theta_a) < 0,\end{aligned}$$

where we have used the facts that $\hat{\pi}(\theta_a,\theta_a,\theta_a) = \pi(\theta_a)$, that $|y'(\theta_a)| < \infty$ and that $\pi(\theta_a) > \bar{u}_2^p$. Therefore, by continuity of Δ, there exists $t < \theta_a$ such that $\Delta(t) > 0$, i.e., there is a nontrivial profitable pooling region. ■

Proof of Proposition 13

Consider pooling over the interval $[t,y(t))$, where $y(\cdot)$ is as given in the statement of the proposition. Let $\hat{\pi}(q_1,t)$ be the principal's continuation payoffs over the pool. By continuity of $R(\cdot)$, $e_2\,(q_1)$ is also continuous, and so is $\hat{\pi}$. Therefore $\hat{\pi}(q_a,q_a) = \pi(q_a)$.

Let $\Pi(t)$ be the expected net profit of the principal under pooling over $[t,y(t))$, and let $e_1(t)$ be the agent's choice of first-period effort, under this pooling:

$$\begin{aligned}\Pi(t) = &\int_0^t \bar{u}_2^p g(q_1 - e_1(t))dq_1 + \int_t^{y(t)} \hat{\pi}(q_1,t)g(q_1 - e_1(t))dq_1 \\ &+ \int_{y(t)}^{\infty} \pi(q_1)g(q_1 - e_1(t))dq_1.\end{aligned}$$

Observe that a principal will pool about q_a if $\Pi'(q_a) < 0$. Now:

$$\begin{aligned}\Pi'(q_a) = &\, g(q_a - e_1(q_a))[\bar{u}_2^p - \hat{\pi}(q_a,q_a) + y'(q_a)(\hat{\pi}(q_a,q_a) - \pi(q_a))] \\ &- e_1'(q_a)\left[\int_0^{q_a} g'(q_1 - e_1(q_a))\bar{u}_2^p dq_1 + \int_{q_a}^{\infty} g'(q_1 - e_1(q_a))\pi(q_1)dq_1\right].\end{aligned}$$

Substituting in $\hat{\pi}(q_a,q_a) = \pi(q_a)$ and integrating by parts:

$$\begin{aligned}\Pi'(q_q) = &\, g(q_a - e_1(q_a))(\bar{u}_2^p - \pi(q_a)) \\ &- e_1'(q_a)[\bar{u}_2^p g(q_a - e_1(q_a)) - \pi(q_a)g(q_a - e_1(q_a)) \\ &- \int_{q_a}^{\infty} \pi'(q_1)g(q_1 - e_1(q_q))dq_1] \\ = &\, g(q_a - e_1(q_a))(\bar{u}_2^p - \pi(q_a)) - e_1'(q_a)[(u_2^{-p} - \pi(q_a))g(q_a - e_1(q_a)) \\ &- \int_{q_a}^{\infty} \pi'(q_1)g(q_1 - e_1(q_a))dq_1] < 0,\end{aligned}$$

where we have used the facts that $\pi'(q_1) > 0$, $\pi(q_a) > \pi(q_p) = \bar{u}_2^p$, and $e_1'(q_a) \leq 0$.

Since $\Pi'(q_a) < 0$, pooling will occur about q_a. ■

REFERENCES

Arvey, R. D., and A. P. Jones. 1985. The use of discipline in organizations: A framework for future research. In *Research in Organizational Behavior*, edited by L. Cummings, and B. Staw, 7. Greenwich, CT: JAI Press.

Bagnoli, M., and T. Bergstrom. 2005. Log-concave probability and its applications. *Economic Theory* 26 (2): 445–469.

Baiman, S. 1990. Agency research in managerial accounting: A second look. *Accounting, Organizations and Society* 15 (4): 341–371.

Bernardin, H. J., and R. W. Beatty. 1984. *Performance Appraisal: Assessing Human Behavior at Work*. Boston, MA: Kent Publishing Company.

Bjerke, D., J. Cleveland, R. Morrison, and W. Wilson. 1987. *Officer Fitness Report Evaluation Study*. Navy Personnel Research and Development Center Report TR 88-4. San Diego, CA: Navy Personnel Research and Development Center.

Cleveland, J. N., R. Morrison, and D. Bjerke. 1986. Rater intentions in appraisal ratings: Male-violent manipulation or functional fudging. Working paper, Society for Industrial and Organizational Psychology, Chicago, IL.

Darrough, M. N., and N. M. Stoughton. 1989. Financial disclosure policy in an entry game. *Journal of Accounting and Economics* 12: 219–243.

Dye, R. A. 1985. Disclosure of nonproprietary information. *Journal of Accounting Research* 23 (1): 123–145.

Gibbons, R., and K. J. Murphy. 1992. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy* 100 (3): 468–505.

Grossman, S. J. 1981. The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics* 24 (3): 461–483.

———, and O. D. Hart. 1980. Disclosure laws and takeover bids. *Journal of Finance* 35 (2): 323–344.

Holmstrom, B. 1999. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies* 66 (1): 169–182.

Jones, A. P., M. Tait, and M. C. Butler. 1983. Perceived punishment and reward values of supervisory actions. *Motivation and Emotion* 1 (September): 313–329.

Jung, W. O., and Y. K. Kwon. 1988. Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research* 26 (1): 146–153.

Kay, E., H. Meyer, and J. French. 1965. Effects of threat in a performance interview. *Journal of Applied Psychology* 49 (December): 311–317.

Lambert, R. A. 2001. Contracting theory and accounting. *Journal of Accounting and Economics* 32 (1–3): 3–87.

Landy, F. J., and J. L. Farr. 1980. Performance rating. *Psychological Bulletin* 87 (January): 72–107.

Lizzeri, A., M. A. Meyer, and N. Persico. 2002. The incentive effects of interim performance evaluations. Working paper, University of Pennsylvania, Philadelphia, PA.

Longenecker, C. O., H. P. Sims, and D. A. Gioia. 1987. Behind the mask: The politics of employee appraisal. *Academy of Management Executive* 1: 183–193.

MacLeod, W. B. 2003. Optimal contracting with subjective evaluation. *American Economic Review* 93 (1): 216–240.

Meyer, M. A., and J. Vickers. 1997. Performance comparisons and dynamic incentives. *Journal of Political Economy* 105 (3): 547–581.

Milgrom, P. R. 1981. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics* 12 (2): 380–391.

Milkovich, G. T., and A. K. Wigdor, eds. 1991. *Pay for Performance: Evaluating Performance Appraisal and Merit Pay*. Washington, D.C.: National Academy Press.

———, and J. M. Newman, eds. 1996. *Compensation*. Chicago, IL: McGraw-Hill/Irwin.

Mitchell, T. R., and C. A. O'Reilly. 1983. Managing poor performance and productivity in organizations. In *Research in Personnel and Human Resources Management*, edited by K. Rowland, and G. Ferris, 1. Greenwich, CT: JAI Press.

Mohrman, A. M., and E. E. Lawler. 1983. Motivation and performance appraisal behavior. In *Performance Measurement and Theory*, edited by F. Landy et al., 173–189. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Murphy, K. R., and J. N. Cleveland. 1991. *Performance Appraisal: An Organizational Perspective*. Needham Heights, MA: Allyn & Bacon.

Newman, P., and R. Sansing. 1993. Disclosure policies with multiple users. *Journal of Accounting Research* 31 (1): 92–112.

Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura. 1990. Strategic information revelation. *The Review of Economic Studies* 57 (1): 25–47.

Prendergast, C., and R. H. Topel. 1996. Favoritism in organizations. *Journal of Political Economy* 104 (5).

Rajan, M. V., and S. J. Reichelstein. 2006. Subjective performance indicators and discretionary bonus pools. *Journal of Accounting Research* 44 (3): 585–617.

Saal, F. E., R. G. Downey, and M. A. Lahey. 1980. Rating the ratings: Assessing the quality of rating data. *Psychological Bulletin* 88: 413–428.

———, and P. A. Knight. 1988. *Industrial/Organizational Psychology: Science and Practice*. Pacific Grove, CA: Brooks/Cole.

Schall, M. S. 1983. A communication-rules approach to organizational culture. *Administrative Science Quarterly* 28 (December): 557–581.

Shin, H. S. 1994. News management and the value of firms. *The RAND Journal of Economics* 25 (1): 58–71.

Verrecchia, R. E. 1983. Discretionary disclosure. *Journal of Accounting and Economics* 5: 179–194.

———. 2001. Essays on disclosure. *Journal of Accounting and Economics* 32: 97–180.