# Performance Feedback with Career Concerns

Stephen E. Hansen*
Department of Economics, Universitat Pompeu Fabra and Barcelona GSE

This article examines the incentive effects of interim performance evaluation when a worker has career concerns and effort is history dependent. Disclosure has two effects: it increases the variance of future effort, and it allows the worker to use current effort to influence his employer's belief about future effort, creating a ratchet effect. The article provides necessary and sufficient conditions for full disclosure to dominate no disclosure; shows that the optimal disclosure policy reveals output realizations in the center of the distribution, but not in the tails; and discusses the potential implications of the results for the analysis of performance appraisal systems. (*JEL* D82, D86, L20)

## 1. Introduction

In many organizational settings, workers have a fixed amount of time in which to convince their employers they are sufficiently talented to deserve a promotion. Examples include junior associates in consulting and law firms and assistant professors in academia. In such organizations, workers' career concerns are a primary source of incentives. At the same time, the evaluators of such workers are often better able to judge workers' performance than the workers themselves. For example, senior professors are probably better placed to judge the potential of new research projects than their junior colleagues. The disclosure of performance information at an interim stage—after the workers' initial performance has been observed but before the end of the evaluation period—is a common feature in these

*Universitat Pompeu Fabra, 25-27 Ramon Trias Fargas, Barcelona 08005, Spain. Email: stephen.hansen@upf.edu

This article is a modified version of the second chapter of my PhD dissertation submitted to the London School of Economics and Political Science and was previously circulated under the title "The Benefits of Limited Feedback in Organizations." I am exceedingly grateful to my supervisors Gilat Levy and Andrea Prat for their support and encouragement throughout my doctoral degree, as well as to the Bank of England for its financial support. Oliver Denk, Selina Gann, Luis Garicano, Mike Gibbs, and Karl Schlag also contributed much of their time in helping to improve the article. In addition, the article (in its current incarnation) has benefited from the advice of Heski Bar-Isaac, Tim Besley, Pablo Casas, Joan de Martí, Wouter Dessein, Nathan Foley-Fisher, Yuk-Fai Fong, Clare Leaver, Matthew Gentzkow, Fabrizio Germano, Ian Jewitt, Jin Li, Michael McMahon, Hannes Müller, Marco Ottaviani, Canice Prendergast, Ronny Razin, and three anonymous referees. The usual disclaimer applies.

*The Journal of Law, Economics, and Organization*, Vol. 29, No. 6
doi:10.1093/jleo/ews032
Advance Access published on October 18, 2012

organizations.[1] While such disclosure most likely plays a variety of roles, this article focuses on a particular one that has not yet been explored: it examines performance evaluation as a tool for modifying career concerns incentives. As such, it uncovers both the effects of interim disclosure in a principal–agent relationship with career concerns, and identifies circumstances under which disclosure increases surplus.

The model is an adaptation of Holmström (1999). A worker (whom the paper refers to as *he*) exerts effort for two periods and his employer (whom the paper refers to as *she*) privately observes his performance. If the employer concludes on the basis of the worker's performance that his ability is higher than a threshold $\theta^*$, then the worker gets a promotion, which is worth $W$ to him. Prior to the first period, the employer can commit[2] to a disclosure policy that either reveals to the worker the exact value of his first period performance before his second period effort choice, or else gives coarser feedback.

If the employer tells the worker that his ability is near the tenure threshold, i.e., his promotion hangs in the balance, then he works hard in the second period; if she tells him that his ability is a way above or below the threshold, he exerts little effort in the second period since any new information is not likely to sway the opinion of the employer. With quadratic effort costs, these effects on second period effort exactly offset each other, and expected effort is the same regardless of the feedback the worker gets. On the other hand, feedback does change expected effort costs. Revealing more information increases the variance of second period effort, which the worker dislikes since effort costs are convex. The article terms this the *effort risk* effect of information disclosure. Looking only at the second period effort decision, the agent would like to remain ignorant and be "insured" against effort variation.

The novel effect of information disclosure comes via a *ratchet effect* that influences first period effort. One can understand it in three steps. First, the worker would like the employer to believe that he is not exerting much effort in the second period. The lower this belief is, the more the employer attributes second period performance to the worker's innate ability rather than his effort, which makes establishing a good reputation easier. Second, when performance is revealed, the employer's belief about second period effort depends on the information revealed. Third, this provides a channel through which the worker can use first period effort, which in turn

1. The 2004 Workplace Employment Relations Survey, a cross-sectional survey of 2295 workplaces in the United Kingdom, illustrates this [see Kersley et al. (2006) for full details]. Establishments were asked whether or not they formally assessed the performance of the largest nonmanagerial occupational group. Tabulating the answers by five-digit industry code reveals that 100% of the establishments in the "business and managements consultancy activities" and "legal services" industries answered in the affirmative.

2. The commitment assumption is crucial in the model because it rules out the employer lying to the worker ex post, even though she has an incentive to do so. Section 2 provides justifications for the assumption.

determines first period performance, to manipulate the amount of effort the employer expects him to exert in the second. Information disclosure can work to both raise and lower first period effort: revealing that expected ability is above $\theta^*$ (*positive feedback*) increases first period effort; revealing it is below (*negative feedback*) dampens it.

Disclosing information presents the employer with a trade-off between risk and incentives. When first period effort is less than first-best, more positive feedback raises first period effort but also second period effort costs. The article first studies when the employer would like to reveal all performance information versus none, since these policies involve less extreme commitment than more general ones. High effort costs, a high payoff to tenure ($W$), and a high signal-to-noise ratio in observed output all result in full disclosure inducing high variance in second period effort, so that no disclosure always dominates full disclosure. On the other hand, when ex ante ability is sufficiently high, full disclosure dominates because feedback is positive—and therefore motivating—with a high probability. When the employer can choose among more general disclosure policies, the optimal policy lumps together everyone below $\theta^*$ and high achievers, but reveals performance in an intermediate interval, where incentives are strong and risk is small.

Rather than provide a complete theory of feedback in organizations, the model is useful for clarifying some potentially false intuitions about feedback and motivation in organizations. First, it shows that the *anticipation* of good (bad) news boosts (dampens) effort. This is in contrast to accounts that suggest the incentive effects of good and bad news arise primarily from reactions to feedback. Second, it shows that coarse feedback can be optimal, since neither full disclosure nor no disclosure maximizes surplus in the relationship.[3] Human resource management policies that seek to maximize the flow of information from managers to workers may be ill-advised in some cases. Third, the optimal disclosure policy reveals strictly less information than the effort-maximizing policy, which always reveals ability above $\theta^*$. So, if one observes firm $A$ revealing more feedback than firm $B$ and workers in $A$ exerting more effort, one cannot necessarily conclude that $A$ has a better appraisal system.

The main contribution of the article is to point out new ways in which interim performance disclosure interacts with reputational incentives when effort is history-dependent. Given the importance of both career concerns and appraisal systems in many organizations, and the lack of theoretical literature combining them, the findings are potentially important in informing a more complete view of how and why firms evaluate workers. Of course, in the real world, disclosure affects other variables as well, such

3. While the rationale for this finding is novel, the idea of the optimality of coarse feedback has appeared in the literature on tournaments (Goltsman and Mukherjee 2010); relational contracting (MacLeod 2003; Fuchs 2007); status incentives (Dubey and Geanakoplos 2010); and static moral hazard with output contracts (Ray 2007).

as the quality of worker training and the ability of upper management to sort lower level workers in the organizational hierarchy. The key challenge going forward is to quantify the relative importance of these other effects compared to the ratchet effect in environments with career concerns.

## 1.1 Literature Review

This article relates to two strands in career concerns literature. First, Kovrijnykh (2007) and Martinez (2009) both analyze career concerns models with history-dependent effort and point out that, in such environments, current effort influences the market's beliefs about future effort;[4] at the same time, neither analyzes how disclosing or withholding information from the worker about his performance affects the strength of the ratchet effect. Second, several papers (Dewatripont et al. 1999; Kovrijnykh 2007; Mukherjee 2008; Koch and Peyrache 2010) examine how varying the amount of information available to the *labor market* about worker performance affects the strength of signal-jamming incentives (the incentives that underlie effort provision in career concerns models). A common theme is that limiting the amount of performance information available to the labor market can be optimal. This article instead holds fixed the amount of performance information available to the labor market, and, indeed, the strength of signal-jamming incentives is independent of the disclosure policy. Instead, it shows that limiting the amount of information available to the *worker* is optimal due to the interaction of the ratchet and effort risk effects.

While the analysis of performance appraisal and career concerns is new, there is a growing literature that studies performance feedback within other contracting frameworks. Several papers (Aoyagi 2010; Ederer 2010; Goltsman and Mukherjee 2010) analyze information disclosure in dynamic tournaments.[5] The main insight that carries over to this article is that, under quadratic effort costs, expected second period effort is independent of the disclosure policy (Aoyagi 2010; Ederer 2010). With the exception of Ederer (2010), none of these papers feature heterogeneity in worker ability, so the ratchet effect does not arise.[6] Ederer (2010) shows that, when workers have unknown ability, and effort and ability are

4. Martinez (2009) also terms this phenomenon the ratchet effect.

5. Other related papers in the dynamic contest literature are Yildirim (2005), in which agents themselves decide whether to commit to revealing performance, and Gershkov and Perry (2009), in which the principal chooses whether to use interim performance to decide the final prize allocation (not whether to disclosure interim performance).

6. More generally, the tournament literature views workers as competing with each other for a promotion on the basis of their output, whereas this article views workers as competing with an exogenous promotion standard that depends on their ability. This framework is more appropriate when the firm cares about the talent of the people it promotes, and when talented workers are hard to find. In this case, there would not be job shortages at senior levels, as the tournament literature implicitly assumes, but a shortage of talented individuals to take up the jobs that are available.

complements in production, full disclosure raises first period effort since workers want to signal high ability to discourage their opponents from exerting effort in the second. In contrast, in this article, disclosure allows the worker to signal to the employer the ability type believed to exert low effort, and full disclosure may or may not raise first period effort.

Analysis of the disclosure of performance information to workers has also appeared in the literature on relational contracting (MacLeod 2003; Fuchs 2007) and moral hazard with output contracts (Lizzeri et al. 2002; Ray 2007; Nafziger 2009). Again, these papers do not feature talent heterogeneity,[7] so their mechanics are quite different from those here. Cremer (1995) takes a complementary perspective to this article: he shows that limiting the information of the *principal* about the worker's ability can improve incentives since it makes commitment to firing rules easier.

Finally, Kamenica and Gentzkow (2011) consider a much more general model of persuasion. A sender selects a signal distribution whose realizations are correlated with a payoff relevant state variable and are seen by a receiver before he takes an action that affects the utility of both parties. In their model, for a given realization, the action of the receiver—and so the payoff to the sender—depends only on the posterior belief formed by the receiver on the state variable. In this article, the worker's second period effort depends not only on his own belief on his ability, but also on the employer's belief about his second period effort choice, which in turn depends on the employer's belief about the worker's ability. This provides the channel through which the ratchet effect operates in the first period, and is absent in Kamenica and Gentzkow (2011).[8]

Section 2 lays out a model of career concerns and information disclosure. Section 3 discusses the effects of feedback, whereas Section 4 examines the value of disclosure in the presence of career concerns. Section 5 discusses the results and provides concluding comments. Proofs of all results in the main text are in appendix B.

## 2. Model

A worker works for an employer during Periods $t = 1, 2$. The production function in period $t$ is $y_t = \theta + a_t + \varepsilon_t$ where $\theta \in \mathbb{R}$ is talent, $a_t \in \mathbb{R}_+$ is effort, and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ is an i.i.d. output shock. Neither the worker nor the employer knows $\theta$ ex ante, but they share a common prior distribution $\theta \sim N(\bar{\theta}, \sigma_\theta^2)$, where $\bar{\theta} > 0$ is initial ability. The cost to the worker of

7. In Ray (2007), a principal privately observes an agent's ability prior to their commencing a project together, and can decide whether or not to reveal it. In some cases, revealing performance information on the tails of the ability distribution and withholding it in the middle is optimal. However, the model does not have career concerns since the principal already knows the ability.

8. Another difference is that, in the current article, the employer already knows the worker's performance and simply decides whether or not to reveal it; in Kamenica and Gentzkow (2011), the sender has no private information before choosing the signal that both observe.

exerting effort is $c(a_t) = Ca_t^2/2$, where $C \in \mathbb{R}_+ \backslash \{0\}$ scales the marginal cost of effort. The worker has an outside option that gives utility 0. The article makes the standard informational assumption that the worker privately observes his effort. To this, it adds the nonstandard assumption that the employer privately observes his output. The employer's updated belief on the worker's ability after observing $y_1$ is $\hat{\theta}_1$, and her updated belief after observing $(y_1, y_2)$ is $\hat{\theta}_2$. In the tradition of career concerns models, explicit incentives are not feasible, and all incentives arise from the desire to acquire reputation; that is, to increase $\hat{\theta}_2$.

The private observability of output creates two potential informational asymmetries: one between the employer and outside labor market, and another between the employer and the worker. Since the focus of the article is feedback within organizations, it will take as exogenous the former and assume that the outside market does not observe any direct signal of output. As such, the worker's wage following Period 2 is the outcome of an asymmetric information bidding game between the employer and outside market, which the article sets up along the lines of Waldman (1984). The game is fully solved in Appendix A. Its key feature is that the payoff to reputation for the worker is given by[9]

$$w\left(\hat{\theta}_2\right) = \begin{cases} \overline{W} \text{ if } \hat{\theta}_2 \geq \theta^* \\ \underline{W} \text{ if } \hat{\theta}_2 < \theta^*. \end{cases} \tag{1}$$

Here, $W = \overline{W} - \underline{W} > 0$ is a "prize" the worker earns if the employer's belief about his ability surpasses the threshold $\theta^*$ after Period 2. More generally, the prize can represent the higher wages or status that accompanies promotion.

While the employer cannot reveal output information to the outside market, she can commit to revealing information to the worker between Periods 1 and 2.[10] A *disclosure policy* is a partition $P$ of the first period output space $Y_1 = \mathbb{R}$. Before the second period, the worker learns that $y_1 \in P(y_1)$, the element of the partition into which his first period output fell. An important characteristic of a disclosure policy is the set of output realizations that it directly reveals to the worker. We denote this set $D(P) = \{ y_1 | P(y_1) = y_1 \}$. It is also useful to distinguish between disclosed output realizations in terms of the posterior beliefs they induce. *Positive* (*negative*) feedback consists of all disclosed output realizations that inform the worker that the employer's assessment of his ability is above (below)

9. The key intuition is that the employer's wage offers cannot signal its private information in equilibrium. If the employer paid two different retained types, two different wages, and the outside market believed these signals credibly communicated private information, the employer would have an incentive to "lie" to the market and tell that the worker of higher talent was the one of lower talent through offering it a lower wage. So in equilibrium, all retained workers are paid the same wage.

10. Disclosing information to the worker after Period 2 is payoff irrelevant to all actors in the model.

the promotion threshold. More formally, positive feedback is the set $S(P) = \left\{ y_1 | y_1 \in D(P), \hat{\theta}_1(y_1) > \theta^* \right\}$, and negative feedback $N(P) = D(P) \backslash S(P)$.[11] To avoid measure theoretic technicalities, the article assumes that $D\,(P)$ is either empty or is a union of positive measure intervals, and that every nonsingleton element of a disclosure policy is a union of positive measure intervals.

This definition of a disclosure policy is quite flexible. It can accommodate full disclosure by taking $P(y_1) = y_1$ for all $y_1$, hereafter denoted $P^F$. It can also accommodate no disclosure by taking $P(y_1) = Y_1$ for all $y_1$, hereafter denoted $P^N$. In between these two extremes, many intermediate cases are possible: reveal whether output is above or below a certain threshold, disclose output over some interval, and group all other realizations together, etc.

The commitment assumption is both important and strong. Once a disclosure policy is in place, the employer cannot deviate from it and lie to the worker about his performance. Without commitment, this is not credible because the employer will want to communicate to the worker whatever message is consistent with his exerting the most effort in the second period. While this might seem extreme, it has several justifications.

First, it sets a welfare benchmark for the maximum amount of surplus that communication can achieve. This allows one to compute the loss in surplus associated with whatever amount of information is revealed in a strategic communication game. Second, as Aoyagi (2010) points out, there may be ways of designing appraisal systems to ensure commitment. For example, the employer could delegate the task of carrying out an evaluation to a manager with no monetary stake in the effort of the worker. Third, as Aoyagi (2010) and Kamenica and Gentzkow (2011) both point out, the principal's reputation plays an important role in sustaining commitment to disclosure. Most firms put in place appraisal criteria for years at a time, so one might expect systematic lying to be detectable and to result in a loss of credibility. Fourth, if a firm employs a mass of identical workers, forced ranking systems—which some firms, such as General Electric under CEO Jack Welch, have recently adopted—can serve as commitment devices. The reason is that a disclosure policy implies a distribution of feedback messages, and deviations from this distribution are detectable if the distribution is observable. Finally, even if the firm cannot commit to *any* disclosure policy, it may well be able to commit to either $P^N$ or $P^F$. Deviations from $P^N$ are easily detectable since it only contains one element. Commitment to $P^F$ could be achieved even with strategic

11. Positive and negative feedbacks are defined in terms of how the employer's interim belief on the worker's ability compares to the promotion threshold rather than how it compares to the worker's ex ante ability $\bar{\theta}$. In other words, if the worker learns that $\hat{\theta}_1(y_1) \in (\bar{\theta}, \theta^*)$, then he receives negative feedback (in terms of his promotion chances) even though his assessment of his ability increases relative to his prior. See the equilibrium definition below for how the interim belief is formed.

communication by assigning the worker to a task that yielded verifiable performance information; unravelling (Grossman 1981; Milgrom 1981) would then imply full disclosure. Accordingly, the article first analyzes the trade-off between full and no disclosure before considering optimal disclosure.

The article solves the model with the Perfect Bayesian Equilibrium solution concept. Let $a_1^W : P \rightarrow \mathbb{R}_+$ be the worker's first period strategy and $a_2^W : (a_1, P(y_1)) \rightarrow \mathbb{R}_+$ his second. Furthermore, let $a_1^*$ and $a_2^*(P(y_1), a_1^*)$ be the employer's conjectures on these strategies.

*Definition 1.* Strategies $a_1^*(P)$ and $a_2^*(P(y_1), a_1^*)$ constitute a Perfect Bayesian Equilibrium if they satisfy the following conditions, where $\lambda_t = \sigma_\theta^2 / t\sigma_\theta^2 + \sigma_\varepsilon^2$:

$$a_1^W(P) \in \arg\max_{a_1} W\mathbb{E}_{y_1}\left[\Pr\left[\hat{\theta}_2 > \theta^* \middle| P, a_2^W, y_1\right] - \frac{C}{2}\left(a_2^W\right)^2\right] - \frac{C}{2}a_1^2 \tag{2}$$

$$a_2^W(P(y_1), a_1) \in \arg\max_{a_2} W\mathbb{E}_{y_1}\left[\Pr\left[\hat{\theta}_2 > \theta^* \middle| a_1, y_1\right] \middle| y_1 \in P(y_1)\right] - \frac{C}{2}a_2^2 \tag{3}$$

$$a_1^*(P) = a_1^W(P) \tag{4}$$

$$a_2^*(P(y_1), a_1^*) = \left[a_2^W(P(y_1), a_1)\right]_{a_1 = a_1^*} \tag{5}$$

$$\hat{\theta}_1(y_1) = \lambda_1(y_1 - a_1^*) + (1 - \lambda_1)\bar{\theta} \tag{6}$$

$$\hat{\theta}_2(y_1, y_2) = \lambda_2(y_2 - a_2^*) + (1 - \lambda_2)\hat{\theta}_1. \tag{7}$$

Conditions (2) and (3) require the worker to maximize expected third period wages net of effort costs in the first and the second periods, respectively. Conditions (4) and (5) require the worker's strategies to coincide with the employer's conjectures, and the final two conditions require the employer to update her beliefs on ability using Bayes' Rule, where $\lambda_t$ measures the responsiveness of the posterior to new information.[12]

## 3. Effects of Information Disclosure

The first step in the theoretical analysis is to examine what effect a given disclosure policy has on the worker's effort incentives. The worker's ultimate goal in the model is to push the employer's belief on his expected ability above the retention threshold. So, the relationship between effort and the distribution of $\hat{\theta}_2$ is a natural starting point. The expected value of this belief conditional on effort variables and first period output can be expressed as:

$$\mu(y_1, a_1, a_1^*, a_2, a_2^*) = \left[\lambda_1(y_1 - a_1) + (1 - \lambda_1)\bar{\theta}\right] + \lambda_2(a_1 - a_1^* + a_2 - a_2^*). \tag{8}$$

12. The result that Bayes' Rule takes this form is standard (see DeGroot 1970).

The term in square brackets would be the employer's estimate of the worker's ability after observing $y_1$ if she knew the true value of $\alpha_1$. Of course, the employer does not observe effort, but conjectures their values as $a_1^*$ and $a_2^*$. The second term reflects any difference between the worker's actual effort and the conjectured effort. When the worker puts in more effort than the employer expects in period $t$, he increases his expected payoff. This signal-jamming incentive lies at the heart of all career concerns models. Here, we are concerned with how this incentive is mediated by information disclosure. The full distribution of $\hat{\theta}_2$ is

*Lemma 1.* $\hat{\theta}_2 | y_1, a_1, a_1^*, a_2, a_2^* \sim N\big(\mu(y_1, a_1, a_1^*, a_2, a_2^*), \sigma_2^2\big)$.

Now one can make a more precise statement concerning the payoff to effort in the model. Increasing effort is valuable to the worker to the extent that it increases the mean of $\hat{\theta}_2$. This in turn increases the probability of capturing the prize $W$. The following presents the resulting implications for effort, where $\phi$ is the standard normal pdf.

*Proposition 1.* There exists a $\bar{C}$ such that, for all $C \geq \bar{C}$, there exist unique and positive equilibrium first- and second period efforts levels given by:

$$Ca_2^*(P(y_1), a_1^*) = \mathbb{E}_{y_1}\left[ W\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \hat{\theta}_1(y_1)}{\sigma_2}\right) \middle| y_1 \in P(y_1)\right] \tag{9}$$

$$\begin{aligned} Ca_1^*(P) = {} & \mathbb{E}_{y_1}\left[ W\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \hat{\theta}_1(y_1)}{\sigma_2}\right)\right] + \\ & \mathbb{E}_{y_1}\left[ \frac{W^2}{C}\frac{\lambda_2^2\lambda_1}{\sigma_2^3}\left(\frac{\hat{\theta}_1(y_1) - \theta^*}{\sigma_2}\right)\phi^2\left(\frac{\hat{\theta}_1(y_1) - \theta^*}{\sigma_2}\right) \middle| y_1 \in D(P)\right] \times \\ & \Pr[y_1 \in D(P)] \end{aligned} \tag{10}$$

While these expressions appear rather complex, they simply equate the marginal cost of effort on the left-hand side with the marginal benefit on the right. The rest of this section breaks down in detail what this marginal benefit is. The condition on $C$ ensures the worker's objective functions in equations (2) and (3) are globally concave.[13] Finally, information disclosure affects effort only via the information the worker receives about his expected ability. Accordingly, much of the subsequent analysis discusses feedback in terms of disclosed interim expected ability rather than disclosed output. From this point on, $\hat{\theta}_1$ should be understood to represent the employer's belief on the equilibrium path.

13. One might worry that the worker supplies no effort if $C$ is sufficiently large. In fact, for $C$ large, the marginal cost of supplying zero effort is zero, whereas the marginal benefit is positive. So, zero effort provision can never be an equilibrium outcome for large values of $C$.

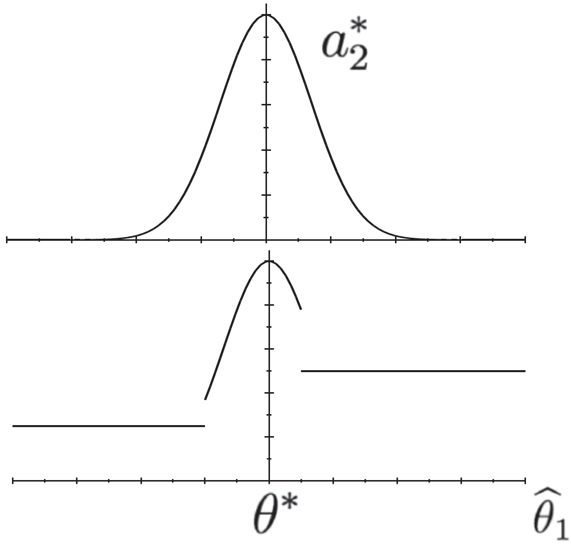


Figure 1. Second Period Effort and Information Disclosure.

### 3.1 Second Period Effort: Effort Risk

The shape of the marginal benefit of effort in the second period is quite intuitive. Since the reputational prize $W$ is earned with probability $\Pr\left[\hat{\theta}_2 > \theta^*\right]$ and $\hat{\theta}_2$ is normally distributed, the change in the probability when effort increases is proportional to the normal pdf. To see how second period effort depends on information, consider Figure 1. The top portion shows equilibrium effort under the full disclosure policy. In this situation, second period effort is highest when the disclosed interim ability is near the promotion threshold $\theta^*$, and monotonically decreasing as ability moves into either tail of the talent distribution. Given normality, $\hat{\theta}_2$ reacts less and less to new information when interim ability is very high and low. Informally speaking, in these regions, the employer's mind is already made up that the worker is either good or bad. In contrast, when $\hat{\theta}_1$ is near the threshold $\theta^*$, second period performance is still important for determining the promotion, so the worker exerts effort.

The bottom portion of Figure 1 shows equilibrium second period effort under a disclosure policy that reveals the value of interim ability between two points $\theta^* - 2b$ and $\theta^* + b$, and otherwise reports only whether it lies in $(-\infty, \theta^* - 2b)$ or in $(\theta^* + b, \infty)$ (here $b$ is some positive constant). Clearly, effort remains the same as under the full disclosure policy for interim ability realizations that lie in $(\theta^* - 2b, \theta^* + b)$. However, effort changes in the tails, where the worker now exerts an effort level formed by taking the expectation over all ability levels contained in the associated partition element. Consider the case in which the worker learns his interim ability lies in$(-\infty, \theta^* - 2b)$. Worker types for whom $\hat{\theta}_1$ lies close to $\theta^* - 2b$ now exert less effort than under full disclosure, because they are pooled with types further away from the promotion threshold. On the other hand, worker types for whom $\hat{\theta}_1$ is very low will now work harder than under full disclosure, since they believe their ability to be closer to $\theta^*$ than is actually the case.

In general, then, moving from one disclosure policy to another will cause some ability types to exert more effort and others to exert less. The first interesting property of a disclosure policy is that in expectation these changes cancel out.

*Corollary 1.* $\mathbb{E}\left[ a_2^* \right]$ is independent of the disclosure policy.

Since expected second period effort does not depend on the disclosure policy, the employer does not have to consider the incentive effects of the worker's reaction to feedback when choosing $P$. This result relies on the quadratic effort cost assumption, since this gives linear marginal costs, allowing one to use the law of total probability to compute expected second period effort. Any deviation from quadratic costs will mean that expected second period effort depends on the disclosure policy. The article maintains the quadratic cost assumption to isolate the trade-off between the risk and the ratchet effects without introducing a third effect as well. We return to this point in Section 5.

While a disclosure policy does not affect expected second period effort, it does affect expected second period effort *costs*. Providing more information to the worker about his performance increases the variance of second period effort, which, because his preferences over effort are given by a convex cost function, increases his disutility. In short, the worker prefers to exert a given effort level with certainty than to do so in expectation.

*Corollary 2.* Let P′ be a refinement of P. Then $E\left[\left(a_2^*\right)^2\right]$ is higher under P′ than P.

This is the first substantive point about feedback in organizations. It implies that no information provision is optimal if workers only exert effort for one period.

### 3.2 First Period Effort: Ratchet Effect

The marginal benefit of the first period effort arises from two sources, corresponding to the two summands in Equation (10). The first is the standard signal jamming. By putting in effort above $a_1^*$, the worker can convince the employer he is more talented. The first term of Equation (10) captures this incentive, which is independent of the disclosure policy and equal to second period effort under the no disclosure policy $P^N$. In both cases, the worker has no additional information beyond the prior distribution on which to base his effort choice.

The second source of incentives is more subtle. As one can see from equation (8), the mean of $\hat{\theta}_2$ is decreasing in $a_2^*$. In other words, signaling a high ability is easier in the second period, if the employer believes the worker is exerting low effort—she will then attribute more of his output to his raw talent rather than his effort. A key point of the article is that information disclosure affects whether the worker can influence $a_2^*$ through his choice of $\alpha_1$. Return again to the bottom half of Figure 1. This not only represents the worker's equilibrium second period effort, but also the employer's *belief* about second period effort. Suppose that $\hat{\theta}_1$ falls in the lower tail $(-\infty, \theta^* - 2b)$. Since the worker is unable to distinguish his interim ability within this range, the employer expects his effort to be the same for all interim ability levels within it. So, the employer's belief on his second period effort is unresponsive to a marginal increase in $y_1$. A similar argument holds for interim ability levels falling in $(\theta^*+b, \infty)$: $a_2^*$ is again unresponsive to marginal changes in $y_1$.

The situation is different when one looks at intermediate ability levels that fall in $(\theta^* - 2b, \theta^*+b)$, in which the worker directly learns his interim ability. Now, since the worker is able to distinguish among interim ability levels locally, the employer expects his second period effort to respond to marginal changes in his first period output. But this means the worker can use his first period effort to reduce the employer's belief about his second period effort. When interim ability lies in $(\theta^*, \theta^*+b)$, the worker does this by exerting more effort in the first period. This pushes $\hat{\theta}_1$ away from the promotion threshold, where maximum effort is expected, and into the upper tail, where less effort is expected. In contrast, when interim ability lies in $(\theta^* - 2b, \theta^*)$, the worker wants to exert *less* effort since lower values of $\hat{\theta}_1$ are associated with lower values of $a_2^*$ in this range.

This example illustrates two more general points. First, the employer's belief about the worker's second period effort always depends on the feedback the worker receives via $P\ (y_1)$. But it only depends on $y_1$ *on the margin* within the set of disclosed output realizations $D\ (P)$ because only within this set can the worker distinguish locally among ability levels. Second, depending on whether the worker receives positive or negative feedback, his first period effort either increases or decreases.

*Corollary 3.* Suppose two disclosure policies P and P′ are such that $S(P') \subset S(P)$ and $N(P) \subset N(P')$. Then $a_1^*$ is higher under P than under P′.

Thus information disclosure creates a ratchet effect, whereby the worker's current effort affects the employer's belief about future effort. Here, though, the effect can work to both decrease and increase first period effort, whereas the ratchet effect has usually been seen to discourage effort in the dynamic moral hazard literature with explicit contracting. More importantly, the strength and the direction of the effect are endogenous to the disclosure policy chosen by the employer. The next section combines the two effects identified in this section to examine the value of disclosure.

## 4. The Value of Disclosure

We assume the employer chooses $P$ to maximize joint surplus[14]

$$a_1^*(P) - g(a_1^*(P)) + \mathbb{E}\left[\, a_2^*(P(y_1), a_1^*) - g(a_2^*(P(y_1), a_1^*)) \,\right]. \tag{11}$$

This assumption captures a situation in which the employer competes with others firms prior to Period 1 to hire the worker, which is typical in career concerns models. In this case, each firm would offer a disclosure policy that maximized the worker's utility subject to a zero profit condition, which is equivalent to maximizing joint surplus. Since the usual worry in career concerns models is the under provision of effort, the article assumes that $Ca_1^*(P^N) < 1,$ so that first period effort is less than first best under no disclosure.

One useful way of thinking about the employer's choice of disclosure policy is in terms of a risk–incentive trade-off. Providing positive feedback not only raises first period effort, but also increases uncertainty in the second period. The key question is which effect dominates. This is quite distinct from the well-known risk–incentive trade-off in the static moral hazard literature (Holmström 1979). There, the instrument for inducing effort is output pay and the associated risk is over wealth levels, not effort levels.

As discussed in Section 2, if commitment is limited, comparing no disclosure $P^N$ with full disclosure $P^F$ may be more relevant than solving for the optimal disclosure policy. So, this section first compares these two extreme policies before deriving the optimal $P$.[15]

### 4.1 Full Versus no Disclosure

Moving from $P^N$ to $P^F$ lowers second period surplus by

$$\frac{C}{2} V\left[\, a_2^*(\hat{\theta}_1(y_1)) \,\right] = \frac{1}{2C}\left(\frac{W\lambda_2}{\sigma_2}\right)^2 V\left[\phi\left(\frac{\theta^* - \hat{\theta}_1(y_1)}{\sigma_2}\right)\right], \tag{12}$$

or a term proportional to the variance of second period effort under full disclosure. It also changes first period effort by

$$\Delta = \mathbb{E}\left[\left(\frac{W}{C}\right)^2 \frac{\lambda_2^2\lambda_1}{\sigma_2^3}\left(\frac{\hat{\theta}_1(y_1) - \theta^*}{\sigma_2}\right)\phi^2\left(\frac{\hat{\theta}_1(y_1) - \theta^*}{\sigma_2}\right)\right], \tag{13}$$

which changes first period surplus by an amount[16]

14. The expression only includes those quantities in surplus that the choice of disclosure policy affects.

15. The approach of this article parallels a similar distinction in the delegation literature, in which some papers study full versus no delegation (e.g., Dessein 2002), whereas others study partial, optimal delegation (e.g., Alonso and Matouschek 2008).

16. This follows from manipulating the expression: $\left(a_1^*(P^N)+\Delta\right) - \frac{C}{2}\left(a_1^*(P^N)+\Delta\right)^2 - a_1^*(P^N)+\frac{C}{2}\left(a_1^*(P^N)\right)^2.$

$$\Delta\left(1 - Ca_1^*(P^N) - \frac{C}{2}\Delta\right). \tag{14}$$

There are two situations where one can immediately establish that the increase in effort variance from full disclosure wipes out more than any gains in first period effort. First, when the marginal cost of effort parameter $C$ is high, variance in second period effort is costly and equation (12) is clearly larger than equation (14). Also, when $Ca_1^*(P^N)$ is close to its first-best value 1, the gains from disclosure are small since first period surplus is close to being maximized even without additional incentives from the ratchet effect. Moreover, first period effort is close to first-best when the promotion prize $W$ is sufficiently high, since signal-jamming incentives are strong. There is also a third, more subtle, situation in which $P^N$ outperforms $P^F$: when the signal-to-noise ratio $\sigma_\theta^2/\sigma_\varepsilon^2$ is high. In this case, $y_1$ contains a large amount of information about ability. So, from an ex ante perspective, there is greater uncertainty about the realization of interim expected ability $\hat{\theta}_1$ and thus about second period effort.

One might then wonder whether $P^F$ ever dominates $P^N$. Ignoring the expectation terms for the moment, one can conjecture that equation (13) is larger than equation (12), if $\hat{\theta}_1$ is large. For high realizations of interim ability, the *rate of change* in the employer's belief about second period effort with respect to $\hat{\theta}_1$ is much larger, relatively speaking, than the value of equilibrium effort at $\hat{\theta}_1$. Since this rate of change determines the strength of the ratchet effect, one can conclude that when $\hat{\theta}_1$ is likely to be high, full disclosure will dominate no disclosure. Moreover, $\hat{\theta}_1$ is likely to be high when initial ability $\bar{\theta}$ is sufficiently high.

*Proposition 2.* $\mathrm{P^N}$ yields higher surplus than $\mathrm{P^F}$ as

1. $C \to \infty$
2. $W \to \frac{\sigma_2}{\lambda_2}\mathbb{E}\left[\phi\left(\frac{\hat{\theta}_1(y_1)-\theta^*}{\sigma_2}\right)\right]^{-1}$, and
3. $\frac{\sigma_\theta^2}{\sigma_\varepsilon^2} \to \infty$.

*$P^F$ yields higher surplus than $P^N$ as $\bar{\theta} \to \infty$.*

The first important message from Proposition 2 is that when information disclosure cannot be targeted toward certain output realizations, but simply designed to release all output information, the risk effect dominates the incentive effect in many situations. This is perhaps not surprising given that $P^F$, in addition to increasing risk, includes negative feedback that actually worsens the already existing inefficiency in the first period effort. In spite of the fact that the model stacks the odds against $P^F$ dominating $P^N$, it nevertheless does so when the worker is expected to easily meet the promotion threshold. Of course, in general, one would expect the employer to do even better by moving away from both extremes and by instead adopting partial disclosure, the situation to which the paper now turns.

### 4.2 Optimal Disclosure

Denote by $P^S$ the surplus-maximizing disclosure policy. To begin, one can place some minimal structure on it using the insights so far. Since the ratchet effect arises only over the set of output realizations directly revealed to the worker, $P^S$ only contains one nonsingleton element. If a disclosure policy contained two nonsingleton elements, one could combine them without changing the first period effort. At the same time, one would decrease expected second period effort costs (by Corollary 2) through reducing effort risk. The following result provides the full structure of $P^S$.

*Proposition 3.* The surplus maximizing disclosure policy $\mathrm{P}^{\mathrm{S}}$ takes the form

$$P^S = \begin{cases} y_1 & \text{if } y_1 \in [y^*, y^{**}] \\ (-\infty, y^*) \cup (y^{**}, \infty) & \text{if } y_1 \in (-\infty, y^*) \cup (y^{**}, \infty) \end{cases}$$

where $\theta^* < \hat{\theta}_1(y^*) < \hat{\theta}_1(y^{**}) < \infty$.

One can build up the intuition for the result step-by-step. Since, by assumption $Ca_1^*(P^N) < 1$, negative feedback is doubly bad. First, it further reduces first period effort from its already inefficiently low level, and second it exposes the worker to effort risk. Therefore, $P^S$ provides no negative feedback. So, one only needs to consider disclosure policies that can be fully described by the amount of positive feedback they contain; let $D_\theta(P) = \left\{ \hat{\theta}_1(y_1) | y_1 \in S(P) \right\}$ be the resulting set of disclosed interim expected ability levels. Now consider moving from disclosure policy $P$ to disclosure policy $P'$ that satisfies $D_\theta(P') = D_\theta(P) \cup [t, t+\varepsilon]$, where $\varepsilon$ is small and $t \in (\theta^*, \infty)$. This raises first period surplus by increasing first period effort and lowers second period surplus by increasing expected effort costs. Figure 2 illustrates the associated marginal benefit and cost curves for all $t \in (\theta^*, \infty)$. It is important to keep in mind that this figure is drawn for *any* $P$ that has one nonsingleton element and provides no negative feedback, not just $P^S$.

The additional risk generated in moving from $P$ to $P'$ is proportional to the squared difference between the effort the worker exerts under $P$ conditional on not learning his interim ability and the effort he exerts conditional on learning $\hat{\theta}_1 = t$. This risk is highest when $t$ is large and when $t$ is near $\theta$*. In the first case, the worker's effort conditional on learning $t$ is substantially lower than the effort level he would exert conditional on not learning his type. In the second, his effort conditional on learning $t$ is substantially higher. On the other hand, there exists some $\tilde{t}$, such that the worker's effort conditional on learning $\hat{\theta}_1 = \tilde{t}$ is exactly equal to the effort level he exerts under ignorance. Disclosing the additional ability realizations $\left[\tilde{t}, \tilde{t}+\varepsilon\right]$ to the worker thus exposes him to almost no additional risk.

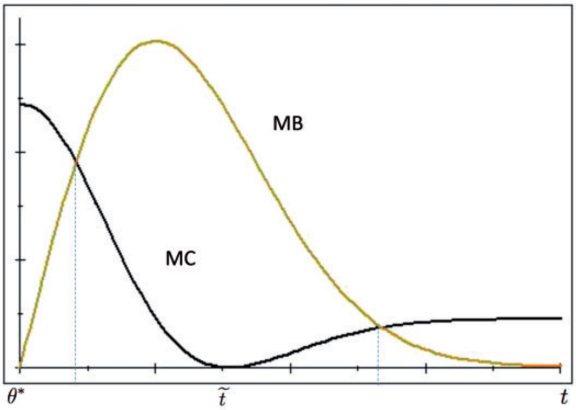


Figure 2. Marginal Cost and Benefit of Information Disclosure.

The marginal benefit of additional disclosure depends on the strength of the ratchet effect, which is proportional to $\partial a_2^*/\partial a_1$. When $\hat{\theta}_1$ is near $\theta$* and very large, $a_2^*$ is essentially flat, as Figure 1 shows, whereas it is positive for intermediate ability realizations. The key point is that the additional incentives generated by disclosing $[t, t{+}\varepsilon]$ are small in exactly the cases in which the additional risk is large, and vice versa. Thus the surplus maximizing policy conceals beliefs near $\theta$* and large beliefs and reveals intermediate beliefs.[17]

## 5. Discussion and Conclusion

The article concludes by discussing the wider implications of the results, and how the ratchet effect would manifest itself in appraisal systems compared to other effects.

### 5.1 Potential Implications for Design of Performance Appraisal Systems

The effects uncovered in the article provide alternative explanations for findings about the feedback in the human resources literature. First, Jackman and Strober (2003) document that a large fraction of workers actively avoid feedback, but emphasize psychological mechanisms (e.g., anxiety about criticism) as an explanation. In contrast, this article predicts

17. One can contrast this argument with the logic above on full disclosure versus no disclosure, in which high realizations of expected ability were claimed to favor incentives over risk. Here the argument is that, holding fixed any disclosure policy, and thus the expected effort under no feedback, the marginal benefit of removing an infinitely small measure of positive feedback in the extreme upper tail of the performance distribution always raises surplus. However, one cannot find the total cost of moving from no disclosure to full disclosure by simply summing up the incremental differences in marginal benefits and costs in a diagram like Figure 2; expected effort cannot be treated as constant as strictly positive measures of disclosed output realizations are added to or taken away from a disclosure policy.

that workers want to avoid feedback in order to minimize effort risk.[18] Second, Meyer et al. (1965) argue that positive feedback stimulates hard work through encouragement, whereas negative feedback reduces effort through discouragement. The ratchet effect predicts the same relationship between feedback and effort, but shows that positive and negative feedback have different ex ante effects. Ex post, positive, and negative feedback actually have symmetric effects: the worker exerts the same effort after learning that his interim ability is $x$ units above the promotion threshold as he does after learning it is $x$ units below. Moreover, while positive feedback increases effort, it reduces utility since effort is costly. If the employer offers more positive feedback to the worker, it must compensate him with higher wages.

Another interesting point is that positive feedback is never used to fully eliminate the inefficiency in the first period effort provision. Suppose, the employer uses a disclosure policy $P$ for which $Ca_1^*(P) = 1$. Then, since effort is already at its first-best level, removing a small interval of positive feedback reduces first period surplus very little when compared to the reduction in risk. So, the trade-off between incentives and risk is never resolved by fully eliminating the first period inefficiency.[19]

While the effort risk effect is perhaps less theoretically interesting than the ratchet effect, it has important welfare implications. The effort maximizing disclosure policy ($P^E$) provides maximum positive feedback, whereas the optimal disclosure policy ($P^S$) provides partial positive feedback. Suppose, one compared two firms with workers of similar talent, one of which used $P^E$ and the other used $P^S$. The former would provide more information to the workers and would have higher productivity levels. Nevertheless, one could not conclude that it used an unambiguously better disclosure policy. The risk faced by the workers in the firm using $P^E$ outweighs the associated gain in productivity.

Of course, this observation only matters if there are convincing reasons for observing $P^E$ in practice. One such reason arises from the nature of some industries in which career concerns operate. In professional services, firms often have bargaining power with respect to entry-level workers, whereas competition for talented senior workers is strong, leading to high promotion returns (Maister 1993). One can interpret high promotion returns in the model as a high $W$. If this were high, the employer had period 0 bargaining power, and the worker were wealth constrained, the employer could offer $P^E$ and still satisfy the worker's participation

18. Another explanation for information avoidance is that workers are threatened by losing status in their organization if they learn—or others learn—that they have not done well. These effects are likely to be relevant when appraisal scores are made public, which is yet another design element of an appraisal system.

19. A more formal argument is given in the proof of Proposition 3. This is similar in flavor to the well-known result that in a moral hazard problem with a risk neutral principal, an agent with CARA preferences, a production function with a normal error term, and linear contracts, the optimal contract never implements first-best effort due to the agent's risk aversion.

constraint. In this case, she would not internalize effort risk and would provide too much information.[20] Finally, managers might provide effort-maximizing feedback if their pay depended on their workers' productivity. In this case, they would also not internalize the increased disutility of the effort they imposed by using $P^E$.

The more general message is that too much disclosure is as legitimate a concern as too little disclosure in the presence of career concerns. A stylized fact about real world feedback (Prendergast 1999) is that ratings are concentrated. In many cases, workers with different actual performance are lumped together and given the same performance rating.[21] This article shows that such feedback is consistent with optimality. This does not imply that concentrated ratings are always efficient (indeed, firms often go to great lengths to force additional disclosure). The point is that implementing a system with *full* disclosure may be ill-advised, especially if incentive considerations are primary.

### 5.2 Generality

We derive results within a quite specific model. The main concern with generality is not the existence of the ratchet and effort risk effects (the logic underlying them should transfer to more complicated environments), but their relative magnitudes, which are used to prove Propositions 2 and 3. The single-peakedness of the second period effort in interim ability is key for establishing Proposition 3. This allows one both (a) to divide interim expected ability realizations into an upper part over which the ratchet effect increases effort and a lower part over which it decreases effort; and, (b) to find an interim ability realization whose disclosure generates no risk (marked $\tilde{t}$ in Figure 2). The shape of the second period effort is itself linked to the shape of the wage function, which derives from the assumption that the outside market does not observe performance. What if one instead assumed that the outside market always observed performance? If the worker enjoyed some nonmonetary benefit $F$ of remaining in the industry, such as enhanced pride, a larger office, etc., wages become[22]

20. In occupations characterized by high returns to promotion, previous research has found evidence of over provision of effort (Landers et al. 1996). These industries' feedback policies may thus worsen an already existing "rat race" brought on by large reputational rewards.

21. Prendergast (1999) and others also discuss the prevalence of a *leniency bias*, whereby low ratings are under-represented relative to the true performance distribution. This article (and most likely any paper with full commitment) cannot account for the leniency bias because the frequency of observed ratings corresponds exactly to what one would expect given the underlying performance distribution.

22. Workers whose expected ability is negative cannot be profitably employed, so they exit the industry and take their outside option; workers who remain in the industry are paid their expected output, which is equal to their expected ability given they will exert no effort in the last period as career concerns no longer exist.

$$w\left(\hat{\theta}_2\right) = \begin{cases} F+\hat{\theta}_2 & \text{if } \hat{\theta}_2 \geq 0 \\ 0 & \text{if } \hat{\theta}_2 < 0. \end{cases}$$

Here, second period effort is again single-peaked in expected ability, meaning most of the results go through; for example, a disclosure policy that lumps some performers together always generates more surplus than full disclosure.

Another concern for generality is the cost of effort function. Aoyagi (2010) and Ederer (2010) have shown that, when the cost function has convex (concave) marginal costs, no (full) disclosure maximizes expected second period effort. The same carries over to this model. So, with general cost functions, information disclosure must trade-off the ratchet and effort risk effects with a third effect on second period incentives. While a general treatment of this problem has not been attempted, one can partially characterize optimal disclosure when the worker's cost of effort function is $c(a_t) = (C/\beta{+}1)a_t^{\beta+1}$, where $\beta > 1$ so that marginal costs are convex. In this case, a disclosure policy that groups good and bad performers together and reveals intermediate performance does better than either no disclosure or disclosing performance in the tails. In fact, the intuition is quite straightforward: disclosing a tiny measure of interim ability realizations around the "riskless" realization discussed in Section 4.2 changes neither risk nor expected second period effort.

Finally, one might think about introducing some complementarity between talent and effort in the production function. Ederer (2010) has a model with complementarity, and shows that expected second period effort is equal under full and no disclosure. On this basis, one can conjecture that complementarity here would introduce no new incentive effects. Ederer (2010) also shows that expected second period *output* does depend on disclosure since ability and effort interact in the production function. In his paper, full disclosure maximizes expected second period output because more able agents work harder. With the wage function in this article, there is not a monotonic relationship between ability and effort, so the effect of disclosure on expected output is less clear.

### 5.3 Empirical Identification

The article also left out other channels through which feedback might operate within performance appraisal systems. Training and sorting are two common ones that are relevant in the presence of unknown ability. The information a worker receives in appraisals often teaches him how to do his job better. Also, the information recorded in the appraisals often serves as an input in promotion decisions. Is the ratchet effect distinguishable from these other forces? One would expect the training value of information to arise in the second period once the learning has taken place. Moreover, one would expect the sorting value of information to arise in the (unmodeled) third period: with better information, better workers will be promoted. On the other hand, the ratchet effect should arise in the first period, before the worker (or employer) has observed any performance history. Thus, by exploiting the timing of information release, one could potentially test the empirical relevance of the ratchet effect.[23]

This article provides a clean and tractable way of thinking about interim performance disclosure within a career concerns model with history-dependent effort. It shows how information disclosure generates a novel ratchet effect that can both increase and decrease effort, and characterizes the value of disclosure. The model produces several insights that complement and extend the growing literature in economics on feedback and motivation. Given the importance of career concerns in many organizations, it also provides a potential framework for analyzing some dimensions of performance appraisal systems hitherto unexplored.

## Appendix A: Labor Market Bidding Game

This section sets up and solves an asymmetric information labor market game whose equilibrium outcome is consistent with the wage schedule given by equation (1) in the text. It is essentially a modified version of the model of Waldman (1984).[24] Suppose that after Period 2, the employer competes with two outside firms $M1$ and $M2$ to hire the worker. Suppose further that $M1$ and $M2$ do not directly observe any performance information of the worker. The timing of the game is the following:

1. The employer makes the worker a wage offer $w^E \in W^E = \mathbb{R}^+$,
2. Both market firms observe $w^E$ and simultaneously make wage offers $w^m \in W^m = \mathbb{R}^+$ for $m = 1, 2,$ and
3. The worker observes all wage offers and chooses which firm to join in Period 3

If the worker remains with the employer, his output is $y_3^E = \kappa + \theta + \varepsilon_3$ where $\varepsilon_3 \sim N(0, \sigma_\varepsilon^2)$ is an output shock and $\kappa > 0$ reflects firm-specific human capital accumulation. If the worker instead moves to another firm, his output is $y_3^m = \theta + \varepsilon_3$.[25] If the worker receives no positive wage offer, he

23. There are two additional caveats. First, the model assumes *all* incentives arise from career concerns. If there is some formula that links feedback and bonuses, feedback would also have first period effects. Second, the model assumes that the worker does not learn about the feedback received by other workers. If this were not the case, status incentives might affect first period effort.

24. The reason one cannot simply apply the Waldman model is because here the employer has imperfect private information on worker ability, ability is normally (as opposed to uniformly) distributed, the outside market is modeled as two separate firms to adhere to the spirit of career concerns models (as opposed to one entity), and firms can only choose wages (as opposed to wages and job assignments).

25. The fact that third period output does not depend on effort is without loss of generality as the worker will exert zero effort in the last period since career concerns cease to exist.

leaves the market. Otherwise, he moves to the firm that offers him the highest wage. If the employer matches the highest wage offered by the market firms, the worker remains with the employer. If $M1$ and $M2$ jointly offer the highest wage, he joins each with probability 0.5. Thus the worker's third period wage is $w_3 = \max\{0, w^E, w^1, w^2\}$. Finally, if any firm makes a positive wage offer, it incurs an arbitrarily small cost $\delta$, which could for example be the legal costs from drafting a wage contract.[26] These assumptions together imply that third period profit for $E$ is given by

$$\pi_3^E = \begin{cases} \kappa+\theta - w^E - \delta & \text{if } w^E > 0, w^E \geq \max\{w^1, w^2\} \\ -\delta & \text{if } w^E > 0, w^E < \max\{w^1, w^2\} \\ 0 & \text{if } w^E = 0; \end{cases}$$

and that third period profit for market firm $m \in \{1, 2\}$ is given by (where $n \in \{1, 2\}\backslash\{m\}$)

$$\pi_3^m = \begin{cases} \theta - w^m - \delta & \text{if 1. } w^m > 0 \text{ and } w^m > \max\{w^n, w^E\}, \text{or} \\ & \quad \text{2. } w^m > 0, w^m = w^n > w^E, \text{and worker joins firm m} \\ -\delta & \text{if 1. } w^m > 0 \text{ and } w^m < \max\{w^n, w^E\}, \text{or} \\ & \quad \text{2. } w^m > 0, w^m = w^n > w^E, \text{and worker joins firm n} \\ 0 & \text{if } w^m = 0. \end{cases}$$

Let $a_1^W : P \to R^+$ be the worker's strategy in Period 1 and $a_2^W : (P(y_1), a_1) \to R^+$ be the worker's strategy in Period 2, and let $a_1^*(P)$ and $a_2^*(P(y_1), a_1^*)$ be firms' beliefs about these strategies. Prior to making a wage offer $w^E$, the employer has private information on the worker's ability in the form of the signals $y_1 - a_1^*$ and $y_2 - a_2^*$ on which it can condition its wage offers. Let $\hat{\theta}_1^E$ be its updated belief on the worker's ability after observing the first signal, and $\hat{\theta}_2^E$ be its updated belief after observing both. Denote the strategy of the employer as $\bar{w}^E : (y_1, y_2) \to W^E$ and the strategy of market firm $m$ as $\bar{w}^m : w^E \to W^m$. Denote by $\hat{\theta}^m$, the market firms' (common) updated belief on the worker's ability after observing $w^E$. The article solves the model via Perfect Bayesian Equilibrium.

*Definition 2.* A Perfect Bayesian Equilibrium is a set of strategies $\left(\bar{w}^{E*}, \bar{w}^{1*}, \bar{w}^{1*}\right)$ and beliefs $\left(\hat{\theta}_1^E, \hat{\theta}_2^E, \hat{\theta}^m\right)$ that satisfy the following conditions for $m \in \{1, 2\}$ and $n \in \{1, 2\}\backslash\{m\}$, where $\lambda_t = \sigma_\theta^2 / t\sigma_\theta^2 + \sigma_\varepsilon^2$:

$$\bar{w}^{m*} \in \arg\max_{w^m} \mathbb{E}\left[\pi_3^m | \bar{w}^{E*}, \bar{w}^{n*}\right] \forall w^E \tag{A1}$$

$$\bar{w}^{E*} \in \arg\max_{w^E} \mathbb{E}\left[\pi_3^E | \bar{w}^{m*}\right] \forall y_1, y_2 \tag{A2}$$

$$\hat{\theta}_1^E = \lambda_1(y_1 - a_1^*) + (1 - \lambda_1)\bar{\theta} \tag{A3}$$

$$\hat{\theta}_2^E = \lambda_2(y_2 - a_2^*) + (1 - \lambda_2)\hat{\theta}_1^E \tag{A4}$$

26. While this assumption is rather ad hoc, it plays a role in equilibrium selection; see below for discussion.

$$\hat{\theta}^m = \mathbb{E}\left[\hat{\theta}_2^E|(y_1, y_2) \in \left(\bar{w}^{E*}\right)^{-1}\left(w^E\right)\right] \forall w^E \in \bar{w}^{E*}. \tag{A5}$$

Conditions (A1) and (A2) require market firms and the employer to best respond to each other's strategies. Conditions (A3) and (A4) require the employer to update its beliefs using Bayes' Rule. Condition (A5) requires market firms to both use Bayes' Rule to update their beliefs on worker talent and to correctly infer the information conveyed about the employer's private information from observing $w^E$. Notice that no restrictions are placed on $\hat{\theta}^m$ following observations of $w^E$ not on the equilibrium path.[27]

In fact, there are a continuum of equilibria that satisfy Definition 2, but share all the same essential properties.

*Proposition 4.* In every pure strategy equilibrium, the worker remains with the employer if and only if $\hat{\theta}_2^E \geq \theta^*$ where $\theta^*$ satisfies

$$\kappa + \theta^* - \mathbb{E}\left[\hat{\theta}_2^E|\hat{\theta}_2^E \geq \theta^*\right] \geq 0.$$

*Furthermore,*

$$w_3 = \begin{cases} \overline{W} & \text{if } \hat{\theta}_2^E \geq \theta^* \\ \underline{W} & \text{if } \hat{\theta}_2^E < \theta^* \end{cases}$$

*where*

$$\overline{W} \geq \mathbb{E}\left[\hat{\theta}_2^E|\hat{\theta}_2^E \geq \theta^*\right] - \delta \text{ and } \underline{W} = \max\left\{\mathbb{E}\left[\hat{\theta}_2^E|\hat{\theta}_2^E < \theta^*\right] - \delta, 0\right\}.$$

*Proof.* Let $W^{E*}$ be the set of equilibrium actions defined by $\bar{w}^{E*}$. For all $w^E \in W^{E*}$, $M1$ and $M2$ engage in a Bertrand bidding game whose solution is standard. Each firm offers the worker the maximum between the surplus of market employment and zero so that

$$\bar{w}^{m*} = \begin{cases} \hat{\theta}^m(w^E) - \delta & \text{if } \hat{\theta}^m(w^E) - w^E - \delta > 0 \\ \hat{\theta}^m(w^E) - \delta \text{ or } 0 & \text{if } \hat{\theta}^m(w^E) - w^E - \delta = 0 \\ 0 & \text{if } \hat{\theta}^m(w^E) - w^E - \delta < 0. \end{cases} \tag{A6}$$

Let $Y' = \left\{(y_1, y_2)|\bar{w}^{E*}(y_1, y_2) > 0\right\}$ be the set of output pairs after which the employer makes a positive wage offer to the worker. In equilibrium, it must be the case that

$$\hat{\theta}^m(\bar{w}^{E*}(y_1, y_2)) - \bar{w}^{E*}(y_1, y_2) - \delta \leq 0 \;\forall (y_1, y_2) \in Y'. \tag{A7}$$

27. One could allow the two market firms to have different beliefs following off-equilibrium observations of $w^E$ without altering the results.

That is, the employer cannot make a positive wage offer that it knows the market will better, since otherwise the employer would be better off offering $w^E\!=\!0$ and saving the bidding cost $\delta$. Also it must be that

$$\kappa+\hat{\theta}_2^E(y_1,y_2) - \bar{w}^{E*}(y_1,y_2) - \delta \geq 0 \; \forall (y_1,y_2) \in Y'. \tag{A8}$$

That is, the employer must make non-negative profit to all workers to whom it makes a positive wage offer. Otherwise, it would again improve profit by offering $w^E\!=\!0$.

Now suppose there exists some pair of outputs $(y_1^1, y_2^1) \subset Y'$ and $(y_1^2, y_2^2) \subset Y'$ such that

$$\bar{w}^{E*}(y_1^1, y_2^1) = w^{E1} > w^{E2} = \bar{w}^{E*}(y_1^2, y_2^2)$$

Then, from the arguments above, it must be the case that

$$\hat{\theta}^m(w^{Ei}) - w^{Ei} - \delta \leq 0 \text{ for } i = 1, 2$$

as well as

$$\kappa+\hat{\theta}_2^E(y_1^i, y_2^i) - w^{Ei} - \delta \geq 0 \text{ for } i = 1, 2$$

But then the employer strictly improves profit by offering the wage $w^{E2}$ after observing $(y_1^1, y_2^1)$ instead of $w^{E1}$: it continues to retain the worker while paying strictly lower wage costs. So the employer can only make one positive wage offer $\overline{W}$ in equilibrium.

Since the wage offered to workers retained in equilibrium cannot vary with $\hat{\theta}_2^E$, it must be the case that the employer retains all workers for whom $\hat{\theta}_2^E \geq \theta^*$ where $\theta^*$ satisfies

$$\kappa+\theta^* - \overline{W} - \delta = 0.$$

That is, the employer retains all workers on whom it makes non-negative profit. In equilibrium, the market firms must correctly infer this rule, so that their estimate on worker talent after observing $\overline{W}$ is

$$\hat{\theta}_2^m(\overline{W}) = \mathbb{E}\left[\, \hat{\theta}_2^E \mid \hat{\theta}_2^E \geq \theta^* \,\right]$$

Thus, for an equilibrium to exist, it must be the case that the pair $(\overline{W}, \theta^*)$ satisfies the following two conditions:

$$\kappa+\theta^* - \overline{W} - \delta = 0 \tag{A9}$$

$$\mathbb{E}\left[\, \hat{\theta}_2^E | \hat{\theta}_2^E \geq \theta^* \,\right] - \delta \leq \overline{W}, \tag{A10}$$

which in turn imply that $\theta^*$ must satisfy

$$\kappa+\theta^* - \mathbb{E}\left[\, \hat{\theta}_2^E | \hat{\theta}_2^E \geq \theta^* \,\right] \geq 0. \tag{A11}$$

One now needs to establish the existence of a $\theta^*$ that satisfies equation (A11). First, note that because $\hat{\theta}_2^E$ is a linear combination of normal random variables, it is itself normally distributed with mean

$$\mathbb{E}\left[\frac{\sigma_\theta^2}{2\sigma_\theta^2+\sigma_\varepsilon^2}(y_1-a_1^*+y_2-a_2^*)+\frac{\sigma_\varepsilon^2}{2\sigma_\theta^2+\sigma_\varepsilon^2}\bar{\theta}\right]=$$
$$\mathbb{E}\left[\frac{\sigma_\theta^2}{2\sigma_\theta^2+\sigma_\varepsilon^2}(\theta+\varepsilon_1+\theta+\varepsilon_2)+\frac{\sigma_\varepsilon^2}{2\sigma_\theta^2+\sigma_\varepsilon^2}\bar{\theta}\right]=\frac{2\sigma_\theta^2+\sigma_\varepsilon^2}{2\sigma_\theta^2+\sigma_\varepsilon^2}\bar{\theta}=\bar{\theta}$$

and variance that one can denote by $\sigma^2$. Now consider the function

$$f(x)=x-\mathbb{E}\left[\hat{\theta}_2^E|\hat{\theta}_2^E\geq x\right] \tag{A12}$$

Two helpful results from distribution theory (Greene 2003: 759) are the following, where $\gamma$ is the normal hazard rate:

$$\mathbb{E}\left[\hat{\theta}_2^E|\hat{\theta}_2^E\geq x\right]=\bar{\theta}+\sigma\gamma\left(\frac{x-\bar{\theta}}{\sigma}\right) \tag{A13}$$

$$\gamma'(a)=\gamma(a)(\gamma(a)-a)\in(0,1)\ \forall a\in\mathbb{R} \tag{A14}$$

Together, these imply that $f'(x)>0$. Now by observation $\lim_{x\to-\infty}f(x)=-\infty$. From equation (A14)

$$\frac{\gamma'\left(x-\bar{\theta}/\sigma\right)}{\gamma\left(x-\bar{\theta}/\sigma\right)}=\gamma\left(\frac{x-\bar{\theta}}{\sigma}\right)-\left(\frac{x-\bar{\theta}}{\sigma}\right).$$

Observe that

$$\lim_{x\to\infty}\frac{\gamma'\left(x-\bar{\theta}/\sigma\right)}{\gamma\left(x-\bar{\theta}/\sigma\right)}=0$$

since

$$\gamma\left(\frac{x-\bar{\theta}}{\sigma}\right)=\frac{\mathbb{E}\left[\hat{\theta}_2^E|\hat{\theta}_2^E\geq x\right]-\bar{\theta}}{\sigma}\geq\frac{x-\bar{\theta}}{\sigma}$$

and $\lim_{x\to\infty}\frac{x-\bar{\theta}}{\sigma}=\infty$. So

$$\lim_{x\to\infty}\gamma\left(\frac{x-\bar{\theta}}{\sigma}\right)-\left(\frac{x-\bar{\theta}}{\sigma}\right)=0,$$

implying that $\lim_{x\to\infty}f(x)=0$.

The above arguments show that equation (A11) is satisfied for any $\theta^*\geq x^*$, where $x^*$ uniquely satisfies $f(x^*)=-\kappa$. Therefore, if an equilibrium exists, the employer retains all workers for whom $\hat{\theta}_2^E\geq\theta^*\geq x^*$ at a wage $\overline{W}$ that satisfies equation (A9). If $\hat{\theta}_2^E<\theta^*$, the employer sets $w^E=0$ and the worker's wage is implied by equation (A6).

Finally, the proposed equilibrium exists as long as the employer can make no profitable deviation to some $w^E$ other than 0 or $\overline{W}$. In order to rule out this possibility, one can set

$$\hat{\theta}^m(w^E) \geq \hat{\theta}^m(\overline{W})\ \forall w^E \neq \{0, \overline{W}\}$$

so that market firms infer the worker to have a higher ability after observing an out-of-equilibrium wage offer that after observing $\overline{W}$. ■

So every equilibrium wage schedule contains only two wages: one paid to worker types whose expected ability exceeds the threshold $\theta^*$ and who stay with the employer; and another paid to worker types whose expected ability falls short of $\theta^*$ and who separate from the employer. The intuition for wage pooling for retained workers is contained in footnote 9. Wage pooling for released workers emerges from costly bidding. Without costly bidding, the employer could still only retain workers above $\theta^*$ at a constant wage, but could credibly communicate private information to the market for worker types below $\theta^*$. Providing information for these worker types would be costless and would not affect third period profits. Assuming costly wage offers avoids the problem of solving for the optimal disclosure from the employer to the market at the same time as solving for the optimal disclosure from the employer to the worker.

Note that any worker type $\theta^*$ for whom (a) the employer earns zero profit while incurring total labor costs $\overline{W} - \delta$ and (b) whom the outside labor market cannot profitably bid away at a total cost larger than $\overline{W} - \delta$ gives rise to an equilibrium. In other words, any pair $(\overline{W}, \theta^*)$ that satisfies the following two conditions constitutes an equilibrium:

$$\kappa + \theta^* - \overline{W} - \delta = 0 \tag{A15}$$

$$\mathbb{E}\left[\hat{\theta}_2^E | \hat{\theta}_2^E \geq \theta^*\right] - \overline{W} - \delta \leq 0. \tag{A16}$$

From these conditions, one can observe that $\kappa > 0$ is a necessary (and, as the proof shows, sufficient) condition for there to exist an equilibrium in which the employer retains *any* worker types. If $\kappa = 0$ and the employer valued all worker types the same as the market firms did, it could never make zero profit on a worker type $\theta^*$ while paying it a wage equal to the expected market output of all types above it. Also, one can easily establish that the set of values of $\theta^*$ that satisfies equations (A15) and (A16) is unbounded above. However, career concerns exist whenever there is a positive probability of meeting the performance standard $\theta^*$. So as long as $\theta^*$ is finite, equilibrium multiplicity is not problematic.

Finally, note that $\theta^*$, $\overline{W}$, and $\underline{W}$ are all independent of the worker's first and second period effort choices as well as firms' beliefs about these effort choices. So, their values do not depend on the disclosure policy.

## Appendix B: Omitted Proofs in Text

### B.1 Proof of Lemma 1

*Proof.* $\hat{\theta}_2|y_1, a_1, a_1^*, a_2, a_2^*$ is a linear combination of normal random variables so is itself normal with variance $\sigma_2^2$ and mean

$$\begin{aligned}\mathbb{E}\left[\hat{\theta}_2|y_1, a_1, a_1^*, a_2, a_2^*\right] &= \mathbb{E}\left[\lambda_2(y_2 - a_2^*) + (1-\lambda_2)\hat{\theta}_1^E|y_1, a_1, a_1^*, a_2, a_2^*\right] \\ &= \lambda_2\left(\lambda_1(y_1 - a_1) + (1-\lambda_1)\bar{\theta} + (a_2 - a_2^*)\right) + (1-\lambda_2) \\ &\quad \left(\lambda_1(y_1 - a_1) + (1-\lambda_1)\bar{\theta} + \lambda_1\left(a_1 - a_1^*\right)\right) \\ &= \lambda_1(y_1 - a_1) + (1-\lambda_1)\bar{\theta} + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)\end{aligned}$$

and variance[28]

$$V\left[\lambda_2(\theta + \varepsilon_2)|y_1, a_1, a_1^*, a_2, a_2^*\right] = \lambda_2^2\left(\frac{\sigma_\varepsilon^2\sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2} + \sigma_\varepsilon^2\right) = \lambda_2\frac{\sigma_\varepsilon^2\sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2} \equiv \sigma_2^2.$$

### B.2 Proof of Proposition 1

*Proof.* Let $\tilde{\theta}_1(y_1) = \lambda_1(y_1 - a_1) + (1-\lambda_1)\bar{\theta}$. From Lemma 1, the worker's second period objective function is

$$W\mathbb{E}\left[1 - \Phi\left(\frac{\theta^* - \tilde{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}\right)\middle| y_1 \in P(y_1)\right] - \frac{C}{2}a_2^2 \tag{B1}$$

where $\Phi$ is the standard normal cumulative distribution function and the expectation is with respect to $y_1$. The first derivative is

$$\mathbb{E}\left[W\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \tilde{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}\right)\middle| y_1 \in P(y_1)\right] - Ca_2. \tag{B2}$$

As $a_2 \to 0$, (B2) $> 0$, while as $a_2 \to \infty$, (B2) $< 0$. So an interior solution to the optimization problem exists. The second derivative is

$$-\mathbb{E}\left[W\frac{\lambda_2^2}{\sigma_2^2}\phi'\left(\frac{\theta^* - \tilde{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}\right)\middle| y_1 \in P(y_1)\right] - C, \tag{B3}$$

28. The expression for the conditional variance of $\theta$ comes from DeGroot (1970).

which limits to $-\infty$ as $C \to \infty$. So there exists a $\bar{C}_1$ such that, for all $C > \bar{C}_1$, equation (B1) is globally concave. From the first order condition, the global maximum is

$$a_2^W[P(y_1), a_1] = \mathbb{E}\left[ \frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \tilde{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2^W - a_2^*)}{\sigma_2}\right) \middle| y_1 \in P(y_1) \right]. \tag{B4}$$

For $a_2^*$ to be consistent with the worker's strategy, we must have

$$a_2^*\big[P(y_1), a_1^*\big] = \mathbb{E}\left[ \frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \hat{\theta}_1(y_1)}{\sigma_2}\right) \middle| y_1 \in P(y_1) \right]. \tag{B5}$$

Suppose, there are $n \in \mathbb{N}$ nonsingleton elements of the disclosure policy $P$ and that nonsingleton element $Y_1^i$ for $i \in \{1, \cdots, n\}$ is made up of $m_i \in \mathbb{N}$ intervals. Denote by $\underline{y}_{ij}$ and $\bar{y}_ij,$ the left and right endpoints of the $j$th such interval. Also, suppose that $D$ $(P)$ is made up of $m_d \in \mathbb{N}$ intervals and denote by $\underline{y}_{dj}$ and $y_{dj},$ the left and right endpoints of the $j$th such interval. Finally, let

$$B(P) = \left\{ y^B | \lim_{\varepsilon \to 0} P\big(y^B + \varepsilon\big) \neq P\big(y^B - \varepsilon\big) \right\}.$$

be the set of boundary points between the elements of $P$. Each finite interval endpoint described above is a member of $B$ $(P)$. By applying the change of variables $z = y_1 - a_1,$ one can express the worker's first period objective function as

$$\sum_{j=1}^{m_d} \int_{\underline{y}_{dj} - \mathbb{1}\left(\underline{y}_{dj} \in B(P)\right) a_1}^{\bar{y}_{dj} - \mathbb{1}\left(\bar{y}_{dj} \in B(P)\right) a_1} \psi\big[a_1, a_1^*, a_2^W(z + a_1, a_1), a_2^*(z + a_1, a_1)\big] f_z(z) dz +$$
$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} \int_{\underline{y}_{ij} - \mathbb{1}\left(\underline{y}_{ij} \in B(P)\right) a_1}^{\bar{y}_{ij} - \mathbb{1}\left(\bar{y}_{ij} \in B(P)\right) a_1} \psi\big[a_1, a_1^*, a_2^W(Y_1^i, a_1), a_2^*(Y_1^i, a_1)\big] f_z(z) dz - \frac{C}{2} a_1^2. \tag{B6}$$

Here,

$$\psi\big[a_1, a_1^*, a_2^W(\cdot, a_1), a_2^*(\cdot, a_1)\big] =$$
$$W\left\{ 1 - \Phi\left[ \frac{\theta^* - \tilde{\theta}_1(z) - \lambda_2(a_1 - a_1^* + a_2^W(z + a_1, a_1) - a_2^*(z + a_1, a_1^*))}{\sigma_2} \right] \right\},$$

$f_z$ is the pdf of $z \sim N(\bar{\theta}, \sigma_\theta^2 + \sigma_\varepsilon^2),$ $\tilde{\theta}_1(z) = \lambda_1 z + (1 - \lambda_1)\bar{\theta},$ and $a_2^W(Y_1^i, a_1)$ and $a_2^*(Y_1^i, a_1^*)$ are constants independent of $z$. The indicator functions in the limits of integration take account of the fact that one need not transform the two infinite interval endpoints by subtracting $\alpha_1$. Differentiating with respect to $\alpha_1$ gives

$$\sum_{j=1}^{m_d}\int_{\underline{y}_{dj}-\mathbb{1}\left(\underline{y}_{dj}\in B(P)\right)a_1}^{\bar{y}_{dj}-\mathbb{1}\left(\bar{y}_{dj}\in B(P)\right)a_1}\begin{pmatrix}\frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^*-\tilde{\theta}_1(z)-\lambda_2(a_1-a_1^*+a_2^W(z+a_1,a_1)-a_2^*(z+a_1,a_1^*))}{\sigma_2}\right)\times\\ \left(1+\frac{\partial a_2^W(z+a_1,a_1)}{\partial a_1}-\frac{\partial a_2^*(z+a_1,a_1^*)}{\partial a_1}\right)\\ -Ca_2^W(z+a_1,a_1)\frac{\partial a_2^W(z+a_1,a_1)}{\partial a_1}\end{pmatrix}f_z(z)dz+$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_{\underline{y}_{ij}-\mathbb{1}\left(\underline{y}_{ij}\in B(P)\right)a_1}^{\bar{y}_{ij}-\mathbb{1}\left(\bar{y}_{ij}\in B(P)\right)a_1}\begin{pmatrix}\frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^*-\tilde{\theta}_1(z)-\lambda_2(a_1-a_1^*+a_2^W(Y_1^i,a_1)-a_2^*(Y_1^i,a_1^*))}{\sigma_2}\right)\times\\ \left(1+\frac{\partial a_2^W(Y_1^i,a_1)}{\partial a_1}\right)\\ -Ca_2^W(Y_1^i,a_1)\frac{\partial a_2^W(Y_1^i,a_1)}{\partial a_1}\end{pmatrix}f_z(z)dz+$$

$$\sum_{y^B\in B(P)}W\begin{pmatrix}\Phi\left(\frac{\theta^*-\hat{\theta}_1(y^B-a_1)-\lambda_2(a_1-a_1^*+a_2^W(\lim_{\varepsilon\to 0}P(y^B-\varepsilon),a_1)-a_2^*(\lim_{\varepsilon\to 0}P(y^B-\varepsilon),a_1^*))}{\sigma_2}\right)-\\ \Phi\left(\frac{\theta^*-\hat{\theta}_1(y^B-a_1)-\lambda_2(a_1-a_1^*+a_2^W(\lim_{\varepsilon\to 0}P(y^B+\varepsilon),a_1)-a_2^*(\lim_{\varepsilon\to 0}P(y^B+\varepsilon),a_1^*))}{\sigma_2}\right)\\ +\frac{C}{2}\left(\left(a_2^W(\lim_{\varepsilon\to 0}P(y^B-\varepsilon),a_1)\right)^2-\left(a_2^W(\lim_{\varepsilon\to 0}P(y^B+\varepsilon),a_1)\right)^2\right)\end{pmatrix}\times$$

$$f_z(y^B-a_1)-Ca_1. \tag{B7}$$

The first two terms simplify to

$$\sum_{j=1}^{m_d}\int_{\underline{y}_{dj}-\mathbb{1}\left(\underline{y}_{dj}\in B(P)\right)a_1}^{\bar{y}_{dj}-\mathbb{1}\left(\bar{y}_{dj}\in B(P)\right)a_1}\begin{pmatrix}\frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^*-\tilde{\theta}_1(z)-\lambda_2(a_1-a_1^*+a_2^W(z+a_1,a_1)-a_2^*(z+a_1,a_1^*))}{\sigma_2}\right)\times\\ \left(1-\frac{\partial a_2^*(z+a_1,a_1^*)}{\partial a_1}\right)\end{pmatrix}f_z(z)dz+$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_{\underline{y}_{ij}-\mathbb{1}\left(\underline{y}_{ij}\in B(P)\right)a_1}^{\bar{y}_{ij}-\mathbb{1}\left(\bar{y}_{ij}\in B(P)\right)a_1}\left(\frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^*-\tilde{\theta}_1(z)-\lambda_2(a_1-a_1^*+a_2^W(Y_1^i,a_1)-a_2^*(Y_1^i,a_1^*))}{\sigma_2}\right)\right)f_z(z)\mathrm{d}z \tag{B8}$$

For the sake of space, we omit the second derivative. One can show that it tends to $-\infty$ as $C\to\infty$ so that there exists a $\bar{C}_2$, such that equation (B7) is globally concave for all $C>\bar{C}_2$. In this case, the first order condition is sufficient for a maximum.

It remains to be shown that the first order condition has an interior solution.[29] The first step is to derive conditions under which the first three terms of equation (B7) are positive as $a_1\to 0$. As one can see from equation (B8), the second term is always positive. From equations (B4) and (B5), one obtains $\lim_{C\to\infty}a_2^W=0$, $\lim_{C\to\infty}a_2^*=0$, and $\lim_{C\to\infty}\frac{\partial a_2^*}{\partial a_1}=0$. So the third term of the equation (B7) limits to 0 as $C\to\infty$, whereas the first term limits to is

29. The first order condition for $\alpha_1$ is obtained by setting equation (B7) equal to zero.

$$\sum_{j=1}^{m_d} \int_{\underline{y}_{dj} - \mathbb{1}\left(\underline{y}_{dj} \in B(P)\right) a_1}^{\bar{y}_{dj} - \mathbb{1}\left(\bar{y}_{dj} \in B(P)\right) a_1} \left( \frac{\lambda_2 W}{\sigma_2} \phi \left( \frac{\theta^* - \tilde{\theta}_1(z) - a_1 + a_1^*}{\sigma_2} \right) \right) f_z(z) dz > 0.$$

So there exists a $\bar{C}_3$, such that the first and the third terms of the equation (B7) are positive at $\alpha_1 = 0$ for all $C \geq \bar{C}_3$. By observation, equation (B7) tends to $-\infty$ as $a_1 \to \infty$. So for all $C \geq \bar{C}_3$, the first order condition has a unique interior solution given by $a_1^W$. Thus, whenever $C > \max\{\bar{C}_1, \bar{C}_2, \bar{C}_3\}$, the worker's first and second period optimization problems have unique interior solutions. One then obtains expressions for equilibrium effort by imposing the condition $a_1^* = a_1^W$ and noting from equations (B4) and (B5) that $a_2^* = a_2^W$ whenever $a_1^* = a_1^W$.

It remains to be shown that $a_1^*$ is unique. Consider the function

$$h(a_1^*) = a_1^* - \int_{-\infty}^{\infty} \frac{\lambda_2 W}{\sigma_2} \phi \left( \frac{\theta^* - \tilde{\theta}_1(z)}{\sigma_2} \right) f_z(z) dz -$$
$$\sum_{j=1}^{m_d} \int_{\underline{y}_{dj} - \mathbb{1}\left(\underline{y}_{dj} \in B(P)\right) a_1^*}^{\bar{y}_{dj} - \mathbb{1}\left(\bar{y}_{dj} \in B(P)\right) a_1^*} \left( \frac{W}{C} \right)^2 \frac{\lambda_2^2 \lambda_1}{\sigma_2^3} \left( \frac{\tilde{\theta}_1(z) - \theta^*}{\sigma_2} \right) \phi^2 \left( \frac{\theta^* - \tilde{\theta}_1(z)}{\sigma_2} \right) f_z(z) \mathrm{d}z. \tag{B9}$$

As $a_1^* \to 0$, $h(a_1^*) \to \bar{h} > 0$ and as $a_1^* \to \infty$, $h(a_1^*) \to \infty$. So as long as $h$ is strictly increasing, there exists a unique solution. Differentiating gives

$$1 - \sum_{j=1}^{m_d} \left( \frac{W}{C} \right)^2 \frac{\lambda_2^2 \lambda_1}{\sigma_2^3} \left[ \begin{array}{l} \left( \frac{\hat{\theta}_1(\underline{y}_{dj} - a_1^*) - \theta^*}{\sigma_2} \right) \phi^2 \left( \frac{\theta^* - \hat{\theta}_1(\underline{y}_{dj} - a_1^*)}{\sigma_2} \right) f_z(\underline{y}_{dj} - a_1^*) dz - \\ \left( \frac{\hat{\theta}_1(\bar{y}_{dj} - a_1^*) - \theta^*}{\sigma_2} \right) \phi^2 \left( \frac{\theta^* - \hat{\theta}_1(\bar{y}_{dj} - a_1^*)}{\sigma_2} \right) f_z(\bar{y}_{dj} - a_1^*) dz \end{array} \right]. \tag{B10}$$

Clearly, there exists a $\bar{C}_4$ such that equation (B10) is positive for all $C \geq \bar{C}_4$. The proposition is established by setting $\bar{C} = \max\{\bar{C}_1, \bar{C}_2, \bar{C}_3, \bar{C}_4\}$. ■

### B.3 Proof of Corollary 1

*Proof*. Suppose the elements of $P$ are $\{Y_1^i\}_{i=1}^n \cup D(P)$ and let $\tilde{a}(y_1) = \frac{W}{C} \frac{\lambda_2}{\sigma_2} \phi \left( \frac{\theta^* - \hat{\theta}_1(y_1)}{\sigma_2} \right)$.

$$\mathbb{E}[a_2^*] = \sum_i \mathbb{E}[\tilde{a}(y_1) | y_1 \in Y_1^i] \Pr[y_1 \in Y_1^i] + \int_{y_1 \in D(P)} \tilde{a}(y_1) f_y(y_1) dy_1 =$$
$$\int_{-\infty}^{\infty} \tilde{a}(y_1) f_y(y_1) dy_1,$$

which is independent of $P$. ■

B.4 Proof of Corollary 2

*Proof.* First, let $\tilde{a}(y_1)$ be defined as in Section B.3. Now suppose the elements of $P$ are $\{Y_1^i\}_{i=1}^n \cup D(P)$ and the elements of $P'$ are $Y_1^{11} \cup Y_1^{12} \cup \{Y_1^i\}_{i=2}^n \cup D(P')$. $\mathbb{E}\left[(a_2^*)^2 | P'\right] > \mathbb{E}\left[(a_2^*)^2 | P\right]$ holds if

$$\left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{11}\right]\right)^2 \Pr\left[y_1 \in Y_1^{11}\right] + \left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{12}\right]\right)^2 \Pr\left[y_1 \in Y_1^{12}\right]$$
$$> \left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{1}\right]\right)^2 \Pr\left[y_1 \in Y_1^{1}\right] \Rightarrow$$
$$\left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{11}\right]\right)^2 \frac{\Pr\left[y_1 \in Y_1^{11}\right]}{\Pr\left[y_1 \in Y_1^{1}\right]} + \left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{12}\right]\right)^2 \frac{\Pr\left[y_1 \in Y_1^{12}\right]}{\Pr\left[y_1 \in Y_1^{1}\right]}$$
$$> f\left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{11}\right] \frac{\Pr\left[y_1 \in Y_1^{11}\right]}{\Pr\left[y_1 \in Y_1^{1}\right]} + \mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{12}\right] \frac{\Pr\left[y_1 \in Y_1^{12}\right]}{\Pr\left[y_1 \in Y_1^{1}\right]}\right)^2,$$

which is satisfied by the discrete version of Jensen's inequality.

Second, suppose the elements of $P$ are $\{Y_1^i\}_{i=1}^n \cup D(P)$ and the elements of $P'$ are $Y_1^{1'} \cup \{Y_1^i\}_{i=2}^n \cup D(P')$ where $Y_1^{1'} \subset Y_1^1$ and $D(P) \subset D(P')$. $\mathbb{E}\left[(a_2^*)^2 | P'\right] > \mathbb{E}\left[(a_2^*)^2 | P\right]$ holds if

$$\left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{1'}\right]\right)^2 \Pr\left[y_1 \in Y_1^{1'}\right] +$$
$$\mathbb{E}\left[(\tilde{a}(y_1))^2 | y_1 \in (D(P') \backslash D(P))\right] \Pr[y_1 \in (D(P') \backslash D(P))]$$
$$> \left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{1}\right]\right)^2 \Pr\left[y_1 \in Y_1^{1}\right]$$

Now,

$$\left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{1'}\right]\right)^2 \Pr\left[y_1 \in Y_1^{1'}\right] +$$
$$\mathbb{E}\left[(\tilde{a}(y_1))^2 | y_1 \in (D(P') \backslash D(P))\right] \Pr[y_1 \in (D(P') \backslash D(P))]$$
$$> \left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{1'}\right]\right)^2 \Pr\left[y_1 \in Y_1^{1'}\right] +$$
$$(\mathbb{E}[\tilde{a}(y_1)|y_1 \in (D(P') \backslash D(P))])^2 \Pr[y_1 \in (D(P') \backslash D(P))]$$

by the probability version of Jensen's inequality. Moreover, by the arguments above, the last expression is strictly bigger than

$$\left(\mathbb{E}\left[\tilde{a}(y_1)|y_1 \in Y_1^{1}\right]\right)^2 \Pr\left[y_1 \in Y_1^{1}\right].$$

To complete the proof, note that every refinement of $P$ can be generated by a stepwise repetition of the above two simple refinements. Thus, by applying the above arguments sequentially, one arrives at the conclusion. ■

B.5 Proof of Proposition 2

*Proof.* $\hat{\theta}_1(y_1) = \lambda_1(\theta+\varepsilon) + \lambda_1\bar{\theta}$ so $\mathbb{E}\left[\hat{\theta}_1(y_1)\right] = \bar{\theta}$ and $V\left[\hat{\theta}_1(y_1)\right] = \lambda_1^2\left(\sigma_\theta^2 + \sigma_\varepsilon^2\right) = \sigma_\theta^2\lambda_1 \equiv \sigma_1^2$. So $\mathbb{E}\left[\phi\left(\frac{\hat{\theta}_1(y_1)-\theta^*}{\sigma_2}\right)\right] = \mathbb{E}[\phi(x)]$ where

$x \sim N(\mu, \sigma^2)$ and $\mu = \bar{\theta} - \theta^*/\sigma_2$ and $\sigma = \sigma_1^2/\sigma_2^2$, where, from Section B.1, $\sigma_2^2 \equiv \lambda_2 \frac{\sigma_\varepsilon^2 \sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2}$. This expectation becomes

$$\begin{aligned}\mathbb{E}[\phi(x)] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right] \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \mathrm{d}x = \\ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}x^2 - \frac{1}{2\sigma^2}(x-\mu)^2\right] \mathrm{d}x = \\ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(x^2 + \frac{1}{\sigma^2}(x-\mu)^2\right)\right] \mathrm{d}x.\end{aligned}$$

Expanding and completing the square gives the last expression as

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\sigma^2+1}{\sigma^2}\left(x - \frac{\mu}{\sigma^2+1}\right)^2 + \frac{\mu^2}{\sigma^2+1}\right)\right] \mathrm{d}x$$

which in turn becomes

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2+1}} \exp\left[-\frac{1}{2}\left(\frac{\mu^2}{\sigma^2+1}\right)\right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left[-\frac{1}{2\tilde{\sigma}^2}\left(x - \frac{\mu}{\sigma^2+1}\right)^2\right] \mathrm{d}x$$

where $\tilde{\sigma}^2 = \sigma^2/\sigma^2{+}1$. Since the integrand in the above expression is a normal density, it integrates to 1, and so

$$\mathbb{E}[\phi(x)] = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2+1}} \exp\left[-\frac{1}{2}\left(\frac{\mu^2}{\sigma^2+1}\right)\right].$$

Using similar arguments, one can show that

$$\mathbb{E}\left[\phi^2\left(\frac{\hat{\theta}_1(y_1) - \theta^*}{\sigma_2}\right)\right] = \frac{1}{\sqrt{2\sigma^2+1}} \frac{1}{2\pi} \exp\left(-\frac{\mu^2}{2\sigma^2+1}\right) \text{ and}$$

$$\mathbb{E}\left[\left(\frac{\hat{\theta}_1(y_1) - \theta^*}{\sigma_2}\right)\phi^2\left(\frac{\hat{\theta}_1(y_1) - \theta^*}{\sigma_2}\right)\right] = \frac{1}{\sqrt{2\sigma^2+1}} \frac{1}{2\pi}\left(\frac{\mu}{2\sigma^2+1}\right) \exp\left(-\frac{\mu^2}{2\sigma^2+1}\right).$$

The loss in surplus in the second period from full disclosure [equation (12) in the text] can be divided through by $\frac{\sqrt{2\sigma^2+1}}{\exp(-(1/2)\mu^2/2\sigma^2+1)} \frac{2C\sigma_2^2}{(W\lambda_2)^2}$ and expressed as

$$\frac{1}{2\pi} - \frac{1}{2\pi} \frac{\sqrt{2\sigma^2+1}}{\sigma^2+1} \exp\left(-\frac{\sigma^2\mu^2}{(\sigma^2+1)(2\sigma^2+1)}\right), \tag{B11}$$

while the change in first period surplus from full disclosure [equation (14) in the text] can be similarly divided to become

$$\frac{\lambda_1}{C\sigma_2}\frac{1}{\pi}\frac{\mu}{2\sigma^2+1}-\frac{2}{C}\frac{\lambda_1}{\sigma_2}\frac{W\lambda_2}{\sigma_2}\frac{1}{(2\pi)^{1.5}}\frac{1}{\sqrt{\sigma^2+1}}\frac{\mu}{2\sigma^2+1}\exp\left(-\frac{1}{2}\frac{\mu^2}{\sigma^2+1}\right)-$$
$$\frac{\lambda_1}{\sigma_2}\left(\frac{W}{C}\right)^2\frac{\lambda_2^2\lambda_1}{\sigma_2^3}\frac{1}{(2\pi)^2}\frac{1}{\sqrt{2\sigma^2+1}}\frac{\mu^2}{(2\sigma^2+1)^2}\exp\left(-\frac{\mu^2}{2\sigma^2+1}\right). \tag{B12}$$

As $\bar{\theta}\to\infty$, $\mu\to\infty$, and equation (B11) limits to $1/2\pi$ and equation (B12) limits to $\infty$. Now $\sigma_1^2/\sigma_2^2=2(\sigma_\theta^2/\sigma_\varepsilon^2)+1$, so as $\sigma_\theta^2/\sigma_\varepsilon^2\to\infty$, $\sigma\to\infty$. Moreover, as $\sigma\to\infty$, equation (B11) limits to $1/2\pi$ and equation (B12) limits to 0.[30] ■

## B.6 Proof of Proposition 3

*Proof*. Let $D_\theta\,(P)$ be as defined in the text. The proof proceeds in two stages. First, it establishes the form that $D_\theta\,(P^S)$ takes. It then shows that there exists a $P^S$ that induces $D_\theta\,(P^S)$.

Let $P$ and $P'$ be disclosure policies that satisfy $D_\theta(P')=D_\theta(P)\cup(t,t+\varepsilon)$. By Corollary 2, moving from $P$ to $P'$ increases the expected second period effort costs, and the following gives the amount by which it does so when $\varepsilon$ is small. In the proof, $f_\theta$ denotes the probability density function of $\hat{\theta}_1\sim N(\bar{\theta},\sigma_1^2)$.

*Lemma 2*.

$$\lim_{\varepsilon\to0}\left(\mathbb{E}\left[\left(a_2^*\right)^2|P'\right]-\mathbb{E}\left[\left(a_2^*\right)^2|P\right]\right)=$$
$$\frac{1}{2C}\left(\frac{W\lambda_2}{\sigma_2}\right)^2\left(\phi\left(\frac{\theta^*-t}{\sigma_2}\right)-\mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\notin D_\theta(P)\right]\right)^2 f_\theta(t) \tag{B13}$$

*Proof*. One can express $\mathbb{E}\left[\phi^2\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|P'\right]$ as

$$\Pr\left[\hat{\theta}_1\in D_\theta(P)\cup(t,t+\varepsilon)\right]\mathbb{E}\left[\phi^2\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\in D_\theta(P)\cup(t,t+\varepsilon)\right]+$$
$$\Pr\left[\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]\mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]^2$$

30. An implicit assumption here is that $\sigma_2^2$ is kept constant as $\sigma_\theta^2/\sigma_\varepsilon^2\to\infty$.

which can be further expanded as

$$\left(\begin{array}{c}\int_{\hat{\theta}_1\in D_\theta(P)}\phi^2\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1+\int_t^{t+\varepsilon}\phi^2\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1+\\ \left[\int_{\hat{\theta}_1\notin D_\theta(P)}f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1-\int_t^{t+\varepsilon}f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1\right]\times\\ \left[\frac{\int_{\hat{\theta}_1\notin D_\theta(P)}\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1-\int_t^{t+\varepsilon}\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1}{\int_{\hat{\theta}_1\notin D_\theta(P)}f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1-\int_t^{t+\varepsilon}f_\theta\left(\hat{\theta}_1\right)\mathrm{d}\hat{\theta}_1}\right]^2\end{array}\right)$$

The derivative of the term in brackets with respect to $\varepsilon$ is

$$\phi^2\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right)f_\theta(t+\varepsilon)-f_\theta(t+\varepsilon)\left(\mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]\right)^2$$

$$+\Pr\left[\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]2\mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]\times$$

$$\frac{\left(\begin{array}{c}-\phi\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right)f_\theta(t+\varepsilon)\Pr\left[\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]+\\ \mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]\Pr\left[\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]f_\theta(t+\varepsilon)\end{array}\right)}{\left(\Pr\left[\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]\right)^2}$$

which reduces to

$$\phi^2\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right)f_\theta(t+\varepsilon)+\left(\mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]\right)^2f_\theta(t+\varepsilon)$$

$$-\phi\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right)2\mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\hat{\theta}_1\notin D_\theta(P)\cup(t,t+\varepsilon)\right]f_\theta(t+\varepsilon).$$

Taking the limit as $\varepsilon\to 0$ gives the result. ■

Now $a_1^*(P')$ can be written as

$$\mathbb{E}\left[\frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\right]+\int_{\hat{\theta}_1\in D_\theta(P)}\left(\frac{W}{C}\right)^2\frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{\hat{\theta}_1-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)f_\theta(\hat{\theta}_1)d\hat{\theta}_1+$$

$$\int_t^{t+\varepsilon}\left(\frac{W}{C}\right)^2\frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{\hat{\theta}_1-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)f_\theta(\hat{\theta}_1)d\hat{\theta}_1 \tag{B14}$$

Taking the derivative respect to $\varepsilon$ and letting $\varepsilon\to 0$ gives

$$\left(\frac{W}{C}\right)^2\frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{t-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-t}{\sigma_2}\right)f_\theta(t). \tag{B15}$$

So the change in first period welfare from adding beliefs $(t, t+\varepsilon)$ to $D_\theta(P)$ for small $\varepsilon$ is approximately

$$\left(\frac{W}{C}\right)^2 \frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{t-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-t}{\sigma_2}\right) f_\theta(t)\left(1-Ca_1^*(P)\right). \tag{B16}$$

Let

$$L(t,P) = A\left(\frac{t-\theta^*}{\sigma_2}\right) \tag{B17}$$

where $A = \frac{2\lambda_1}{C\sigma_2}\left(1-Ca_1^*(P)\right)$ and

$$R(t,P) = \left(\frac{\mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\,\middle|\,\hat{\theta}_1 \notin D_\theta(P)\right]}{\phi\left(\frac{\theta^*-t}{\sigma_2}\right)} - 1\right)^2. \tag{B18}$$

From equations (B13) and (B16), one can conclude that $P^S$ must satisfy

$$L\left(\hat{\theta}_1, P^S\right) \geq R\left(\hat{\theta}_1, P^S\right)\ \forall \hat{\theta}_1 \in D_\theta\left(P^S\right)$$

and

$$L\left(\hat{\theta}_1, P^S\right) < R\left(\hat{\theta}_1, P^S\right)\ \forall \hat{\theta}_1 \notin D_\theta\left(P^S\right).$$

Now, $P^S$ must also satisfy $Ca_1^*\left(P^S\right) \leq 1$ since otherwise one could improve social welfare by removing a small measure of positive feedback. Moreover, $a_1^*\left(P^N\right) \leq a_1^*\left(P^S\right)$ since otherwise $P^N$ would yield a higher surplus than $P^S$. This implies that $P^S$ contains no negative feedback since any disclosure policy $P^1$ that provides negative feedback and for which $Ca_1^*\left(P^N\right) \leq Ca_1^*\left(P^1\right)$ can be replaced by a disclosure policy $P^2$ that contains no negative feedback and for which $a_1^*\left(P^1\right) = a_1^*\left(P^2\right)$. Since $P^2$ implements the same first period effort while reducing risk, it provides higher surplus than $P^1$.

Fix a disclosure policy $P'$ for which $D_\theta(P') \cap (-\infty, \theta^*) = \emptyset$. This implies that

$$0 < \mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\,\hat{\theta}_1 \notin D_\theta(P')\right] < \phi(0).$$

So, since $\phi(\theta^* - t/\sigma_2)$ is strictly decreasing on $t \in (\theta^*, \infty)$ from $\phi(0)$ to 0, there exists a unique point $\tilde{t}$ such that

$$\phi\left(\frac{\theta^*-\tilde{t}}{\sigma_2}\right) = \mathbb{E}\left[\phi\left(\frac{\theta^*-\hat{\theta}_1}{\sigma_2}\right)\middle|\,\hat{\theta}_1 \notin D_\theta(P')\right].$$

Moreover, $R(t, P') > 0$ for all $t \neq \tilde{t}$. Suppose $Ca_1^*(P^S) = 1$. Then $L(t, P^S) = 0$ for all $t \in D_\theta(P^S)$ while there exists some $t'$ for which $R(t', P^S) > 0$. Thus, replacing $P^S$ with a disclosure policy $\tilde{P}$ satisfying $D_\theta(\tilde{P}) = D_\theta(P^S) \setminus [t', t'+\varepsilon]$ improves social welfare for small enough $\varepsilon$, contradicting the optimality of $P^S$. So $Ca_1^*(P^S) < 1$. So further assume that $Ca_1^*(P') < 1$.

From the above arguments, one can also conclude that $R\,(t,\ P')$ is strictly decreasing on $t \in (\theta^*, \tilde{t})$. Since $L(t, P')$ is strictly increasing on $t \in (\theta^*, \tilde{t})$, there exists a unique point $t_L(P') < \tilde{t}$ at which $L(t_L(P'), P') = R(t_L(P'), P')$. Clearly, $L(t, P') < R(t, P')\ \forall t \in (\theta^*, t_L(P'))$ and $L(t, P') > R(t, P')\ \forall t \in (t_L(P'), \tilde{t})$.

Let $\mathrm{d} = \mathbb{E}\left[\phi\left(\frac{\theta^* - \hat{\theta}_1}{\sigma_2}\right) \middle| \hat{\theta}_1 \notin D_\theta(P')\right]$.

$$\frac{\partial R}{\partial t} = 2\left[\mathrm{d}\sqrt{2\pi}\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right) - 1\right]\mathrm{d}\sqrt{2\pi}\left(\frac{t - \theta^*}{\sigma_2^2}\right)\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right), \tag{B19}$$

which is strictly bigger than zero on $t \in (\tilde{t}, \infty)$, since $\mathrm{d}\sqrt{2\pi}\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right) > 1$ on the same domain.

$$\begin{aligned}\frac{\partial^2 R}{\partial t^2} &= 2\left[\mathrm{d}\sqrt{2\pi}\left(\frac{t - \theta^*}{\sigma_2^2}\right)\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right)\right]^2 + \\ &2\left[\mathrm{d}\sqrt{2\pi}\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right) - 1\right] \times \\ &\left[\frac{\mathrm{d}\sqrt{2\pi}}{\sigma_2^2}\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right) + \mathrm{d}\sqrt{2\pi}\left(\frac{t - \theta^*}{\sigma_2^2}\right)^2\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right)\right],\end{aligned} \tag{B20}$$

which is also strictly bigger than zero on $t \in (\tilde{t}, \infty)$. Thus, $R$ is strictly convex and increasing on $t \in (\tilde{t}, \infty)$ while $L$ is linear on the same domain. Moreover, $\lim_{t\to\infty} \frac{\partial R}{\partial t} = \infty$ while $\lim_{t\to\infty} \frac{\partial L}{\partial t} = A < \infty$. So there exists a unique point $t_H(P') > t_L(P')$ at which $L(t_H(P'), P') = R(t_H(P'), P')$ and at which

$$\left[\frac{\partial L(t, P')}{\partial t}\right]_{t=t_H(P')} < \left[\frac{\partial R(t, P')}{\partial t}\right]_{t=t_H(P')}. \tag{B21}$$

Clearly, $L(t, P') > R(t, P')\ \forall t \in (\tilde{t}, t_H(P'))$ and $L(t, P') < R(t, P')\ \forall t \in (t_H(P'), \infty)$. These arguments establish that $L(t, P') > R(t, P')\ \forall t \in (t_L(P'), t_H(P'))$ and $L(t, P') < R(t, P')\ \forall t \in (\theta^*, t_L(P')) \cup (t_H(P'), \infty)$.

Now one can show that $t_H$ is bounded above. By the implicit function theorem

$$\frac{\partial t_H}{\partial A} = \frac{\partial L/\partial A}{\partial R/\partial t_H - \partial L/\partial t_H} = \frac{t_H - \theta^*/\sigma_2}{\partial R/\partial t_H - \partial L/\partial t_H} > 0 \tag{B22}$$

and

$$\frac{\partial t_H}{\partial \mathrm{d}} = -\frac{\partial R/\partial \mathrm{d}}{\partial R/\partial t_H - \partial L/\partial t_H} = -\frac{2\left(\mathrm{d}\phi^{-1}(\theta^* - t_H/\sigma_2) - 1\right)\phi^{-1}(\theta^* - t_H/\sigma_2)}{\partial R/\partial t_H - \partial L/\partial t_H} < 0. \tag{B23}$$

Now $A$ is bounded above by $2\lambda_1/C\sigma_2$ and as argued above d has some lower bound $\underline{\mathrm{d}}$. By the arguments in the previous paragraph, the equation

$$\frac{2\lambda_1}{C\sigma_2}\left(\frac{\theta^* - t}{\sigma_2}\right) = \left(\underline{\mathrm{d}}\phi^{-1}\left(\frac{\theta^* - t}{\sigma_2}\right) - 1\right)^2$$

has two solutions $\underline{t}$ and $\bar{t}$, where $\theta^* < \underline{t} < \bar{t} < \infty$. Finally, by equations (B22) and (B23) $t_H \leq \bar{t}$.

Consider social surplus as a function of disclosure policies of the form $D_\theta = (\theta_L, \theta_H)$ where $\theta^* \leq \theta_L \leq \theta_H \leq \bar{t}$. The first section of the proof established the continuity of surplus in $\theta_L$ and $\theta_H$, so since $\theta^* \leq \theta_L \leq \theta_H \leq \bar{t}$ is a compact set, one can use the Weierstrass Maximum Theorem to establish the existence of a surplus maximizing disclosure policy $P^S$ for which $D_\theta(P^S) = (\theta'_L, \theta'_H)$ and $\theta^* < \theta'_L < \theta'_H \leq \bar{t}$.

Thus, if one can find a disclosure policy $P$ for which $D(P) = (y^*, y^{**})$ and where $y^*$ and $y^{**}$ satisfy the following, $P$ is optimal:

$$\begin{pmatrix} y^* - a_1^*(P) \\ y^{**} - a_1^*(P) \end{pmatrix} = \begin{pmatrix} \theta'_L - (1-\lambda_1)\bar{\theta}/\lambda_1 \\ \theta'_H - (1-\lambda_1)\bar{\theta}/\lambda_1 \end{pmatrix}. \tag{B24}$$

Satisfying equation (B24) is equivalent to finding a disclosure policy $P$ with $D(P) = \left[y^*, y^* + \frac{\theta'_H - \theta'_L}{\lambda_1}\right]$ where $y^*$ satisfies[31]

$$y^* - a_1^*(y^*) = \frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1}. \tag{B25}$$

Now when $y^* = \frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1}$ the LHS of equation (B25) is smaller than the RHS since $a_1^*(y^*) > 0$ and when $y^* = \theta'_L - (1-\lambda_1)\bar{\theta}/\lambda_1 + 2\mathbb{E}[a_2^*]$, the LHS is bigger than the RHS since $a_1^*(y^*) < 2\mathbb{E}[a_2^*]$. Since the LHS of equation (B25) is continuous is $y^*$, there exists some

$$y^* \in \left[\frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1}, \frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1} + 2\mathbb{E}[a_2^*]\right]$$

for which equation (B25) holds and an optimal disclosure policy with the stated form exists. ■

31. Here, we make an abuse of notation by making the dependence of $a_1^*$ on $y^*$ rather than the disclosure policy $P$.

## References


Alonso, R., and N. Matouschek. 2008. "Optimal Delegation," 75 *Review of Economic Studies* 259–93.

Aoyagi, M. 2010. "Information Feedback in a Dynamic Tournament," 70 *Games and Economic Behavior* 242–60.

Cremer, J. 1995. "Arm's Length Relationships," 110 *Quarterly Journal of Economics* 275–95.

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. New York, NY: McGraw-Hill.

Dessein, W. 2002. "Authority and Communication in Organizations," 69 *Review of Economic Studies* 811–38.

Dewatripont, M., I. Jewitt, and J. Tirole. 1999. "The Economics of Career Concerns, Part I: Comparing Information Structures," 66 *Review of Economic Studies* 183–98.

Dubey, P., and J. Geanakoplos. 2010. "Grading Exams: 100, 99,. . .,1 or A, B, C? Incentives in Games of Status," 69 *Games and Economic Behavior* 72–94.

Ederer, F. 2010. "Feedback and Motivation in Dynamic Tournaments," 19 *Journal of Economics and Management Strategy* 733–69.

Fuchs, W. 2007. "Contracting with Repeated Moral Hazard and Private Evaluations," 97 *American Economic Review* 1432–48.

Gershkov, A., and M. Perry. 2009. "Tournaments with Midterm Reviews," 66 *Games and Economic Behavior* 162–90.

Goltsman, M., and A. Mukherjee. 2010. "Interim Performance Feedback in Multistage Tournaments: The Optimality of Partial Disclosure," 29 *Journal of Labor Economics* 229–65.

Greene, W. H. 2003. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Grossman, S. J. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality," 24 *Journal of Law & Economics* 461–83.

Holmström, B. 1979. "Moral Hazard and Observability," 10 *Bell Journal of Economics* 74–91.

———. 1999. "Managerial Incentive Problems: A Dynamic Perspective," 66 *Review of Economic Studies* 169–82.

Jackman, J. M., and M. H. Strober. 2003. "Fear of Feedback," 81 *Harvard Business Review* 101–07.

Kamenica, E., and M. Gentzkow. 2011. "Bayesian Persuasion," 101 *American Economic Review* 2590–2615.

Kersley, B., C. Alpin, J. Forth, A. Bryson, H. Bewley, G. Dix, and S. Oxenbridge. 2006. *Inside the Workplace: Findings from the 2004 Workplace Employment Relations Survey*. Abingdon, Oxfordshire: Routledge.

Koch, A., and E. Peyrache. 2011. "Aligning Ambition and Incentives," 27 *Journal of Law, Economics and Organization* 655–688.

Kovrijnykh, A. 2007. "Career Uncertainty and Dynamic Incentives," Working paper, University of Chicago.

Landers, R. M., J. B. Rebitzer, and L. J. Taylor. 1996. "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms," 86 *American Economic Review* 329–48.

Lizzeri, A., M. A. Meyer, and N. Persico. 2002. "The Incentive Effects of Interim Performance Evaluations," Penn CARESS Working Papers 02-09, University of Pennsylvania.

MacLeod, W. B. 2003. "Optimal Contracting with Subjective Evaluation," 93 *American Economic Review* 216–40.

Maister, D. H. 1993. *Managing the Professional Service Firm*. New York, NY: Free Press Paperbacks.

Martinez, L. 2009. "Reputation, Career Concerns, and Job Assignments," 9 *B.E. Journal of Theoretical Economics*.

Meyer, H. H., E. Kay, and J. R. P. French. 1965. "Split Roles in Performance Appraisal," 43 *Harvard Business Review* 123–29.

Milgrom, P. R. 1981. “Good News and Bad News: Representation Theorems and Applications,” 12 *Bell Journal of Economics* 380–91.

Mukherjee, A. 2008. “Career Concerns, Matching, And Optimal Disclosure Policy,” 49 *International Economic Review* 1211–50.

Nafziger, J. 2009. “Timing of Information in Agency Problems with Hidden Actions,” 45 *Journal of Mathematical Economics* 751–66.

Prendergast, C. 1999. “The Provision of Incentives in Firms,” 37 *Journal of Economic Literature* 7–63.

Ray, K. 2007. “The Retention Effect of Withholding Performance Information,” 82 *The Accounting Review* 389–425.

Waldman, M. 1984. “Job Assignments, Signalling, and Efficiency,” 15 *RAND Journal of Economics* 255–67.

Yildirim, H. 2005. “Contests with Multiple Rounds,” 51 *Games and Economic Behavior* 213–27.