The Benefits of Limited Feedback in Organizations

Stephen Hansen*
Universitat Pompeu Fabra

July 25, 2010

Abstract

In most firms, managers periodically assess workers' performance. Evidence suggests that managers withhold information during these reviews, and some observers argue that this necessarily reduces surplus. This paper assesses the validity of this argument when workers have career concerns. Disclosure has two effects: it exposes the worker to uncertainty about future effort levels, but allows him to use current effort to influence his employer's beliefs about future effort. The surplus-maximizing disclosure policy reveals output realizations in the center of the distribution, but not in the tails. Thus, it is efficient for firms to reveal some but not all performance information.

Keywords: Performance Appraisal, Career Concerns, Incentives, Risk

JEL Codes: D82, D86, L20, M12

*Address: Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain. Email: `stephen.hansen@upf.edu`. This paper is a modified version of the second chapter of my PhD dissertation submitted to the London School of Economics and Political Science. 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 Garciano, Mike Gibbs, and Karl Schlag also contributed much of their time to helping improve the paper. In addition, the paper (in its current incarnation) has benefited from the advice of Heski Bar-Isaac, Tim Besley, Pablo Casas, Joan de Martí, Nathan Foley-Fisher, Yuk-Fai Fong, Clare Leaver, Fabrizio Germano, Ian Jewitt, Jin Li, Michael McMahon, Hannes Müller, Marco Ottaviani, Canice Prendergast, Ronny Razin. The usual disclaimer applies.

1 Introduction

Performance appraisal systems within organizations are an ancient and common institution.[1] They also consume many firm resources through the demands they place on human-resource offices and managerial time.[2]

At the same time, performance appraisals are often not fully informative. A typical firm conducts periodic reviews in which supervisors give numerical ratings to the workers they oversee. Figure 1 displays a rating distribution from a medium-sized service firm in the United States, where 1 is associated with highest performance, and 5 the worst.[3]

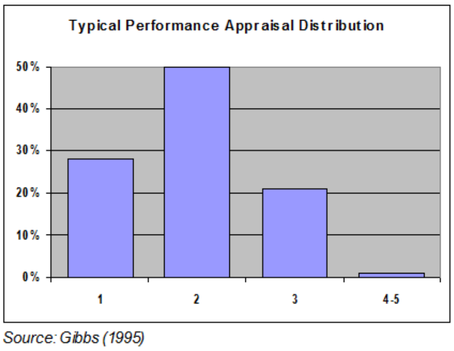

Figure 1: Rating Distribution Example

At the very least, one can conclude that managers in this firm do not differentiate among performance levels as much as the rating scale allows: 4 and 5 make up just one percent of the sample, and fifty percent of workers receive the rating 2. While one might argue that this distribution reflects true performance, there is suggestive evidence that it reflects information hiding. Several studies have shown that the ratings that supervisors report to workers are significantly higher and more skewed than the ratings they report to independent researchers (see Murphy and Cleveland 1991, p.79, and references therein). Also, the same patterns emerge when rating categories have labels such as "average" and "below average" (Gibbs 1991).[4] Finally, workers and managers themselves report that

[1]Performance appraisal systems were in place by 300 AD in the Chinese state bureaucracy. As of the early 1980's, between seventy four and eighty nine per cent of American businesses used them (Murphy and Cleveland 1991).

[2]The Chief Human Resource Counsel for International Paper recently noted that "...few tasks occupy as much time by human resource professionals as designing, implementing, monitoring, and defending performance appraisal systems" (Murphy and Margulies 2004).

[3]This figure is taken from Lazear and Gibbs (2008). The larger dataset on which it is based was analyzed in Baker, Gibbs, and Holmström (1994a) and Baker, Gibbs, and Holmström (1994b). Although this is evidence from one firm, other firms' rating distributions exhibit similar patterns (Medoff and Abraham 1980, Murphy 1992).

[4]In one particularly stark example, Milkovich, Newman, and Milkovich (2007) report a ten-year study of a thousand-member social service department in which only three of the possible ten thousand ratings were "below average".

managers do not distinguish among workers.[5]

For the most part, researchers have attributed limited feedback to organizational dysfunction in one way or another, implicitly assuming that a departure from full information disclosure is a pathology to be explained.[6] However, economic analysis of performance appraisal systems is very limited, and without first understanding the precise effects of feedback, it is impossible to comment on what is the surplus maximizing amount of information provision. The aim of this paper is to partially fill this gap in understanding. Its central finding is that there are reasonable circumstances under which firms would like to *commit* to disclosing some, but not full, information about performance. This has two important implications: (1) the fact that firms invest in performance appraisal systems that they then use to provide limited feedback is compatible with efficiency and (2) the welfare loss from distortions in feedback (of which there are surely many) should not necessarily be computed from a full-disclosure benchmark.

The argument is the following. In period 0, ex ante symmetric firms compete for a worker (he) who must be retained in periods 1 and 2, after which firms again compete to hire the worker for period 3. The first and second period employer privately observes the worker's performance, which it cannot credibly disclose to the outside labor market. After period 2, the employer retains the worker if and only if his expected ability crosses a threshold. More importantly, due to the informational asymmetry between his employer and outside firms, the worker earns a constant wage if retained and a constant, lower, wage if released, meaning that he earns a fixed reputational reward for meeting the employer's (endogenous) retention standard.[7] The worker's first and second period work incentives come solely through career concerns, and he exerts effort to signal an ability level that surpasses the retention standard.

While the initial employer cannot commit to disclosing the worker's performance to the outside labor market, it can commit to a disclosure policy that gives the worker information about his first period performance before he chooses second period effort. In particular, a disclosure policy partitions the first period output space and reports to the worker into which element of the partition his output lies.[8] Unlike in the standard

[5] In a case study of Merck, Murphy (1992) reports such sentiments as "Tell me this, how in the world can 83 per cent of the people be exceeding job expectations while the company, as a whole, is doing just average?" and "How can I rate my people objectively when the other directors are giving all their people 4s? A 3 isn't acceptable. I wouldn't mind if everyone played by the same rules, but they don't."

[6] For example, Baron and Kreps (1999) argue that managers are not rewarded for providing accurate appraisals so do not exert the required effort to obtain performance information; Longenecker, Sims, and Gioia (1987) emphasize managers' consideration of organizational politics when providing feedback; Jackman and Strober (2003) suggest that worker's psychological reactions to feedback may inhibit managerial truth telling; and Prendergast and Topel (1996) show that managers bias feedback as a result of favoritism for their workers.

[7] The reason for wage pooling is similar to that in Waldman (1984).

[8] The paper assumes that the set of possible feedback messages is rich enough to describe all per-

career concerns model (Holmström 1999), second period effort is history dependent since information about first period performance allows the worker to update his belief on how close his ability is to the retention threshold. More specifically, second period effort is highest when expected ability lies on the retention threshold and is monotonically decreasing as it moves away from it. In this setup, information disclosure has two primary effects:

1. *Effort risk.* Whenever two feedback messages induce two different effort levels, expected second period effort costs increase compared to a disclosure policy that combines them. More information increases the variance of second period effort, which the worker dislikes since his preferences are given by his convex cost of effort function.

2. *Coasting incentive.* When the worker finds out the exact value of his first period output, he can use first period effort to reduce the amount of effort his employer expects him to exert in the second (i.e. convince the employer he will coast) because the employer uses first period output to form its beliefs on second period effort. Whenever the disclosed output realization leads to an updated belief about expected ability that lies above (below) the retention threshold, exerting higher (lower) effort in the first period decreases the employer's belief on second period effort by moving its belief on expected ability further away from the threshold. [9]

The surplus maximizing disclosure policy (offered by firms to the worker in period 0) must balance these two effects. Assuming that signal jamming incentives alone are insufficient to support first-best first period effort, there is scope for increasing surplus by disclosing output realizations that lead to expected ability beliefs above the retention threshold. However, disclosing output realizations that lead to expected ability beliefs both near the retention threshold and that are very high expose the worker to the most risk: in these cases, actual second period effort differs substantially from its expected value. The efficient disclosure policy therefore discloses a convex, bounded set of output realizations all of which lead to inferred ability above the retention threshold, and combines all other output realizations in a single message.

The basic setup holds fixed one information asymmetry (that between the employer and outside firms) and endogenizes another (that between the employer and the worker)

formance levels, but that the principal endogenously opts to coarsen feedback. In the real world, it is unclear whether firms' giving mangers a finite set of messages to deliver to workers reflects a discrete number of distinguishable performance levels or an exogenous source of coarsening.

[9]The important incentive effect in the model comes from the anticipation of feedback, not the reaction to it. In fact, with quadratic effort costs, expected second period effort is independent of the disclosure policy. Two recent empirical papers, Blanes i Vidal and Nossol (2009) and Azmat and Iriberri (2010) have found an anticipatory effect on effort of feedback, although in environments without career concerns.

in order to keep the focus squarely on feedback within organizations. Before concluding, the paper examines an alternative informational environment: one in which outside firms perfectly observe the worker's performance prior to the third period. As long as there is some fixed payoff component to remaining in the industry, second period effort remains single peaked in expected ability, and the qualitative features of the efficient disclosure policy remain unchanged if the prior belief on worker talent is high enough. Thus the basic message of the paper is not necessarily tied to a particular assumption on what outside firms can observe about worker performance.

Related Literature There is a small but growing literature that analyzes feedback in organizations within a variety of contracting frameworks. Several recent papers solve for the effort maximizing disclosure policy in tournaments with an exogenous prize. Aoyagi (2007) finds that information disclosure increases expected effort in the presence of concave marginal costs and reduces expected effort in the presence of convex marginal costs. This paper assumes linear marginal costs, so these effects are not present. Closer in spirit is Ederer (2008), who introduces complementarity between ability and effort into the production function so that workers' effort depends on their beliefs about ability. He also finds that interim feedback gives rise to first period effort incentives, since each worker wants to signal a high ability to his opponent to discourage him from exerting effort in the second period. However, in this paper, the worker is not competing with someone else, but seeking to push the employer's belief about his ability above an absolute standard. Feedback generates incentives because the worker wants to manipulate the belief of his employer to make achieving the standard easier.[10]

Lizzeri, Meyer, and Persico (2002) study optimal disclosure in a two period moral hazard problem with explicit contracting. Information disclosure increases expected second period effort costs in all cases due to effort risk, and increases first period effort whenever wages are non-linear in output.[11] If the worker's wage schedule is fixed, there are circumstances under which full disclosure increases surplus relative to no disclosure, but no analysis is made of partial disclosure.

MacLeod (2003) analyzes how to sustain effort in a principal-agent model in which the principal privately observes the agent's subjective output. The optimal contract pays the

[10]Goltsman and Mukherjee (2008) also derive the optimality of partial feedback in a tournament framework due to the following tension: on the one hand, the designer wants to maintain competition in the second stage of the tournament by not informing workers when their first period outputs differ; on the other, workers would like to exert effort in the first period to avoid ending up in competitive second stage.

[11]The first period incentive effects arise because of the difference in the marginal product of second period effort. Although ability is in the production function, it is a fixed productivity parameter that the agent cannot influence through effort exertion.

agent a constant wage for all output realizations except the one most informative about low effort, following which the principal "burns" money, reducing the agent's surplus.[12] He interprets feedback as synonymous with wage payments, so that the rating distribution is coarser than the output distribution. By contrast, in this paper, feedback messages are not connected to wage payments; indeed, the worker's third period wage schedule is independent of the disclosure policy. Instead, feedback changes the worker's belief about where along the wage schedule his expected ability lies, and complements the underlying signal jamming incentives.

This paper relates to several papers in the career concerns literature initiated by Holmström (1999). Kovrijnykh (2007) and Martinez (2009) point out that, in the presence of history dependence, the worker's current effort affects the market's belief about his future effort, which is the basis for this paper's coasting incentive. However, neither paper explores the relationship between information disclosure and the risk-incentive trade-off this paper identifies.[13] Koch and Peyrache (2010) and Mukherjee (2008) both examine the effect of information release to the labor market in the presence of worker career concerns. The former examines a situation in which the market can back out worker ability from wage payments. The optimal contract does not fully reveal ability to the outside market because doing so weakens the worker's reputational incentives. In the latter, a commitment by the employer to full information disclosure to the outside market eliminates adverse selection and increases the up-front surplus it can extract from the worker, as long as it can insure him with a long-term wage contract. This paper instead takes as given the amount of information the outside market receives about performance, and considers the effect of information disclosure just to the worker. It also rules out incentive pay by assuming output is purely subjective.

Finally, two recent paper examine other environments in which the principal has private information on workers' abilities and can provide feedback. Ray (2007) considers a situation in which a principal privately observes an agent's ability prior to their commencing a project together. A trade-off arises between disclosing information to induce the worker to tailor effort to ability and withholding it to retain him. In some cases, revealing performance information on the tails of the ability distribution and withholding it in the middle is optimal. Crutzen, Swank, and Visser (2010) model communication between the principal and two agents as a cheap talk game and show that some information can be transmitted in equilibrium, although less than the efficient amount. Neither paper is a

[12]Fuchs (2007) extends this model to multiple periods and finds that the optimal contract pays a constant wage unless the principal observes the lowest output realization for all periods.

[13]Kovrijnykh (2007) asks a different question about information release: at what point in time do all actors in the model want to become aware of worker performance? He shows that by delaying the release of information, overexertion in early periods can be mitigated.

career-concerns model as such since the prinicpal already knows ability; in particular they do not shed light on the relationship between signal jamming and information disclosure.

The organization is as follows. The next section lays out the model; the third derives the equilibrium of period three labor market competition; the fourth discusses the effects of feedback; and the fifth derives the equilibrium disclosure policy. The sixth section examines the situation where market firms observe worker performance, and the seventh section concludes. The appendix contains all proofs.

2 Model

There are four time periods $t = 0, 1, 2, 3$. In period 0 three identical firms indexed by i compete to hire a single risk-neutral worker (he) for periods 1 and 2 by simultaneously offering employment contracts whose components are defined below. Once a firm hires the worker, they are matched for two periods. The firm that hires the worker is the employer E while the other two firms in the market are $M1$ and $M2$.

In periods 1 and 2, the worker produces $y_t = \theta + a_t + \varepsilon_t$ in whichever firm he joins, where θ is talent, a_t is effort, and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ is an output shock. Neither the worker nor the employer knows θ at $t = 0$, but they share a common prior distribution $\theta \sim N\left(\overline{\theta}, \sigma_\theta^2\right)$, where $\overline{\theta} > 0$. The cost to the worker of exerting effort is $g(a_t) = \frac{C}{2}a_t^2$, and he has an outside option of 0.

After period 2 firms again compete for the worker, and he has the opportunity to move to another firm. First the employer makes the worker a wage offer in $w_3^E \in W_3^E = \mathbb{R}^+$. Both market firms observe w_3^E and then simultaneously make wage offers $w_3^m \in W_3^m = \mathbb{R}^+$ for $m = 1, 2$. If the worker remains with the employer in period 3, 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 $y_3^M = \theta + \varepsilon_3$.[14] The role of κ is discussed in the following section.

If the worker receives no positive wage offer, he 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 $w_3 = \max\{0, w_3^E, w_3^1, w_3^2\}$. Finally, if any firm makes a positive wage offer, it incurs an arbitrarily small cost δ, which could for example be the legal costs from drafting a wage contract. These assumptions together imply that third period profit for

[14] 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.

E is given by

$$\pi_3^E = \begin{cases} \kappa + \theta - w_3^E - \delta & \text{if } w_3^E > 0, w_3^E \geq \max\{w_3^1, w_3^2\} \\ -\delta & \text{if } w_3^E > 0, w_3^E < \max\{w_3^1, w_3^2\} \\ 0 & \text{if } w_3^E = 0; \end{cases}$$

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

$$\pi_3^m = \begin{cases} \theta - w_3^m - \delta & \text{if 1. } w_3^m > 0 \text{ and } w_3^m > \max\{w_3^n, w_3^E\}, \text{ or} \\ & \quad \text{2. } w_3^m > 0, w_3^m = w_3^n > w_3^E, \text{ and worker joins firm m} \\ -\delta & \text{if 1. } w_3^m > 0 \text{ and } w_3^m < \max\{w_3^n, w_3^E\}, \text{ or} \\ & \quad \text{2. } w_3^m > 0, w_3^m = w_3^n > w_3^E, \text{ and worker joins firm n} \\ 0 & \text{if } w_3^m = 0. \end{cases}$$

The paper makes the standard informational assumption that the worker privately observes his effort. However, the paper also makes the non-standard assumption that the employer privately observes his output. This creates two distinct potential informational asymmetries: one between the employer and market firms and one between the employer and the worker. Since the focus of the paper is feedback within organizations, it will take as exogenous the informational asymmetry between the employer and market and assume that outside firms do not observe any direct signal of the worker's output. This assumption also rules out contracting on output since output is necessarily unverifiable. As a result, in period 0 firm i can only offer the worker a fixed compensation w_0^i equal to the present discounted value of wages in periods 1 and 2, and effort incentives arise in periods 1 and 2 solely to increase expected third period wages.

In period 0 firms can also commit to revealing information to the worker between periods 1 and 2 through choosing a disclosure policy P^i.[15] Thus a contract is the pair (w_0^i, P^i). A *disclosure policy* is a partition P of the first period output space $Y_1 = \mathbb{R}$ with the interpretation that the worker learns that $y_1 \in P(y_1)$, the element of the partition into which his first period output falls. One important subset of these elements is the set of disclosed output realizations $Y_1^D(P) = \{\, y_1 \mid P(y_1) = y_1 \,\}$. To avoid measure-theoretic technicalities, the paper puts some additional structure on disclosure policies.

Assumption 1 *Y_1^D is either empty or is a union of positive measure intervals. Moreover, whenever $Y_1^D \neq \mathbb{R}$, every non-singleton element of a disclosure policy is a union of positive measure intervals.*

[15]Disclosing information to the worker after period 2 is payoff irrelevant to all actors in the model.

While this definition is compatible with full disclosure ($P(y_1) = y_1$ for all y_1) and no disclosure ($P(y_1) = Y_1$ for all y_1), it can also accommodate many intermediate case (reveal whether output is above or below a certain threshold, disclose output over some interval and hide all other realizations, etc.). The no disclosure policy is used in the statement of some results, so the paper will denote it by P^N.

There are two important implicit assumptions that are important to clarify. The first is that, while it can obscure information, firms cannot lie to the worker about his first period performance. Indeed, the solution to the model will show that the employer has an incentive to lie to the worker if given the opportunity to do so. Why then assume commitment? First, it sets the benchmark for what the employer's preferred disclosure policy would be in the absence of communication (or other) frictions. One can then use this benchmark to compute the welfare loss of deviations from the optimal policy that may arise for whatever reason. Second, in a more complex model in which the employer and worker interact for multiple periods or in which there are multiple workers hired by the employer that can compare the information they receive, the employer may not have an incentive to lie.

The second important assumption is that the information that workers receive about their performance is not verifiable. Otherwise, there would be an indirect channel through which outside firms could discover worker performance: through asking workers themselves. The paper maintains this assumption primarily to isolate one informational asymmetry from the other. One can keep in mind a situation in which the valuable information in performance appraisals is the verbal assessment and the information in written reports is essentially meaningless.

While the model is straightforward to describe, writing down the equilibrium of the full game is rather complicated since the contracts that firms offer the worker in period 0 depend on how his equilibrium effort choices in periods 1 and 2 depend on information disclosure, which in turn depend on the equilibrium of the game played between the employer and market firms after period 2. So, rather than describe the full equilibrium here, the paper instead solves the game via backward induction and provides a definition of equilibrium for each sub-game.

3 Rewards to Talent

The equilibrium of the labor market competition game played between periods 2 and 3 is important to understand in its own right since it determines how the worker's performance in periods 1 and 2 is rewarded. 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_3^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. Essentially, this environment is a straightforward modification of that in Waldman (1984), in which an employer with a perfect signal of worker ability chooses the worker's job assignment and wage before the outside market makes a counter-offer. Although this paper is different enough to warrant a separate equilibrium derivation,[16] they are mainly technical; the economic reasoning underlying the result below is nearly identical to Waldman (1984).

Denote the strategy of the employer as $\overline{w}_3^E : (y_1, y_2) \to W_3^E$ and the strategy of market firm m as $\overline{w}_3^m : w_3^E \to W_3^m$. Denote by $\widehat{\theta}_1^E$ the employer's updated belief on the worker's ability after observing y_1, by $\widehat{\theta}_2^E$ the employer's updated belief on the worker's ability after observing (y_1, y_2), and by $\widehat{\theta}^M$ the market firms' updated belief on the worker's ability after observing w_3^E. As this is a straightforward signalling game, the paper uses the Perfect Bayesian Equilibrium solution concept to solve it.

Definition 1 *A Perfect Bayesian Equilibrium is a set of strategies* $\left(\overline{w}_3^{E*}, \overline{w}_3^{1*}, \overline{w}_3^{1*}\right)$ *and beliefs* $\left(\widehat{\theta}_1^E, \widehat{\theta}_2^E, \widehat{\theta}^M\right)$ *that satisfy the following conditions for* $m \in \{1,2\}$ *and* $n \in \{1,2\} \setminus \{m\}$ $\left(\textit{where } \lambda_t = \frac{\sigma_\theta^2}{t\sigma_\theta^2 + \sigma_\varepsilon^2}\right)$*:*

$$\overline{w}_3^{m*} \in \arg\max_{w_3^m} \mathbb{E}\left[\pi_3^m \;\middle|\; \overline{w}_3^{E*}, \overline{w}_3^{n*}\right] \;\forall w_3^E \tag{1}$$

$$\overline{w}_3^{E*} \in \arg\max_{w_3^E} \mathbb{E}\left[\pi_3^E \;\middle|\; \overline{w}_3^{m*}\right] \;\forall y_1, y_2 \tag{2}$$

$$\widehat{\theta}_1^E = \lambda_1(y_1 - a_1^*) + (1 - \lambda_1)\overline{\theta} \tag{3}$$

$$\widehat{\theta}_2^E = \lambda_2(y_2 - a_2^*) + (1 - \lambda_2)\widehat{\theta}_1^E \tag{4}$$

$$\widehat{\theta}^M = \mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; (y_1, y_2) \in \left(\overline{w}_3^{E*}\right)^{-1}\left(w_3^E\right)\right] \;\forall w_3^E \in \overline{w}_3^{E*} \tag{5}$$

Conditions (1) and (2) require market firms and the employer to best respond to each other's strategies. Conditions (3) and (4) require the employer to update its beliefs using Bayes' Rule. The result that this rule is linear in the prior and signal, along with the result that the weights take the above form are standard (see DeGroot 1970). Condition (5) 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_3^E. Notice that no restrictions are placed on $\widehat{\theta}^M$ following observations of w_3^E not on the equilibrium path.[17]

[16] In particular, 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 (as opposed to one entity), and firms can only choose wages (as opposed to wages and job assignments).

[17] Technically speaking, one could allow the two market firms to have different beliefs following off-

In fact, there are a continuum of equilibria that satisfy Definition 1, but all share the same essential properties.

Proposition 1 *In every pure strategy equilibrium, the worker remains with the employer if and only if* $\widehat{\theta}_2^E \geq \theta^*$ *where* θ^* *satisfies*

$$\kappa + \theta^* - \mathbb{E}\Big[\widehat{\theta}_2^E \,\Big|\, \widehat{\theta}_2^E \geq \theta^*\Big] \geq 0.$$

Furthermore,

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

where

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

The key feature of every equilibrium wage schedule is that they contain only two wages: one paid to worker types whose expected ability crosses the threshold θ^* and who stay with the employer; and another paid to worker types whose expected ability falls short of θ^* and who separate from the employer. If the employer paid two different retained types two different wages in equilibrium, and the outside market believed these signals credibly communicated private information, the employer would have an immediate incentive to "lie" to the market and tell it the worker of higher talent was the one of lower talent through offering it a lower wage. Thus, within the set of retained workers, credible communication between the employer and the outside market is impossible. The reason that only one wage is paid to released workers is because of costly bidding. Without costly bidding, the employer could still only retain workers above θ^* at a constant wage, but could credibly communicate private information to the market for worker types below θ^*. Providing information for these worker types would be costless and would not affect third period profits. However, as discussed previously, the purpose of the paper is to isolate the effects of employer-worker communication from employer-market communication. Assuming costly wage offers not only avoids the problem of solving for the optimal disclosure from the employer to the market, but has a realistic interpretation.

In order to discuss the equilibrium in more detail, note that any worker type θ^* for whom (1) the employer earns zero profit while incurring total labor costs $\overline{W} - \delta$ and (2) 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

equilibrium observations of w_3^E, but this would not alter the results.

two conditions constitutes an equilibrium:

$$\kappa + \theta^* - \overline{W} - \delta = 0 \tag{6}$$

$$\mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq \theta^* \right] - \overline{W} - \delta \leq 0 \tag{7}$$

From these conditions one can observe that $\kappa > 0$ is a necessary (and, as the proof of Proposition 1 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 θ^* while paying it a wage equal to the expected market output of all types above it.[18] Also, one can easily establish that the set of values of θ^* that satisfies (6) and (7) is unbounded above. However, the qualitative features of the optimal disclosure policy only depend on the existence of career concerns, which in turn exist whenever there is a positive probability of meeting the performance standard θ^*. So as long as θ^* is finite, equilibrium multiplicity is not problematic.

Another important point to note is that θ^*, $\overline{W}$, 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. The expectations computed over $\widehat{\theta}_2^E$ in this section are market firms' expectations over the employer's posterior belief on the worker's expected ability. Because the employer and market firms share the same beliefs on the worker's effort choices, these expectations depend only on $\overline{\theta}$, σ_θ^2, and σ_ε^2, and so these primitives alone determine the equilibrium values of θ^*, $\overline{W}$, and $\underline{W}$. Therefore, the rest of the paper takes as given constants the employer's retention threshold θ^* as well as the payoff to retention $W = \overline{W} - \underline{W}$.

Effort incentives for the worker arise because increasing output in the first and second periods increases $\widehat{\theta}_2^E$, which in turn increases the probability of earning the reputational reward W. The next section explores the relationship between these incentives and information disclosure.

[18] Note also that whenever $\kappa > 0$, turnover is inefficiently high. The efficient turnover rule would be for the employer to retain any worker type whose expected ability exceeded $-\kappa$, but since

$$\mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq -\kappa \right] \geq \overline{\theta} > 0$$

the employer would make negative profit on retained worker types near $-\kappa$ in an equilibrium with the efficient turnover rule. An interesting extension of the current model would be to examine the optimal retention threshold θ^* that balanced the trade-off between incentive provision and worker turnover, but this paper does not consider this.

4 Effects of Information Disclosure

Before discussing the relationship between first and second period effort and information disclosure, one first must derive equilibrium effort levels as a function of a given disclosure policy P. The following definition determines their solution.

Definition 2 *Equilibrium efforts level $a_1^*(P)$ and $a_2^*(P(y_1), a_1^*)$ satisfy the following conditions:*

$$a_2^W(P(y_1), a_1) \in \arg\max_{a_2} \mathbb{E}\Big[W \Pr\Big[\widehat{\theta}_2^* > \theta^* \;\Big|\; a_1 \Big] \;\Big|\; y_1 \in P(y_1) \Big] - \frac{C}{2} a_2^2 \tag{8}$$

$$a_1^W(P) \in \arg\max_{a_1} \mathbb{E}\Bigg[W \Pr\Big[\widehat{\theta}_2^* > \theta^* \;\Big|\; P, a_2^W \Big] - \frac{C}{2}\left(a_2^W\right)^2 \Bigg] - \frac{C}{2} a_1^2 \tag{9}$$

$$a_2^*(P(y_1), a_1^*) = \left[a_2^W(P(y_1), a_1) \right]_{a_1 = a_1^*} \tag{10}$$

$$a_1^W[P] = a_1^*[P] \tag{11}$$

$$\widehat{\theta}_1^* \equiv \widehat{\theta}_1^E = \lambda_1(y_1 - a_1^*) + (1 - \lambda_1)\overline{\theta} \tag{12}$$

$$\widehat{\theta}_2^* \equiv \widehat{\theta}_2^E = \lambda_2(y_2 - a_2^*) + (1 - \lambda_2)\widehat{\theta}_1^* \tag{13}$$

Condition (8) requires the worker in the second period to maximize the expected probability of earning the reputational reward minus his cost of effort, given a_1 and $y_1 \in P(y_1)$. Condition (9) requires him in the first period to maximize the expected probability of earning the reputational reward minus his expected second period effort costs minus his first period effort costs, given P and his second period strategy.[19] Conditions (10) and (11) require the worker's strategies to coincide with firms' conjectures, and the final two conditions require the employer to update its beliefs on worker ability using Bayes' Rule. The paper denotes by $\widehat{\theta}_t^*$ the employer's equilibrium belief on the worker's expected ability after observing his output realizations through period t. One does not need to introduce an expression for how the worker's beliefs evolve because his expected payoff in the third period depends solely on his employer's beliefs about his ability.

Clearly the distribution of $\widehat{\theta}_2^E$ is key for determining effort incentives since its realized value determines w_3. It is given by

Lemma 1 $\widehat{\theta}_2^E \mid y_1, a_1, a_2 \sim N\left(\lambda_1(y_1 - a_1) + (1 - \lambda_1)\overline{\theta} + \lambda_2(a_1 - a_1^* + a_2 - a_2^*), \sigma_2^2\right)$.

If the employer knew the true value of a_1 he would estimate the worker's ability to be $\lambda_1(y_1 - a_1) + (1 - \lambda_1)\overline{\theta}$. The second term in the expression for the mean of $\widehat{\theta}_2^E \mid y_1, a_1, a_2$ represents "fooling": the difference between the effort levels expected by the employer and the ones exerted by the worker.

[19]The expectations in these expressions are taken with respect to y_1.

A concern for equilibrium existence is the potential non-concavity of (8) and (9). However, concavity obtains when the cost of effort function is sufficiently convex to offset any non-concavities in the other terms, so equilibrium efforts are stated assuming that C is large enough for this to be the case.[20]

Proposition 2 *There exists a $\overline{C}$ such that, for all $C \geq \overline{C}$, there exist unique and positive equilibrium first and second period efforts levels given by*

$$a_2^*[P(y_1), a_1^*] = \mathbb{E}\left[\frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1^*(y_1)}{\sigma_2}\right) \,\middle|\, y_1 \in P(y_1) \right] \tag{14}$$

$$\begin{aligned} a_1^*[P] =& \mathbb{E}\left[\frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1^*(y_1)}{\sigma_2}\right) \right] + \\ & \mathbb{E}\left[\left(\frac{W}{C}\right)^2 \frac{\lambda_2^2\lambda_1}{\sigma_2^3}\left(\frac{\widehat{\theta}_1^*(y_1) - \theta^*}{\sigma_2}\right)\phi^2\left(\frac{\widehat{\theta}_1^*(y_1) - \theta^*}{\sigma_2}\right) \,\middle|\, y_1 \in Y_1^D(P) \right] \times \\ & \Pr\left[y_1 \in Y_1^D(P) \right] \end{aligned} \tag{15}$$

A first implication of Proposition 2 is that the channel through which information disclosure matters for effort comes via the information the worker receives about the employer's beliefs about his expected ability: a_1^* and a_2^* are only affected by the realization of y_1 through the realization of $\widehat{\theta}_1^*(y_1)$. Accordingly, much of the subsequent analysis discusses information disclosure in terms of $\widehat{\theta}_1^*$ rather than in terms of output.

Before proceeding, two more definitions are needed. *Positive feedback* is the set

$$Y_1^P(P) = \left\{ y_1 \mid y_1 \in Y_1^D(P), \widehat{\theta}_1^*(y_1) > \theta^* \right\}$$

and *negative feedback* is the set $Y_1^N(P) = Y_1^D(P) \backslash Y_1^P(P)$. Positive (negative) feedback consists of all output realizations that are disclosed to the worker and that inform him that his expected ability is above (below) the retention threshold. This distinction is important because each type of disclosure has quite distinct effects.

4.1 Second period effort: effort risk

Effort incentives arise in the second period because the mean of $\widehat{\theta}_2^E \mid y_1, a_1, a_2$ is increasing in the gap $(a_2 - a_2^*)$. So, by exerting more effort, the worker increases the probability of capturing the reputational reward in the third period. This is simply a particular version

[20] One might worry that the worker may supply no effort if C is sufficiently large. In fact, for C large, the marginal cost of supplying zero effort is zero, while the marginal benefit is positive. So, zero effort provision can never be an equilibrium outcome for large values of C.

of the signal jamming incentive that lies at the heart of effort provision in all career concerns models.

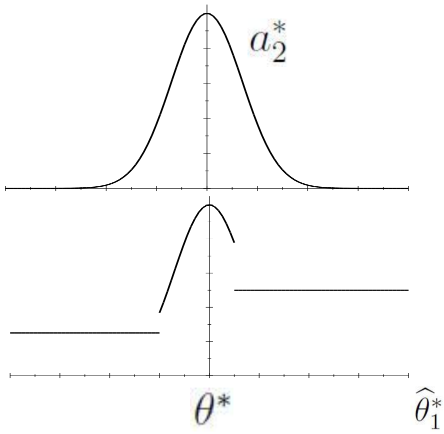

Figure 2: Second Period Effort and Information Disclosure

To see how second period effort depends on information, consider Figure 2. The top portion shows equilibrium effort under a disclosure policy in which the worker always learns $\widehat{\theta}_1^*$. In this situation, second period effort is highest when $\widehat{\theta}_1^* = \theta^*$ and monotonically decreasing as $\widehat{\theta}_1^*$ moves away from θ^*.[21] This is because when $\widehat{\theta}_1^* = \theta^*$ there is still a large amount of uncertainty about whether the worker will remain with the employer in period 3, so effort is important on the margin for determining future payoffs. In contrast, as $\widehat{\theta}_1^*$ moves into the upper tail of the distribution, it becomes increasingly certain that the worker will remain with the employer, while as $\widehat{\theta}_1^*$ moves into the lower tail, it becomes increasingly certain that he will leave. In these regions, a change in effort changes the probability of receiving W very little, so equilibrium effort falls correspondingly.

An interesting observation is that positive and negative feedback have symmetric effects on second period effort. A worker who is told that $\widehat{\theta}_1^*$ lies some distance x above θ^* exerts just as much effort as a worker who is told that $\widehat{\theta}_1^*$ lies a distance x below θ^*. This is in contrast to behavioral perspectives on feedback that emphasize the encouraging effects of positive feedback and the discouraging effects of negative feedback (Meyer, Kay, and French 1965).

The bottom portion of Figure 2 shows equilibrium second period effort under a disclosure policy that reveals the value of interim expected ability between two points $-2b$ and b, and otherwise reports whether it lies in $(-\infty, -2b)$ or in (b, ∞). Clearly effort remains the same as under the full disclosure policy for interim ability realizations that lie in $(-2b, b)$. However, effort changes in the tails, where the worker now exerts an effort

[21]This also shows that the employer has an incentive to renege on his commitment to a disclosure policy and simply tell the worker that $\widehat{\theta}_1^* = \theta^*$.

level formed by taking the expectation over all expected ability levels contained in the associated partition element. Consider the case in which the worker learns his expected ability lies in $(-\infty, -2b)$. Workers for whom $\widehat{\theta}_1^*$ lies close to $-2b$ now exert less effort than under full disclosure because they are pooled with types further away from the retention threshold. On the other hand, workers for whom $\widehat{\theta}_1^*$ is very low will now work harder than under full disclosure since they expect their ability to be closer to θ^* 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 $\mathbb{E}[\, a_2^* \,]$ *is independent of the disclosure policy.*

The fact that expected second period effort does not depend on the disclosure policy means that firms do not have to take into account the incentive effects of the worker's reaction to feedback. 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 does depend on the disclosure policy, but the paper maintains the quadratic assumption to isolate the trade-off between risk and coasting without introducing a third effect as well.[22]

While a disclosure policy does not affect expected second period effort, it does affect expected second period effort costs.

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

Providing more information to the worker about his performance increases the variance of his 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. This is the first substantive point about feedback in organizations. It implies that if workers only exert effort for one period, the surplus maximizing policy provides no information.

[22]Nevertheless, one can show that when the worker's cost of effort function is $g(a_t) = \frac{C}{\beta+1} a_t^{\beta+1}$ where $\beta > 1$, the essential findings of the paper are robust in the following sense. For any disclosure policy that (1) reveals no information or (2) reveals an unbounded set of output realizations, there exists some alternative disclosure policy that reveals a bounded set of output realizations and that yields higher surplus.

4.2 First period effort: coasting incentive

While feedback has no second period incentive effects, it does have first period incentive effects. As given in Proposition 2, first period equilibrium effort consists of two parts. These derive from the two ways the worker has of fooling the employer in period 1. The first is to increase the gap $a_1 - a_1^*$ by increasing a_1. This is again signal jamming, and is captured by the term

$$\mathbb{E}\left[\frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1^*(y_1)}{\sigma_2}\right)\right]$$

which is independent of the disclosure policy. This is 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 way in which the worker can fool the employer in the first period is more subtle and relates to the gap $a_2 - a_2^*$. Recall that a_2^* is equal not only to the worker's equilibrium second period effort level, but also his employer's *belief* about the amount of effort he will exert in the second period. When the worker learns the value of y_1,

$$a_2^*(P(y_1), a_1^*) = \frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1^E(y_1)}{\sigma_2}\right)$$

which depends on directly on y_1. In contrast, when the disclosure policy does not reveal y_1 directly,

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

which is a constant independent of y_1. Thus the worker can use first period effort to influence the amount of effort that the employer expects him to exert in the second, *but only when the disclosure policy reveals y_1 directly.*

Now, the worker wants to reduce the amount of effort his employer expects of him so that it attributes more of his second period output to his innate ability rather than to his effort. In other words, he wants the employer to believe he is "coasting" in the second period and not exerting high effort. Whenever the worker learns y_1, $\frac{\partial \widehat{\theta}_1^E(y_1)}{\partial a_1} > 0$. However, how this translates into a change in a_2^* is asymmetric. When $\widehat{\theta}_1^E(y_1) < \theta^*$, increasing a_1 increases a_2^* by pushing $\widehat{\theta}_1^E(y_1)$ closer to θ^*, the point at which the employer expects maximum effort. In this case, the worker wants to reduce his first period effort. On the other hand, when $\widehat{\theta}_1^E(y_1) > \theta^*$, increasing a_1 decreases a_2^* by pushing $\widehat{\theta}_1^E(y_1)$ further away from θ^*. In this case, the worker wants to raise his first period effort. Thus, the coasting incentive can either work to increase first period effort or lower it.

Corollary 3 *Suppose two disclosure policies P and P' are such that $Y_1^P(P') \subset Y_1^P(P)$ and $Y_1^N(P) \subset Y_1^N(P')$. Then a_1^* is higher under P than under P'.*

The effort maximizing disclosure policy P^E thus maximizes the amount of positive feedback and minimizes the amount of negative feedback. In a sense, the coasting incentive is related to the ratchet effect studied previously in the dynamic moral hazard literature. In both cases, the agents's current effort affects the principal's expectations about future performance. Here, though, coasting can work to both increase and decrease first period effort, whereas the ratchet effect is usually seen as discouraging effort. Moreover, the employer can endogenously choose the strength of the coasting incentive by altering the amount of information the worker receives about his performance, whereas feedback is not an instrument for influencing the ratchet effect.

Now that the paper has identified the two channels through which information disclosure operates, it can finally turn to answering its basic question: what disclosure policies do firms offer the worker in period 0?

5 Equilibrium Disclosure Policy

The previous section solved for the equilibrium effort levels conditional on the worker already having joined one of the firms under some fixed disclosure policy. The analysis now moves to the beginning of the game (period 0) in which the worker is not yet matched with any firm and in which firms simultaneously offer contracts to the worker. The equilibrium of the game is for each firm $i \in \{1, 2, 3\}$ to offer the contract (w_0^i, P^i) that satisfies

$$\max_{w_0^i, P^i} w_0^i - g(a_1^*(P^i)) - \mathbb{E}\big[\, g(a_2^*(P^i(y_1), a_1^*)) \,\big] + \overline{W} \Pr\left[\widehat{\theta}_2^E \geq \theta^* \right] + \underline{W} \Pr\left[\widehat{\theta}_2^E < \theta^* \right]$$

$$\text{s.t. } 2\overline{\theta} + a_1^*(P^i) + \mathbb{E}\big[\, a_2^*(P^i(y_1), a_1^*) \,\big] = 0. \tag{16}$$

That is, firms maximize the worker's utility subject to earning zero profit. Substituting for w_0^i and using the fact that $\overline{\theta}$ and period 3 compensation are independent of P^i, one can conclude that each firm offers the worker the surplus maximizing disclosure policy P^S given by

$$P^S = \arg\max_P a_1^*(P) - g(a_1^*(P)) + \mathbb{E}[\, a_2^*(P(y_1), a_1^*) - g(a_2^*(P(y_1), a_1^*)) \,]. \tag{17}$$

Of course P^S depends on the effort level with no information disclosure. Since the usual worry in career concerns models is the under provision of effort, the paper assumes that $Ca_1^*\left(P^N\right) < 1$ so that first period effort is less than first best under no disclosure. In this case, P^S must satisfy a risk-incentive trade-off, but one that is quite distinct from the one familiar from the moral hazard literature (Holmström 1979). There, the instrument for inducing effort is output pay and the associated risk is over wealth levels. Here, the instrument for inducing (first period) effort is positive feedback, and the associated risk is over (second period) effort levels. The following is the central result of the paper.

Proposition 3

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

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

One can build up the intuition for the result step-by-step. As discussed in the previous section, the coasting incentive only arises over the set of output realizations directly revealed to the worker. Thus P^S should only contain only one non-singleton element. If a disclosure policy contained two non-singleton elements, one could combine them without changing first period effort. At the same time, one would decrease expected second period effort costs (by Corollary 2) through reducing effort risk. So disclosure policies of the form "reveal whether output is above or below 0" or "reveal whether output lies in $(-10, 10)$ or $(-\infty, 10) \cup (10, \infty)$" can never maximize surplus.

Since by assumption $Ca_1^*\left(P^N\right) < 1$, 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. Combining this insight with the one that P^S can only contain one non-singleton element means that in the search for P^S one only needs to consider disclosure policies that can be fully described by the set of (equilibrium) beliefs on expected ability that the employer reveals to the worker, which one can denote with the set $\Theta_1^{d*}(P) = \left\{ \widehat{\theta}_1^*(y_1) \mid y_1 \in Y_1^D(P) \right\}$. Consider moving from disclosure policy P to disclosure policy P' that satisfies $\Theta_1^{d*}(P') = \Theta_1^{d*}(P) \cup [t, t+\varepsilon]$ where ε is small and $t \in (\theta^*, \infty)$. This raises first period surplus on the margin by increasing first period effort and lowers second period surplus on the margin by increasing expected effort costs. The following figure plots out 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 non-singleton element, provides no negative feedback, and for which $Ca_1^*(P) < 1$, not just P^S.

One can show that the marginal cost of disclosing additional beliefs $[t, t+\varepsilon]$ is proportional to

$$\left(\mathbb{E}\left[\phi\left(\frac{\theta^* - \widehat{\theta}_1^*}{\sigma_2} \right) \middle| \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \right] - \phi\left(\frac{\theta^* - t}{\sigma_2} \right) \right)^2. \tag{18}$$

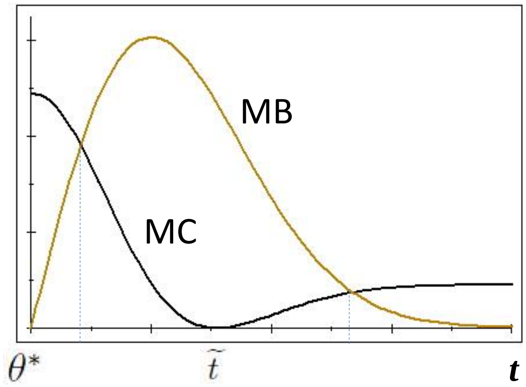

Figure 3: Marginal Cost and Benefit of Information Disclosure

In words, the additional risk to which the worker is exposed when one moves from a disclosure policy that keeps $[t, t+\varepsilon]$ hidden and one that reveals $[t, t+\varepsilon]$ is proportional to the squared difference between the effort level that the worker exerts under the original disclosure policy conditional on not learning his expected ability and the effort level the worker exerts conditional on learning that $\widehat{\theta}_1^* = t$. This risk is highest when t is large and when t is near θ^*. In the first case the worker's effort conditional on learning t is much lower than the effort level he would exert conditional on not learning his type, and in the second case his effort conditional on learning t is much higher than he would exert under ignorance of his type. On the other hand there exists some $\widetilde{t}$ such that the worker's effort conditional on learning $\widehat{\theta}_1^* = \widetilde{t}$ is exactly equal to the effort level he exerts under ignorance. Disclosing the additional beliefs $\left[\widetilde{t}, \widetilde{t} + \varepsilon\right]$ to the worker thus exposes him to almost no additional risk.

On the other hand, the marginal benefit of additional disclosure depends on the strength of the coasting incentive, which itself is proportional to the sensitivity of the employer's conjecture on the worker's second period effort to first period effort. When $\widehat{\theta}_1^*$ is near θ^*, this conjecture is flat since it is near is maximum value. When $\widehat{\theta}_1^*$ is very large, it is also flat since it is near its minimum. The key point is that the additional coasting 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 θ^* and large beliefs and reveals intermediate beliefs.[23]

Discussion. The first, and perhaps most important, implication of Proposition 3 is that the surplus maximizing disclosure policy features partial information disclosure.

[23]Using the exact same logic, one can conclude that if $Ca_1^*\left(P^N\right) > 1$, P^S reveals a bounded interval of negative feedback. So the only case in which firms do not disclose information to workers is when $Ca_1^*\left(P^N\right) = 1$; that is, when signal jamming incentives alone give rise to first-best first period effort.

The possibility of information disclosure from managers to workers raises social welfare, and as long as implementing such systems is not too costly, it is efficient to invest in them. However, in environments in which workers receive reputational rewards such as professional service firms, the most efficient way to use feedback systems is not necessarily to provide full information. Prima facie, one cannot conclude that it is inefficient for companies to set up performance appraisal systems and for managers to subsequently use these systems to deliver limited feedback.

The model also shows that the efficient disclosure policy is characterized by a particular type of limited feedback. If the worker's first period output lies between y^* and y^{**}, he learns the precise value of his expected ability. If his first period output lies outside of this set, he simply learns that his expected ability is $\mathbb{E}\left[\widehat{\theta}_1^*(y_1) \,\middle|\, y_1 \in Y_1 \backslash [y^*, y^{**}]\right]$. The interesting point here is that the worker never learns that his expected ability lies in the tails of the distribution, and that the rating distribution is concentrated. Broadly speaking, this is consistent with stylized facts about real world rating distributions.[24]

The model also clarifies other, more subtle, potentially false intuitions about feedback and motivation. Although positive feedback is a way of alleviating inefficiently low effort provision, it is never used to fully eliminate the inefficiency. Suppose that the employer is using a disclosure policy P at 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 compared to the reduction in risk.[25] So, the trade-off between incentives and risk is never resolved by fully eliminating the first period inefficiency.[26]

Recall from section 4.2 that the effort maximizing disclosure policy P^E takes the form $\Theta_1^{d*}(P^E) = (\theta^*, \infty)$; that is, it provides maximum positive feedback by revealing all expected ability realizations above the retention standard. This section shows that the output realizations disclosed under P^S are a strict subset of the output realizations disclosed under P^E. Thus the worker receives more information about his performance

[24]Prendergast (1999) writes on page 30 of his oft-cited overview of incentives in organizations that

> There is considerable evidence in the personnel literature that supervisors distort subjective performance ratings by not sufficiently differentiating good from bad performance in their ratings...Two relevant forms of compression are noted in this literature: "centrality bias" and "leniency bias." Centrality bias refers to a practice where supervisors offer all workers ratings that differ little from a norm. Leniency bias implies that supervisors simply overstate the performance of the poor performers. Such compression is well documented in the personnel literature...

Note, however, that the model cannot speak to the phenomenon of grade inflation in which supervisors overstate performance. By assumption, disclosure policies can hide information from workers but cannot deceive them.

[25]A more formal argument is given in the proof of Proposition 3

[26]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.

under the effort maximizing policy than is efficient. Suppose one compared two firms with workers of similar talent, one of which used P^E and one of which used P^S. One would observe that the former provided more information to workers and enjoyed higher productivity levels. It would be tempting to conclude that it was also pursuing an unambiguously superior policy disclosure policy, but this is false. The risk that workers in the firm using P^E face outweighs the associated gain in productivity.

Of course, this observation only matters if there are convincing reasons for why one might observe P^E in practice. In many occupations in which career concerns operate, firms have bargaining power with respect to entry level workers while competition for talented experienced workers is strong, with a correspondingly high wage differential between junior and senior members of a firm (Maister 1993). If entry level workers are wealth constrained and future expected reputational rewards are high compared to initial effort costs, firms can offer P^E and still satisfy the worker's participation constraint. In this case, firms would not internalize effort risk and would provide too much information.[27] Also, if line managers are rewarded on the basis of their workers' productivity alone, they would also not internalize the increased disutility of effort they impose on workers by using P^E. The main message here is that the possibility of too much information disclosure is as legitimate a concern as too little information in the presence of career concerns.

6 Disclosure with Symmetric Information in the Labor Market

One of the most important assumptions in the model is that outside firms cannot observe worker output. While effort risk and coasting are presumably general effects in career concerns models with history dependent effort, their direction and magnitude are linked to the payoff to reputation, which in turn is determined by the amount of information labor market participants have about worker performance. Ultimately, Proposition 3 relies on the reward schedule derived in Proposition 1. This section therefore explores the robustness of Proposition 3 to an alternative assumption about market firms' information; namely, that they perfectly observe y_1 and y_2.[28] It continues to assume that the worker remains with the employer for periods 1 and 2 and that the employer alone controls what

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

[28]One concern is that observability opens up the possibility of contracting on output. However, the paper adopts the approach in Holmström (1999) and continues to assume that career concerns remain the only source of effort incentive in spite of observability. Of course, nothing about observability implies verifiability.

the worker observes about y_1.

Under symmetric labor market information, solving for the rewards schedule is much easier than before. While the bidding cost and human capital accumulation assumptions from the baseline model are important for characterizing equilibria, here they are not, so one can set $\delta = \kappa = 0$ for simplicity. After they observe y_1 and y_2, market firms share the same belief as the employer on worker ability, so that $\widehat{\theta}^M = \widehat{\theta}_2^E$. Firms then engage in a standard Bertrand bidding game in which the worker stays in the industry at a wage equal to his expected output if his expected output is non-negative, and otherwise leaves the industry and earns his outside option 0. The paper makes the additional assumption that there is some fixed reward component F of remaining in the industry. This could represent many situations. For example, there could be some non-monetary compensation from remaining with the employer like moving to a bigger office or there could be psychological benefits deriving from enhanced pride or greater professional responsibility. The resulting rewards schedule is

$$w_3 = \begin{cases} F + \widehat{\theta}_2^E & \text{if } \widehat{\theta}_2^E \geq 0 \\ 0 & \text{if } \widehat{\theta}_2^E < 0. \end{cases} \tag{19}$$

The main difference between this wage schedule and the one in the asymmetric information case is that the worker earns a higher wage when his reputation increases, conditional on remaining in the industry. Before, more talented workers did not earn higher wages because market participants had no way of distinguishing talented workers from mediocre ones. Now, since output information is available to outside firms, they are willing to pay more to workers whose performance is better. The resulting equilibrium effort levels are given by the following.

Proposition 4 *There exists a $\overline{C}$ such that, for all $C \geq \overline{C}$, there exist unique and positive equilibrium first and second period efforts levels given by*

$$a_2^*[P(y_1), a_1^*] = \mathbb{E}\left[\frac{F}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\widehat{\theta}_1^*(y_1)}{\sigma_2}\right) + \frac{\lambda_2}{C}\Phi\left(\frac{\widehat{\theta}_1^*(y_1)}{\sigma_2}\right) \,\middle|\, y_1 \in P(y_1) \right] \tag{20}$$

$$\begin{aligned} a_1^*[P] =& \mathbb{E}\left[\frac{F}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\widehat{\theta}_1^*(y_1)}{\sigma_2}\right) + \frac{\lambda_2}{C}\Phi\left(\frac{\widehat{\theta}_1^*(y_1)}{\sigma_2}\right) \right] + \\ & \mathbb{E}\left[\left(\begin{array}{c} \frac{\lambda_1\lambda_2}{C\sigma_2}\left[\phi\left(\frac{\widehat{\theta}_1^*(y_1)}{\sigma_2}\right)\left(\widehat{\theta}_1^*(y_1)\frac{F}{\sigma_2^2} - 1\right)\right] \times \\ \left[\frac{F}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\widehat{\theta}_1^*(y_1)}{\sigma_2}\right) + \frac{\lambda_2}{C}\Phi\left(\frac{\widehat{\theta}_1^*(y_1)}{\sigma_2}\right)\right] \end{array} \right) \,\middle|\, y_1 \in Y_1^D(P) \right] \times \\ & \Pr\left[y_1 \in Y_1^D(P) \right]. \end{aligned} \tag{21}$$

Effort incentives now come from two sources. First, the worker wants to signal an ability level above 0 to remain in the industry and earn the fixed reward F. This source of effort incentives works just like the asymmetric information case: the first term of (20) is equal to the expression for a_2^* in Proposition 2 taking $\theta^* = 0$ and $W = F$. Second, the worker wants to signal a high ability in order to increase his third period wage conditional on remaining in the industry. This is reflected in the second term of (20), which is absent from the expression for a_2^* in Proposition 2. Figure 4 plots second period effort under full disclosure.

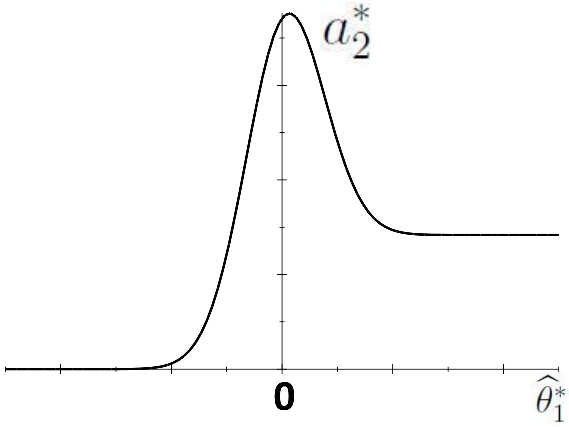

Figure 4: Second Period Effort Under Full Disclosure

Even though the wage schedule is now increasing in expected ability above 0, the presence of F continues to make effort single peaked; one can easily show that this peak occurs at $\frac{\sigma_2^2}{F}$. The intuition is the same as before: for high and low levels of $\widehat{\theta}_1^*(y_1)$, effort at the margin matters very little for changing the probability of remaining in the industry. However, there is an important difference with the asymmetric information case. Rather than dropping to 0 as $\widehat{\theta}_1^*(y_1)$ becomes large, a_2^* now limits to $\frac{\lambda_2}{C}$ because increases in $\widehat{\theta}_2^E$ translate into higher wages even as earning F becomes a near certainty. The next result shows the associated implications for the efficient disclosure policy, where $\Theta_1^{d*}(P)$ is defined as in the previous section.

Proposition 5 *Suppose that $Ca_1^*(P^N) < 1$. For every P for which $\Theta_1^{d*}(P)$ is not a strict subset of $\left(\frac{\sigma_2^2}{F}, \infty\right)$ or $\emptyset$, one can find a P' for whcih $\Theta_1^{d*}(P')$ is a strict subset of $\left(\frac{\sigma_2^2}{F}, \infty\right)$ and that yields a higher surplus than P.*

Moreover, there exists a $\overline{\theta}'$ such that, for all $\overline{\theta} > \overline{\theta}'$, for every P for which $\Theta_1^{d*}(P) = \emptyset$ or for which $\Theta_1^{d*}(P)$ is unbounded above, one can find a P'' for which $\Theta_1^{d*}(P'')$ is a bounded subset of $\left(\frac{\sigma_2^2}{F}, \infty\right)$ and that yields a higher surplus than P.

One can think about feedback here the same way as before. First, coasting incentives will reduce first period effort whenever the worker learns that y_1 is such that $\widehat{\theta}_1^*(y_1) < \frac{\sigma_2^2}{F}$ and increase it whenever he learns that $\widehat{\theta}_1^*(y_1) > \frac{\sigma_2^2}{F}$. Second, all feedback will increase effort risk. So the efficient policy has only one non-singleton element that contains all output realizations for which $\widehat{\theta}_1^*(y_1) < \frac{\sigma_2^2}{F}$ and must resolve a risk-incentive trade-off for all output realizations for which $\widehat{\theta}_1^*(y_1) > \frac{\sigma_2^2}{F}$.

The fact that reputational incentives do not decline to 0 for high realizations of expected ability presents two complications for characterizing the efficient policy. To see them, note that the marginal cost of replacing disclosure policy P with a disclosure policy P' for which $\Theta_1^{d*}(P') = \Theta_1^{d*}(P) \cup [t, t+\varepsilon]$ where ε is small is proportional to

$$\left(A(t) - \mathbb{E}\left[A(\widehat{\theta}_1^*(y_1)) \,\middle|\, \widehat{\theta}_1^*(y_1) \notin \Theta_1^{d*}(P) \right]\right)^2 \tag{22}$$

where

$$A(x) = \frac{F}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{x}{\sigma_2}\right) + \frac{\lambda_2}{C}\Phi\left(\frac{x}{\sigma_2}\right). \tag{23}$$

The first complication is that one cannot always find a measure of "riskless" beliefs to disclose to the worker. While the MC curve plotted in Figure 3 always had a minimum point of 0, here this is not the case. In particular, under the no disclosure policy P^N, the worker's effort level is $\mathbb{E}\left[A(\widehat{\theta}_1^*(y_1)) \right]$. Since $A(t) \geq \frac{\lambda_2}{C}$ on $t \in \left[\frac{\sigma_2^2}{F}, \infty\right)$, riskless beliefs only exist if $\mathbb{E}\left[A(\widehat{\theta}_1^*(y_1)) \right] > \frac{\lambda_2}{C}$. Thus, depending on the strength of the coasting incentive, P^N might actually maximize surplus. The second complication is that one cannot necessarily rule out the possibility that the efficient policy reveals a set of output realizations that is unbounded above. Consider a disclosure policy for which $\mathbb{E}\left[A(\widehat{\theta}_1^*(y_1)) \,\middle|\, \widehat{\theta}_1^*(y_1) \notin \Theta_1^{d*}(P) \right] = \frac{\lambda_2}{C}$. The risk to which workers are exposed from adding beliefs $[t, t+\varepsilon]$ tends to zero as $t \to \infty$; moreover, one can show that the marginal benefit of adding beliefs $[t, t+\varepsilon]$ to $\Theta_1^{d*}(P)$ exceeds the marginal cost as $t \to \infty$. Both of these complications are not relevant if $\overline{\theta}$ is large enough because this ensures that $\mathbb{E}\left[A(\widehat{\theta}_1^*(y_1)) \,\middle|\, \widehat{\theta}_1^*(y_1) \notin \Theta_1^{d*}(P) \right]$ is bound away from $\frac{\lambda_2}{C}$ for the relevant set of disclosure policies. Thus a crucial property of the rewards schedule in the baseline case is that reputational incentives disappear for high output realizations.

While all of the qualitative features of the efficient policy are only carried through to the symmetric information case if $\overline{\theta}$ is large, important messages emphasized in the previous section are preserved regardless of the value of $\overline{\theta}$. First, the efficient amount of information for the worker to receive is a coarsening of the underlying performance distribution. Second, when the efficient disclosure policy reveals some information, it will never fully eliminate the inefficiency in first period effort. Finally, the effort maximizing disclosure policy provides strictly more information than the surplus maximizing policy for all parameter values.

7 Conclusion

This paper has argued that limited feedback in organizations is compatible with efficiency in the presence of career concerns. Thus, one must take great care in drawing any welfare conclusions from empirical evidence on limited feedback. Without going deeper into the reasons for why an organization provides limited feedback, one cannot determine whether more feedback would increase surplus. Indeed, if firms do not internalize workers' effort risk, they may even provide too much feedback.

Of course feedback in the real world operates through many channels that are not present in this model. For example, the ratings a manager gives a worker often serve as an input into bonus and promotion decisions. Also, performance appraisals often serve a training purpose by informing workers about how to do their jobs better. A more complete model that incorporated these other effects may well favor more information disclosure.

Still, given the absence of a well-developed economic literature on feedback in organizations, the model provides a starting point for thinking about feedback in environments where reputation matters for promotion. At the very least it provides answers to the questions of whether performance appraisals can increase firm value and whether more feedback is always better. The answer to the first is yes and to the second is no. Answers to more detailed questions with more empirical content await future research.

A Proofs

A.1 Section 3

A.1.1 Proof of Proposition 1

Proof. Let W_3^{E*} be the set of equilibrium actions defined by $\overline{w}_3^{E*}$. For all $w_3^E \in W_3^{E*}$, $M1$ 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

$$\overline{w}_3^{m*} = \begin{cases} \widehat{\theta}^M(w_3^E) - \delta & \text{if } \widehat{\theta}^M(w_3^E) - w_3^E - \delta > 0 \\ \widehat{\theta}^M(w_3^E) - \delta \text{ or } 0 & \text{if } \widehat{\theta}^M(w_3^E) - w_3^E - \delta = 0 \\ 0 & \text{if } \widehat{\theta}^M(w_3^E) - w_3^E - \delta < 0. \end{cases} \tag{24}$$

Let $Y' = \left\{ (y_1, y_2) \mid \overline{w}_3^{E*}(y_1, y_2) > 0 \right\}$ 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

$$\widehat{\theta}^M(\overline{w}_3^{E*}(y_1, y_2)) - \overline{w}_3^{E*}(y_1, y_2) - \delta \leq 0 \ \forall (y_1, y_2) \in Y'. \tag{25}$$

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 $w_3^E = 0$ and saving the bidding cost δ. Also it must be that

$$\kappa + \widehat{\theta}_2^E(y_1, y_2) - \overline{w}_3^{E*}(y_1, y_2) - \delta \geq 0 \ \forall (y_1, y_2) \in Y'. \tag{26}$$

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 $w_3^E = 0$.

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

$$\overline{w}_3^{E*}(y_1^1, y_2^1) = w_3^{E1} > w_3^{E2} = \overline{w}_3^{E*}(y_1^2, y_2^2).$$

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

$$\widehat{\theta}^M(w_3^{Ei}) - w_3^{Ei} - \delta \leq 0 \text{ for } i = 1, 2$$

as well as

$$\kappa + \widehat{\theta}_2^E(y_1^i, y_2^i) - w_3^{Ei} - \delta \geq 0 \text{ for } i = 1, 2$$

But then the employer strictly improves profit by offering the wage w_3^{E2} after observing (y_1^1, y_2^1) instead of w_3^{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.

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

$$\widehat{\theta}_2^M(\overline{W}) = \mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq \theta^* \right]$$

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

$$\kappa + \theta^* - \overline{W} - \delta = 0 \tag{27}$$

$$\mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq \theta^*\right] - \delta \leq \overline{W}, \tag{28}$$

which in turn imply that θ^* must satisfy

$$\kappa + \theta^* - \mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq \theta^*\right] \geq 0. \tag{29}$$

One now needs to establish the existence of a θ^* that satisfies (29). First note that because $\widehat{\theta}_2^E$ is a linear combination of normal random variables, it is itself normally distributed with mean

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

and variance that one can denote by σ^2. Now consider the function

$$f(x) = x - \mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq x\right] \tag{30}$$

Two helpful results from distribution theory (Greene 2003, p.759) are the following, where γ is the normal hazard rate:

$$\mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq x\right] = \overline{\theta} + \sigma\gamma\left(\frac{x - \overline{\theta}}{\sigma}\right) \tag{31}$$

$$\gamma'(a) = \gamma(a)(\gamma(a) - a) \in (0, 1)\ \forall a \in \mathbb{R} \tag{32}$$

Together these imply that $f'(x) > 0$. Now by observation $\lim_{x\to-\infty} f(x) = -\infty$. From (32)

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

Observe that

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

since

$$\gamma\left(\frac{x - \overline{\theta}}{\sigma}\right) = \frac{\mathbb{E}\left[\widehat{\theta}_2^E \;\middle|\; \widehat{\theta}_2^E \geq x\right] - \overline{\theta}}{\sigma} \geq \frac{x - \overline{\theta}}{\sigma}$$

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

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

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

The above arguments show that (29) 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 $\widehat{\theta}_2^E \geq \theta^* \geq x^*$ at a wage $\overline{W}$ that satisfies (27). If $\widehat{\theta}_2^E < \theta^*$ the employer sets $w_3^E = 0$ and the worker's wage is implied by (24).

Finally, the proposed equilibrium exists as long as the employer can make no profitable deviation to some w_3^E other than 0 or $\overline{W}$. In order to rule out this possibility, one can set

$$\widehat{\theta}^M(w_3^E) \geq \widehat{\theta}^M(\overline{W}) \ \forall w_3^E \neq \{0, \overline{W}\}$$

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

A.2 Section 4

A.2.1 Proof of Lemma 1

Proof. $\widehat{\theta}_2^E \mid y_1, a_1, a_2$ is a linear combination of normal random variable so is itself normal with variance σ_2^2 and mean

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

■

A.2.2 Proof of Proposition 2

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

$$W\mathbb{E}\left[1 - \Phi\left(\frac{\theta^* - \widehat{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}\right) \;\middle|\; y_1 \in P(y_1)\right] - \frac{C}{2}a_2^2 \tag{33}$$

where Φ is the standard normal CDF and the expectation is with respect to y_1. The first derivative is

$$\mathbb{E}\left[W\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}\right) \;\middle|\; y_1 \in P(y_1)\right] - Ca_2. \tag{34}$$

As $a_2 \to 0$, (34) > 0, while as $a_2 \to \infty$, (34) < 0. So an interior solution to the optimization problem exists. The second derivative is

$$-\mathbb{E}\left[W\frac{\lambda_2^2}{\sigma_2^2}\phi'\left(\frac{\theta^* - \widehat{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}\right) \;\middle|\; y_1 \in P(y_1)\right] - C, \tag{35}$$

which limits to $-\infty$ as $C \to \infty$. So there exists a $\overline{C}_1$ such that, for all $C > \overline{C}_1$, (33) is globally concave and the first order condition gives as a global maximum

$$a_2^W\left[P(y_1), a_1\right] = \mathbb{E}\left[\frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2^W - a_2^*)}{\sigma_2}\right) \;\middle|\; y_1 \in P(y_1)\right]. \tag{36}$$

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

$$a_2^*\left[P(y_1), a_1^*\right] = \mathbb{E}\left[\frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1^E}{\sigma_2}\right) \,\middle|\, y_1 \in P(y_1)\right]. \tag{37}$$

Suppose there are $n \in \mathbb{N}$ non-singleton elements of the disclosure policy P and that non-singleton 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 $\overline{y}_{ij}$ the left and right endpoints of the jth such interval. Also, suppose that Y_1^D is made up of $m_d \in \mathbb{N}$ intervals and denote by $\underline{y}_{dj}$ and $\overline{y}_{dj}$ the left and right endpoints of the jth such interval. Finally, let

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

be the set of boundary points between the elements of P. Each finite interval endpoint described above is a member of Y^B.

By applying the change of variables $z = y_1 - a_1$, one can express the worker's first period objective function as

$$\begin{aligned}
&\sum_{j=1}^{m_d} \int_{\underline{y}_{dj} - \mathbb{1}\left(\underline{y}_{dj} \in Y^B\right)a_1}^{\overline{y}_{dj} - \mathbb{1}\left(\overline{y}_{dj} \in Y^B\right)a_1} \left(\begin{array}{l} W\left(1 - \Phi\left(\frac{\theta^* - \widehat{\theta}_1(z) - \lambda_2(a_1 - a_1^* + a_2^W(z+a_1, a_1) - a_2^*(z+a_1, a_1^*))}{\sigma_2}\right)\right) \\ -\frac{C}{2}\left(a_2^W(z + a_1, a_1)\right)^2 \end{array} \right) f_z(z)dz + \\
&\sum_{i=1}^{n}\sum_{j=1}^{m_i} \int_{\underline{y}_{ij} - \mathbb{1}\left(\underline{y}_{ij} \in Y^B\right)a_1}^{\overline{y}_{ij} - \mathbb{1}\left(\overline{y}_{ij} \in Y^B\right)a_1} \left(\begin{array}{l} W\left(1 - \Phi\left(\frac{\theta^* - \widehat{\theta}_1(z) - \lambda_2(a_1 - a_1^* + a_2^W(Y_1^i, a_1) - a_2^*(Y_1^i, a_1^*))}{\sigma_2}\right)\right) - \\ \frac{C}{2}\left(a_2^W(Y_1^i, a_1)\right)^2 \end{array} \right) f_z(z)dz \\
&- \frac{C}{2}a_1^2.
\end{aligned} \tag{38}$$

Here f_z is the pdf of $z \sim N(\overline{\theta}, \sigma_\theta^2 + \sigma_\varepsilon^2)$, $\widehat{\theta}_1(z) = \lambda_1 z + (1-\lambda_1)\overline{\theta}$, 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 takes account of the fact that one need not transform the two infinite interval endpoints by subtracting a_1.

Differentiating with respect to a_1 gives

$$\begin{aligned}
&\sum_{j=1}^{m_d} \int_{\underline{y}_{dj} - \mathbb{1}\left(\underline{y}_{dj} \in Y^B\right)a_1}^{\overline{y}_{dj} - \mathbb{1}\left(\overline{y}_{dj} \in Y^B\right)a_1} \left(\begin{array}{l} \frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1(z) - \lambda_2(a_1 - a_1^* + a_2^W(z+a_1, a_1) - a_2^*(z+a_1, a_1^*))}{\sigma_2}\right) \times \\ \left(1 + \frac{\partial a_2^W(z+a_1, a_1)}{\partial a_1} - \frac{\partial a_2^*(z+a_1, a_1^*)}{\partial a_1}\right) \\ -Ca_2^W(z + a_1, a_1)\frac{\partial a_2^W(z+a_1, a_1)}{\partial a_1} \end{array} \right) f_z(z)dz + \\
&\sum_{i=1}^{n}\sum_{j=1}^{m_i} \int_{\underline{y}_{ij} - \mathbb{1}\left(\underline{y}_{ij} \in Y^B\right)a_1}^{\overline{y}_{ij} - \mathbb{1}\left(\overline{y}_{ij} \in Y^B\right)a_1} \left(\begin{array}{l} \frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1(z) - \lambda_2(a_1 - a_1^* + a_2^W(Y_1^i, a_1) - a_2^*(Y_1^i, a_1^*))}{\sigma_2}\right) \times \\ \left(1 + \frac{\partial a_2^W(Y_1^i, a_1)}{\partial a_1}\right) \\ -Ca_2^W(Y_1^i, a_1)\frac{\partial a_2^W(Y_1^i, a_1)}{\partial a_1} \end{array} \right) f_z(z)dz + \\
&\sum_{y^B \in Y^B} W \left(\begin{array}{l} \Phi\left(\frac{\theta^* - \widehat{\theta}_1(y^B - a_1) - \lambda_2(a_1 - a_1^* + a_2^W(\lim_{\varepsilon\to 0} P(y^B - \varepsilon), a_1) - a_2^*(\lim_{\varepsilon\to 0} P(y^B - \varepsilon), a_1^*))}{\sigma_2}\right) - \\ \Phi\left(\frac{\theta^* - \widehat{\theta}_1(y^B - a_1) - \lambda_2(a_1 - a_1^* + a_2^W(\lim_{\varepsilon\to 0} P(y^B + \varepsilon), a_1) - a_2^*(\lim_{\varepsilon\to 0} P(y^B + \varepsilon), a_1^*))}{\sigma_2}\right) \\ +\frac{C}{2}\left(\left(a_2^W(\lim_{\varepsilon\to 0} P(y^B - \varepsilon), a_1)\right)^2 - \left(a_2^W(\lim_{\varepsilon\to 0} P(y^B + \varepsilon), a_1)\right)^2\right) \end{array} \right) \times \\
&f_z(y^B - a_1) - Ca_1.
\end{aligned} \tag{39}$$

The first two terms simplify to

$$\sum_{j=1}^{m_d}\int_{\underline{y}_{dj}-\mathbb{1}\left(\underline{y}_{dj}\in Y^B\right)a_1}^{\overline{y}_{dj}-\mathbb{1}\left(\overline{y}_{dj}\in Y^B\right)a_1}\left(\begin{array}{l}\frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^*-\widehat{\theta}_1(z)-\lambda_2(a_1-a_1^*+a_2^W(z+a_1,a_1)-a_2^*(z+a_1,a_1^*))}{\sigma_2}\right)\times\\ \left(1-\frac{\partial a_2^*(z+a_1,a_1^*)}{\partial a_1}\right)\end{array}\right)f_z(z)dz+$$
$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int_{\underline{y}_{ij}-\mathbb{1}\left(\underline{y}_{ij}\in Y^B\right)a_1}^{\overline{y}_{ij}-\mathbb{1}\left(\overline{y}_{ij}\in Y^B\right)a_1}\left(\frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^*-\widehat{\theta}_1(z)-\lambda_2(a_1-a_1^*+a_2^W(Y_1^i,a_1)-a_2^*(Y_1^i,a_1^*))}{\sigma_2}\right)\right)f_z(z)dz \qquad (40)$$

For the sake of space the paper does not compute the second derivative. One can show that it tends to $-\infty$ as $C\to\infty$ so that there exists a $\overline{C}_2$ such that (39) is globally concave for all $C>\overline{C}_2$. 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 (39) are positive as $a_1\to 0$. As one can see from (40) the second term is always positive. From (36) and (37) 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 (39) limits to 0 as $C\to\infty$, while the first term limits to is

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

So there exists a $\overline{C}_3$ such that the first and third terms of (39) are positive at $a_1=0$ for all $C\geq\overline{C}_3$. By observation (39) tends to $-\infty$ as $a_1\to\infty$. So for all $C\geq\overline{C}_3$, the first order condition has a unique interior solution given by a_1^W. Thus whenever $C>\max\{\overline{C}_1,\overline{C}_2,\overline{C}_3\}$ 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 (36) and (37) 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

$$h(a_1^*)=a_1^*-\int_{-\infty}^{\infty}\frac{\lambda_2 W}{\sigma_2}\phi\left(\frac{\theta^*-\widehat{\theta}_1(z)}{\sigma_2}\right)f_z(z)dz-$$
$$\sum_{j=1}^{m_d}\int_{\underline{y}_{dj}-\mathbb{1}\left(\underline{y}_{dj}\in Y^B\right)a_1^*}^{\overline{y}_{dj}-\mathbb{1}\left(\overline{y}_{dj}\in Y^B\right)a_1^*}\left(\frac{W}{C}\right)^2\frac{\lambda_2^2\lambda_1}{\sigma_2^3}\left(\frac{\widehat{\theta}_1(z)-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-\widehat{\theta}_1(z)}{\sigma_2}\right)f_z(z)dz. \qquad (41)$$

As $a_1^*\to 0$, $h(a_1^*)\to\overline{h}>0$ 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

$$1-\sum_{j=1}^{m_d}\left(\frac{W}{C}\right)^2\frac{\lambda_2^2\lambda_1}{\sigma_2^3}\left[\begin{array}{l}\left(\frac{\widehat{\theta}_1(\underline{y}_{dj}-a_1^*)-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-\widehat{\theta}_1(\underline{y}_{dj}-a_1^*)}{\sigma_2}\right)f_z(\underline{y}_{dj}-a_1^*)dz-\\ \left(\frac{\widehat{\theta}_1(\overline{y}_{dj}-a_1^*)-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-\widehat{\theta}_1(\overline{y}_{dj}-a_1^*)}{\sigma_2}\right)f_z(\overline{y}_{dj}-a_1^*)dz\end{array}\right]. \qquad (42)$$

Clearly there exists a $\overline{C}_4$ such that (42) is positive for all $C\geq\overline{C}_4$. The proposition is established by setting $\overline{C}=\max\{\overline{C}_1,\overline{C}_2,\overline{C}_3,\overline{C}_4\}$. ■

[29] The first order condition for a_1 is obtained by setting (39) equal to zero.

A.2.3 Proof of Corollary 1

Proof. Suppose the elements of P are $\left\{Y_1^i\right\}_{i=1}^n \cup Y_1^D$ and let

$$\widetilde{a}(y_1) = \frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^* - \widehat{\theta}_1^*(y_1)}{\sigma_2}\right)$$

$$\begin{aligned}
\mathbb{E}\left[\, a_2^* \,\right] &= \sum_i \mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^i \,\right] \Pr\left[\, y_1 \in Y_1^i \,\right] + \int_{y_1 \in Y_1^D} \widetilde{a}(y_1) f_y(y_1) dy_1 \\
&= \int_{-\infty}^{\infty} \widetilde{a}(y_1) f_y(y_1) dy_1,
\end{aligned}$$

which is independent of P. ■

A.2.4 Proof of Corollary 2

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

$$\begin{aligned}
&\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{11} \,\right]\right)^2 \Pr\left[\, y_1 \in Y_1^{11} \,\right] + \left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{12} \,\right]\right)^2 \Pr\left[\, y_1 \in Y_1^{12} \,\right] \\
> &\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{1} \,\right]\right)^2 \Pr\left[\, y_1 \in Y_1^{1} \,\right] \Longrightarrow \\
&\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{11} \,\right]\right)^2 \frac{\Pr\left[\, y_1 \in Y_1^{11} \,\right]}{\Pr\left[\, y_1 \in Y_1^{1} \,\right]} + \left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{12} \,\right]\right)^2 \frac{\Pr\left[\, y_1 \in Y_1^{12} \,\right]}{\Pr\left[\, y_1 \in Y_1^{1} \,\right]} \\
> &f\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{11} \,\right] \frac{\Pr\left[\, y_1 \in Y_1^{11} \,\right]}{\Pr\left[\, y_1 \in Y_1^{1} \,\right]} + \mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{12} \,\right] \frac{\Pr\left[\, y_1 \in Y_1^{12} \,\right]}{\Pr\left[\, y_1 \in Y_1^{1} \,\right]}\right)^2,
\end{aligned}$$

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

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

$$\begin{aligned}
&\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{1\prime} \,\right]\right)^2 \Pr\left[\, y_1 \in Y_1^{1\prime} \,\right] + \mathbb{E}\left[\, (\widetilde{a}(y_1))^2 \,\middle|\, y_1 \in \left(Y_1^{D\prime} \backslash Y_1^D\right) \,\right] \Pr\left[\, y_1 \in \left(Y_1^{D\prime} \backslash Y_1^D\right) \,\right] \\
> &\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{1} \,\right]\right)^2 \Pr\left[\, y_1 \in Y_1^{1} \,\right]
\end{aligned}$$

Now,

$$\begin{aligned}
&\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{1\prime} \,\right]\right)^2 \Pr\left[\, y_1 \in Y_1^{1\prime} \,\right] + \mathbb{E}\left[\, (\widetilde{a}(y_1))^2 \,\middle|\, y_1 \in \left(Y_1^{D\prime} \backslash Y_1^D\right) \,\right] \Pr\left[y_1 \in \left(Y_1^{D\prime} \backslash Y_1^D\right)\right] \\
> &\left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in Y_1^{1\prime} \,\right]\right)^2 \Pr\left[\, y_1 \in Y_1^{1\prime} \,\right] + \left(\mathbb{E}\left[\, \widetilde{a}(y_1) \mid y_1 \in \left(Y_1^{D\prime} \backslash Y_1^D\right) \,\right]\right)^2 \Pr\left[\, y_1 \in \left(Y_1^{D\prime} \backslash Y_1^D\right) \,\right]
\end{aligned}$$

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

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

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

A.3 Section 5

A.3.1 Proof of Proposition 3

Proof. Let $\Theta_1^{d*}(P)$ be as defined in the text. The proof proceeds in two stages. First it establishes the form that $\Theta_1^{d*}(P^S)$ takes. It then shows that there exists a P^S that induces $\Theta_1^{d*}(P^S)$.

Let P and P' be disclosure policies that satisfy $\Theta_1^{d*}(P') = \Theta_1^{d*}(P) \cup (t, t+\varepsilon)$. By Corollary 2 moving from P to P' increases expected second period effort costs, and the following gives the amount by which it does so when ε is small. In the proof f_θ denotes the probability density function of $\widehat{\theta}_1^* \sim N\left(\overline{\theta}, \sigma_1^2\right)$.

Lemma 2

$$\lim_{\varepsilon\to 0}\left(\mathbb{E}\left[(a_2^*)^2 \,\middle|\, P'\right] - \mathbb{E}\left[(a_2^*)^2 \,\middle|\, P\right]\right) =$$
$$\frac{1}{2C}\left(\frac{W\lambda_2}{\sigma_2}\right)^2\left(\phi\left(\frac{\theta^*-t}{\sigma_2}\right) - \mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}(P)\right]\right)^2 f_\theta(t) \tag{43}$$

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

$$\Pr\left[\widehat{\theta}_1^* \in \Theta_1^{d*}(P)\cup(t,t+\varepsilon)\right]\mathbb{E}\left[\phi^2\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\,\middle|\,\widehat{\theta}_1^* \in \Theta_1^{d*}(P)\cup(t,t+\varepsilon)\right]+$$
$$\Pr\left[\widehat{\theta}_1^* \notin \Theta_1^{d*}(P)\cup(t,t+\varepsilon)\right]\mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\,\middle|\,\widehat{\theta}_1^* \notin \Theta_1^{d*}(P)\cup(t,t+\varepsilon)\right]^2$$

which can be further expanded as

$$\left(\begin{array}{c}\int_{\widehat{\theta}_1^*\in\Theta_1^{d*}(P)}\phi^2\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^* + \int_t^{t+\varepsilon}\phi^2\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^* + \\ \left[\int_{\widehat{\theta}_1^*\notin\Theta_1^{d*}(P)}f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^* - \int_t^{t+\varepsilon}f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^*\right]\times \\ \left[\frac{\int_{\widehat{\theta}_1^*\notin\Theta_1^{d*}(P)}\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^* - \int_t^{t+\varepsilon}\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^*}{\int_{\widehat{\theta}_1^*\notin\Theta_1^{d*}(P)}f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^* - \int_t^{t+\varepsilon}f_\theta\left(\widehat{\theta}_1^*\right)d\widehat{\theta}_1^*}\right]^2\end{array}\right)$$

The derivative of the term in brackets with respect to ε is

$$\phi^2\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right)f_\theta(t+\varepsilon) - f_\theta(t+\varepsilon)\left(\mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\,\middle|\,\widehat{\theta}_1^* \notin \Theta_1^{d*}(P)\cup(t,t+\varepsilon)\right]\right)^2$$
$$+\Pr\left[\widehat{\theta}_1^* \notin \Theta_1^{d*}(P)\cup(t,t+\varepsilon)\right]2\mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\,\middle|\,\widehat{\theta}_1^* \notin \Theta_1^{d*}(P)\cup(t,t+\varepsilon)\right]\times$$

$$\frac{\left(\begin{array}{c} -\phi\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right) f_\theta\left(t+\varepsilon\right)\Pr\left[\widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P\right)\cup\left(t,t+\varepsilon\right)\right]+ \\ \mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\middle|\widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P\right)\cup\left(t,t+\varepsilon\right)\right]\Pr\left[\widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P\right)\cup\left(t,t+\varepsilon\right)\right] f_\theta\left(t+\varepsilon\right)\end{array}\right)}{\left(\Pr\left[\widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P\right)\cup\left(t,t+\varepsilon\right)\right]\right)^2}$$

which reduces to

$$\phi^2\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right) f_\theta\left(t+\varepsilon\right)+\left(\mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\middle|\widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P\right)\cup\left(t,t+\varepsilon\right)\right]\right)^2 f_\theta\left(t+\varepsilon\right)$$
$$-\phi\left(\frac{\theta^*-t-\varepsilon}{\sigma_2}\right) 2\mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\middle|\widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P\right)\cup\left(t,t+\varepsilon\right)\right] f_\theta\left(t+\varepsilon\right).$$

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

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

$$\mathbb{E}\left[\frac{W}{C}\frac{\lambda_2}{\sigma_2}\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\right]+\int_{\widehat{\theta}_1^*\in\Theta_1^{d*}(P)}\left(\frac{W}{C}\right)^2\frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{\widehat{\theta}_1^*-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right) f_\theta(\widehat{\theta}_1^*)d\widehat{\theta}_1^*+$$
$$\int_t^{t+\varepsilon}\left(\frac{W}{C}\right)^2\frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{\widehat{\theta}_1^*-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right) f_\theta(\widehat{\theta}_1^*)d\widehat{\theta}_1^* \tag{44}$$

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

$$\left(\frac{W}{C}\right)^2\frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{t-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-t}{\sigma_2}\right) f_\theta(t). \tag{45}$$

So the change in first period welfare from adding beliefs $(t, t+\varepsilon)$ to $\Theta_1^{d*}\left(P\right)$ for small ε is approximately

$$\left(\frac{W}{C}\right)^2\frac{\lambda_1\lambda_2^2}{\sigma_2^3}\left(\frac{t-\theta^*}{\sigma_2}\right)\phi^2\left(\frac{\theta^*-t}{\sigma_2}\right) f_\theta(t)\left(1-Ca_1^*\left(P\right)\right). \tag{46}$$

Let

$$L\left(t,P\right)=A\left(\frac{t-\theta^*}{\sigma_2}\right) \tag{47}$$

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

$$R\left(t,P\right)=\left(\frac{\mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right)\middle|\widehat{\theta}_1 \notin \Theta_1^{d*}(P)\right]}{\phi\left(\frac{\theta^*-t}{\sigma_2}\right)}-1\right)^2. \tag{48}$$

From (43) and (46) one can conclude that P^S must satisfy

$$L\left(\widehat{\theta}_1^*, P^S\right)\geq R\left(\widehat{\theta}_1^*, P^S\right)\ \forall\widehat{\theta}_1^*\in\Theta_1^{d*}\left(P^S\right)$$

and

$$L\left(\widehat{\theta}_1^*, P^S\right)< R\left(\widehat{\theta}_1^*, P^S\right)\ \forall\widehat{\theta}_1^*\notin\Theta_1^{d*}\left(P^S\right).$$

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 $\Theta_1^{d*}\left(P'\right) \cap (-\infty, \theta^*) = \emptyset$. This implies that

$$0 < \mathbb{E}\left[\phi\left(\frac{\theta^* - \widehat{\theta}_1^*}{\sigma_2}\right) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P'\right)\right] < \phi(0).$$

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

$$\phi\left(\frac{\theta^* - \widetilde{t}}{\sigma_2}\right) = \mathbb{E}\left[\phi\left(\frac{\theta^* - \widehat{\theta}_1^*}{\sigma_2}\right) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P'\right)\right].$$

Moreover, $R(t, P') > 0$ for all $t \neq \widetilde{t}$. Suppose $Ca_1^*\left(P^S\right) = 1$. Then $L(t, P^S) = 0$ for all $t \in \Theta_1^{d*}(P^S)$ while there exists some t' for which $R(t', P^S) > 0$. Thus replacing P^S with a disclosure policy $\widetilde{P}$ satisfying $\Theta_1^{d*}\left(\widetilde{P}\right) = \Theta_1^{d*}\left(P^S\right) \setminus [t', t' + \varepsilon]$ improves social welfare for small enough ε, contradicting the optimality of P^S. So $Ca_1^*\left(P^S\right) < 1$. So further assume that the P' considered above satisfies $Ca_1^*\left(P'\right) < 1$.

From the above arguments one can also conclude that $R(t, P')$ is strictly decreasing on $t \in \left(\theta^*, \widetilde{t}\right)$. Because $L(t, P')$ is strictly increasing on $t \in \left(\theta^*, \widetilde{t}\right)$, there exists a unique point $t_L(P') < \widetilde{t}$ 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 $L(t, P') > R(t, P')$ $\forall t \in \left(t_L(P'), \widetilde{t}\right)$.

Let $d = \mathbb{E}\left[\phi\left(\frac{\theta^*-\widehat{\theta}_1^*}{\sigma_2}\right) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}\left(P'\right)\right]$.

$$\frac{\partial R}{\partial t} = 2\left[d\sqrt{2\pi}\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right) - 1\right] d\sqrt{2\pi}\left(\frac{t - \theta^*}{\sigma_2^2}\right)\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right), \tag{49}$$

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

$$\begin{aligned}\frac{\partial^2 R}{\partial t^2} &= 2\left[d\sqrt{2\pi}\left(\frac{t - \theta^*}{\sigma_2^2}\right)\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right)\right]^2 + \\ &2\left[d\sqrt{2\pi}\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right) - 1\right] \times \\ &\left[\frac{d\sqrt{2\pi}}{\sigma_2^2}\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right) + d\sqrt{2\pi}\left(\frac{t - \theta^*}{\sigma_2^2}\right)^2\exp\left(\frac{1}{2}\left(\frac{\theta^* - t}{\sigma_2}\right)^2\right)\right],\end{aligned} \tag{50}$$

which is also strictly bigger than zero on $t \in \left(\widetilde{t}, \infty\right)$. Thus R is strictly convex and increasing on $t \in \left(\widetilde{t}, \infty\right)$ 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

$$\left[\frac{\partial L\left(t,P'\right)}{\partial t}\right]_{t=t_H(P')} < \left[\frac{\partial R\left(t,P'\right)}{\partial t}\right]_{t=t_H(P')}. \tag{51}$$

Clearly $L(t,P') > R(t,P')$ $\forall t \in \left(\widetilde{t}, t_H(P')\right)$ 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

$$\frac{\partial t_H}{\partial A} = \frac{\frac{\partial L}{\partial A}}{\frac{\partial R}{\partial t_H} - \frac{\partial L}{\partial t_H}} = \frac{\frac{\theta^* - t_H}{\sigma_2}}{\frac{\partial R}{\partial t_H} - \frac{\partial L}{\partial t_H}} > 0 \tag{52}$$

and

$$\frac{\partial t_H}{\partial d} = -\frac{\frac{\partial R}{\partial d}}{\frac{\partial R}{\partial t_H} - \frac{\partial L}{\partial t_H}} = \frac{2\left(d\phi^{-1}\left(\frac{\theta^* - t_H}{\sigma_2}\right) - 1\right)\phi^{-1}\left(\frac{\theta^* - t_H}{\sigma_2}\right)}{\frac{\partial R}{\partial t_H} - \frac{\partial L}{\partial t_H}} < 0. \tag{53}$$

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

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

has two solutions $\underline{t}$ and $\bar{t}$ where $\theta^* < \underline{t} < \bar{t} < \infty$. Finally, by (52) and (53) t_H is never greater than $\bar{t}$.

Consider social surplus as a function of disclosure policies of the form $\Theta_1^{d*} = (\theta_L, \theta_H)$ where $\theta^* \leq \theta_L \leq \theta_H \leq \bar{t}$. The first section of the proof established the continuity of surplus in θ_L and θ_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 $\Theta_1^{d*}(P^S) = (\theta'_L, \theta'_H)$ and $\theta^* < \theta'_L < \theta'_H \leq \bar{t}$.

Thus if one can find a disclosure policy P for which $Y_1^D(P) = (y^*, y^{**})$ and where y^* and y^{**} satisfy the following, P is optimal:

$$\begin{pmatrix} y^* - a_1^*(P) \\ y^{**} - a_1^*(P) \end{pmatrix} = \begin{pmatrix} \frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1} \\ \frac{\theta'_H - (1-\lambda_1)\bar{\theta}}{\lambda_1} \end{pmatrix}. \tag{54}$$

Satisfying (54) is equivalent to finding a disclosure policy P with $Y_1^D(P) = \left[y^*, y^* + \frac{\theta'_H - \theta'_L}{\lambda_1}\right]$ where y^* satisfies[30]

$$y^* - a_1^*(y^*) = \frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1}. \tag{55}$$

Now when $y^* = \frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1}$ the LHS of (55) is smaller than the RHS since $a_1^*(y^*) > 0$ and when $y^* = \frac{\theta'_L - (1-\lambda_1)\bar{\theta}}{\lambda_1} + 2\mathbb{E}[a_2^*]$ the LHS is bigger than the RHS since $a_1^*(y^*) < 2\mathbb{E}[a_2^*]$. Since the LHS of (55) 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 (55) holds and an optimal disclosure policy with the stated form exists. ■

[30] Here the paper makes an abuse of notation by making the dependence of a_1^* on y^* rather than the disclosure policy P.

A.4 Section 6

A.4.1 Proof of Proposition 4

Proof.

$$
\begin{aligned}
&\mathbb{E}[\, w_3 \mid y_1, a_1, a_2 \,] \\
&= \mathbb{E}\left[F + \widehat{\theta}_2^E \,\middle|\, \widehat{\theta}_2^E \geq 0, y_1, a_1, a_2 \right] \Pr\left[\widehat{\theta}_2^E \geq 0 \,\middle|\, y_1, a_1, a_2 \right] + 0 \Pr\left[\widehat{\theta}_2^E < 0 \,\middle|\, y_1, a_1, a_2 \right] \\
&= F \left[1 - \Phi\left(\frac{-\widehat{\theta}_1(y_1) - \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2} \right) \right] + \int_0^\infty \widehat{\theta}_2^E f_{\widehat{\theta}}\left(\widehat{\theta}_2^E \right) d\widehat{\theta}_2^E
\end{aligned} \tag{56}
$$

Where $f_{\widehat{\theta}}$ represents the probability density function for $\widehat{\theta}_2^E \mid y_1, a_1, a_2$ derived in Lemma 1. The last term of (56) can be transformed as

$$
\int_{-\frac{\widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}}^{\infty} \left(\sigma_2 v + \widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*) \right) \phi(v) dv. \tag{57}
$$

So, by Leibnitz's Rule,

$$
\begin{aligned}
&\frac{\partial \mathbb{E}[\, w_3 \mid y_1, a_1, a_2 \,]}{\partial a_2} \\
=& F \frac{\lambda_2}{\sigma_2} \phi\left(\frac{\widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2} \right) + \int_{-\frac{\widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2}}^{\infty} \lambda_2 \phi(v) dv \\
=& F \frac{\lambda_2}{\sigma_2} \phi\left(\frac{\widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2} \right) + \lambda_2 \Phi\left(\frac{\widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2} \right)
\end{aligned} \tag{58}
$$

So the first order condition characterizing a_2^W is

$$
\mathbb{E}\left[\left(\begin{array}{c} F \frac{\lambda_2}{\sigma_2} \phi\left(\frac{\widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2} \right) + \\ \lambda_2 \Phi\left(\frac{\widehat{\theta}_1(y_1) + \lambda_2(a_1 - a_1^* + a_2 - a_2^*)}{\sigma_2} \right) \end{array} \right) \,\middle|\, y_1 \in P(y_1) \right] = C a_2^W. \tag{59}
$$

From here one can proceed using equivalent arguments from Proposition 2. ■

A.4.2 Proof of Proposition 5

Proof. From arguments identical to those in Lemma 2, the change in second period surplus from adding disclosed beliefs $(t, t + \varepsilon)$ to $\Theta_1^{D*}(P)$ is $MC(t, P) =$

$$
\frac{C}{2} \left(\frac{F}{C} \frac{\lambda_2}{\sigma_2} \phi\left(\frac{t}{\sigma_2} \right) + \frac{\lambda_2}{C} \Phi\left(\frac{t}{\sigma_2} \right) - \mathbb{E}\left[\frac{F}{C} \frac{\lambda_2}{\sigma_2} \phi\left(\frac{\widehat{\theta}_1^*}{\sigma_2} \right) + \frac{\lambda_2}{C} \Phi\left(\frac{\widehat{\theta}_1^*}{\sigma_2} \right) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \right] \right)^2 f_\theta(t) \tag{60}
$$

while the change in first period surplus is $MB(t, P) =$

$$
\left(\frac{\lambda_1 \lambda_2}{C \sigma_2} \left[\phi\left(\frac{t}{\sigma_2} \right) \left(t \frac{F}{\sigma_2^2} - 1 \right) \right] \left[\frac{F}{C} \frac{\lambda_2}{\sigma_2} \phi\left(\frac{t}{\sigma_2} \right) + \frac{\lambda_2}{C} \Phi\left(\frac{t}{\sigma_2} \right) \right] \right) \left(1 - C a_1^*[P] \right) f_\theta(t). \tag{61}
$$

By the same reasoning as on page 34, one can restrict attention to a disclosure policy

P for which $Ca_1^*(P) < 1$ and for which $\Theta_1^{d*}(P) \subseteq \left(\frac{\sigma_2^2}{F}, \infty\right)$. Now suppose that P satisfies $\Theta_1^{d*}(P) = \left(\frac{\sigma_2^2}{F}, \infty\right)$. By observation, $MC\left(\frac{\sigma_2^2}{F}, P\right) > MB\left(\frac{\sigma_2^2}{F}, P\right) = 0$, establishing the first part of the claim.

Before continuing, the following lemma is helpful.

Lemma 3 *There exists a $\overline{\theta}'$ such that $\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \right] > \frac{\lambda_2}{C}$ for all $\overline{\theta} > \overline{\theta}'$ whenever $\Theta_1^{d*}(P) \subset \left(\frac{\sigma_2^2}{F}, \infty\right)$.*

Proof. Let $A(t)$ be defined as in the text.

$$A'(t) = \frac{F}{C}\frac{\lambda_2}{\sigma_2^2}\phi'\left(\frac{t}{\sigma_2}\right) + \frac{\lambda_2}{C\sigma_2}\phi\left(\frac{t}{\sigma_2}\right) = \frac{\lambda_2}{C\sigma_2}\phi\left(\frac{t}{\sigma_2}\right)\left(1 - t\frac{F}{\sigma_2^2}\right).$$

So $A'(t) \gtreqless 0$ for $t \lesseqgtr \frac{\sigma_2^2}{F}$. Moreover $\lim_{t\to-\infty} A(t) = 0$ and $\lim_{t\to\infty} A(t) = \frac{\lambda_2}{C}$. This implies that there exists some t' for which $A(t) \gtreqless \frac{\lambda_2}{C}$ for $t \gtreqless t'$.

Now observe that

$$\begin{aligned}
&\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \right] \\
=\; &\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \leq \frac{\sigma_2^2}{F} \right] \Pr\left[\widehat{\theta}_1^* \leq \frac{\sigma_2^2}{F}\right] + \\
&\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \in \left(\frac{\sigma_2^2}{F}, \infty\right) \backslash \Theta_1^{d*}(P) \right] \Pr\left[\widehat{\theta}_1^* \in \left(\frac{\sigma_2^2}{F}, \infty\right) \backslash \Theta_1^{d*}(P)\right] \\
=\; &\begin{pmatrix} \mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* < t' \right] \Pr\left[\widehat{\theta}_1^* < t' \,\middle|\, \widehat{\theta}_1^* \leq \frac{\sigma_2^2}{F}\right] + \\ \mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \in \left[t', \frac{\sigma_2^2}{F}\right] \right] \Pr\left[\widehat{\theta}_1^* \in \left[t', \frac{\sigma_2^2}{F}\right] \,\middle|\, \widehat{\theta}_1^* \leq \frac{\sigma_2^2}{F}\right] \end{pmatrix} \Pr\left[\widehat{\theta}_1^* \leq \frac{\sigma_2^2}{F}\right] + \\
&\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \in \left(\frac{\sigma_2^2}{F}, \infty\right) \backslash \Theta_1^{d*}(P) \right] \Pr\left[\widehat{\theta}_1^* \in \left(\frac{\sigma_2^2}{F}, \infty\right) \backslash \Theta_1^{d*}(P)\right].
\end{aligned}$$

Since $\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \in \left[t', \frac{\sigma_2^2}{F}\right] \right]$ and $\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \in \left(\frac{\sigma_2^2}{F}, \infty\right) \backslash \Theta_1^{d*}(P) \right]$ and both strictly greater than $\frac{\lambda_2}{C}$, a sufficient condition for $\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \right] > \frac{\lambda_2}{C}$ is for

$$\Pr\left[\widehat{\theta}_1^* \in \left[t', \frac{\sigma_2^2}{F}\right] \,\middle|\, \widehat{\theta}_1^* \leq \frac{\sigma_2^2}{F}\right] = \frac{\Phi\left(\frac{\frac{\sigma_2^2}{F} - \overline{\theta}}{\sigma_2}\right) - \Phi\left(\frac{t' - \overline{\theta}}{\sigma_2}\right)}{\Phi\left(\frac{\frac{\sigma_2^2}{F} - \overline{\theta}}{\sigma_2}\right)}$$

to be sufficiently close to one. This limits to 1 as $\overline{\theta} \to \infty$ since

$$\lim_{\overline{\theta}\to\infty} \frac{\Phi\left(\frac{t'-\overline{\theta}}{\sigma_2}\right)}{\Phi\left(\frac{\frac{\sigma_2^2}{F}-\overline{\theta}}{\sigma_2}\right)} = \lim_{\overline{\theta}\to\infty} \frac{\phi\left(\frac{t'-\overline{\theta}}{\sigma_2}\right)}{\phi\left(\frac{\frac{\sigma_2^2}{F}-\overline{\theta}}{\sigma_2}\right)} = \lim_{\overline{\theta}\to\infty} \exp\left(\left(\frac{\sigma_2^2}{F}\right)^2 - \left(t'\right)^2 - 2\overline{\theta}\left(\frac{\sigma_2^2}{F} - t'\right)\right) = 0.$$

So there exists a $\overline{\theta}'$ such that for all $\overline{\theta} > \overline{\theta}'$, $\mathbb{E}\left[A(\widehat{\theta}_1^*) \,\middle|\, \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \right] > \frac{\lambda_2}{C}$. ∎

Now suppose that $\overline{\theta} > \overline{\theta}'$. This implies that $\mathbb{E}\left[A(\widehat{\theta}_1^*) \right]$ lies strictly between the maximum and minimum of $A(t)$ so that there exists a $\widetilde{t} \in \left(\frac{\sigma_2^2}{F}, \infty\right)$ for which $MC(\widetilde{t}, P^N) = 0$ and $MB(\widetilde{t}, P^N) > 0$. So a disclosure policy P' for which $\Theta_1^{d*}(P') = \left(\widetilde{t}, \widetilde{t} + \varepsilon\right)$ yields higher surplus than P^N for small enough epsilon.

Now consider a disclosure policy P for which (1) $\Theta_1^{d*}(P) \subset \left(\frac{\sigma_2^2}{F}, \infty\right)$ and (2) $\Theta_1^{d*}(P)$ is unbounded above. Whenever $\overline{\theta} > \overline{\theta}'$ one obtains

$$\lim_{t \to \infty} \frac{MC(t, P)}{f_\theta(t)} = \frac{C}{2} \left(\frac{\lambda_2}{C} - \mathbb{E}\Big[A\left(\widehat{\theta}_1^*\right) \Bigm| \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \Big] \right)^2 > 0$$

while $\lim_{t\to\infty} \frac{MB(t,P)}{f_\theta(t)} = 0$. So there exists a finite point $t_H(P)$ for which

$$MC(t_H(P), P) > MB(t_H(P), P) \; \forall t > t_H(P).$$

One can show that t_H is bounded above. First note that because there exists a $\widetilde{t}$ at which $MC(\widetilde{t}, P) = 0 < MB(\widetilde{t}, P)$, $t_H > \widetilde{t}$ and $\frac{\partial MC(t_H,P)}{\partial t_H} > \frac{\partial MB(t_H,P)}{\partial t_H}$ by continuity. Let $B = 1 - Ca_1^*(P)$ and $D = \mathbb{E}\Big[A\left(\widehat{\theta}_1^*\right) \Bigm| \widehat{\theta}_1^* \notin \Theta_1^{d*}(P) \Big]$. By the implicit function theorem

$$\frac{\partial t_H}{\partial B} = \frac{\frac{\lambda_1 \lambda_2}{C\sigma_2} \phi\left(\frac{t_H}{\sigma_2}\right) \left(t_H \frac{F}{\sigma_2^2} - 1\right) A(t_H)}{\frac{\partial MC(t_H,P)}{\partial t_H} - \frac{\partial MB(t_H,P)}{\partial t_H}} > 0 \tag{62}$$

and

$$\frac{\partial t_H}{\partial D} = \frac{C\left(A(t_H) - D\right)}{\frac{\partial MC(t_H,P)}{\partial t_H} - \frac{\partial MB(t_H,P)}{\partial t_H}} < 0. \tag{63}$$

Note that B has an upper bound of 1 and D has a lower bound of $\frac{\lambda_2}{C} + \delta$ for some $\delta > 0$. By limit arguments similar to those above there exists a finite number $\overline{t}$ for which

$$\frac{C}{2}\left(A(t) - \frac{\lambda_2}{C} + \delta\right)^2 > \frac{\lambda_1\lambda_2}{C\sigma_2}\phi\left(\frac{t}{\sigma_2}\right)\left(t\frac{F}{\sigma_2^2} - 1\right)A(t)$$

for all $t > \overline{t}$. By (62) and (63) $t_H \leq \overline{t}$.

To complete the proof note that replacing P with a disclosure policy P' that satisfies $\Theta_1^{d*}(P') = \Theta_1^{d*}(P) \backslash \left[\overline{t}, \infty\right)$ strictly increases surplus. ■

References

Aoyagi, M. (2007): "Information Feedback in a Dynamic Tournament," mimeo, The Institute of Social and Economic Research, Osaka University.

Azmat, G., and N. Iriberri (2010): "The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness," mimeo, Universitat Pompeu Fabra.

Baker, G., M. Gibbs, and B. Holmström (1994a): "The Internal Economics of the Firm: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 109(4), 881–919.

——— (1994b): "The Wage Policy of a Firm," *The Quarterly Journal of Economics*, 109(4), 921–55.

Baron, J. N., and D. M. Kreps (1999): *Strategic Human Resources: Frameworks for General Managers*. John Wiley, New York, New York.

Blanes i Vidal, J., and M. Nossol (2009): "Tournaments without Prizes: Evidence from Personnel Records," mimeo, London School of Economics.

Crutzen, B., O. Swank, and B. Visser (2010): "Confidence Management: on Interpersonal Comparisons in Teams," mimeo, Erasmus School of Economics.

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

Ederer, F. (2008): "Feedback and Motivation in Dynamic Tournaments," mimeo, MIT.

Fuchs, W. (2007): "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, 97(4), 1432–1448.

Gibbs, M. (1991): "An Economic Approach to Process in Pay and Performance Appraisals," mimeo, University of Chicago GSB.

Goltsman, M., and A. Mukherjee (2008): "Information Disclosure in Multistage Tournaments," mimeo, University of Western Ontario.

Greene, W. H. (2003): *Econometric Analysis*. Prentice Hall, Upper Saddle River, New Jersey.

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

——— (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66(1), 169–82.

Jackman, J. M., and M. H. Strober (2003): "Fear of Feedback," *Harvard Business Review*, 81(4), 101–07.

Koch, A., and E. Peyrache (2010): "Aligning Ambition and Incentives," *Journal of Law, Economics and Organization*, forthcoming.

Kovrijnykh, A. (2007): "Career Uncertainty and Dynamic Incentives," mimeo, 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," *American Economic Review*, 86(3), 329–48.

Lazear, E. P., and M. Gibbs (2008): *Personnel Economics in Practice*. Wiley, Hoboken, NJ, second edn.

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

Longenecker, C. ., H. P. Sims, and D. A. Gioia (1987): "Behind the Mask: The Politics of Employee Appraisal," *The Academy of Management Executive*, 1(3), 183–93.

MacLeod, W. B. (2003): "Optimal Contracting with Subjective Evaluation," *American Economic Review*, 93(1), 216–240.

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

Martinez, L. (2009): "Reputation, Career Concerns, and Job Assignments," *B.E. Journal of Theoretical Economics*, 9(1).

Medoff, J. L., and K. G. Abraham (1980): "Experience, Performance, and Earnings," *The Quarterly Journal of Economics*, 95(4), 703–36.

Meyer, H. H., E. Kay, and J. R. P. French (1965): "Split Roles in Performance Appraisal," *Harvard Business Review*, 43(1), 123–29.

Milkovich, G. T., J. M. Newman, and C. Milkovich (2007): *Compensation*. McGraw-Hill, New York, New York, ninth edn.

Mukherjee, A. (2008): "Career Concerns, Matching, And Optimal Disclosure Policy," *International Economic Review*, 49(4), 1211–1250.

Murphy, K. J. (1992): "Performance Measurement and Appraisal: Motivating Managers to Identify and Reward Performance," in *Performance Measurement, Evaluation, and Incentives*, ed. by W. J. Bruns, pp. 37–62, Boston, Massachusetts.

Murphy, K. M., and J. N. Cleveland (1991): *Performance Appraisal: An Organizational Persepctive.* Allyn and Bacon, Needham Heights, Massachusetts.

Murphy, T. H., and J. Margulies (2004): "Performance Appraisals," Presentation, ABA Labor and Employment Law Section, Equal Employment Opportunity Committee, Mid-Winter Meeting.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37(1), 7–63.

Prendergast, C., and R. H. Topel (1996): "Favoritism in Organizations," *Journal of Political Economy*, 104(5), 958–78.

Ray, K. (2007): "The Retention Effect of Withholding Performance Information," *The Accounting Review*, 82(2), 389–425.

Waldman, M. (1984): "Job Assignments, Signalling, and Efficiency," *RAND Journal of Economics*, 15(2), 255–267.