ISSN 0956-8549-737

The Value of Informativeness for Contracting

By

Pierre Chaigneau
Alex Edmans
Daniel Gottlieb

FINANCIAL MARKETS GROUP
DISCUSSION PAPER No 737

October 2014

Pierre Chaigneau is an Assistant Professor of Finance at HEC Montreal, and a Research Associate at the Financial Markets Group (FMG). His research interests include the structure of executive compensation, information economics, and financial stability. He got his PhD in Finance from the London School of Economics.

Alex Edmans is a Professor of Finance at London Business School (LBS). He got his PhD in Financial Economics from MIT. He is currently on leave from Wharton, where he is an Associate Professor of Finance. He is a Research Fellow at the National Bureau of Economic Research (NBER) and at the Centre for Economic Policy Research (CEPR), and a Research Associate at the European Corporate Governance Institute (ECGI). His research interests include corporate governance, CEO compensation, information disclosure, and blockholder ownership.

Daniel Gottlieb is an Assistant Professor of Business Economics and Public Policy at Wharton. He got his PhD in Economics from MIT, and was a visiting scholar at Princeton before joining Wharton. His research interests include contract theory, behavioral economics, and insurance economics.

Any opinions expressed here are those of the authors and not necessarily those of the FMG. The research findings reported in this paper are the result of the independent research of the authors and do not necessarily reflect the views of the LSE.

The Value of Informativeness for Contracting*

Pierre Chaigneau
HEC Montreal

Alex Edmans
LBS, Wharton, NBER, CEPR, and ECGI

Daniel Gottlieb
Wharton

September 23, 2014

Abstract

The informativeness principle demonstrates qualitative benefits to increasing signal precision. However, it is difficult to quantify these benefits – and compare them against the costs of precision – since we typically cannot solve for the optimal contract and analyze how it changes with informativeness. We consider a standard agency model with risk-neutrality and limited liability, where the optimal contract is a call option. The direct effect of reducing signal volatility is a fall in the value of the option, benefiting the principal. The indirect effect is a change in the agent's effort incentives. If the original option is sufficiently out-of-the-money, the agent can only beat the strike price if he exerts effort and there is a high noise realization. Thus, a fall in volatility reduces effort incentives. As the agency problem weakens, the gains from precision fall towards zero, potentially justifying pay-for-luck.

Keywords: Contract theory, principal-agent model, limited liability, pay-for-luck, relative performance evaluation, options, informativeness principle.
JEL Classification: D86, J33

*pierre.chaigneau@hec.ca, aedmans@london.edu, dgott@wharton.upenn.edu. We thank seminar participants at Wharton and at the Risk Theory Society meeting for helpful comments, and Shiying Dong for excellent research assistance.

A major result in contract theory is the informativeness principle (Holmstrom (1979), Shavell (1979), Gjesdal (1982), Grossman and Hart (1983), Kim (1995)). It argues that the principal should maximize the precision of the performance measure used to evaluate the agent. Greater precision (in the sense of second-order stochastic dominance) allows the principal to use a cheaper contract to implement at least the same effort level. However, increasing informativeness is costly in practice. Investing in a superior monitoring technology involves direct costs. Engaging in relative performance evaluation ("RPE") involves the indirect costs of forgoing the benefits of pay-for-luck documented by prior research (e.g. Oyer (2004), Raith (2008), Axelson and Baliga (2009), and Gopalan, Milbourn, and Song (2010)). Potentially for this reason, numerous violations of RPE have been found in practice. Aggarwal and Samwick (1999) and Murphy (1999) show that CEO pay is determined by absolute, rather than relative performance. Jenter and Kanaan (2013) find an absence of RPE in CEO firing decisions. Whether these violations are an efficient response to the indirect costs of RPE is unclear. Bertrand and Mullainathan (2001) show pay-for-luck is strongest in poorly-governed firms, consistent with the view that it is inefficient. Indeed, Bebchuk and Fried (2004) argue that the absence of RPE is a key piece of evidence that CEO compensation results from rent extraction by CEOs rather than efficient contracting with shareholders.[1]

The informativeness principle argues that there are qualitative benefits to increasing signal precision. However, for a principal to decide whether to invest in greater precision, she must quantify these benefits – in particular, relate them to the underlying parameters of the contracting problem – so that she can compare them against the cost of precision. Similarly, to evaluate whether the general absence of RPE is efficient, it is useful to understand under which settings the benefits of informativeness are smallest, and compare them against the cases in which RPE is particularly absent in reality. Such quantification is difficult under the general framework in which the informativeness principle was derived. As is well-known (e.g. Grossman and Hart (1983)), in a general setting it is not possible to solve for the optimal contract. We cannot analyze precisely how the contract changes in response to increased informativeness, and thus quantify the cost savings from contract redesign.

This paper addresses this open question. We consider the standard setting of risk

[1] In contrast, Brookman and Thistle (2013) find that luck is a relatively unimportant determinant of managerial pay, compared to skill and labor market conditions.

neutrality and limited liability on the agent, which allows us to take an optimal contracting approach. These restrictions lead to optimal contracts that we observe in practice – as shown by Innes (1990), the agent has a call option on the signal. A fall in the strike price increases the option's delta and thus the agent's effort incentives, but also augments the value of the option and thus his expected wage. Thus, the strike price is the minimum possible to satisfy the agent's incentive constraint.

We start by considering general signal distributions. We show that an increase in informativeness has two effects, each of which has a clear economic interpretation. First, ignoring the incentive constraint, a fall in volatility reduces the value of the option and thus the agent's expected wage: the direct effect. Second, the increase in precision changes the agent's incentives. The heart of the paper analyzes this incentive effect and shows how its direction depends on the model's underlying parameters.

The agent's effort incentives stem from the difference in value between two options – the (less valuable) option that he receives when he shirks ("option-when-shirking"), and the (more valuable) option that he receives when he works and improves the signal distribution ("option-when-working"). Changes in signal precision affect the values of these options differentially. If the option satisfies *increasing differences*, i.e. effort and precision are *complements* (an increase in precision raises the sensitivity of the option's value to effort), then a rise in informativeness augments effort incentives. The principal can thus increase the strike price of the option, i.e. reduce its delta, without violating the agent's incentive constraint. This strike price increase further reduces the expected wage, and reinforces the first, direct effect. In contrast, if the option satisfies *decreasing differences*, i.e. effort and precision are *substitutes*, an increase in informativeness weakens effort incentives, offsetting the first effect. In the limit, it fully offsets it, rendering the total benefit of precision zero. The key result from the general model is that we derive a simple condition, which holds for any signal distribution and is easy to verify, that governs whether the option satisfies increasing or decreasing differences and thus whether a rise in informativeness raises or lowers effort incentives.

We then consider signal distributions with a location parameter, i.e. where effort shifts the signal distribution rightwards but does not change its shape. Doing so allows us to relate whether the option satisfies increasing or decreasing differences to its initial strike price, which in turn depends on the severity of the agency problem: a severe agency problem (i.e. a high cost of effort) requires the principal to set a low strike price to induce effort. We can thus relate the effect of informativeness on incentives to the underlying agency problem.

When the initial strike price is below a threshold, i.e. the agency problem is sufficiently strong, then effort and informativeness are complements. The intuition is as follows. A decrease in informativeness, i.e. an increase in signal volatility that augments both tails of the distribution, raises the value of an option. The magnitude of the gain is thus increasing in the asymmetry of the option's payoff. For a low initial strike price, the signal distribution when the agent shirks is centered around the strike price. Thus, the option's payoff is highly asymmetric: high signal realizations lie to the right of the strike price and lead to the option being exercised, and low signal realizations lie to the left and lead to no exercise. As a result, when volatility increases, the agent benefits from the greater probability of very high signal realizations and does not lose from the greater probability of very low signal realizations, and so the value of the option-when-shirking rises significantly. Put differently, for a low initial strike price, the option-when-shirking is relatively close to at-the-money, and thus has a relatively high vega (sensitivity to volatility).

In contrast, if the agent works, the signal distribution lies largely to the right of the initial strike price, and so the payoff function is less asymmetric: most signal realizations lead to the option being exercised. Thus, the agent benefits less from increases in volatility – he gains from the growth in the right tail, but loses from the growth in the left tail. Put differently, for a low initial strike price, the option-when-working is significantly in-the-money, and thus has a low vega. Thus, a fall in informativeness increases the value of the option-when-shirking more than the option-when-working, and lowers the agent's incentives. Intuitively, when volatility is high, the agent's effort incentives are weak because, even if he shirked, he would still earn a high wage if he received a positive shock. The agent is not worried about the fact that, if he shirks and receives a negative shock, the signal will be very low, because his payoff cannot fall below zero due to limited liability.

When the initial strike price is above a second (higher) threshold, i.e. the agency problem is sufficiently weak, then effort and informativeness are substitutes due to the reverse intuition. The option-when-shirking is deeply out-of-the-money, and the option-when-working is closer to at-the-money. Thus, the vega of the latter option is greater, and its value increases with volatility faster than the option-when-shirking, raising incentives. Intuitively, when the strike price is high, the agent will only receive a positive wage if he exerts effort *and* receives a sufficiently positive shock. When volatility rises, such shocks are more likely, and so the agent is more likely to be paid if he works. Thus, his effort incentives increase.

For initial strike prices between the two thresholds, effort and informativeness can either be complements or substitutes. This is because, for general distributions, a decrease in informativeness may not have a consistent effect on the signal distribution. While it is clear that it increases the left and right tails, thus leading to unambiguous results for low or high initial strike prices, a fall in informativeness can have any effect on the center of the distribution. For example, it could shift some mass *away* from a tail, as long as it also moves mass towards a more extreme tail point. Under a simple regularity condition which guarantees that decreases in informativeness consistently shift mass from the center of the distribution to the tails, the two thresholds now coincide at a single point and there is no ambiguous intermediate region. A sufficient (although not necessary) condition for regularity is that the signal distribution has a scale as well as a location parameter – i.e. can be characterized by its mean and standard deviation, as with the Normal, uniform, and logistic distributions. Intuitively, when volatility can be fully characterized by a scale parameter, changes in this parameter have a consistent effect on the shape of the distribution, moving mass towards its tails.

Under regular distributions, the effect of informativeness on incentives depends on whether the initial strike price of the option is above or below a single threshold. Thus, when incentives are strong (weak) to begin with, e.g. for CEOs (rank-and-file workers), an increase in informativeness further increases (reduces) incentives, amplifying (lowering) the gains from informativeness. In contrast, an analysis focusing only on the direct effect of informativeness, and ignoring the incentive constraint, would suggest that the gains from informativeness are highest when the option is at-the-money – i.e. a moderate initial strike price and a moderate agency problem.

In addition to studying whether a firm should endogenously choose to increase informativeness, our analysis also investigates the impact of exogenous changes in informativeness. An exogenous increase in volatility (see Gormley, Matsa, and Milbourn (2013) and DeAngelis, Grullon, and Michenaud (2013) for natural experiments) will increase (reduce) the effort incentives of agents with out-of-the-money (in-the-money) options. Thus, if firms recontract in response to these exogenous shocks, firms with in-the-money options should increase their CEOs' incentives relative to firms with out-of-the-money options, either by granting additional options, or reducing the strike price of new grants or existing options.[2]

For tractability, the analysis features a binary effort level. In the continuous-effort analog, in order to implement a given effort level, the contract must ensure that the agent will not deviate to a slightly lower or a slightly higher effort level (i.e. the incentive constraint will be "local"). This situation resembles a binary model in which the low effort level is very close to the (implemented) high effort level. In this case, the threshold for the initial strike price – that determines whether informativeness increases or decreases effort – is the expected value of the signal. If the initial strike price is above (below) this threshold, increases in informativeness lower (raise) the strike price towards the threshold. Thus, improvements in informativeness (e.g. increases in stock market efficiency) move the strike price closer to the expected value of the signal, and thus closer to at-the-money. Bebchuk and Fried (2004) argue that the almost universal practice of granting at-the-money options is suboptimal and that out-of-the-money options are more effective because the agent is paid only if performance is very high (see also Rappaport (1999)). Our analysis suggests that at-the-money options can be close to optimal if informativeness is high. This result also suggests that accounting or taxation considerations that favor at-the-money options need not induce suboptimal contracting.

Dittmann, Maug, and Spalt (2013) also consider the incentive constraint when assessing the benefits of a specific form of increased informativeness – indexing stock and options – and similarly show that indexation may weaken incentives. They use a quite different setting, which reflects the different aims of each paper. Their primary goal is to calibrate real-life contracts, and so their model incorporates risk aversion to allow them to input risk aversion parameters into the calibration. However, under risk aversion, it is difficult to solve for the optimal contract. They therefore restrict the contract to comprising salary, stock, and options, and hold stock constant when changing the contract to restore the agent's incentives upon indexation. They acknowledge that the actual savings from indexation will be different if the principal uses an initially optimal contract and responds optimally to changes in incentives. In contrast, our primary goal is theoretical. We incorporate risk neutrality and limited liability, allowing us to take an optimal contracting approach and to achieve analytical solutions. In addition,

[2] Acharya, John, and Sundaram (2000) also study the repricing of stock options theoretically, although in response to changes in the mean rather than volatility of the signal. Brenner, Sundaram, and Yermack (2000) analyze repricing empirically.

our model allows the analysis of reductions in volatility through other means than indexation, for example investing in a superior monitoring technology.

Other explanations for pay-for-luck have been proposed in the literature, partially reviewed by Edmans and Gabaix (2009). Oyer (2004) shows that pay-for-luck may be optimal if the value of workers' outside options vary with economic conditions and if re-contracting is costly. Raith (2008) shows that it may be preferable to base compensation on measures of output rather than input when the agent has private information on the production technology. Axelson and Baliga (2009) argue that, for contracts to be renegotiation-proof, the manager must have private information that causes him to have a different view from the board on the value of his long-term pay. Industry performance is an example of such information, and so it may be efficient not to filter it out. Gopalan, Milbourn, and Song (2010) show that tying the CEO's pay to industry performance induces him to choose the firm's industry exposure correctly.

This paper proceeds as follows. Section 1 presents the model. Section 2 shows that the optimal contract takes the form of a call option. Section 3 derives the gains from a reduction in the variance of the performance measure. Section 4 concludes. Appendix A contains all proofs not in the main text.

1 The Model

We consider a standard principal-agent model with risk neutrality and limited liability, similar to Innes (1990). The timing is as follows. At time $t = -1$, the principal (firm) offers a contract W to the agent (manager). At $t = 0$, the agent chooses his effort level $e \in \{0, \overline{e}\}$. Effort of $e = 0$ is of zero cost to the agent, and $e = \overline{e}$ costs him $C > 0$. We will sometimes refer to $e = \overline{e}$ as "high effort" or "working", and $e = 0$ as "low effort" or "shirking".

At $t = 1$, the agent's contribution to firm value ("output") q is realized. As in the literature on performance measurement (e.g. Baker (1992)), output is not contractible: for example, it is difficult to measure an employee's contribution independently of his colleagues'. Instead, contracts can depend on a performance measure ("signal")

$$s = q + \eta,$$

where η is a mean-zero random variable that is uncorrelated with effort: $\mathbb{E}[\eta|e] = 0$.

For example, η may be a market or industry shock, the contribution of other workers, or measurement error.

We assume that output q is not contractible and the contract depends on a separate signal s, so that we can change signal precision without affecting output volatility. However, the model allows for the case in which output is contractible, which corresponds to the degenerate distribution concentrated at $\eta = 0$ (i.e. signal equals output). Our results also hold in the case in which output equals the signal, and so the analysis considers changes in the variance of output.

Conditional on effort e, the signal s is continuously distributed on the real line according to the probability density function ("PDF") $f_\theta(s|e)$, although it needs not have full support over the real line. Let $F_\theta(s|e)$ denote the cumulative distribution function ("CDF") of s. A high signal is good news about effort in the sense of the strict monotone likelihood ratio property ("MLRP"). Formally, for all θ and for all signals s_1 and s_0 with $s_1 > s_0$,

$$\frac{f_\theta(s_1|\bar{e})}{f_\theta(s_1|0)} > \frac{f_\theta(s_0|\bar{e})}{f_\theta(s_0|0)}.$$

Strict MLRP implies that the distribution of the signal is ordered according to strict first-order stochastic dominance ("FOSD"): $F_\theta(s|0) > F_\theta(s|\bar{e}) \; \forall \; s, \theta$. Combined with $\mathbb{E}[\eta|e] = 0$, the above also implies that effort improves output in terms of FOSD.

The real-valued parameter θ, which lies in an interval Θ, captures the informativeness or precision of the signal, and orders the distributions in terms of second-order stochastic dominance. Formally, the mean of the signal is independent of θ, and

$$\theta \geq \theta' \implies \int_{-\infty}^{t} F_\theta(s|e)\, ds \leq \int_{-\infty}^{t} F_{\theta'}(s|e)\, ds, \tag{1}$$

$\forall \; t \in (-\infty, \infty)$. Thus, increases in θ generate more precise signal distributions in the sense of mean-preserving spreads.

Our analysis solves for the optimal contract for each given level of precision θ. This approach applies to settings in which signal precision depends on exogenous forces (such as technological change); the analysis derives empirical predictions on how these changes affect the optimal contract. Our approach can easily be extended to settings in which the principal can choose the level of precision θ at a cost $\kappa(\theta)$. Under the interpretation that η arises from measurement error, removing the shock corresponds to improving the monitoring technology at cost $\kappa(\theta)$. For example, Cornelli, Kominek, and Ljungqvist (2013) show that boards of directors engage in extensive (and thus costly) monitoring to gather soft information on the CEO's competence, strategic choice, and effort. Under the interpretation that η is a market or industry shock, increasing θ corresponds to relative performance evaluation (RPE), in which case the cost $\kappa(\theta)$ stems from two sources. First, it can arise from the literal cost of implementing RPE. While the actual calculation of industry performance, given a peer group, is relatively costless, the determination of the peer group may involve the hiring of compensation consultants. Second, the cost can also represent the loss of the benefits of pay-for-luck highlighted by prior work, e.g. Oyer (2004), Raith (2008), Axelson and Baliga (2009), and Gopalan, Milbourn, and Song (2010).

The discount rate is normalized to zero. Given a contract $W(\cdot)$ and a level of effort e, the agent's expected wage is

$$\mathbb{E}\left[W(s)\,|e\right] = \int_{-\infty}^{\infty} W(s)\, f_\theta(s|e)\, ds.$$

The agent is risk-neutral and so maximizes his expected wage, less the cost of effort. His reservation utility is zero. The principal is also risk-neutral and chooses a contract $W(\cdot)$ and an effort level e to maximize expected output $\mathbb{E}[q]$ less the expected wage $\mathbb{E}[W]$.

Following Innes (1990), we make two assumptions on the set of feasible contracts. First, the agent is protected by limited liability: $W(s) \geq 0 \;\forall\; s$. Second, pay-performance sensitivity lies between 0 and 1:

$$W(s+\epsilon) \geq W(s) \text{ and } s+\epsilon - W(s+\epsilon) \geq s - W(s)$$

for all s and all $\epsilon \geq 0$. These constraints must be satisfied if the agent can freely borrow to artificially increase output and the principal can freely destroy output. If the constraint on the left did not hold, the agent would artificially increase output, increasing the signal and thus his payoff. If the constraint on the right did not hold, the principal would exercise her control rights to "burn" output, reducing the signal and increasing her payoff. These constraints can be expressed as

$$1 \geq \frac{W(s+\epsilon) - W(s)}{\epsilon} \geq 0 \tag{2}$$

$\forall\ \epsilon$. It thus follows that $W(\cdot)$ is Lipschitz continuous and, therefore, differentiable almost everywhere. Hence, without loss of generality, we can assume that the contract $W(\cdot)$ is a cadlag function satisfying $0 \leq W'(s) \leq 1$ at all points of differentiability.[3]

In the first best, effort is verifiable. There is no incentive constraint and only a participation constraint. If the principal wishes to induce high effort, the participation constraint is given by:

$$\mathbb{E}[W(s)|\bar{e}] - C \geq 0. \tag{3}$$

To satisfy (3), the principal pays an expected wage $\mathbb{E}[W(s)|\bar{e}]$ that equals the agent's cost of effort C. Thus, if

$$\mathbb{E}[q|\bar{e}] - \mathbb{E}[q|0] > C, \tag{4}$$

high effort is optimal for the principal. We assume (4) throughout, else even under the first-best, the principal would not want to induce effort.

In the second best, the agent's effort is unverifiable and so the contract must satisfy an incentive constraint. The agent will exert effort if and only if:

$$\mathbb{E}[W(s)|\bar{e}] - \mathbb{E}[W(s)|0] \geq C. \tag{5}$$

Following standard arguments, this incentive constraint will bind, in which case the participation constraint will be slack and can be ignored in the analysis that follows.

We define X_θ implicitly by

$$\int_{X_\theta}^{\infty} (s - X_\theta) [f_\theta(s|\bar{e}) - f_\theta(s|0)] \, ds = C. \tag{6}$$

We will show in Lemma 1 that X_θ exists and is unique. The intuition behind (6) is that, if the agent is given a call option on s, X_θ is the strike price such that working increases the value of the agent's option by an amount equal to the cost of effort, so that the incentive constraint is satisfied with equality.

[3] A cadlag function is everywhere right-continuous and has left limits everywhere.

We make the following assumption to ensure that $e = \overline{e}$ is second-best optimal:

$$\mathbb{E}\left[q|\overline{e}\right] - \mathbb{E}\left[q|0\right] - \int_{X_\theta}^{\infty} (s - X_\theta) f_\theta(s|\overline{e}) ds \geq 0. \tag{7}$$

The first term, $\mathbb{E}\left[q|\overline{e}\right] - \mathbb{E}\left[q|0\right]$, is the benefit to the principal of inducing $e = \overline{e}$. The second term is the cost of the contract required to do so. If (7) did not hold, the principal would allow the agent to shirk, in which case the problem would be trivial and the contract would involve a constant wage.

Given informativeness θ, the principal's problem is to choose a contract $W(\cdot)$ to minimize the expected wage $\mathbb{E}\left[W(s)|\overline{e}\right]$ subject to the agent's incentive and limited liability constraints, plus the constraints on the slope of the contract:

$$\mathbb{E}\left[W(s)|\overline{e}\right] \geq \mathbb{E}\left[W(s)|0\right] + C, \tag{8}$$

$$0 \leq W(s) \ \forall\, s, \quad \text{and} \tag{9}$$

$$0 \leq W'(s) \leq 1 \text{ at all points of differentiability of } W. \tag{10}$$

Our contracting problem is the dual to the one in Innes (1990). In his model, the agent (entrepreneur) chooses a financing contract subject to his own incentive constraint and the participation constraint of the principal (investor). Here, the principal (firm) chooses an employment contract subject to the incentive and participation constraints of the agent (manager). In both models, the optimal contract has the same form; the only difference is in the division of the rents. Since Innes studies a financing setting, the optimal contract for the principal is debt. Thus, the agent has equity, which is a call option on the firm's assets; here, we will show that the agent receives a call option on the signal.

Another difference is that Innes features a continuous action set. His focus was to derive the form of the optimal contract and thus he wishes to do so in the most general setting. Our goal is different: given that the optimal contract is a call option, we study how changes in informativeness affect the agent's incentives and thus the strike price. We thus specialize to a binary effort level. With a continuous effort level, a change in informativeness may alter the optimal effort level. It is well known that solving for the optimal effort level in addition to the cheapest contract that induces a given effort level is extremely complex (see, e.g., Grossman and Hart (1983)), and thus many papers focus on the implementation of a given effort level. For example, Biais et al. (2010) consider a binary effort choice and conditions to guarantee that high effort is optimal, as in our setting, and Dittmann and Maug (2007) and Dittmann, Maug, and Spalt (2010, 2013) consider the implementation of a given effort level out of a continuum. (Indeed, Innes (1990) does not solve for the optimal effort level or study how it is affected by the parameters of the setting, but shows that an optimum exists.) Edmans and Gabaix (2011) show that, if the benefits of effort are multiplicative in firm size and the firm is sufficiently large, it is always optimal for the principal to implement the highest effort level and so the optimal effort level is indeed fixed. We thus consider a binary effort setting where high effort is optimal.

2 The Optimal Contract

This section solves for the optimal contract for a given level of informativeness θ. The analysis is similar to Innes (1990). Our main results will come in Section 3, which analyzes the gains from increasing θ.

Let $W_\theta(\cdot)$ denote the optimal contract for a given θ. Lemma 1 establishes that $W_\theta(\cdot)$ is a call option on s, where the strike price X_θ is chosen to satisfy the incentive constraint (6).

Lemma 1 *(Optimal contract): For each given θ, there exists a unique optimal contract, characterized by $e = \overline{e}$, and*

$$W_\theta(s) = \max\{0, s - X_\theta\}, \tag{11}$$

where X_θ is determined by the unique solution of (6).

The setting is slightly different from Innes (1990), since the principal is contracting on a signal rather than output. We show that the Innes (1990) result of the optimality of a call option extends to this case, and the intuition is the same. The absolute value of the likelihood ratio is highest in the tails of the distribution of s, so the signal is most informative about effort in the tails. The left tail cannot be used for incentive purposes due to limited liability, and so incentives are concentrated in the right tail. This maximizes the likelihood that positive payments are received by a working agent. With an upper bound on the slope, the optimal contract involves call options on s with the maximum feasible slope, i.e. $W'(s) = 1$.

Lemma 2 below shows that the strike price falls with the cost of effort, which parameterizes the severity of the agency problem.

Lemma 2 *(Effect of effort cost on strike price): Let X_θ be the strike price in the optimal contract for a given θ. Then, X_θ is strictly decreasing in the cost of effort C.*

3 The Value of Informativeness

This section calculates the gains from increasing informativeness. Section 3.1 considers general signal distributions and provides a condition under which increases in informativeness raise effort incentives. Section 3.2 studies distributions with a location parameter and shows that whether this condition is satisfied depends on the initial strike price and thus the severity of the agency problem. Section 3.3 graphically illustrates the benefits of informativeness for the Normal distribution. It also proves analytically that, for this distribution, the benefits from informativeness are monotonically increasing in the cost of effort, and thus monotonically decreasing in the initial strike price.

3.1 General Distributions

The total effect of increasing informativeness on the expected wage can be decomposed as follows:

$$\frac{d}{d\theta}\mathbb{E}\left[W(s)|\overline{e}\right] = \underbrace{\frac{\partial}{\partial\theta}\mathbb{E}\left[W(s)|\overline{e}\right]}_{\textit{direct effect}} + \underbrace{\frac{\partial}{\partial X_\theta}\mathbb{E}\left[W(s)|\overline{e}\right]\frac{dX_\theta}{d\theta}}_{\textit{incentive effect}}. \tag{12}$$

The first component is the *direct effect*, $\frac{\partial}{\partial\theta}\mathbb{E}\left[W(s)|\overline{e}\right]$. Holding constant the strike price, an increase in signal precision changes the value of the option; we will later prove that this effect is negative. This reduction in the cost of compensation is the benefit of informativeness highlighted by Bebchuk and Fried (2004) in their argument that the lack of RPE is inefficient. In the Holmstrom (1979) setting of a risk-averse agent, an increase in informativeness reduces the risk borne by the agent and thus allows the principal to lower the expected wage without violating the agent's participation constraint. In our setting of risk neutrality and limited liability, an increase in precision reduces the value of the option.

The second component is the *incentive effect*, $\frac{\partial}{\partial X_\theta}\mathbb{E}\left[W(s)|\overline{e}\right]\frac{dX_\theta}{d\theta}$, which arises because the increase in precision requires the strike price to rise by $\frac{dX_\theta}{d\theta}$ to maintain incentive compatibility. $\frac{\partial}{\partial X_\theta}\mathbb{E}\left[W(s)|\overline{e}\right]$ is negative – any increase in the strike price reduces the value of the option – but the sign of $\frac{dX_\theta}{d\theta}$ is unclear. We thus seek to derive conditions under which an increase in precision raises or lowers the optimal strike price. The following definition will be useful:

Definition 1 *(Increasing differences): Let* $(\Theta, E) \subseteq \mathbb{R}^2$. *A function* $g(\theta, e) : \Theta \times E \rightarrow \mathbb{R}$ *satisfies increasing differences if, for all* $\theta_L < \theta_H$ *and* $e_L \leq e_H$,

$$g(\theta_H, e_H) - g(\theta_H, e_L) \geq g(\theta_L, e_H) - g(\theta_L, e_L). \tag{13}$$

It satisfies decreasing differences if $-g$ *satisfies increasing differences.* θ *and* e *are* complements *if* $g(\theta, e)$ *satisfies increasing differences,* substitutes *if* $g(\theta, e)$ *satisfies decreasing differences, and* neutral *if* $g(\theta, e)$ *satisfies both increasing and decreasing differences.*

The increasing differences condition (13) means that the incremental gain (i.e., increase in g) from effort, $g(\theta, e_H) - g(\theta, e_L)$, is increasing in θ: effort and informativeness are complements in terms of their effect on g. Conversely, decreasing differences means that the incremental gain from effort is decreasing in θ. Thus, effort and informativeness are substitutes. Indeed, increasing (decreasing) differences is the most common definition of complementarity (substitutability).[4] In our setting, if g is differentiable, increasing differences is equivalent to the single-crossing condition:

$$\frac{\partial g}{\partial \theta}(\theta, \bar{e}) - \frac{\partial g}{\partial \theta}(\theta, 0)\, ds \geq 0.$$

We are concerned with how changes in precision affect the incentive constraint (5). The agent's incentives stem from the fact that exerting effort increases the value of his option. If he works, his option is worth $\mathbb{E}\left[W(s)|\overline{e}\right]$; we refer to this as an "option-when-working." If he shirks, he receives an "option-when-shirking" worth $\mathbb{E}\left[W(s)|0\right]$. His effort incentives are given by the difference in the values of these options, i.e.

$$\mathbb{E}\left[W(s)|\overline{e}\right] - \mathbb{E}\left[W(s)|0\right]. \tag{14}$$

[4] There is a very large literature using these concepts for understanding outcomes of games – e.g. Bulow, Geanakoplos, and Klemperer (1985) and Milgrom and Roberts (1990).

Since a change in precision θ affects the option-when-working and the option-when-shirking to different degrees, it affects the agent's effort incentives (14). When precision and effort are complements, i.e. $\mathbb{E}[W(s)|e]$ satisfies increasing differences, increases in precision augment the agent's effort incentives:

$$\frac{\partial}{\partial\theta}\left\{\mathbb{E}\left[W\left(s\right)|\bar{e}\right]-\mathbb{E}\left[W\left(s\right)|0\right]\right\}>0, \tag{15}$$

We thus wish to understand the conditions under which $\mathbb{E}[W(s)|e]$ satisfies increasing differences. We do so by using integration by parts (see Appendix A) to rewrite the agent's expected wage as

$$\mathbb{E}\left[W(s)|e\right]=\mathbb{E}\left[s|e\right]-X_\theta+\int_{-\infty}^{X_\theta}F_\theta\left(s|e\right)ds. \tag{16}$$

The area under the CDF for signals below X_θ (the third term) is the value of a put option with a strike price of X_θ:

$$\Pr\left(s<X_\theta|e\right)\mathbb{E}\left[\left(X_\theta-s\right)|s<X_\theta,e\right]=\int_{-\infty}^{X_\theta}-\left(s-X_\theta\right)f(s|e)ds=\int_{-\infty}^{X_\theta}F_\theta\left(s|e\right)ds,$$

where the last equality follows from integration by parts. Therefore, equation (16) can be interpreted as the put-call parity equation. The agent's call option equals the expected value of the signal, minus the strike price, plus the value of a put option. Let $\pi_X\left(\theta,e\right)\equiv\int_{-\infty}^{X}F_\theta\left(s|e\right)ds$ denote the value of a put option with a strike price of X. By second-order stochastic dominance (equation (1)), the value of the put option is decreasing in θ ($\frac{\partial\pi_{X_\theta}}{\partial\theta}\left(\theta,e\right)\leq 0$).[5]

To study whether $\mathbb{E}[W(s)|e]$ satisfies increasing differences, we examine each of the three terms on the right-hand side ("RHS") of (16). While $\mathbb{E}[s|e]$ depends on e, it is independent of θ since changes in θ represent mean-preserving spreads. In addition, X_θ depends on θ but not e. Thus, θ and e are neutral in their effect on both of these terms, and non-neutral only in their effect on the third term $\int_{-\infty}^{X_\theta}F_\theta\left(s|e\right)ds$. This observation leads to the following Lemma:

[5] In the Black-Scholes model, we have a strict inequality. This is because the Black-Scholes model assumes a lognormal distribution for stock returns, so an increase in precision (which corresponds to a decrease in volatility) affects the whole distribution. However, in our setting with a general distribution, a change in θ may only affect the part of the distribution above the strike price X_θ, where the put option has zero payoff, and so its value does not change.

Lemma 3 *(Condition for increasing differences): The agent's expected pay* $\mathbb{E}[W(s)|e]$ *satisfies increasing differences if and only if the area under the CDF,* $\int_{-\infty}^{X_\theta} F_\theta(s|e)\,ds$, *satisfies increasing differences.*

The usefulness of Lemma 3 lies in the fact that, while the value of the call option (16) involves several terms, it is relatively easy to verify whether the area under the CDF $\int_{-\infty}^{X_\theta} F_\theta(s|e)\,ds$ satisfies increasing differences. While it may seem intuitive that the value of the call option satisfies increasing differences if and only if the value of the put option satisfies increasing differences, the value of Lemma 3 is that we can check whether expected pay satisfies increasing differences by studying a single term $\int_{-\infty}^{X_\theta} F_\theta(s|e)\,ds$ – not that this term can be interpreted as the value of a put option. The condition in Lemma 3 is simple to check and general: it holds for all signal distributions that satisfy MLRP (so that the optimal contract is a call option). Definition 1 then allows us to determine the effect of informativeness on the strike price X_θ:

Proposition 1 *(Effect of informativeness on strike price): Let* $\pi_X(\theta, e) \equiv \int_{-\infty}^{X} F_\theta(s|e)\,ds$. *The optimal strike price* X_θ *is increasing in informativeness* θ *if* $\pi_X(\theta, e)$ *satisfies increasing differences at* X_θ, *decreasing in informativeness if* $\pi_X(\theta, e)$ *satisfies decreasing differences at* X_θ, *and constant if it satisfies both increasing and decreasing differences at* X_θ.

When precision and effort are complements, exerting effort augments the value of the call option by a greater amount when precision is high. As a result, the agent's marginal benefit from effort (equation (14)) is increasing in informativeness. Increases in precision loosen the incentive constraint and allow the principal to increase the strike price while still inducing effort. Thus, in addition to the direct benefit of informativeness (it reduces the expected wage, holding constant the strike price X_θ), the principal further benefits from its incentive effect (it allows the strike price X_θ to increase, further reducing the expected wage). Proposition 1 in turn leads to Corollary 1 below.

Corollary 1 *(Partial and total effects of informativeness on expected wage):*

$$\left|\frac{d}{d\theta}\mathbb{E}[W(s)|\bar{e}]\right| > \left|\frac{\partial}{\partial\theta}\mathbb{E}[W(s)|\bar{e}]\right| \qquad \text{if and only if } \frac{dX_\theta}{d\theta} > 0. \tag{17}$$

Proof. From equation (12), we have:

$$\frac{d}{d\theta}\mathbb{E}\left[W(s)|\bar{e}\right] = \underbrace{\frac{\partial}{\partial\theta}\mathbb{E}\left[W(s)|\bar{e}\right]}_{\textit{direct effect}} + \underbrace{\frac{\partial}{\partial X_\theta}\mathbb{E}\left[W(s)|\bar{e}\right]\frac{dX_\theta}{d\theta}}_{\textit{incentive effect}}$$
$$= \frac{\partial\pi_X(\theta,e)}{\partial\theta} - \left[1 - \frac{\partial\pi_X(\theta,e)}{\partial X_\theta}\right]\frac{dX_\theta}{d\theta} = \frac{\partial\pi_X(\theta,e)}{\partial\theta} - \left[1 - F_\theta\left(X_\theta|e\right)\right]\frac{dX_\theta}{d\theta}.$$

$\frac{d}{d\theta}\mathbb{E}\left[W(s)|\bar{e}\right]$ and $\frac{\partial}{\partial\theta}\mathbb{E}\left[W(s)|\bar{e}\right]$ are both negative, and the former is more negative (i.e. its absolute value is higher) if and only if $\frac{dX_\theta}{d\theta} > 0$, i.e. effort and precision are complements. ■

The direct effect, $\frac{\partial}{\partial\theta}\mathbb{E}\left[W(s)|\bar{e}\right]$, is negative. An increase in precision decreases the value of the put option ($\frac{\partial\pi_{X_\theta}}{\partial\theta} \leq 0$) and thus the expected wage. Turning to the incentive effect, higher precision augments the strike price by $\frac{dX_\theta}{d\theta}$, which in turn requires the principal to pay an additional $\frac{dX_\theta}{d\theta}$ dollars whenever the price exceeds X_θ, which occurs with probability $1 - F_\theta(X_\theta|e)$. The sign of $\frac{dX_\theta}{d\theta}$ in turn depends on whether informativeness and effort are substitutes or complements. When they are complements, then $\frac{dX_\theta}{d\theta} > 0$. The strike price increases, further reducing the expected wage and reinforcing the direct effect. When they are substitutes, then $\frac{dX_\theta}{d\theta} < 0$, opposing the direct effect.

Even when $\frac{dX_\theta}{d\theta} < 0$ and the incentive effect counteracts the direct effect, it can never outweigh it. The total effect $\frac{d}{d\theta}\mathbb{E}\left[W(s)|\bar{e}\right]$ is always weakly negative, i.e. increasing precision weakly reduces the expected wage. This result arises from revealed preference. If reducing precision reduced the expected wage, the principal would have added in randomness to the contract, and so the initial contract would not have been optimal. Even though the incentive effect cannot outweigh the direct effect, it is still important to consider as it affects the optimal level of precision θ that the principal should choose, since increasing precision is costly. Indeed, it is possible that the incentive effect exactly offsets the direct effect, and so that the total gains from informativeness equal exactly zero: see Appendix B for an example.[6]

[6]While we consider the effect of changes in signal informativeness, Chaigneau, Edmans, and Gottlieb (2014) derive a general condition for whether the addition of a new signal has strictly positive value for contracting.

3.2 Distributions in the Location Family

Section 3.1 shows that, with general distributions, the effect of precision on incentives depends on whether effort and precision are complements or substitutes. We now gradually add more structure to the signal distribution, which allows us to relate whether we have complements or substitutes – and thus the effect of precision on the agent's incentives – to the underlying parameters of the agency problem.

Since the expected signal is strictly increasing in effort, we can always normalize effort to be measured in units of the average signal: $e = \mathbb{E}[s|e]$. Hence, for any signal distribution, we can write the signal as $s = e + \varepsilon$, where we define the noise term ε as the difference between effort and the signal. Thus, if $G_\theta(\varepsilon)$ denotes the CDF of the noise, we have $G_\theta(\varepsilon) = G_\theta(s - e) = F_\theta(s|e)$. While the conditional mean of noise is zero ($\mathbb{E}[\varepsilon|e] = 0$), higher moments of ε may, in general, depend on the effort e. We now assume that the whole distribution of noise G_θ (not only its first moment) is independent of e. This assumption is equivalent to specifying that the distribution of s has a location parameter – changes in effort shift the location of the distribution but do not change its shape.

To ensure that $\lim_{\varepsilon \searrow -\infty} G_\theta(\varepsilon)$ and $\lim_{\varepsilon \nearrow \infty} G_\theta(\varepsilon)$ are differentiable with respect to θ, we make the technical assumptions that G_θ is differentiable with respect to θ and that the sequences $\{G_\theta(-n)\}_{n\in\mathbb{N}}$, $\left\{\frac{\partial G_\theta}{\partial\theta}(-n)\right\}_{n\in\mathbb{N}}$, $\{G_\theta(n)\}_{n\in\mathbb{N}}$, $\left\{\frac{\partial G_\theta}{\partial\theta}(n)\right\}_{n\in\mathbb{N}}$ are uniformly convergent. These assumptions are automatically satisfied if the noise has bounded support and are also satisfied under standard unbounded distributions (such as the Normal, logistic, Cauchy, and Laplace distributions).

With a slight abuse of terminology, we say that s belongs to a *location family* if the two above conditions – independence between e and θ and the technical differentiability condition – are satisfied. When the signal belongs to the location family, we can show that whether effort and informativeness are complements or substitutes depends on the level of the initial strike price:

Lemma 4 *(Effect of informativeness on the strike price): Suppose that the signal distribution belongs to a location family. Then there exist $\widehat{X}_1$ and $\widehat{X}_2 \geq \widehat{X}_1$ such that $\frac{dX_\theta}{d\theta} \geq 0$ if $X_\theta < \widehat{X}_1$ and $\frac{dX_\theta}{d\theta} \leq 0$ if $X_\theta > \widehat{X}_2$.*

Effort and precision are complements when the initial strike price X_θ is below a threshold $\widehat{X}_1$, substitutes when X_θ exceeds a higher threshold $\widehat{X}_2$, and may be either complements or substitutes for $\widehat{X}_1 \leq X_\theta \leq \widehat{X}_2$.

The intuition is as follows. Recall from inequality (15) that the effect of informativeness on effort depends on its differential effects on the values of the option-when-working and the option-when-shirking. We thus study how changes in informativeness affect the value of each option.

A decrease in informativeness (from θ to θ') increases both the left and right tails of the signal distribution. If the initial strike price of the option is sufficiently low ($X_\theta < \widehat{X}_1$), then the signal distribution upon shirking lies on both sides of X_θ. The agent benefits from high signal realizations ($s > X_\theta$), since the option-when-shirking is in-the-money ("ITM") and so he exercises it, but does not lose from low signals ($s < X_\theta$) as he does not exercise the option. Thus, when informativeness falls, a shirking agent benefits from the growth in the right tail, but does not lose from the growth in the left tail, and so the value of the option-when-shirking increases significantly.

Since the signal distribution has a location parameter, working shifts it rightwards. Thus, for $X_\theta < \widehat{X}_1$, the signal distribution upon working is mostly to the right of X_θ, and remains this way after informativeness falls. Since the option usually ends up ITM, the agent usually exercises it. Thus, a working agent benefits from the growth in the right tail *and* loses from the growth in the left tail, and so the value of the option-when-working is relatively unchanged. Put differently, reductions in informativeness increase the value of an option due to its asymmetric payoff: the agent benefits from $s > X_\theta$, but does not lose from $s < X_\theta$. When X_θ is low and the agent shirks, the mean signal 0 is close to the kink X_θ and the agent benefits from the asymmetry. When the agent works, the mean signal $\overline{e}$ is far from the kink X_θ, and so he enjoys little asymmetry. Overall, a fall in θ raises the value of the option-when-shirking more than the option-when-working and reduces effort incentives: effort and informativeness are complements. In simple language, the agent thinks "I'm not going to bother working hard, because even if I do, I might be unlucky and so profits will be low. I might as well shirk, because even if I get unlucky and profits become very low, that doesn't matter, because I can't get paid less than zero no matter how low profits get."

The existence of a location parameter is central to the above intuition, as it means that effort shifts the signal distribution rightward and reduces the probability of very low signals ($s < X_\theta$). If, in contrast, effort increased the dispersion of the signal in addition to its mean, it could increase the probability of very low signals. Thus, the effect on effort incentives would be ambiguous.

For a sufficiently high strike price ($X_\theta > \widehat{X}_2$), the signal distribution upon shirking is mostly to the left of X_θ – and remains this way even after informativeness falls and the right tail expands. Thus, the option-when-shirking remains usually OTM and its value is little changed. In contrast, if the agent works, this shifts the signal distribution rightward and so decreases in informativeness now push the right tail above X_θ. Thus, when informativeness falls, a working agent benefits from the growth in the right tail (since he can now exercise the option) but does not lose from the growth in the left tail (since he does not exercise the option). Put differently, when X_θ is high and the agent works, the expected signal $\bar{e}$ is close to the kink X_θ and the agent benefits from the asymmetry. When the agent shirks, the expected signal 0 is far from the kink X_θ, and so he enjoys little asymmetry. Overall, a fall in informativeness raises the value of the option-when-working more than the option-when-shirking, and increases the agent's effort incentives: effort and informativeness are substitutes. In simple language, the agent thinks "If informativeness were high, I wouldn't bother working because the target X_θ is so high that I wouldn't meet it, even if I did work. But, now that the signal is more noisy, I will work – because if I do, and I get lucky, I'll meet the target."

From Lemma 2, the initial strike price X_θ is decreasing in the cost of effort, and thus the severity of the agency problem. When the agency problem is mild (severe), the initial strike price is high (low); increases in informativeness reduce (increase) effort incentives, causing the strike price to fall (rise).

However, without any restrictions on the distribution, it is unclear how changes in θ affect the tails of the distribution between $\widehat{X}_1$ and $\widehat{X}_2$. The source of the ambiguity is that, for arbitrary distributions, decreases in precision in a SOSD sense need not consistently shift mass from the center of the distribution towards the tails. It may be that a fall in θ shifts some mass away from a tail, as long as it also moves mass towards a more extreme tail point. Figure 1 shows that, while a fall in θ increases $G_\theta(s)$ for low s below a threshold s_1 (i.e. increases the left tail) and increases $1 - G_\theta(s)$ for high s above a threshold $s_2 > s_1$ (i.e. increases the right tail), the effect of θ on $G_\theta(s)$ is unclear for intermediate s. The CDFs G_θ and $G_{\theta'}$ could cross many times between $\widehat{X}_1$ and $\widehat{X}_2$.

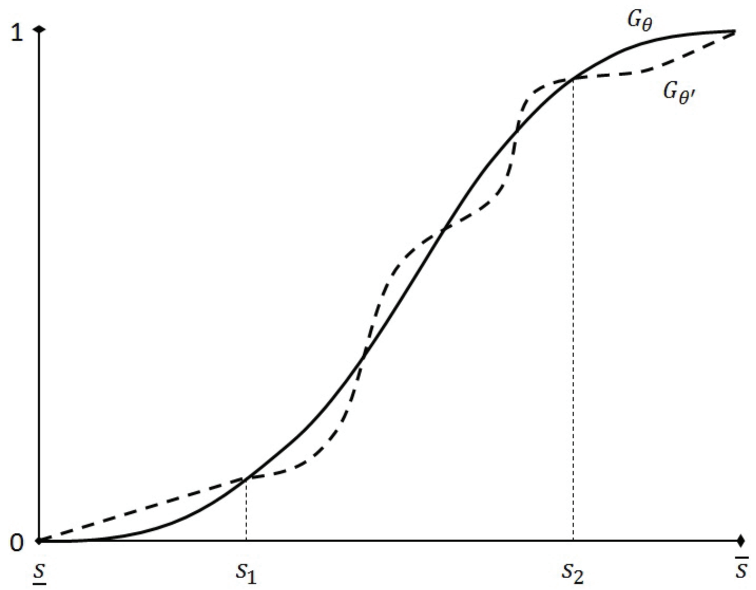

Figure 1: Distributions in the location family.

Definition 2 below introduces a simple regularity condition that guarantees that falls in informativeness have a "consistent" effect on the distribution – they shift mass towards the tails away from the center. Proposition 2 shows that, for regular noise distributions, $\widehat{X}_1 = \widehat{X}_2 = \widehat{X}$, and so there is a single threshold below (above) which effort and precision are complements (substitutes): the CDFs cross at a single point $\widehat{X}$, as shown in Figure 2. There is no intermediate range in which changes in informativeness have an ambiguous effect.

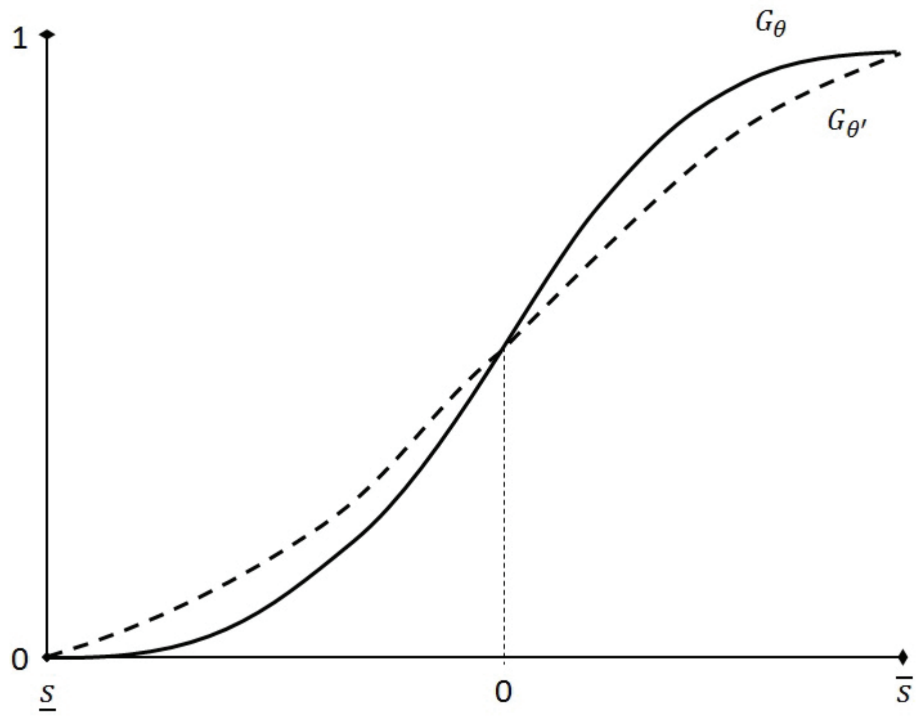

Figure 2: Regular signal distributions.

Definition 2 *The distribution G_θ is regular if there exists $\widehat{X}$ such that*

$$s\left\{\begin{array}{c}<\\>\end{array}\right\}\widehat{X}\implies\frac{\partial G_\theta}{\partial\theta}(s)\left\{\begin{array}{c}\leq\\\geq\end{array}\right\}0.$$

Proposition 2 *(Effect of informativeness with regular distributions): Suppose that the signal distribution F belongs to a location family and that the noise distribution G_θ is regular. Then there exists $\widehat{X}$ such that $\frac{dX_\theta}{d\theta}\geq 0$ if $X_\theta<\widehat{X}$, and $\frac{dX_\theta}{d\theta}\leq 0$ if $X_\theta>\widehat{X}$.*

Regularity is not automatically implied by SOSD, but is satisfied by most standard distributions. Indeed, Corollary 2 shows that regularity is satisfied by any signal distribution that has a scale parameter (in addition to a location parameter).

Corollary 2 *(Distributions with a scale parameter): If the signal distribution F has a scale parameter, i.e. its CDF and PDF can be written as $F_\sigma(s|e) = G\left(\frac{s-e}{\sigma}\right)$ and $f_\sigma(s|e) = \frac{1}{\sigma}g\left(\frac{s-e}{\sigma}\right)$, then the noise distribution G is regular and so there exists $\widehat{X}$ such that $\frac{dX_\theta}{d\theta} \geq 0$ if $X_\theta < \widehat{X}$, and $\frac{dX_\theta}{d\theta} \leq 0$ if $X_\theta > \widehat{X}$.*

A distribution with a location and scale parameter can be fully characterized by its mean e and standard deviation σ. Since the volatility of a signal is the inverse of its precision, we have $\sigma = \frac{1}{\sqrt{\theta}}$ and so:

$$\frac{\partial}{\partial\sigma}G\left(\frac{s-e}{\sigma}\right) = -\frac{s-e}{\sigma^2}g\left(\frac{s-e}{\sigma}\right)\begin{cases} < 0 & \text{if } s > e \\ > 0 & \text{if } s < e \end{cases} \tag{18}$$

as required by Definition 2. Intuitively, the existence of a scale parameter means that informativeness is characterized by this parameter, and so changes in this parameter have a consistent effect on the shape of the distribution, moving mass towards its tails, thus satisfying the regularity condition.

While regularity guarantees a single cutoff $\widehat{X}$, for general distributions we do not know where this cutoff lies. Indeed, Claim 1 in Appendix C shows that, for distributions with a scale parameter, $\widehat{X}$ may lie anywhere between 0 and $\bar{e}$. Proposition 3 shows that, when the distribution is not only regular (for which a scale parameter is sufficient but not necessary) but also symmetric (as with the Normal, uniform, logistic, Cauchy, and Laplace distributions), $\widehat{X}$ lies half-way between 0 and $\bar{e}$, i.e. $\widehat{X} = \frac{\bar{e}}{2}$, as is intuitive. Thus, we can compare the initial strike price, which depends on the underlying parameters of the agency problem (see Lemma 2) to the threshold $\frac{\bar{e}}{2}$. Hence, we can relate whether effort and informativeness are complements or substitutes to model primitives.

Proposition 3 *(Symmetric regular distributions): Suppose that the signal distribution F belongs to a location family and that the noise distribution G_θ is regular and symmetric. Then, $\frac{dX_\theta}{d\theta} \geq 0$ if $X_\theta < \widehat{X}$, and $\frac{dX_\theta}{d\theta} \leq 0$ if $X_\theta > \widehat{X}$, where $\widehat{X} \equiv \frac{\bar{e}}{2}$.*

In addition to being sufficient for regularity, the presence of a scale parameter also clarifies the intuition because we can fully parameterize changes in precision by changes in volatility σ. We can thus examine how changes in σ affect the values of the two options using the familiar concept of the option "vega": the sensitivity of its value to

σ. In turn, the vega of each option will depend on the initial strike price, and thus the underlying parameters of the agency problem. With a scale parameter, equation (15) now becomes

$$\frac{\partial}{\partial\sigma}\left\{\mathbb{E}\left[W\left(s\right)|\bar{e}\right]-\mathbb{E}\left[W\left(s\right)|0\right]\right\}<0. \tag{19}$$

The left-hand side ("LHS") of inequality (19) – the vega of the option-when-working minus the vega of the option-when-shirking – represents the effect of changes in σ on incentives. The vega of an option is always positive, highest for an at-the-money ("ATM") option (see Claim 2 in Appendix C[7]), and declines when the option moves either ITM or OTM. Thus, the vega of the option-when-working is highest at $X=\bar{e}$, and so if it has a strike price of $\widehat{X}=\frac{\bar{e}}{2}$, then it is ITM by $\frac{\bar{e}}{2}$. The vega of the option-when-shirking is highest at $X=0$, and so if it has a strike price of $\widehat{X}=\frac{\bar{e}}{2}$, then it is OTM by $\frac{\bar{e}}{2}$. Overall, at a strike price of $\widehat{X}=\frac{\bar{e}}{2}$, both options are equally away-from-the-money and have the same vega (see Claim 3 in Appendix C), and so effort incentives are independent of σ. We thus have $\frac{dX_\sigma}{d\sigma}=0$ for $X=\widehat{X}$.

When $X<\widehat{X}$, then $Y\left(0,X\right)$ is closer to being ATM than $Y\left(\bar{e},X\right)$, and so it has a higher vega. Thus, an increase in σ reduces effort incentives, and so $\frac{dX_\sigma}{d\sigma}<0$. When $X<\widehat{X}$, then $Y\left(\bar{e},X\right)$ is closer to being ATM than $Y\left(0,X\right)$, and so it has a higher vega. Thus, an increase in σ lowers effort incentives, and so $\frac{dX_\sigma}{d\sigma}>0$.

Proposition 3 implies that, for all symmetric regular distributions, regardless of the initial strike price X, improvements in informativeness draw X towards $\widehat{X}=\frac{\bar{e}}{2}$. When the initial strike price is high (low), a fall in volatility reduces (increases) effort incentives, causing the strike price to fall (rise). In the current discrete model, there are two effort levels, $\bar{e}$ and 0. In a continuous-effort analog, where the principal wishes to implement effort of $\bar{e}$, the contract must ensure induce the agent to exert effort of $\bar{e}$ rather than $\bar{e}+\varepsilon$ or $\bar{e}-\varepsilon$, i.e. the incentive constraint must be "local". In our discrete model, a local incentive constraint resembles the case in which the (implemented) high effort level ($\bar{e}$) is very close to the low effort level (0). If $\bar{e}\simeq 0$, then $\widehat{X}\simeq 0$. Moreover, since the contract induces the agent to exert effort of $\bar{e}$, the mean value of the signal is $\bar{e}$ and so an ATM option will have a strike price of $\bar{e}\simeq 0$. Thus, if the initial strike price is higher (lower) than $\widehat{X}\simeq 0$, improvements in informativeness (e.g. increases in stock market efficiency) will lower (raise) the optimal strike price towards 0, i.e.

[7]It is well-known that for lognormal distributions, the vega is highest for ATM options. Claim 2 in Appendix C extends this result to all distributions with a location and scale parameter.

bring the option closer to ATM. (Indeed, Appendix E sketches a continuous effort model which shows that increases in informativeness bring the option closer to ATM.) Bebchuk and Fried (2004) argue that the almost universal practice of granting ATM options is inefficient and that OTM options would be cheaper for the firm. Similarly, Rappaport (1999) advocates OTM options as they reward the agent only for exceptional performance. However, such views ignore the incentive effect: OTM options have lower deltas and so more would be required to achieve incentive compatibility. Murphy (2002) notes that ITM options would provide the strongest incentives, but that the tax code discourages such options. One interpretation is that the tax code leads to firms choosing ATM options when ITM options may be more efficient. Our analysis instead suggests that increases in informativeness lead to options optimally being close to ATM.[8]

In addition, Proposition 2 suggests that exogenous increases in informativeness will have different effects on the incentives of agents depending on the moneyness of their options. In particular, it will reduce (increase) the incentives of agents with OTM (ITM) options. Thus, firms may wish to reduce the strike prices of OTM options to restore incentives. Option repricing is documented empirically by Brenner, Sundaram, and Yermack (2000), although they do not study if it is prompted by falls in volatility. Acharya, John, and Sundaram (2000) also study the repricing of options theoretically, although in responses to changes in the mean rather than volatility of the signal.

Finally, note that the above analysis takes an optimal contracting approach, so the slope of the contract is the maximum possible without violating the constraint (10). We thus have $W'(s) = 1$: the agent is the residual "claimant" of any increase in the signal (as long as $s \geq X_\theta$). Thus, the principal changes X_θ to ensure that the incentive constraint binds. An alternative approach is to restrict the contract to comprising ATM options, e.g. for accounting or tax reasons[9], and instead meet the incentive constraint by varying the slope of the contract. Appendix D demonstrates an analogous result for this case. With ATM options, we have $X = \overline{e} \geq \widehat{X} = \frac{\overline{e}}{2}$ and so effort and precision are substitutes. An increase in informativeness requires the number of options granted to

[8] Hall and Murphy (2000) restrict the contract to consist of options, rather than taking an optimal contracting approach, and calibrate the optimal strike price depending on the CEO's risk aversion, the proportion of his wealth in stock, and the proportion of his wealth in options. They show that, in most cases, the range of optimal strike prices includes the current stock price, i.e. corresponds to an ATM option. In contrast, Dittmann and Yu (2011) feature a risk-taking as well as effort decision, and restrict the contract to consisting of fixed salary, stock, and options. They show that ITM options are typically optimal.

[9] Murphy (1999) documents that ATM options are almost universally granted.

increase to maintain incentive compatibility. This augments the expected wage, just like a decrease in the strike price, and so the total effect of informativeness on expected pay is less than the direct effect. Thus, the results of the core model, where $X > \widehat{X}$, extend to the case of ATM options.

3.3 Normal Distribution

We now demonstrate graphically the direct and incentive effects. We need to assume a specific distribution to enable us to calculate the derivatives, and so we consider the common case of a Normal distribution. Let φ and Φ denote the PDF and CDF of the standard Normal distribution, respectively. As we show in Appendix A, the total effect and the direct effect are respectively given by:

$$\frac{d\mathbb{E}\left[W\left(s\right)|\overline{e}\right]}{d\sigma} = \varphi\left(\frac{X_\sigma - \overline{e}}{\sigma}\right) - \left[1 - \Phi\left(\frac{X_\sigma - \overline{e}}{\sigma}\right)\right] \frac{\varphi\left(\frac{X_\sigma - \overline{e}}{\sigma}\right) - \varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right) - \Phi\left(\frac{X_\sigma - \overline{e}}{\sigma}\right)}, \text{ and} \tag{20}$$

$$\frac{\partial\mathbb{E}\left[W\left(s\right)|\overline{e}\right]}{\partial\sigma} = \varphi\left(\frac{X_\theta - \overline{e}}{\sigma}\right). \tag{21}$$

Figure 3 illustrates how these effects change as we vary the severity of the moral hazard problem (parameterized by the the cost of effort C). As is standard for graphs of option values, the figure contains the strike price X on the x-axis; since X is strictly decreasing in C (equation (31)), there is a one-to-one mapping between X and C.

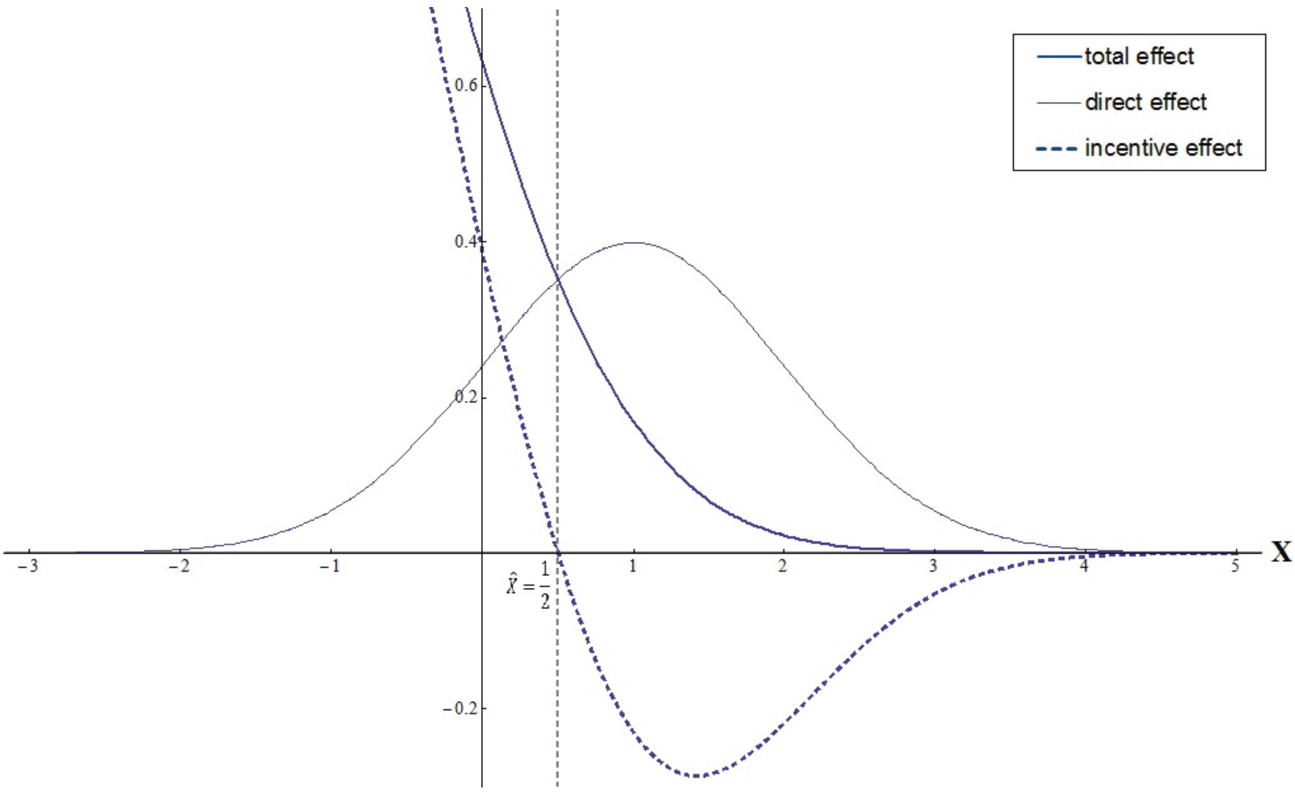

Figure 3: Total and partial derivative of expected pay with respect to σ for a range of values of X, for $\bar{e} = 1$ and $\sigma = 1$.

To understand Figure 3, recall from (12) that the total effect is given by $\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma} = \frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial\sigma} + \frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial X_\theta}\frac{dX_\theta}{d\sigma}$. The direct effect, $\frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial\sigma}$, tends to zero as the strike price approaches either $-\infty$ or ∞. The vega of an option is greatest when the option is ATM, i.e. $X = 1$. An ATM option benefits most from the asymmetry in an option's payoff: a high noise realization leads to a large increase in the option's payoff, but a low noise realization has no effect as the agent will not exercise the option.

The incentive effect, $\frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma}\frac{dX_\sigma}{d\sigma}$, consists of two components. The first is the change in strike price required to maintain incentive compatibility, $\frac{dX_\sigma}{d\sigma}$. From Proposition 3 and using $\sigma = \frac{1}{\sqrt{\theta}}$, $\frac{dX_\sigma}{d\sigma} > 0$ if and only if $X > \widehat{X} = \frac{1}{2}$. Indeed, for the Normal distribution, not only does $\frac{dX_\sigma}{d\sigma}$ turn from negative to positive as X crosses above $\widehat{X}$, but it is also monotonically increasing in X, i.e. monotonically decreasing in the cost of effort. This result is stated in Lemma 5 below:

Lemma 5 *(Normal distribution, change in strike price): Suppose ε is Normally distributed. Then, the benefits of informativeness are decreasing in the cost of effort, i.e.*

$$\frac{d^2 X_\sigma}{d\sigma dC} < 0. \tag{22}$$

The second is the change in the value of the option when the strike price increases, $\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma}$. This change is always negative, and so the sign of the incentive effect is the opposite of the sign of $\frac{dX_\sigma}{d\sigma}$: indeed, in Figure 3, the incentive effect is positive (negative) for $X < (>) \widehat{X}$. In addition, the magnitude of $\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma}$ is increasing in the moneyness of the option. Overall, as X falls below $\widehat{X}$, $\frac{dX_\sigma}{d\sigma}$ becomes increasingly negative (see Lemma 5), and the option becomes increasingly in the money so $\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma}$ also becomes increasingly negative (it falls towards -1). Thus, the overall incentive effect $\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma} \frac{dX_\sigma}{d\sigma}$ becomes monotonically more positive as X falls below $\widehat{X}$. However, as X rises above $\widehat{X}$, the two components of the incentive effect move in opposite directions. On the one hand, greater informativeness becomes increasingly detrimental to incentives ($\frac{dX_\sigma}{d\sigma}$ becomes more positive). On the other hand, $\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma}$ falls towards zero: when the option is deeply OTM, its value is small to begin with and thus little affected by changes in the strike price. Overall, the impact of X on the incentive effect is non-monotonic. As X initially rises above $\widehat{X}$, the incentive effect becomes increasingly negative as the option has significant value, and this value is affected by the change in the strike price required to maintain incentives ($\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma}$ is large). However, as X continues to rise, the option's value falls and so is little affected by the strike price ($\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial X_\sigma}$ is small). Thus, the incentive effect tends to zero.

The total effect $\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}$ combines these direct and incentive effects. While the direct effect is initially increasing in X, this is outweighed by the fact that the incentive effect is initially decreasing in X. Thus, in Figure 3, the total gains from increased informativeness are monotonically decreasing in X. Indeed, focusing on the Normal distribution allows us to prove this result analytically: C is the exogenous parameter that drives X, and Proposition 4 shows that the gains from informativeness are monotonically increasing in C.

Proposition 4 *(Normal distribution, effect of cost of effort on gains from informativeness) Suppose* ε *is Normally distributed. Then,* $\frac{d}{dC}\left\{\frac{d\mathbb{E}[W(s)|\overline{e}]}{d\sigma}\right\} > 0$.

An analysis focusing purely on the direct effect would suggest that the gains from informativeness are greatest when the initial option is ATM, which in turn corresponds to a moderate strike price and a moderate cost of effort. In contrast, considering the total effect (which incorporates the incentive effect) shows that the gains from informativeness are monotonically increasing in the severity of the agency problem. Thus, workers with high-powered incentives (such as CEOs) should be evaluated more precisely than those with low-powered incentives (such as rank-and-file workers).

Corollary 3 shows that, as the cost of effort goes to zero, the gains from informativeness also approach zero, as does the total gain relative to the direct effect.

Corollary 3 *(Normal distribution, limiting cases). Suppose* ε *is Normally distributed. Then,*

$$\frac{d\mathbb{E}\left[W\left(s\right)|\overline{e}\right]}{d\sigma} \longrightarrow_{C\rightarrow 0} 0 \qquad \textit{and} \qquad \frac{\frac{d\mathbb{E}[W(s)|\overline{e}])}{d\sigma}}{\frac{\partial\mathbb{E}[W(s)|\overline{e}]}{\partial\sigma}} \longrightarrow_{C\rightarrow 0} 0. \tag{23}$$

As the moral hazard problem becomes weaker, the total effect of informativeness becomes very small relative to the direct effect. Thus, ignoring the incentive effect and considering only the direct effect would substantially overestimate the gains from informativeness. Indeed, in Figure 3, the direct effect significantly overestimates the total gains for sufficiently large X. For example, for $\sigma = 1$ and $X = 2$ (which is only one standard deviation away from the expected performance of $\overline{e} = 1$), the gains from a marginal change in σ are 10.8 times larger with the direct effect than with the total effect. Thus, even for non-extreme parameter values, gains from improved informativeness can be much lower if the incentive effect is taken into account. This ratio becomes much greater for higher X, because the total benefits of informativeness fall towards zero.

4 Conclusion

This paper studies the principal's benefits from increasing the informativeness of the signal used to evaluate the agent. The direct effect is that higher signal precision reduces the value of the agent's option and thus expected pay. The core focus of the paper is on the indirect effect – how changes in precision affect the agent's incentives. By taking an optimal contracting approach, we can be specific on how the contract changes in response to increases in informativeness. With general signal distributions, we show that, if effort and informativeness are substitutes, increases in precision weaken the agent's incentives. Thus, the principal must reduce the strike price of the option to preserve effort incentives, increasing the cost of compensation and offsetting the direct effect. Indeed, we derive a simple condition that governs whether effort and informativeness are substitutes or complements, that holds for all distributions that satisfy the monotone likelihood ratio property.

Focusing on signal distributions with a location parameter allows us to relate whether effort and informativeness are substitutes or complements to the initial strike price of the option, and thus the severity of the underlying agency problem. When the initial strike price is above a threshold, i.e. incentives are low-powered to begin with, an increase in informativeness reduces the agent's effort incentives and thus the benefits of informativeness. The principal therefore optimally invests less in improving informativeness, potentially rationalizing the scarcity of relative performance evaluation for some agents in reality.

In contrast, if the initial strike price is below a second (lower) threshold, i.e. incentives are high-powered to begin with, an increase in informativeness augments the agent's effort incentives. This provides an additional gain from informativeness over and above the direct effect of reducing volatility traditionally focused upon. Thus, the gains from informativeness depend critically on the strength of incentives, and thus the magnitude of the moral hazard problem to begin with. For regular signal distributions, where an increase in informativeness consistently moves mass from the tails of the distribution to the center, both thresholds coincide at a single point. This regularity condition is satisfied by any distribution with a location and scale parameter, such as the Normal, uniform, and logistic distributions. Furthermore, with a Normal distribution, the benefits from informativeness are monotonically increasing in the cost of effort, and thus the severity of the agency problem. Finally, if incentive constraints are local, i.e. the implemented effort level is close to other feasible effort levels, then increases in informativeness cause the strike price to move closer to at-the-money.

References

[1] Acharya, Viral V., Kose John, and Rangarajan K. Sundaram (2000): "On the optimality of resetting executive stock options." *Journal of Financial Economics* 57, 65-101.

[2] Aggarwal, Rajesh K. and Andrew A. Samwick (1999): "Executive compensation, relative performance evaluation, and strategic competition: theory and evidence." *Journal of Finance* 54, 1999–2043.

[3] Axelson, Ulf and Sandeep Baliga (2009): "Liquidity and manipulation of executive compensation schemes." *Review of Financial Studies* 22, 3907–3939.

[4] Baker, George (1992): "Incentive contracts and performance measurement." *Journal of Political Economy* 100, 598–614.

[5] Bebchuk, Lucian and Jesse Fried (2004): *Pay without performance: the unfulfilled promise of executive compensation.* (Harvard University Press, Cambridge).

[6] Bertrand, Marianne and Sendhil Mullainathan (2001): "Are CEOs rewarded for luck? The ones without principals are." *Quarterly Journal of Economics* 116, 901–932.

[7] Biais, Bruno, Thomas Mariotti, Jean-Charles Rochet, and Stéphane Villeneuve (2010): "Large risks, limited liability and dynamic moral hazard." *Econometrica* 78, 73–118.

[8] Brenner, Menachem, Rangarajan K. Sundaram, and David Yermack (2000): "Altering the terms of executive stock options." *Journal of Financial Economics* 57, 103–128.

[9] Brookman, Jeffrey T. and Paul D. Thistle (2013): "Managerial compensation: Luck, skill, or labor markets?" *Journal of Corporate Finance* 21, 252–268.

[10] Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer (1985): "Multimarket oligopoly: Strategic substitutes and complements." *Journal of Political Economy* 93, 488–511.

[11] Chaigneau, Pierre, Alex Edmans and Daniel Gottlieb (2014): "The informativeness principle under limited liability." NBER Working Paper No. 20456.

[12] Cornelli, Francesca, Zbigniew Kominek, and Alexander Ljungqvist (2013): "Monitoring managers: Does it matter?" *Journal of Finance* 68, 431–481.

[13] De Angelis, David, Gustavo Grullon, and Sébastien Michenaud (2013): "Downside risk and the design of CEO incentives: Evidence from a natural experiment." Working Paper, Rice University.

[14] Dittmann, Ingolf and Ernst Maug (2007): "Lower salaries and no options? On the optimal structure of executive pay." *Journal of Finance* 62, 303–343.

[15] Dittmann, Ingolf, Ernst Maug, and Oliver G. Spalt (2010): "Sticks or carrots? On the optimal structure of executive pay." *Journal of Finance* 65, 2015–2050.

[16] Dittmann, Ingolf, Ernst Maug, and Oliver G. Spalt (2013): "Indexing executive compensation contracts." *Review of Financial Studies* 26, 3182–3224.

[17] Dittmann, Ingolf and Ko-Chia Yu (2011): "How important are risk-taking incentives in executive compensation?" Working Paper, Erasmus University Rotterdam.

[18] Edmans, Alex and Xavier Gabaix (2009): "Is CEO pay really inefficient? A survey of new optimal contracting theories." *European Financial Management* 15, 486–496.

[19] Edmans, Alex and Xavier Gabaix (2011): "Tractability in Incentive Contracting." *Review of Financial Studies* 24, 2865–2894.

[20] Gjesdal, Froystein (1982): "Information and incentives: the agency information problem." *Review of Economic Studies* 49, 373–390.

[21] Gopalan, Radhakrishnan, Todd T. Milbourn and Fenghua Song (2010): "Strategic flexibility and the optimality of pay for sector performance." *Review of Financial Studies* 23, 2060–2098.

[22] Gormley, Todd A., David A. Matsa, and Todd T. Milbourn (2013): "CEO compensation and corporate risk-taking: Evidence from a natural experiment." *Journal of Accounting and Economics* 56, 79–101.

[23] Grossman, Sanford J., and Oliver D. Hart (1983): "An analysis of the principal-agent problem." *Econometrica* 51, 7–45.

[24] Hall, Brian J. and Kevin J. Murphy (2000): "Optimal exercise prices for executive stock options." *American Economic Review* 90, 209–214.

[25] Holmstrom, Bengt (1979): "Moral hazard and observability." *Bell Journal of Economics* 10, 74–91.

[26] Innes, Robert D. (1990): "Limited liability and incentive contracting with ex-ante action choices." *Journal of Economic Theory* 52, 45–67.

[27] Jenter, Dirk and Fadi Kanaan (2013): "CEO turnover and relative performance evaluation." *Journal of Finance*, forthcoming.

[28] Kim, Son Ku (1995): "Efficiency of an information system in an agency model." *Econometrica* 63, 89–102.

[29] Milgrom, Paul and John Roberts (1990): "Rationalizability, learning, and equilibrium in games with strategic complementarities." *Econometrica* 58, 1255–1277.

[30] Murphy, Kevin J. (1999): "Executive compensation" in *Handbook of Labor Economics*, O. Ashenfelter and D. Card, eds. Elsevier/North-Holland.

[31] Murphy, Kevin J. (2002): "Explaining executive compensation: Managerial power versus the perceived cost of stock options." *University of Chicago Law Review* 69, 847–869.

[32] Oyer, Paul (2004): "Why do firms use incentives that have no incentive effects?" *Journal of Finance* 59, 1619–1649.

[33] Raith, Michael (2008): "Specific knowledge and performance measurement." *Rand Journal of Economics* 39, 1059–1079.

[34] Rappaport, Alfred (1999): "New thinking on how to link executive pay with performance." *Harvard Business Review* 72, 91–101

[35] Shavell, Steven (1979): "Risk sharing and incentives in the principal and agent relationship." *Bell Journal of Economics* 10, 55–73.

[36] Zang, Israel (1980): "A smoothing-out technique for min-max optimization." *Mathematical Programming* 19, 61–77.

A Proofs

Proof of Lemma 1

Denote the lower bound of the support of s by $\underline{s}$, and the upper bound by $\overline{s}$. We will adopt a two-step approach. First, we solve for the optimal contract for a fixed minimum payment $W_\theta(\underline{s}) = Z \geq 0$. As we will show, the solution involves an "option" with a fixed minimum payment Z. Then, we will show that the floor is zero.

Formally, for a given constant $Z \geq 0$, consider the following "relaxed" program:

$$\min_{W_\theta(\cdot)} \int_{\underline{s}}^{\overline{s}} W_\theta(s) f_\theta(s|\overline{e})\, ds$$

subject to

$$\int_{\underline{s}}^{\overline{s}} W_\theta(s) \left[f_\theta(s|\overline{e}) - f_\theta(s|0)\right] ds \geq C,$$

$$0 \leq \dot{W}_\theta(s) \leq 1, \text{ and } W_\theta(\underline{s}) = Z \text{ fixed.}$$

Note that the monotonicity condition $\dot{W}_\theta(s) \geq 0$ implies that $Z \geq 0$ is both necessary and sufficient for the limited liability constraint to hold.

Introduce the auxiliary variables $y(s) \equiv \dot{W}_\theta(s)$ and set up the Hamiltonian:

$$H(W, y, \lambda, \mu, s) \equiv -W_\theta(s) f_\theta(s|\overline{e}) + \lambda\left[W_\theta(s)\left[f_\theta(s|\overline{e}) - f_\theta(s|0)\right] - C\right] + \mu(s)\, y(s),$$

where W_θ are state variables, y are control variables, μ are co-state variables, and λ is a (state-independent) Lagrange multiplier. The necessary optimality conditions are:

$$y(s) \in \arg\max_{0\leq y\leq 1} \mu(s)\, y \;\therefore\; y(s) = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} \text{ if } \mu(s) \begin{Bmatrix} < \\ > \end{Bmatrix} 0; \tag{24}$$

$$\frac{\partial H}{\partial W_\theta} = -\dot{\mu} \;\therefore\; f_\theta(s|\overline{e}) - \lambda\left[f_\theta(s|\overline{e}) - f_\theta(s|0)\right] = \dot{\mu}(s)\,; \tag{25}$$

and the transversality condition $\mu(\overline{s}) = 0$.

Condition (25) yields:

$$\dot{\mu}(s) > 0 \iff \frac{1}{\lambda} > 1 - \frac{f_\theta(s|0)}{f_\theta(s|\overline{e})} \iff \frac{1}{LR(s)} > \frac{\lambda - 1}{\lambda},$$

where $LR(s) \equiv \frac{f_\theta(s|\overline{e})}{f_\theta(s|0)}$ is the likelihood ratio, which we assumed to be increasing. Thus, the LHS of the last inequality above is decreasing in s while the RHS is constant. Hence, there exists a threshold $s_\theta^* \in [\underline{s}, \overline{s}]$ such that $\dot{\mu}(s) > 0$ for $s < s_\theta^*$ and $\dot{\mu}(s) < 0$ for $s > s_\theta^*$. (Notice that if $s_\theta^* = \underline{s}$ or $s_\theta^* = \overline{s}$, one of these intervals vanishes). Therefore, μ is bell-shaped, with a unique maximum at s_θ^* and (at most) two local minima – one at $\underline{s}$ and and another at $\overline{s}$.

We claim that $s_\theta^* < \overline{s}$, i.e. μ is increasing $\forall\ s$ if $s_\theta^* = \overline{s}$. Then, the transversality condition $\mu(\overline{s}) = 0$ and condition (24) implies that $W_\theta(s)$ is constant ($W_\theta(\overline{s}) = W_\theta(\underline{s}) = Z$ $\forall\ s$), which violates the incentive constraint. Thus, it cannot be a solution.

There are two possible solutions depending on whether $\mu(\underline{s}) \geq 0$ or $\mu(\underline{s}) < 0$. In the former case, we must have $\mu(s) > 0\ \forall\ s \in (\underline{s}, \overline{s})$ (since the only candidates for global minima are $\underline{s}$ and $\overline{s}$ and $\mu(\underline{s}) \geq 0 = \mu(\overline{s})$). In the latter case, there exists a threshold $s_\theta^{**} \in (\underline{s}, s_\theta^*)$ such that $\mu(s) < 0$ if $s < s_\theta^{**}$ and $\mu(s) > 0$ if $s > s_\theta^{**}$. We can combine both cases by letting $s_\theta^{**} \in [\underline{s}, s_\theta^*)$ denote the threshold below which we have $\mu(s) < 0$. The solution is then

$$W_\theta^*(s) = Z + \max\{s - s_\theta^{**},\ 0\} \tag{26}$$

for some $s_\theta^{**} \in [\underline{s}, \overline{s})$. This concludes the first part of the proof.

It remains to be shown that the solution entails $Z = 0$. Substitute the agent's payment from (26) in the principal's expected cost

$$\int_{\underline{s}}^{\overline{s}} W(s) f_\theta(s|\overline{e})ds = Z + \int_{s_\theta^{**}}^{\overline{s}} (s - s_\theta^{**}) f_\theta(s|\overline{e})ds,$$

and in the incentive constraint

$$\int_{s_\theta^{**}}^{\overline{s}} (s - s_\theta^{**}) \left[f_\theta(s|\overline{e}) - f_\theta(s|0)\right] ds \geq C. \tag{27}$$

Note that monotonicity is automatically satisfied under (26) and, as before, the limited liability constraint holds if and only if $Z \geq 0$. The optimal contract must solve:

$$\min_{Z \geq 0,\ s_\theta^{**} \in [\underline{s}, \overline{s}]} Z + \int_{s_\theta^{**}}^{\overline{s}} (s - s_\theta^{**}) f_\theta(s|\overline{e})ds \tag{28}$$

subject to (27). Since Z increases the expression in (28) but does not affect the incentive constraint in (27), the solution entails $Z = 0$. Thus, the solution of the relaxed program is

$$W_{\theta}^{*}(s) = \max\left\{s - s_{\theta}^{**},\ 0\right\},$$

With this contract, and setting $X_{\theta} = s_{\theta}^{**}$, we show in the proof of Lemma 2 that the LHS of the incentive constraint (5) becomes $\int_{X_{\theta}}^{\overline{s}} \left[F_{\theta}(s|0) - F_{\theta}(s|\overline{e})\right] ds$. By FOSD, the LHS of (27) is decreasing in X_{θ}. Since both the expected cost to the principal and the LHS of the incentive constraint are strictly decreasing in X_{θ}, the optimal level of X_{θ} satisfies the incentive constraint with equality:

$$\int_{X_{\theta}}^{\overline{s}} (s - X_{\theta}) \left[f_{\theta}(s|\overline{e}) - f_{\theta}(s|0)\right] ds = C.$$

Finally, we establish that X_{θ} exists. Evaluated at $X = \overline{s}$, the LHS of the incentive constraint is $0 < C$. Evaluated at $X = \underline{s}$, it equals

$$\int_{\underline{s}}^{\overline{s}} \left[F_{\theta}(s|0) - F_{\theta}(s|\overline{e})\right] ds = \int_{\underline{s}}^{\overline{s}} s f_{\theta}(s|\overline{e})\, ds - \int_{\underline{s}}^{\overline{s}} s f_{\theta}(s|0)\, ds, \tag{29}$$

using the same integration by parts as in the proof of Lemma 2. Since high effort is first-best efficient, the expression in (29) exceeds C. Thus, the intermediate value theorem and the monotonicity of the LHS ensure that a unique solution to (30) exists.

In sum, we have established that the optimal contract is

$$W_{\theta}(s) = \max\{s - X_{\theta}, 0\},$$

where the strike price X_{θ} is such that the incentive constraint holds with equality.

Proof of Lemma 2

Denoting the upper bound of the support of s by $\overline{s}$, we first show that the incentive constraint (6) can also be rewritten as

$$\int_{X_{\theta}}^{\overline{s}} \left[F_{\theta}(s|0) - F_{\theta}(s|\overline{e})\right] ds = C. \tag{30}$$

Opening the expressions inside the brackets in equation (5), we obtain:

$$\int_{X_\theta}^{\overline{s}} sf_\theta\left(s|\overline{e}\right)ds - \left[1 - F_\theta\left(X_\theta|\overline{e}\right)\right]X_\theta = \int_{X_\theta}^{\overline{s}} sf_\theta\left(s|0\right)ds - \left[1 - F_\theta\left(X_\theta|0\right)\right]X_\theta + C,$$

which yields

$$\int_{X_\theta}^{\overline{s}} sf_\theta\left(s|\overline{e}\right)ds - \int_{X_\theta}^{\overline{s}} sf_\theta\left(s|0\right)ds = \left[F_\theta\left(X_\theta|0\right) - F_\theta\left(X_\theta|\overline{e}\right)\right]X_\theta + C.$$

Applying integration by parts (for $e \in \{0, \overline{e}\}$), gives

$$\int_{X_\theta}^{\overline{s}} sf_\theta\left(s|e\right)ds = \left[sF_\theta\left(s|e\right) - \int F_\theta\left(s|e\right)ds\right]_{X_\theta}^{\overline{s}} = \overline{s} - X_\theta F_\theta\left(X_\theta|e\right) - \int_{X_\theta}^{\overline{s}} F_\theta\left(s|e\right)ds.$$

Plugging back in the previous expression, we obtain

$$\left[\overline{s} - X_\theta F_\theta\left(X_\theta|\overline{e}\right) - \int_{X_\theta}^{\overline{s}} F_\theta\left(s|\overline{e}\right)ds\right] - \left[\overline{s} - X_\theta F_\theta\left(X_\theta|0\right) - \int_{X_\theta}^{\overline{s}} F_\theta\left(s|0\right)ds\right]$$

$$= \left[F_\theta\left(X_\theta|0\right) - F_\theta\left(X_\theta|\overline{e}\right)\right]X_\theta + C.$$

Canceling terms gives equation (30). Applying the implicit function theorem to (30) yields:

$$\frac{dX_\theta}{dC} = -\frac{1}{F\left(X_\theta|0\right) - F\left(X_\theta|\overline{e}\right)} < 0. \tag{31}$$

Proof of Equation (16)

Denoting the lower bound of the support of s by $\underline{s}$, the agent's expected pay in case of effort e is

$$\int_X^{+\infty}\left(s - X\right)f_\theta\left(s|e\right)ds = \int_X^{+\infty} sf_\theta\left(s|e\right)ds - X\left[1 - F_\theta\left(X|e\right)\right]$$

$$= \int_{\underline{s}}^{+\infty} sf_\theta\left(s|e\right)ds - X\left[1 - F_\theta\left(X|e\right)\right] - \int_{\underline{s}}^{X} sf_\theta\left(s|e\right)ds.$$

Applying integration by parts, we have:

$$\int_{\underline{s}}^{X} s f_\theta(s|e)\, ds = \left[s F_\theta(s|e) - \int F_\theta(s|e)\, ds \right]_{\underline{s}}^{X} = X F_\theta(X|e) - \int_{\underline{s}}^{X} F_\theta(s|e)\, ds.$$

Substituting this into the previous equation yields:

$$\int_{X}^{+\infty} (s - X) f_\theta(s|e)\, ds = \int_{\underline{s}}^{+\infty} s f_\theta(s|e)\, ds - X + \int_{\underline{s}}^{X} F_\theta(s|e)\, ds = \mathbb{E}[s|e] - X + \int_{\underline{s}}^{X} F_\theta(s|e)\, ds.$$

Proof of Lemma 3

Denote the lower bound of the support of s by $\underline{s}$, and the upper bound by $\overline{s}$. The agent's expected pay in the case of effort e is given by

$$\mathbb{E}[W(s)|e] = \int_{X_\theta}^{\overline{s}} (s - X_\theta) f_\theta(s|e)\, ds. \tag{32}$$

Integration by parts yields:

$$\int_{X_\theta}^{\overline{s}} s f_\theta(s|e)\, ds = \overline{s} - X F_\theta(s|e) - \int_{X_\theta}^{\overline{s}} F_\theta(s|e)\, ds$$

and so (32) can be rewritten:

$$\begin{aligned} \mathbb{E}[W(s)|e] &= \overline{s} - X F_\theta(s|e) - \int_{X_\theta}^{\overline{s}} F_\theta(s|e)\, ds - X[1 - F_\theta(s|e)] \\ &= \overline{s} - X - \int_{X_\theta}^{\overline{s}} F_\theta(s|e)\, ds. \end{aligned} \tag{33}$$

Since $\overline{s}$ and X are not functions of either θ or e, it follows that the agent's expected pay satisfies increasing differences if and only if $\int_{X_\theta}^{\overline{s}} F_\theta(s|e)\, ds$ satisfies decreasing differences (and vice-versa). The following Lemma will be useful for this and future proofs:

Lemma 6 *For any θ, $\int_{\underline{s}}^{\overline{s}} F_\theta(s|e)\, ds = \overline{s} - \mathbb{E}[s|e]$, which is not a function of θ.*

Proof. Applying integration by parts to $\int_{\underline{s}}^{\overline{s}} sf_\theta(q|e)\,ds$, we obtain:

$$\int_{\underline{s}}^{\overline{s}} sf_\theta(q|e)\,ds = \left[sF_\theta(q|e) - \int F_\theta(s|e)\,ds\right]_{\underline{s}}^{\overline{s}} = \overline{s} - \int_{\underline{s}}^{\overline{s}} F_\theta(s|e)\,ds.$$

Since θ parameterizes mean-preserving spreads, the expression on the LHS, $\mathbb{E}[s|e]$ is not a function of θ. ■

From Lemma 6, $\int_{X_\theta}^{\overline{s}} F_\theta(s|e)\,ds$ satisfies decreasing differences if and only if $\int_{\underline{s}}^{X_\theta} F_\theta(s|e)\,ds$ satisfies increasing differences (since their sum is independent of θ by second-order stochastic dominance). Thus, the agent's expected pay satisfies increasing differences if and only if $\int_{\underline{s}}^{X_\theta} F_\theta(s|e)\,ds$ satisfies increasing differences.

Proof of Proposition 1

Applying the implicit function theorem to equation (30) gives:

$$\frac{dX_\theta}{d\theta} = \frac{\int_X^\infty \frac{\partial}{\partial\theta}\left[F_\theta(q|0) - F_\theta(q|\overline{e})\right]dq}{F_\theta(X_\theta|0) - F_\theta(X_\theta|\overline{e})}.$$

By FOSD, the denominator is positive. Lemma 6 yields:

$$\frac{\partial}{\partial\theta}\int_X^\infty F_\theta(q|e)dq = -\frac{\partial}{\partial\theta}\int_{-\infty}^X F_\theta(q|e)dq$$

$\forall\ X, \theta, e$. Plugging back:

$$\frac{dX_\theta}{d\theta} = \frac{\int_{-\infty}^X \frac{\partial}{\partial\theta}\left[F_\theta(q|\overline{e}) - F_\theta(q|0)\right]dq}{F_\theta(X_\theta|0) - F_\theta(X_\theta|\overline{e})}.$$

It is straightforward to show that if $\int_{-\infty}^X F_\theta(q|e)\,dq$ satisfies increasing (decreasing) differences and is differentiable with respect to θ, then

$$\frac{\partial}{\partial\theta}\int_{-\infty}^X \left[F_\theta(q|\overline{e}) - F_\theta(q|0)\right]dq \geq (\leq) 0 \tag{34}$$

$\forall\ \overline{e} > 0$. The denominator is positive by FOSD.

Proof of Lemma 4

From Lemma 3, effort and informativeness are complements if and only if

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{X_\theta}\left[F_\theta(s|\overline{e})-F_\theta(s|0)\right]ds\geq 0, \tag{35}$$

i.e. the single-crossing condition holds. Since $F_\theta(s|e)=G_\theta(s-e)$, we have

$$\begin{aligned}\frac{\partial}{\partial\theta}\int_{-\infty}^{X_\theta}\left[F_\theta(s|\overline{e})-F_\theta(s|0)\right]ds &= \frac{\partial}{\partial\theta}\left\{\int_{-\infty}^{X_\theta}G_\theta(s-\overline{e})\,ds-\int_{-\infty}^{X_\theta}G_\theta(s)\,ds\right\}\\ =\frac{\partial}{\partial\theta}\left\{\int_{-\infty}^{X_\theta-\overline{e}}G_\theta(s)\,ds-\int_{-\infty}^{X_\theta}G_\theta(s)\,ds\right\} &= \frac{\partial}{\partial\theta}\left\{-\int_{X_\theta-\overline{e}}^{X_\theta}G_\theta(s)\,ds\right\}\\ &= -\int_{X_\theta-\overline{e}}^{X_\theta}\frac{\partial G_\theta}{\partial\theta}(s)\,ds. \end{aligned}\tag{36}$$

Therefore, effort and informativeness are complements if and only if (36) >0.

Let $\xi(\theta)\equiv\lim_{\epsilon\searrow-\infty}G_\theta(\epsilon)=0$. Since ξ is differentiable at $-\infty$, it follows that $\xi'(\theta)=0$. Similarly, $\hat{\xi}(\theta)\equiv\lim_{\epsilon\nearrow+\infty}G_\theta(\epsilon)=1$ and the differentiability of $\hat{\xi}$ at ∞ implies that $\hat{\xi}'(\theta)=0$. Moreover, it is straightforward to show that SOSD implies[10]

$$\int_{-\infty}^{X_\theta}\frac{\partial G_\theta}{\partial\theta}(s-e)\,ds\leq 0 \tag{37}$$

$\forall\ X_\theta$. Thus, $\frac{\partial G_\theta}{\partial\theta}\leq 0$ for s small enough. As a result, there exists $\widehat{X}_1$ such that $\int_{\widehat{X}_1-\overline{e}}^{\widehat{X}_1}\frac{\partial G_\theta}{\partial\theta}(s)\,ds<0$. Thus, (36) >0 and so effort and informativeness are complements.

In addition, $\frac{\partial G_\theta}{\partial\theta}=0$ for $s\to\infty$. Thus, $\frac{\partial G_\theta}{\partial\theta}$ must eventually turn positive: $\frac{\partial G_\theta}{\partial\theta}\geq 0$ for s large enough. As a result, there exists $\widehat{X}_2$ such that $\int_{\widehat{X}_2-\overline{e}}^{\widehat{X}_2}\frac{\partial G_\theta}{\partial\theta}(s)\,ds>0$. Thus, (36) <0 and so effort and informativeness are substitutes. In sum, there exists $\widehat{X}_1$ such that $\frac{dX_\theta}{d\theta}\geq 0$ if $X_\theta<\widehat{X}_1$, and $\widehat{X}_2\geq\widehat{X}_1$ such that $\frac{dX_\theta}{d\theta}\leq 0$ if $X_\theta>\widehat{X}_2$. However,

[10] Recall that SOSD requires that for all $\theta'\geq\theta$

$$\int_{-\infty}^{X}G_{\theta'}(s-e)\,ds\leq\int_{-\infty}^{X}G_\theta(s-e)\,ds.$$

Taking the limit as $\theta'\searrow\theta$ gives

$$\int_{-\infty}^{X}\frac{\partial G_\theta}{\partial\theta}(s-e)\,ds\leq 0.$$

for $\widehat{X}_1 < X_\theta < \widehat{X}_2$, it is possible for $\frac{\partial G_\theta}{\partial \theta}$ to alternate signs several times, and so we cannot sign (36).

Proof of Proposition 2

From the definition of regular distributions (Definition 2), $\frac{\partial G_\theta}{\partial \theta}$ alternates signs only once. Furthermore, we know from Lemma 4 that $\frac{\partial G_\theta}{\partial \theta} \leq (\geq) \; 0$ for s small (large) enough. Therefore, there exists $\widehat{X}$ such that $-\int_{X_\theta - \overline{e}}^{X_\theta} \frac{\partial G_\theta}{\partial \theta}(s)\, ds$ is nonnegative for $X_\theta < \widehat{X}$, and nonpositive for $X_\theta > \widehat{X}$. It then follows from Lemma 4 that $\frac{dX_\theta}{d\theta} \geq 0$ if $X_\theta < \widehat{X}$ and $\frac{dX_\theta}{d\theta} \leq 0$ if $X_\theta > \widehat{X}$.

Proof of Proposition 3

We know from Lemma 4 that $\frac{dX_\theta}{d\theta} \geq (\leq) 0$ if

$$-\int_{X_\theta - \overline{e}}^{X_\theta} \frac{\partial G_\theta}{\partial \theta}(s)\, ds \geq (\leq)\; 0. \tag{38}$$

If G is regular and symmetric for any θ, then

$$\begin{aligned} G(x) &= 1 - G(-x) \\ \frac{\partial G_\theta}{\partial \theta}(x) &= -\frac{\partial G_\theta}{\partial \theta}(-x) \\ \frac{\partial G_\theta}{\partial \theta}(x) &\geq 0 \;\Leftrightarrow\; x \geq 0. \end{aligned}$$

It follows that, for $X_\theta = \overline{e}/2$, the LHS of equation (38) is

$$\int_{-\overline{e}/2}^{\overline{e}/2} -\frac{\partial G_\theta}{\partial \theta}(s)\, ds = 0. \tag{39}$$

For $X_\theta - \overline{e} \geq 0$,

$$\int_{X_\theta - \overline{e}}^{X_\theta} -\frac{\partial G_\theta}{\partial \theta}(s)\, ds \leq 0, \tag{40}$$

and for $X_\theta \leq 0$,

$$\int_{X_\theta - \overline{e}}^{X_\theta} -\frac{\partial G_\theta}{\partial \theta}(s)\, ds \geq 0. \tag{41}$$

Finally, for $X_\theta \in (0, \bar{e})$,

$$\frac{\partial}{\partial X_\theta}\left\{\int_{X_\theta-\bar{e}}^{X_\theta} -\frac{\partial G_\theta}{\partial \theta}(s)\, ds\right\} = \frac{\partial G_\theta}{\partial \theta}(X_\theta - \bar{e}) - \frac{\partial G_\theta}{\partial \theta}(X_\theta) \leq 0 \tag{42}$$

Combining (39)-(42) shows that $\frac{dX_\theta}{d\theta} \geq 0$ if $X_\theta < \frac{\bar{e}}{2}$, and $\frac{dX_\theta}{d\theta} \leq 0$ if $X_\theta > \frac{\bar{e}}{2}$.

Proof of Equations (20) and (21)

First, with volatility σ instead of precision θ, the decomposition in (12) can be rewritten as

$$\frac{d}{d\sigma}\mathbb{E}\left[W(s)|\bar{e}\right] = \underbrace{\frac{\partial}{\partial \sigma}\mathbb{E}\left[W(s)|\bar{e}\right]}_{\text{direct effect}} + \underbrace{\frac{\partial}{\partial X_\sigma}\mathbb{E}\left[W(s)|\bar{e}\right]\frac{dX_\sigma}{d\sigma}}_{\text{incentive effect}} \tag{43}$$

Second,

$$\begin{aligned}
\frac{\partial \mathbb{E}[W(s)|e]}{\partial \sigma} &= \frac{\partial}{\partial \sigma}\int_{X_\sigma}^{\infty}(s - X_\sigma)\frac{1}{\sigma}\varphi\left(\frac{s-e}{\sigma}\right) ds = \frac{\partial}{\partial \sigma}\int_{X_\sigma - e}^{\infty}\frac{s+e-X_\sigma}{\sigma}\varphi\left(\frac{s}{\sigma}\right) ds \\
&= \frac{\partial}{\partial \sigma}\int_{X_\sigma-e}^{\infty}\frac{s}{\sigma}\varphi\left(\frac{s}{\sigma}\right) ds - (X_\sigma - e)\frac{\partial}{\partial \sigma}\int_{X_\sigma-e}^{\infty}\frac{1}{\sigma}\varphi\left(\frac{s}{\sigma}\right) ds \\
&= \frac{\partial}{\partial \sigma}\left\{\left[-\sigma\varphi\left(\frac{s}{\sigma}\right)\right]_{X_\sigma-e}^{\infty}\right\} - (X_\sigma - e)\frac{\partial}{\partial \sigma}\left\{1 - \Phi\left(\frac{X_\sigma - e}{\sigma}\right)\right\} \\
&= \varphi\left(\frac{X_\sigma - e}{\sigma}\right) - \sigma\frac{X_\sigma - e}{\sigma^2}\varphi'\left(\frac{X_\sigma - e}{\sigma}\right) + (X_\sigma - e)\left(-\frac{X_\sigma - e}{\sigma^2}\right)\varphi\left(\frac{X_\sigma - e}{\sigma}\right) \\
&= \varphi\left(\frac{X_\sigma - e}{\sigma}\right) - \frac{X_\sigma - e}{\sigma}\varphi'\left(\frac{X_\sigma - e}{\sigma}\right) + \frac{X_\sigma - e}{\sigma}\varphi'\left(\frac{X_\sigma - e}{\sigma}\right) = \varphi\left(\frac{X_\sigma - e}{\sigma}\right)
\end{aligned} \tag{44}$$

where the fourth and sixth equalities use the property that $\varphi'(x) = -x\varphi(x)$, and the fifth equality uses $\varphi(x) \to_{x\to\infty} 0$. This establishes (21). In addition, it follows that

$$\frac{\partial}{\partial \sigma}\left\{\mathbb{E}\left[W(s)|\bar{e}\right] - \mathbb{E}\left[W(s)|0\right]\right\} = \varphi\left(\frac{X_\theta - \bar{e}}{\sigma}\right) - \varphi\left(\frac{X_\theta}{\sigma}\right).$$

Third,

$$\begin{aligned}
\frac{\partial \mathbb{E}[W(s)|e]}{\partial X_\theta} &= \frac{\partial}{\partial X_\theta}\int_{X_\theta}^{\infty}(s - X_\theta)\frac{1}{\sigma}\varphi\left(\frac{s-e}{\sigma}\right) ds \\
&= \int_{X_\theta}^{\infty} -\frac{1}{\sigma}\varphi\left(\frac{s-e}{\sigma}\right) ds = -\left(1 - \Phi\left(\frac{X_\theta - e}{\sigma}\right)\right).
\end{aligned} \tag{45}$$

It follows that

$$\frac{\partial}{\partial X_\sigma}\left\{\mathbb{E}\left[W\left(s\right)|\bar{e}\right]-\mathbb{E}\left[W\left(s\right)|0\right]\right\}=-\left(1-\Phi\left(\frac{X_\theta-\bar{e}}{\sigma}\right)\right)+\left(1-\Phi\left(\frac{X_\theta}{\sigma}\right)\right)$$

$$=\Phi\left(\frac{X_\theta-\bar{e}}{\sigma}\right)-\Phi\left(\frac{X_\theta}{\sigma}\right).$$

which is strictly negative because of MLRP, which implies FOSD.

Using the results above, we can rewrite (43) as

$$\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}=\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)+\left[1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right]\frac{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\Phi\left(\frac{X_\sigma}{\sigma}\right)} \tag{46}$$

This establishes (20).

Proof of Lemma 5

As X_σ is strictly decreasing in C (see Lemma 2), inequality (22) holds if and only if $\frac{d\frac{dX_\sigma}{d\sigma}}{dX}>0$. As established in the proof of equations (20) and (21) above,

$$\frac{dX_\sigma}{d\sigma}=-\frac{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\Phi\left(\frac{X_\sigma}{\sigma}\right)}.$$

To simplify notation, define

$$x\equiv\frac{X_\sigma}{\sigma},t\equiv\frac{\bar{e}}{\sigma}.$$

We wish to show that $\forall t>0$,

$$f(x,t)\equiv[\varphi(x)-\varphi(x-t)]^2-[\Phi(x)-\Phi(x-t)][\varphi'(x)-\varphi'(x-t)]>0,\quad\forall x, \tag{47}$$

where

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

$$\Phi(x)=\int_{-\infty}^{x}\varphi(y)\,dy.$$

For $t=0$, $f(x,0)$ is trivially 0. Since $\varphi(x)=\varphi(-x)$, we have $\Phi(x)-\Phi(x-t)=\Phi(-x+t)-\Phi(-x)$ and $\varphi'(x)-\varphi'(x-t)=\varphi'(-x+t)-\varphi'(-x)$. As a consequence, $f(x,t)=f(-x+t,t)$. We thus only have to study $x\geq\frac{t}{2}>0$.

We first analyze the term $\varphi'(x) - \varphi'(x-t)$. Since

$$\varphi'(x) = -\frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$\varphi'(x) - \varphi'(x-t) = \varphi(x-t)(-xe^{-t(x-t/2)} + x - t).$$

When $x \geq t/2$, the function $e^{-t(x-t/2)} - 1 + \frac{t}{x}$ is only equal to zero at one point, since it monotonically decreases from 2 to -1. Let that point be x_0. Then

$$\varphi'(x) - \varphi'(x-t) \begin{cases} < 0 & \frac{t}{2} \leq x < x_0 \\ = 0 & x = x_0 \\ > 0 & x > x_0 \end{cases}.$$

We know that when $x \in [\frac{t}{2}, x_0]$, $f(x,t) > 0$ since $[\varphi(x) - \varphi(x-t)]^2 > 0$ and $\Phi(x) - \Phi(x-t) > 0$ $\forall x$, so that (47) is proven for $x \in [\frac{t}{2}, x_0]$

We now prove (47) for $x > x_0$. In this interval (we omit the argument t in what follows):

$$f(x,t) > 0 \iff g(x) \equiv \frac{f(x,t)}{\varphi'(x) - \varphi'(x-t)} > 0.$$

To prove the latter, we first calculate

$$\begin{aligned}
g'(x) =& \frac{2[\varphi(x) - \varphi(x-t)][\varphi'(x) - \varphi'^2 - [\varphi(x) - \varphi(x-t)]^2[\varphi''(x) - \varphi''(x-t)]}{[\varphi'(x) - \varphi'^2} \\
& - [\varphi(x) - \varphi(x-t)] \\
=& \frac{[\varphi(x) - \varphi(x-t)][\varphi'(x) - \varphi'^2 - [\varphi(x) - \varphi(x-t)]^2[\varphi''(x) - \varphi''(x-t)]}{[\varphi'(x) - \varphi'^2} \\
=& \frac{[\varphi(x) - \varphi(x-t)]\varphi(x-t)^2}{[\varphi'(x) - \varphi'^2} \\
& \left\{ \left(-xe^{-t(x-t/2)} + x - t\right)^2 - \left[(x^2-1)e^{-t(x-t/2)} - (x-t)^2 + 1\right] \left(e^{-t(x-t/2)} - 1\right) \right\} \\
=& \frac{[\varphi(x) - \varphi(x-t)]\varphi(x-t)^2}{[\varphi'(x) - \varphi'^2} \left[\left(e^{-t(x-t/2)} - 1\right)^2 + t^2 e^{-t(x-t/2)} \right] \\
& < 0, \quad x \in (x_0, \infty),
\end{aligned}$$

where in the last step we used the fact that $\varphi(x) < \varphi(x-t)$ when $x > t/2$. Therefore,

$$g(x) > 0 \quad \forall x \in (x_0, \infty) \iff \lim_{x\to\infty} g(x) \geq 0.$$

Since

$$\begin{aligned} g(x) =& \frac{[\varphi(x) - \varphi(x-t)]^2}{\varphi'(x) - \varphi'(x-t)} - \Phi(x) + \Phi(x-t) \\ =& \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} \frac{\left(e^{-t(x-t/2)} - 1\right)^2}{-xe^{-t(x-t/2)} + x - t} - \Phi(x) + \Phi(x-t) \end{aligned},$$

it is clear that

$$\lim_{x\to\infty} g(x) = 0.$$

Proof of Proposition 4

Using the chain rule,

$$\frac{d}{dC}\left\{\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}\right\} = \frac{d}{dX_\sigma}\left\{\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}\right\}\frac{dX_\sigma}{dC}$$

Since $\frac{dX_\sigma}{dC} < 0$ (see Lemma 2), we have $\frac{d}{dC}\left\{\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}\right\} > 0$ if and only if $\frac{d}{dX_\sigma}\left\{\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}\right\} < 0$.

Using (20) and $\varphi'(x) = -x\varphi(x)$ for the Normal distribution, we have

$$\frac{d}{dX_\sigma}\left\{\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}\right\} = \frac{d}{dX_\sigma}\left\{\varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right) - \left[1 - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)\right]\frac{\varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right) - \varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right) - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)}\right\}$$

$$\begin{aligned} =& \frac{1}{\sigma}\Bigg(-\frac{X_\sigma - \bar{e}}{\sigma}\varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right) + \left[\frac{X_\sigma - \bar{e}}{\sigma}\varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right) - \frac{X_\sigma}{\sigma}\varphi\left(\frac{X_\sigma}{\sigma}\right)\right]\frac{1 - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right) - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)} \\ &+ \left[\varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right) - \varphi\left(\frac{X_\sigma}{\sigma}\right)\right]\frac{\varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right) - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)} \\ &- \frac{1 - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)}{\left(\Phi\left(\frac{X_\sigma}{\sigma}\right) - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)\right)^2}\left(\varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right) - \varphi\left(\frac{X_\sigma}{\sigma}\right)\right)^2\Bigg) \end{aligned} \tag{48}$$

Multiplying all terms by $\sigma\left(\Phi\left(\frac{X_\sigma}{\sigma}\right) - \Phi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right)\right) > 0$, the expression in (48) has the same sign as

$$\left[\frac{\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)}-\frac{X_\sigma}{\sigma}\right]\left[\varphi\left(\frac{X_\sigma}{\sigma}\right)\left[1-\Phi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)\right]-\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)\left[1-\Phi\left(\frac{X_\sigma}{\sigma}\right)\right]\right]$$
$$-\frac{\overline{e}}{\sigma}\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)\left[1-\Phi\left(\frac{X_\sigma}{\sigma}\right)\right]. \tag{49}$$

Since the last term in (49) is always negative, the expression in (49) is negative if the first line in (49) is negative. We show this in two steps.

To start with, the hazard rate $\varphi(x)/(1-\Phi(x))$ of the Normal distribution is increasing, which implies that

$$\frac{\varphi\left(\frac{X_\sigma}{\sigma}\right)}{1-\Phi\left(\frac{X_\sigma}{\sigma}\right)}>\frac{\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)}{1-\Phi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)}.$$

Rearranging, we have

$$\varphi\left(\frac{X_\sigma}{\sigma}\right)\left[1-\Phi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)\right]-\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)\left[1-\Phi\left(\frac{X_\sigma}{\sigma}\right)\right]>0 \tag{50}$$

Define

$$g(X_\sigma,\overline{e})\equiv\frac{\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)}.$$

If $g(X_\sigma,\overline{e})<\frac{X_\sigma}{\sigma}$, then combining with (50) establishes that the expression in (49) is negative, as desired. We now show that $g(X_\sigma,\overline{e})<\frac{X_\sigma}{\sigma}$ holds. To this end, we will show in turn that $g(X_\sigma,\overline{e})\longrightarrow_{\overline{e}\to 0}\frac{X_\sigma}{\sigma}$, and that $g(X_\sigma,\overline{e})$ is decreasing in $\overline{e}$.

First,

$$\begin{aligned}\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)&=-\varphi'\left(\frac{X_\sigma}{\sigma}\right)\frac{\overline{e}}{\sigma}+O(\overline{e}^2)\\ \Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)&=\frac{\overline{e}}{\sigma}\varphi\left(\frac{X_\sigma}{\sigma}\right)+O(\overline{e}^2).\end{aligned}$$

Using $\varphi'(x)=-x\varphi(x)$ for the Normal distribution, we have

$$\frac{\varphi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\overline{e}}{\sigma}\right)}\longrightarrow_{\overline{e}\to 0}\frac{\varphi\left(\frac{X_\sigma}{\sigma}\right)\frac{\overline{e}X_\sigma}{\sigma^2}}{\frac{\overline{e}}{\sigma}\varphi\left(\frac{X_\sigma}{\sigma}\right)}=\frac{X_\sigma}{\sigma}.$$

Second,

$$\frac{d}{d\bar{e}}\left\{\frac{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}\right\}=\frac{d}{d\bar{e}}\left\{\frac{\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma} s\exp\left\{-\frac{s^2}{2}\right\}ds}{\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma}\exp\left\{-\frac{s^2}{2}\right\}ds}\right\}$$

$$=\frac{1}{\sigma}\frac{\frac{X_\sigma-\bar{e}}{\sigma}\exp\left\{-\frac{(X_\sigma-\bar{e})^2}{2\sigma^2}\right\}\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma}\exp\left\{-\frac{s^2}{2}\right\}ds-\exp\left\{-\frac{(X_\sigma-\bar{e})^2}{2\sigma^2}\right\}\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma}s\exp\left\{-\frac{s^2}{2}\right\}ds}{\left(\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma}\exp\left\{-\frac{s^2}{2}\right\}ds\right)^2}.$$

This expression has the same sign as

$$\frac{X_\sigma-\bar{e}}{\sigma}\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma}\exp\left\{-\frac{s^2}{2}\right\}ds-\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma}s\exp\left\{-\frac{s^2}{2}\right\}ds$$

$$=\int_{(X_\sigma-\bar{e})/\sigma}^{X_\sigma/\sigma}\left[\frac{X_\sigma-\bar{e}}{\sigma}-s\right]\exp\left\{-\frac{s^2}{2}\right\}ds<0.$$

This establishes that $g(X_\sigma,\bar{e})$ is decreasing in $\bar{e}$, which completes the proof.

Proof of Corollary 3

We have $\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}\geq 0$ for any X_σ: indeed, if $\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}<0$ for a given σ, and considering that increasing σ is costless, the given level of σ would not be an equilibrium for any cost of increasing informativeness (including a zero cost).

Using (46), $\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}\geq 0$ may be rewritten as

$$\frac{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}\leq\frac{\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}{1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}. \tag{51}$$

Define

$$y_N(x_\sigma)\equiv\left(\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)\right)\left(1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right) \tag{52}$$

$$y_D(x_\sigma)\equiv\left(\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right)\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right). \tag{53}$$

where $x_\sigma \equiv \frac{X_\sigma}{\sigma}$. Differentiating $y_N(X_\sigma)$ and $y_D(x_\sigma)$ with respect to x_σ gives

$$\begin{aligned} y_N'(x_\sigma) &= \left(\varphi'\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi'\left(\frac{X_\sigma}{\sigma}\right)\right)\left(1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right) \\ &\quad -\left(\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)\right)\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right) \\ &= \left(-\frac{X_\sigma-\bar{e}}{\sigma}\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)+\frac{X_\sigma}{\sigma}\varphi\left(\frac{X_\sigma}{\sigma}\right)\right)\left(1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right) \\ &\quad -\left(\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)\right)\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right). \end{aligned} \tag{54}$$

$$\begin{aligned} y_D'(x_\sigma) &= \varphi'\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\left(\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right)+\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\left(\varphi\left(\frac{X_\sigma}{\sigma}\right)-\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right) \\ &= -\frac{X_\sigma-\bar{e}}{\sigma}\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\left(\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right)+\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\left(\varphi\left(\frac{X_\sigma}{\sigma}\right)-\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right). \end{aligned} \tag{55}$$

At any given X_σ, (54) is larger than (55) if and only if

$$\begin{aligned} &\left(-\frac{X_\sigma-\bar{e}}{\sigma}\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)+\frac{X_\sigma}{\sigma}\varphi\left(\frac{X_\sigma}{\sigma}\right)\right)\left(1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right) \\ &\quad > -\frac{X_\sigma-\bar{e}}{\sigma}\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\left(\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right). \end{aligned} \tag{56}$$

Or, for $X_\sigma > \bar{e}$,

$$\frac{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}-\frac{\bar{e}}{X_\sigma-\bar{e}}\frac{\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}<\frac{\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}{1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)} \tag{57}$$

As $\frac{\bar{e}}{X_\sigma-\bar{e}}\frac{\varphi\left(\frac{X_\sigma}{\sigma}\right)}{\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)} > 0$ for any $X_\sigma > \bar{e}$, and because of (51), we know that (54) > (55) for any $X_\sigma > \bar{e}$, i.e., $y_N'(X_\sigma) > y_D'(X_\sigma)$ for any $X_\sigma > \bar{e}$.

Using (46), we get

$$\frac{\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma}}{\frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial\sigma}}=1-\frac{\left(1-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right)\left[\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)-\varphi\left(\frac{X_\sigma}{\sigma}\right)\right]}{\left(\Phi\left(\frac{X_\sigma}{\sigma}\right)-\Phi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)\right)\varphi\left(\frac{X_\sigma-\bar{e}}{\sigma}\right)}. \tag{58}$$

Note that the numerator and the denominator in the fraction of the RHS of (58) are

$y_N(x_\sigma)$ and $y_D(x_\sigma)$, respectively. We have shown above that $y_N'(x_\sigma) > y_D'(x_\sigma)$ for any $x_\sigma > 0$. For any two given b and a such that $b > a \geq 0$, we therefore have

$$y_N(b) - y_N(a) > y_D(b) - y_D(a) \tag{59}$$

Rearranging (59) yields

$$\frac{y_D(b) - y_D(a)}{y_N(b)} < \frac{y_N(b) - y_N(a)}{y_N(b)} \tag{60}$$

for any $b > \frac{1}{2\sigma}$, given that $y_N(x_\sigma) > 0$ for $x_\sigma \geq \frac{1}{2\sigma}$. Setting $a = \frac{1}{2\sigma}$, (60) holds (as $b > a$), and we have $y_N(a) = 0$. In addition, we have $\frac{y_D(a)}{y_N(b)} \longrightarrow_{b\to\infty} 0$, so that

$$\frac{y_D(b) - y_D(a)}{y_N(b)} \longrightarrow_{c\to\infty} \frac{y_D(b)}{y_N(b)}. \tag{61}$$

Using (60) with $y_N(a) = 0$, this implies that

$$\lim_{b\to\infty} \frac{y_N(b)}{y_D(b)} \geq 1. \tag{62}$$

In addition, $\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial \sigma} = \varphi\left(\frac{X_\sigma}{\sigma}\right) > 0$, and $\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma} \geq 0$ for any X_σ, as established above. It then follows from (58) that we must have $\frac{y_N(b)}{y_D(b)} \leq 1$. Combining with (62) then implies that

$$\lim_{b\to\infty} \frac{y_N(b)}{y_D(b)} = 1. \tag{63}$$

Given the definitions of $y_N(x_\sigma)$ and $y_D(x_\sigma)$ and (58), this establishes $\frac{\frac{d\mathbb{E}[W(s)|\bar{e}])}{d\sigma}}{\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial \sigma}} \longrightarrow_{X\to 0} 0$. The result that $\frac{d\mathbb{E}[W(s)|\bar{e}])}{d\sigma} \longrightarrow_{X\to 0} 0$ then follows from this result and from $\frac{\partial \mathbb{E}[W(s)|\bar{e}]}{\partial \sigma} = \varphi\left(\frac{X_\sigma - \bar{e}}{\sigma}\right) \longrightarrow_{X_\sigma\to\infty} 0$.

Given the optimal contract derived in Lemma 1, the incentive constraint in (5) for a given X_σ is

$$\int_{X_\sigma}^{\infty} (s - X_\sigma) f(s|\bar{e}) ds - \int_{X_\sigma}^{\infty} (s - X_\sigma) f(s|0) ds = C.$$

As $X_\sigma \to \infty$, this equation is satisfied if and only if $C \to 0$. Combining with the results above yields (23).

B Informativeness Has Zero Value

This section gives an example where the value of informativeness is exactly zero. From (16), the principal's payoff is

$$\mathbb{E}[s|e] - X_\theta + \int_{\underline{s}}^{X_\theta} F_\theta(s|e)ds,$$

where X_θ solves the incentive constraint (30):

$$\int_{X_\theta}^{\overline{s}} \left[F_\theta\left(s|0\right) - F_\theta\left(s|\overline{e}\right)\right] ds = C. \tag{64}$$

Let the lower bound of the support of s be $\underline{s} = 0$ and let the upper bound be $\overline{s} = 2$. Suppose that, under low effort, s is uniformly distributed in $[0, 1]$ for any level of informativeness θ:

$$F_\theta(s|0) = s \times \mathbf{1}\left(0 \leq s \leq 1\right).$$

This assumption is for concreteness only; the example can be generalized to distributions that, conditional on low effort, are not functions of θ: $F_\theta(s|0) = \zeta(s)$.

Assume two possible informativeness levels: θ_L and θ_H. Under high informativeness, s is uniformly distributed in $[0, 2]$:

$$f_H(x|1) = \frac{1}{2}, \quad F_H(x|1) = \frac{x}{2}.$$

Under low informativeness, s has the following density function:

$$f_L(x|1) = \begin{cases} \frac{1}{4} \text{ if } x \leq .25 \text{ or } .75 \leq x < 1 \\ \frac{3}{4} \text{ if } .25 < x < .75 \\ \frac{1}{2} \text{ if } 1 < x \leq 2 \end{cases}.$$

Notice that θ_H is a mean-preserving spread of θ_L. Integrating, we obtain the CDF

$$F_L(x|1) = \begin{cases} \frac{x}{4} \text{ if } x \leq \frac{1}{4} \\ \frac{1}{16} + \frac{3}{4}\left(x - \frac{1}{4}\right) \text{ if } \frac{1}{4} < x < \frac{3}{4} \\ \frac{7}{16} + \frac{1}{4}\left(x - \frac{3}{4}\right) \text{ if } \frac{3}{4} \leq x < 1 \\ \frac{x}{2} \text{ if } x \geq 1 \end{cases}$$

Suppose the parameters are such that $X_\theta \geq 1$. For $x \geq 1$, the CDF are the same under both θ_H and θ_L so that, for $X_\theta \geq 1$, the incentive constraint (64) yields:

$$\int_{X_\theta}^{2} \left(1 - \frac{s}{2}\right) ds = C \therefore (2 - X_\theta) - \frac{1}{2}\left[2 - \left(\frac{X_\theta^2}{2}\right)\right] = C$$

$$\therefore \frac{X_\theta^2}{4} - X_\theta + (1 - C) = 0.$$

The solution to this quadratic equation is

$$X_\theta = \frac{1 \pm \sqrt{C}}{2}.$$

The relevant root is the smallest one, otherwise we can relax the incentive constraint (64) by reducing the strike price X_θ:

$$X_\theta = \frac{1 - \sqrt{C}}{2},$$

so the indirect effect is zero (the strike price is the same for both precision levels). The direct effect is also zero since $\int_0^x F_{\theta_H}(s|e)ds = \int_0^x F_{\theta_L}(s|e)ds \; \forall \; x \geq 1$.[11] Indeed, we can calculate this expression explicitly:

$$\int_0^1 F_{\theta_H}(s|e)ds = \int_0^1 F_{\theta_L}(s|e)ds = \frac{1}{4}.$$

Thus, the expected wage is independent of informativeness.

[11] This follows because, since $s|e$ has the same mean under both θ_H and θ_L, integration by parts gives:

$$\int_0^2 F_{\theta_H}(s|e)ds = \int_0^2 F_{\theta_L}(s|e)ds.$$

Thus, $\int_1^2 F_{\theta_H}(s|e)ds = \int_1^2 F_{\theta_L}(s|e)ds$ implies that

$$\int_0^1 F_{\theta_H}(s|e)ds = \int_0^1 F_{\theta_L}(s|e)ds.$$

C Additional Results for Location-Scale Distributions

Claim 1 states that, if the distribution of s has a location and scale parameter, $\widehat{X} \in (0, \overline{e})$.

Claim 1 *Suppose the distribution of s belongs to the location-scale family. Then, $\frac{dX_\theta}{d\theta} \geq 0$ if $X_\theta < \widehat{X}$, and $\frac{dX_\theta}{d\theta} \leq 0$ if $X_\theta > \widehat{X}$, where $\widehat{X} \in (0, \overline{e})$. Furthermore, when G_θ is symmetric, $\widehat{X} \equiv \frac{\overline{e}}{2}$.*

Proof. Since $F_\theta(s|e) = G\left(\frac{s-e}{\sigma}\right)$, condition (36) from Lemma 4 becomes

$$-\int_{X_\theta - \overline{e}}^{X_\theta} \frac{\partial G\left(\frac{s}{\sigma}\right)}{\partial \theta} ds \geq 0.$$

Using $\sigma = \frac{1}{\sqrt{\theta}}$, this becomes

$$-\int_{X_\theta - \overline{e}}^{X_\theta} \frac{\partial G\left(s\sqrt{\theta}\right)}{\partial \theta} ds \geq 0 \iff -\int_{X_\theta - \overline{e}}^{X_\theta} \frac{s}{2\sqrt{\theta}} g\left(s\sqrt{\theta}\right) ds \geq 0. \tag{65}$$

For a distribution symmetric about its mean of zero, this inequality holds if and only if

$$(X_\theta - \overline{e}) + X_\theta \leq 0, \tag{66}$$

that is, if and only if $X_\theta \leq \frac{\overline{e}}{2}$. Since $\frac{dX_\theta}{d\theta} \geq 0$ if (36) > 0 and $\frac{dX_\theta}{d\theta} \leq 0$ if (36) < 0, we conclude that $\frac{dX_\theta}{d\theta} \geq 0$ if $X_\theta \leq \frac{\overline{e}}{2}$, and $\frac{dX_\theta}{d\theta} \leq 0$ if $X_\theta \geq \frac{\overline{e}}{2}$.

Now consider asymmetric distributions. Since $g \geq 0$, condition (65) is satisfied for $X_\sigma \leq 0$, whereas the LHS of (65) is nonpositive for $X_\sigma \geq \overline{e}$. In addition,

$$\frac{\partial}{\partial X_\theta}\left\{-\int_{X_\theta - \overline{e}}^{X_\theta} \frac{s}{2\sqrt{\theta}} g\left(s\sqrt{\theta}\right) ds\right\} = \frac{X_\theta - \overline{e}}{2\sqrt{\theta}} g\left((X_\theta - \overline{e})\sqrt{\theta}\right) - \frac{X_\theta}{2\sqrt{\theta}} g\left(X_\theta\sqrt{\theta}\right) \tag{67}$$

which is strictly negative for $X_\theta \in (0, \overline{e})$, as both terms on the RHS are negative. We conclude that there exists a unique $\hat{X} \in (0, \overline{e})$ such that condition (36) is satisfied if $X_\theta \leq \hat{X}$, in which case $\frac{dX_\theta}{d\theta} \geq 0$, whereas the LHS of (36) is nonpositive for $X_\theta \geq \overline{e}$, in which case $\frac{dX_\theta}{d\theta} \leq 0$. ■

Under the Black-Scholes assumption that the stock price is lognormally distributed, the vega of a stock option is highest when the option is ATM. Claim 2 shows that this result extends to distributions with location and scale parameters.

Claim 2 *For distributions parameterized with e and σ such that $F_\sigma(s|e) = G\left(\frac{s-e}{\sigma}\right)$, the option vega is highest for an option such that $X_\sigma = e$.*

Proof. Let $\overline{s}$ be the upper bound of the support of s. By definition, for given e and $X_\sigma \leq \overline{s}$, the vega of the associated option is

$$\nu = \frac{\partial}{\partial\sigma}\mathbb{E}\left[W(s)|e\right] = \frac{\partial}{\partial\sigma}\left\{\overline{s} - X - \int_{X_\sigma}^{\overline{s}} F_\sigma(s|e)\,ds\right\} \tag{68}$$

where we use (33) to derive the second equality. Since $F_\sigma(s|e) = G\left(\frac{s-e}{\sigma}\right)$, we have

$$\nu = \frac{\partial}{\partial\sigma}\left\{-\int_{X_\sigma}^{\overline{s}} G\left(\frac{s-e}{\sigma}\right)ds\right\} = \frac{1}{\sigma}\int_{X_\sigma}^{\overline{s}} \frac{s-e}{\sigma} g\left(\frac{s-e}{\sigma}\right)ds \tag{69}$$

Using the change of variable $y = \frac{s-e}{\sigma}$ gives

$$\nu = \int_{\frac{X_\sigma - e}{\sigma}}^{\frac{\overline{s}-e}{\sigma}} yg(y)ds \tag{70}$$

Given that $g(y) > 0$, this expression is maximized for $X_\sigma = e$, i.e., for an ATM option.[12]
■

Claim 3 shows that, for symmetric distributions with unbounded support, the vegas of the option-when-working and option-when-shirking are equal for $X_\sigma = \frac{\overline{e}}{2}$.

Claim 3 *For symmetric distributions with unbounded support parameterized by e and σ such $F_\sigma(s|e) = G\left(\frac{s-e}{\sigma}\right)$, the vegas of the option-when working and the option-when-shirking are equal for $X_\sigma = \frac{\overline{e}}{2}$.*

Proof. We rely on (70) and use the fact that, for a distribution with unbounded support, $\overline{s} = \infty$.

[12] With high effort, $e = \overline{e}$, so the option-when-working is ATM for $X_\sigma = \overline{e}$. With low effort, $e = 0$, so the option-when-shirking is ATM for $X_\sigma = 0$.

For $X_\sigma = \frac{\overline{e}}{2}$, the vega $\nu_{\overline{e}}$ of the option-when-working ($e = \overline{e}$) is

$$\nu_{\overline{e}} = \int_{\frac{X_\sigma - \overline{e}}{\sigma}}^{\infty} yg(y)ds = \int_{-\frac{\overline{e}}{2\sigma}}^{\infty} yg(y)ds. \tag{71}$$

For $X_\sigma = \frac{\overline{e}}{2}$, the vega ν_0 of the option-when-shirking ($e = 0$) is

$$\nu_0 = \int_{\frac{X_\sigma}{\sigma}}^{\infty} yg(y)ds = \int_{\frac{\overline{e}}{2\sigma}}^{\infty} yg(y)ds. \tag{72}$$

In addition,

$$\int_{-\frac{\overline{e}}{2\sigma}}^{\infty} yg(y)ds = \int_{-\frac{\overline{e}}{2\sigma}}^{\frac{\overline{e}}{2\sigma}} yg(y)ds + \int_{\frac{\overline{e}}{2\sigma}}^{\infty} yg(y)ds \tag{73}$$

For a symmetric distribution, we have $\int_{-\frac{\overline{e}}{2\sigma}}^{\frac{\overline{e}}{2\sigma}} yg(y)ds = 0$. Equation (73) then implies that $\nu_{\overline{e}} = \nu_0$. ■

D At-The-Money Options

This Appendix shows that the model's main results continue to hold when the principal is restricted to granting ATM options.

We consider the same problem described in Section 3.1, except that we assume that the contract takes the form of ATM options. Considering ATM options requires that we complement the model by deriving the $t = 0$ stock price. To simplify the exposition, we assume that the firm has a single share outstanding. Denoting the stock price at time 0 by S_0, we have $S_0 = \mathbb{E}[q]$ given the assumptions of a zero discount rate and risk neutrality. In addition, with a symmetric distribution with location parameter e, we have $S_0 = e$.

Since the strike price is fixed (at the stock price), the number $n \leq 1$ of ATM options granted adjusts to satisfy the incentive constraint.[13] It follows that $e = \bar{e}$ in equilibrium, and $S_0 = \bar{e}$. With ATM options, the exercise price is therefore $X = \bar{e}$. Considering the same distributions as in section 3.3, we have the following results:

Lemma 7 *(Effect of volatility on number of options) With ATM options,* $\frac{dn}{d\sigma} < 0$.

[13] We only consider the cases such that there exists an incentive compatible contract with ATM options subject to the constraint $n \leq 1$.

Proof. Totally differentiating the LHS of the incentive constraint in (5) with respect to σ yields

$$\frac{d}{d\sigma}\left\{\mathbb{E}[W(s)|\bar{e}]-\mathbb{E}[W(s)|0]\right\}=\frac{\partial}{\partial\sigma}\left\{\mathbb{E}[W(s)|\bar{e}]-\mathbb{E}[W(s)|0]\right\}+\frac{\partial}{\partial n}\left\{\mathbb{E}[W(s)|\bar{e}]-\mathbb{E}[W(s)|0]\right\}\frac{dn}{d\sigma}=0$$

so that

$$\frac{dn}{d\sigma}=-\frac{\frac{\partial}{\partial\sigma}\left\{\mathbb{E}[W(s)|\bar{e}]-\mathbb{E}[W(s)|0]\right\}}{\frac{\partial}{\partial n}\left\{\mathbb{E}[W(s)|\bar{e}]-\mathbb{E}[W(s)|0]\right\}}. \tag{74}$$

First, if the agent receives n options instead of 1, we have

$$\mathbb{E}[W(s)|\bar{e}]-\mathbb{E}[W(s)|0]=n\int_X^\infty\left[F(s|0)-F(s|\bar{e})\right]ds$$

for any given X. With distributions with a location parameter e and scale parameter σ, the numerator of the fraction on the RHS of (74) is then

$$\frac{\partial}{\partial\sigma}\left\{\mathbb{E}[W(s)|\bar{e}]-\mathbb{E}[W(s)|0]\right\}=n\frac{\partial}{\partial\sigma}\int_X^\infty\left[G\left(\frac{s}{\sigma}\right)-G\left(\frac{s-\bar{e}}{\sigma}\right)\right]ds$$

$$=n\int_X^\infty\left[-\frac{s}{\sigma^2}g\left(\frac{s}{\sigma}\right)+\frac{s-\bar{e}}{\sigma^2}g\left(\frac{s-\bar{e}}{\sigma}\right)\right]ds$$

$$=n\left[\int_{\frac{X}{\sigma}}^\infty -y_L g\left(y_L\right)ds+\int_{\frac{X-\bar{e}}{\sigma}}^\infty y_H g\left(y_H\right)ds\right]=n\int_{\frac{X-\bar{e}}{\sigma}}^{\frac{X}{\sigma}} yg\left(y\right)ds,$$

where we used the changes of variables $y_L=\frac{s}{\sigma}$ and $y_H=\frac{s-\bar{e}}{\sigma}$. Given the symmetry of g, we have $\int_{\frac{X-\bar{e}}{\sigma}}^{\frac{X}{\sigma}} yg\left(y\right)ds\geq 0$ if and only if $\frac{X}{\sigma}>-\frac{X-\bar{e}}{\sigma}$, which is always the case with ATM options, i.e., with $X=\bar{e}$. We conclude that the numerator of the fraction on the RHS of (74) is strictly positive with ATM options.

Second, for an agent who receives n ATM options, the denominator of the fraction on the RHS of (74) is equal to

$$\frac{\partial}{\partial n}\left\{\int_X^\infty n(s-X)f(s|\bar{e})ds-\int_X^\infty n(s-X)f(s|0)ds\right\}$$

$$=\int_X^\infty (s-X)f(s|\bar{e})ds-\int_X^\infty (s-X)f(s|0)ds>0. \tag{75}$$

Since both the numerator and the denominator of the fraction on the RHS of (74) are strictly positive, we have

$$\frac{dn}{d\sigma} < 0 \tag{76}$$

with ATM options. ■

If informativeness is improved (i.e. σ falls), n must increase to maintain incentive compatibility. This is because $X = \bar{e} > \frac{\bar{e}}{2}$ with ATM options: the exercise price is higher than the threshold $\frac{\bar{e}}{2}$, so that an increase in informativeness reduces effort incentives, ceteris paribus (the intuition is the same as in Section 3.3). Incentive compatibility then requires that the agent be given additional options.

Corollary 4 is the analogy of Corollary 1 in the main paper and compares the partial and total effects of changes in informativeness on the expected wage.

Corollary 4 *(Partial and total effects of informativeness on expected wage):*

$$\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma} < \frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial\sigma}$$

Proof. First,

$$\frac{d\mathbb{E}[W(s)|\bar{e}]}{d\sigma} = \frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial\sigma} + \frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial n}\frac{dn}{d\sigma}$$

Second,

$$\frac{\partial\mathbb{E}[W(s)|\bar{e}]}{\partial n} = \frac{\partial}{\partial n}\left\{\int_X^\infty n(s-X)f(s|\bar{e})ds\right\} = \int_X^\infty (s-X)f(s|\bar{e})ds > 0$$

Corollary 4 then follows from this inequality and Lemma 7. ■

With ATM options, the total change in expected pay that follows a change in informativeness is smaller than the partial change: while an improvement in informativeness lowers the value of the agent's options, it also requires that the agent receives more options for incentive compatibility. This incentive effect partially offsets the benefits to the principal.

E Continuous Effort Model

In this section, we sketch a continuous effort analog of the core model. The model remains the same, except for the following assumptions:

(A1) The agent chooses effort in $e \in [0, \infty)$.

(A2) The agent's objective function is $\mathbb{E}[W(s)|e] - c\xi(e)$, with $c > 0$, $\xi > 0, \xi' > 0, \xi'' > 0$.

(A3) MLRP: $\frac{d}{ds}\left\{\frac{f_e(s|e)}{f(s|e)}\right\} > 0$, where $f(\pi|e)$ denotes the PDF of s conditional on e, and $f_e(\pi|e)$ denotes its first derivative with respect to e.

(A4) $\mathbb{E}[\max\{s - X_\theta, 0\}|e] - c\xi(e)$ is concave in e, and $W(s)$ is piecewise smooth with a right derivative, which guarantees that the first-order approach to the effort choice problem applies (see footnote 12 in Innes (1990)).

As in the core model, the principal induces a given level of effort $\bar{e} > 0$. As in Proposition 3, we consider continuously distributed symmetric and regular distributions with a location parameter, denoted by e. This implies that we can write $s = e + \varepsilon$, where $\mathbb{E}[\varepsilon|e] = 0$.

For a given informativeness parameter θ, the principal's problem is to choose a cad-lag function $W(\cdot)$ to minimize $\mathbb{E}[W(s)|\bar{e}]$ subject to the same constraints on contracting as in the core model and the following incentive constraint:

$$\frac{d}{de}\int_{-\infty}^{\infty} W(s)f(s|\bar{e})ds = c\xi'(\bar{e}). \tag{77}$$

Then applying Proposition 1 in Innes (1990) implies that, for a given θ, the optimal contract is characterized by

$$W(s) = \max\{0, s - X_\theta\}. \tag{78}$$

As in the core model, there is a unique X_θ that satisfies the incentive constraint in (77) with equality. Subsequent calculations require that the $W(s)$ function be of class C^2 on the whole domain. This can be achieved by an arbitrarily small change in $W(s)$ on $(X_\theta - u, X_\theta + u)$, where $u \to 0$, which smooths out the kink at X_θ (Zang (1980)) while leaving expected pay conditional on any e approximately unchanged.

Denoting by ψ and Ψ the PDF and CDF of the distribution of ε, respectively, the incentive constraint in (77) can be rewritten as

$$\int_{-\infty}^{\infty} W'(\bar{e}+\varepsilon)\psi(\varepsilon)d\varepsilon = c\xi'(\bar{e}). \tag{79}$$

Denote by $\bar{\psi}$ and $\bar{\Psi}$ the PDF and CDF of ε, respectively, after a decrease in θ. For a distribution which is regular and symmetric, the CDFs corresponding to different levels of θ cross only once, at the mean of zero, so that $\Psi(0) = \bar{\Psi}(0)$.

For a given exercise price, an improvement in informativeness θ reduces the LHS of the incentive constraint in (77) if and only if

$$\int_{-\infty}^{\infty} W'(\bar{e}+\varepsilon)\left(\psi(\varepsilon)-\bar{\psi}(\varepsilon)\right) d\varepsilon < 0. \tag{80}$$

Integrating by parts, this becomes

$$\left[W'(\bar{e}+\varepsilon)\left(\Psi(\varepsilon)-\bar{\Psi}(\varepsilon)\right)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} W''(\bar{e}+\varepsilon)\left(\Psi(\varepsilon)-\bar{\Psi}(\varepsilon)\right) d\varepsilon < 0. \tag{81}$$

The first term is equal to zero since $\Psi(-\infty) = \bar{\Psi}(-\infty) = 0$ and $\Psi(\infty) = \bar{\Psi}(\infty) = 1$. In addition, $W''(s) > 0$ on $(X_\theta - u, X_\theta + u)$, and is equal to zero elsewhere. Since $u \to 0$ and Ψ is continuous, it follows that (81) is satisfied if and only if $\Psi(X_\theta - \bar{e}) > \bar{\Psi}(X_\theta - \bar{e})$. In turn, because of single crossing of the CDFs at zero and the symmetry of Ψ, this is satisfied if and only if $X_\theta > \bar{e}$.

As in the baseline model, the LHS of the incentive constraint is strictly decreasing in X. Therefore, for the incentive constraint to still be satisfied following an improvement in informativeness (i.e., a higher θ), we have

$$\frac{dX_\theta}{d\theta} < 0 \qquad \text{if and only if } X_\theta > \bar{e}. \tag{82}$$

Thus, as informativeness θ increases, X_θ approaches $\bar{e}$ and the option becomes closer to ATM.