THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

"Moral Hazard and Endogenous Monitoring"

Ofer Setty[1]

Discussion Paper No. 10-15

July 2015

I would like to thank The Pinhas Sapir Center for Development at Tel Aviv University for their financial support.

I also thank Yair Antler, Eddie Dekel, Zvika Neeman, Ady Pauzner, Nicola Pavoni, Dotan Persitz, Michael Richter, Dan Simundza and David Weiss for very helpful comments.

[1] Ofer Setty– The Eitan Berglas School of Economics, Tel Aviv University. Email: ofer.setty@gmail.com

Abstract

I study a principal-agent problem where the principal chooses the signal's precision of the agent's action. I use the model to study how the principal's monitoring choice depends on each of the three properties of the agent: her disutility from performing the task, her probability of succeeding in the task and her outside option.

JEL Classification: D81; D82; J33
Keywords: Principal-agent model; Moral hazard; Monitoring: Costly state verification.

1 Introduction

In the canonical principal-agent model, a risk-neutral principal provides to a risk-averse agent a transfer that depends on a noisy signal of the agent's action. Because of the agent's risk aversion, the spread in possible transfers for a given action implies a cost to the principal.

I contribute to this literature by allowing the principal to choose the signal's precision of the agent's action. More precise signals allow for a reduction of the risk associated with the transfers, and hence reduce the average transfer to the agent. Since the precision of the signal is costly, this framework posits a trade-off for the principal's monitoring decision between the cost of monitoring and the cost of imposing risk on the agent. This paper uses a simple model to study how the principal's monitoring choice depends on each of three characteristics of the agent: her disutility from performing the task, her probability of succeeding in the task, and her outside option.

First, how does the principal's monitoring choice depend on the disutility from performing the task? In a principal-agent problem without monitoring an increase in that disutility compels the principal to increase the spread between payoffs in order to maintain the incentive-compatibility constraint that the agent faces. Once monitoring is allowed, the principal is free to mix the two instruments that contribute to holding the incentive-compatibility constraint: increasing the spread or increasing the precision of monitoring. I show that an increase in the disutility from performing the task results in an increase in *both* instruments. The intuition for this result is that the principal balances the operational cost of monitoring with the cost of compensating the risk-averse agent for the spread. When the disutility from performing the task increases the principal continues to maintain this balance by increasing the marginal cost of both instruments.

Second, how does the principal's monitoring choice depend on the probability of succeeding in the task? The result and the intuition for this case is similar to the previous one. An increase in the probability of succeeding in the task implies a *smaller* spread in the problem without monitoring. When monitoring is available the principal maintains the balance between the two instruments by lowering both the spread and the monitoring precision.

Third, how does the principal's monitoring choice depend on the agent's outside option? The signal's cost is assumed to be independent of the outside option. The signal's benefit, however, depends on the outside option in a nontrivial way. I show the condition under which the cost of spreading out utilities is increasing in the outside option. When this is the case, the signal's *value* increases with the outside option, leading to the choice of more precise monitoring. In the optimal contract, the signal's precision increases (decreases) and the dispersion of utilities faced by the agent decreases (increases) with the agent's outside option if the derivative of the inverse utility is convex (concave).

Observed heterogeneity in the agent's characteristics along the three dimensions studied in this paper rises naturally in various principal-agent settings. Consider, for example, the problem of a firm that hires a worker to do some task. This firm may face individuals with different disutility levels of performing the task (e.g., due a to difference in availability constraints), different probabilities of succeeding in the task (e.g., due to a difference in experience) and different levels of the outside option (e.g., due to a difference in wealth). In this context the model dictates how the firm that chooses both the worker's payoff and the quality of monitoring should take into account all those sources of heterogeneity.

Similarly, those sources of heterogeneity rise in the problem of insuring an agent against some adverse event. The task in this context is being precautious. The insurer may face individuals with different disutility of performing the task (e.g., due to a difference in attitudes towards being precautious), different probabilities of succeeding in the task (e.g., due to a difference in probabilities of experiencing the adverse event), and different outside options (e.g., due to a difference in access to alternative insurance schemes).

A few papers study wealth effects in a principal-agent model. Newman (2007) looks at how occupational choice depends on wealth given two occupations that differ in the amount of risk borne. In his model, workers who differ in their initial wealth level choose between entrepreneurship that entails a risky payoff and being a worker with a risk-free payoff. Using a condition on the inverse of marginal utility, Newman concludes that there is a threshold wealth level such that workers with at most that wealth level choose the riskier occupation and vice versa.[1] Thiele and Wambach (1999) generalize Newman's result by allowing for any finite number of effort levels instead of two.[2] I show that the condition I use, Newman's condition and Thiele and Wambach's condition are equivalent.

The literature on *contingent monitoring systems* introduces monitoring into the principal-agent problem as well. The emphasis in this literature is on how monitoring should depend on the outcome of the agent's action rather than the agent's characteristics. This question is therefore complimentary to the questions of this paper. Kim and Suh (1992) show that under some conditions the optimal monitoring investment is decreasing in the outcome. The intuition for this result is

[1]Newman extends his model to allow for monitoring. He shows that the assignment of workers to monitoring technologies follow their outside option. His result, however, is restricted to log utility and he leaves the extension to general utilities for future research.

[2]Newman's result appears in an early draft from 1995.

that the principal is using monitoring more intensively when the outcomes are lower because in those cases the agent is more likely to exert a low effort. Fagart and Sinclair-Desgagné (2007) rank contingent monitoring systems. They extend Kim and Suh's result by showing that when the derivative of the inverse utility is convex (concave) the principal prefers monitoring systems whose precision increases (decreases) with respect to the outcome.

2 The model

A risk-neutral principal contracts with a risk-averse agent. The agent's action $a \in \{0, \overline{a}\}, \overline{a} > 0$ is her private information. This action determines output $o \in \{H, L\}$ (owned by the principal) as follows: $p(H|\overline{a}) = \pi, p(H|0) = 0$. Denote the value for the principal from high output by V, and normalize the value of low output to 0. The agent's utility is $u(w) - a$, where $u(\cdot)$ is strictly increasing, strictly concave, three times differentiable and its inverse is three times differentiable as well; w is the transfer to the agent; and a is normalized such that it is the agent's utility cost for exerting effort a.

The principal can acquire a binary signal $s \in \{G, B\}$ on the agent's action, representing good and bad outcomes, respectively. The good signal outcome can happen only if the agent's action is $a = \overline{a}$, i.e., $p(G|0) = 0$. This means that the accuracy of the signal is determined by $p(G|\overline{a})$. If $p(G|\overline{a}) = 0$, the signal carries no information; if $p(G|\overline{a}) = 1$, the signal perfectly reveals the agent's action. Denote $p(G|\overline{a})$ by θ and let θ be a choice variable of the principal. The signal's cost is a strictly increasing convex function $c(\theta)$.[3] The signal and output probabilities, conditional on effort, are independent.

In general, the contract should include a precision choice for any level of output. However, since acquiring an informative signal is costly, the principal will always set $\theta = 0$ for an agent with outcome H. This is because $p(H|0) = 0$ implies that outcome H reveals the agent's action. This simplifies the contract as there are only three possible outcomes in equilibrium: $\{H, G, B\}$. The contract specifies a recommendation on action, the precision choice of the monitoring technology when low output is realized, and a transfer to the agent for any outcome. The action recommendation must be incentive compatible. In addition, the contract requires that the agent's expected utility will be at least U.

[3]Since the paper focuses on the tradeoff between the spread and the monitoring precision I make the following assumptions to guarantee an internal solution for θ (except when the signal is non informative): $lim_{\theta \to 0} c'(\theta) = 0, lim_{\theta \to 1} c'(\theta) = \infty$.

3 The contract

Denote by w^x the principal's transfer to the agent conditional on outcome x for $x \in \{H, G, B\}$. Let $\widehat{C}$ be the cost for a principal who recommends the action $\overline{a}$. In what follows I assume that the parameters justify creating the costly incentives for the agent to choose action $\overline{a}$, e.g., V is high enough. Otherwise, the problem becomes trivial with a recommendation of $a = 0$ and full insurance. The principal's problem is as follows:

$$\begin{aligned}
&\widehat{C} = \min_{w^H, w^G, w^B, \theta} \left\{ \pi w^H + (1-\pi)\,\theta w^G + (1-\pi)\,(1-\theta)\,w^B + (1-\pi)\,c(\theta) + \pi V \right\} \\
&s.t. \\
&\pi u\left(w^H\right) + (1-\pi)\,\theta u\left(w^G\right) + (1-\pi)\,(1-\theta)\,u\left(w^B\right) - \overline{a} \geq U \\
&\pi u\left(w^H\right) + (1-\pi)\,\theta u\left(w^G\right) + (1-\pi)\,(1-\theta)\,u\left(w^B\right) - \overline{a} \geq u\left(w^B\right) \qquad (1)
\end{aligned}$$

The first constraint is the individual-rationality (IR) constraint. The second constraint is the incentive-compatibility (IC) constraint. The left-hand side of the two constraints is the expected utility for the agent conditional on $a = \overline{a}$. The right-hand side of the IC constraint is the utility for the agent conditional on $a = 0$. Notice that since the IC constraint holds, the objective function assumes the probabilities given action $\overline{a}$.

The following claim determines the ranking of the transfers.

Claim 1 *In the optimal solution* $u(w^H) = u(w^G) > u(w^B) = U$.

All proofs are relegated to the appendix.

This claim is based on several properties of the problem: both the IR and the IC constraints are tight, and w^B must be lower than $\{w^H, w^G\}$ to satisfy the IC constraint (see the appendix for details). Notice that $u(w^H) = u(w^G)$ because the two outcomes have identical informtaion content (high effort for sure).

Rewrite the problem as follows. In the IC, substitute $u(w^B)$ with U and derive $u\left(w^H\right) = U + \frac{\overline{a}}{\pi + (1-\pi)\theta}$. Using the values for $\{w^H, w^G, w^B\}$ in the objective function and omitting the term πV, which is independent of the choice variable, leads to the following convex optimization problem, whose solution for θ is identical to that of Problem (1):

$$C = \min_{\theta} \left\{ \alpha(\theta) u^{-1}\left(U + \frac{\overline{a}}{\alpha(\theta)}\right) + (1-\alpha(\theta)) u^{-1}\left(U\right) + (1-\pi)\,c(\theta) \right\} \qquad (2)$$

where $\alpha \equiv \pi + (1-\pi)\,\theta$.

To understand the role of monitoring precision in this problem, consider the solution to the first best. In the first best, the principal observes the agent's effort, so no monitoring is required. The first-best allocation is then a fixed transfer (independent of output) that is equal to $u^{-1}(U+\overline{a})$. In this case the principal compensates the agent only for her effort.

The principal's cost in (2) differs from the principal's cost in the first best in two aspects. First, in the constrained problem, monitoring may be used upon low output with a cost of $(1-\pi)\,c(\theta)$. Second, in the constrained problem, the principal is required to create a spread in transfers conditional on outcomes. Therefore, the principal delivers to the agent utility as a lottery between $u^{-1}\left(U+\frac{\overline{a}}{\alpha}\right)$ with probability α and $u^{-1}(U)$ with probability $(1-\alpha)$. I refer to the difference between the two utilities $\left(U+\frac{\overline{a}}{\alpha},U\right)$ as the *spread*.

The average utility delivered through the lottery is $U+\overline{a}$. This is, by construction, equal to the utility delivered in the first best. Therefore, the only role of the signal in this problem is to reduce the risk associated with the spread and thus reduce the cost of delivering utility as a lottery rather than as a certainty equivalent. Indeed, if the signal was without cost, the principal would set $\theta=1$, and both the allocation and the principal's cost would be identical to those of the first best. (To see this, substitute $\theta=1$ and $c(\theta)=0$ in Problem (2) and get the first-best cost.)

4 Optimal monitoring

In this section I analyze how optimal monitoring is affected by each of the three characteristics of the agent: the disutility from performing the task $\bar{a}$, the probability of succeeding in the task π, and the outside option U.

Theorem 1 characterizes the contract w.r.t. the disutility from performing the task $\bar{a}$.

Theorem 1 *The solution to Problem (2) has the following characteristics:*
(i) *the optimal signal's precision* (θ) *increases with the task's disutility* $(\bar{a})$*;*
(ii) *the utility spread* $\left(\frac{\overline{a}}{\pi+(1-\pi)\theta}\right)$ *increases with the task's disutility* $(\bar{a})$.

To gain intuition on Theorem 1 consider the principal-agent problem presented here except that monitoring is unavailable. In this environment an increase in the disutility from performing the task compels the principal to increase the spread between payoffs in order to maintain the incentive-compatibility constraint that the agent faces. Once monitoring is allowed the principal is free to mix the two instruments that contribute to holding the incentive-compatibility constraint: increasing the spread or increasing the precision of monitoring. In this case the principal balances the operational cost of monitoring with the cost of compensating the risk-averse agent for the spread. When the disutility from performing the task increases the principal continues to maintain this balance by increasing the marginal cost of both instruments.

Theorem 2 deals with the effect of the probability of succeeding in the task (π) on optimal monitoring:

Theorem 2 *The solution to Problem (2) has the following characteristics:*
(i) *the optimal signal's precision* (θ) *decreases with the success probability* (π)*;*
(ii) *the utility spread* $\left(\frac{\bar{a}}{\pi+(1-\pi)\theta}\right)$ *decreases with the success probability* (π).

The result and the intuition for this case is similar to the previous case. An increase in the probability of succeeding in the task implies a smaller spread in the problem without monitoring. When monitoring is available the principal decreases both the spread and the monitoring precision maintaining the balance between the two instruments.

Theorem 3 characterizes the contract w.r.t. the outside option U. This requires an additional condition on the concavity of $\left(u^{-1}\right)'(\cdot)$.

Theorem 3 *The solution to Problem (2) has the following characteristics if* $\left(u^{-1}\right)'(\cdot)$ *is convex:*
(i) *the optimal signal's precision* (θ) *increases with the outside option* (U)*;*
(ii) *the utility spread* $\left(\frac{\bar{a}}{\pi+(1-\pi)\theta}\right)$ *decreases with the outside option* (U)*;*
(iii) *the cost of spreading out utility increases with the outside option* (U)*;*
(iv) *the converse version of* $(i)-(iii)$ *holds when* $\left(u^{-1}\right)'(\cdot)$ *is concave.*

What is the intuition behind this Theorem? The decision maker who solves problem (2) has a utility function given by $-(u)^{-1}$. When this utility is that of a prudent individual (i.e. $\left(u^{-1}\right)'$ is concave), the principal is more concerned about spreads at low utility levels so he invests relatively more in monitoring at low levels of U to reduce the spread.[4]

Since the decision maker's utility is linked to the agent's utility it is also possible to see the intuition through the agent's perspective. For any level of outside option (utility), the principal weighs the cost of the signal against its benefit of reducing the risk associated with the spread.

[4]Menezes, Geiss, and Tressler (1980) show that a decision maker whose utility function has a positive third derivative is downside risk averse.

The signal's cost does not depend on the outside option. When $(u^{-1})'(\cdot)$ is convex, the cost of spreading out utilities increases with the outside option. In this case, the value of monitoring increases and the principal increases her investment in the signal. This, in turn, results in a smaller spread between utilities.

5 An equivalence result

The condition that $(u^{-1})'(\cdot)$ is convex is related to other conditions that can be found in the literature. I make the following observation:

Proposition 1 *The following conditions are equivalent:*
1: $(u^{-1})'(\cdot)$ *is convex*
2: $\frac{u'''(w)u'(w)}{u''(w)^2} \leq 3$
3: There is a convex function $h : \Re \to \Re_+$ *such that* $\frac{1}{u'(w)} = h(u(w))$ *.*
4: $\frac{\left(\frac{1}{u'}\right)''}{\left(\frac{1}{u'}\right)'} \geq \frac{u''}{u'}$

Condition 1 is the one used in this paper. Condition 2 is used in Thiele and Wambach (1999). Condition 3 is used in Newman (2007). Condition 4 implies that u is more risk averse than $\frac{1}{u'}$ in the sense of Pratt (1964). Examples of utility functions that satisfy those conditions are IARA, CARA, and CRRA with a coefficient of relative risk aversion of at least $\frac{1}{2}$.

As explained in the introduction, Thiele and Wambach generalize Newman's result by allowing any finite number of effort levels instead of two. They also interpret Newman's condition as implying that $\frac{u'''(w)u'(w)}{u''(w)^2} \leq 2$, making their condition weaker than his. However, as Proposition 1 shows, Newman's condition is equivalent to theirs.[5]

[5]Thiele and Wambach describe Newman's condition as "requir[ing] that the inverse of marginal utility is convex in income.", which indeed implies that $\frac{u'u'''}{(u'')^2} \leq 2$. However, requiring the inverse of marginal utility to be convex in income is a stricter condition than what Newman requires. In fact $\frac{1}{u'}$ may even be concave as long as it is *more convex* than $u(x)$, as is the case of CRRA with a coefficient of risk aversion in $[\frac{1}{2}, 1]$. Instead, Newman's condition is that $\frac{1}{u'}$ is convex in *utility* (rather than in *income*). This condition is equivalent, as Proposition 1 shows, to $\frac{u'u'''}{(u'')^2} \leq 3$.

References

FAGART, M.-C., AND B. SINCLAIR-DESGAGNÉ (2007): "Ranking Contingent Monitoring Systems," *Management Science*, 53(9), 1501–1509.

KIM, S. K., AND Y. S. SUH (1992): "Conditional Monitoring Policy Under Moral Hazard," *Management Science*, 38(8), 1106–1120.

MENEZES, C., C. GEISS, AND J. TRESSLER (1980): "Increasing Downside Risk," *American Economic Review*, 70(5), 921–32.

MILGROM, P., AND C. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–80.

NEWMAN, A. F. (2007): "Risk-bearing and Entrepreneurship," *Journal of Economic Theory*, 137(1), 11–26.

PRATT, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122–136.

THIELE, H., AND A. WAMBACH (1999): "Wealth Effects in the Principal Agent Model," *Journal of Economic Theory*, 89(2), 247–260.

APPENDIX

Proof of claim 1

Lemma 1 *In the optimal solution either* $w^H > w^B$ *or* $w^G > w^B$*, or both.*

Proof. Rewrite the IC as:

$$\pi u\left(w^H\right) + (1-\pi)\,\theta u\left(w^G\right) \geq \left[\pi + (1-\pi)\,\theta\right] u\left(w^B\right) + \overline{a}. \tag{3}$$

Since $\overline{a} > 0$ and since the sum of the coefficients of $\left\{u\left(w^H\right), u\left(w^G\right)\right\}$ is equal to the coefficient of $u\left(w^B\right)$ (and positive), if both $w^H \leq w^B$ and $w^G \leq w^B$ then the IC cannot hold. ■

Lemma 2 *In the optimal solution the IR holds with equality.*

Proof. The solution with a slack IR can be improved by decreasing w^B by ε. For a small ε, the IR is still slack. The IC remains slack or becomes slack (see (3)). The objective function increases by ε, which is a contradiction to the solution being optimal. ■

Lemma 3 *In the optimal solution* $w^H = w^G$.

Proof. Assume that $w^H > w^G$. The optimal solution can be improved as follows. Decrease w^H by ε and increase w^G by $\frac{\pi\varepsilon}{(1-\pi)\theta}$. By construction this change does not affect the objective function because $\pi\left(w^H - \varepsilon\right) + (1-\pi)\,\theta\left(w^G + \frac{\pi}{(1-\pi)\theta}\right) = \pi w^H + (1-\pi)\,\theta w^G$. To study the effect on the IR, consider the part of the IR composed of $\pi u\left(w^H\right) + (1-\pi)\,\theta u\left(w^G\right)$. The change makes the IR slack because it is a lottery with the same certainty equivalent but with less risk. Formally, the claim is that:

$$\pi u\left(w^H - \varepsilon\right) + (1-\pi)\,\theta u\left(w^G + \frac{\varepsilon\pi}{(1-\pi)\,\theta}\right) > \pi u\left(w^H\right) + (1-\pi)\,\theta u\left(w^G\right)$$

$$(1-\pi)\,\theta\left(u\left(w^G + \frac{\varepsilon\pi}{(1-\pi)\,\theta}\right) - u\left(w^G\right)\right) > \pi\left(u\left(w^H\right) - u\left(w^H - \varepsilon\right)\right) \tag{4}$$

Divide both sides by ε and rearrange to get:

$$\frac{u\left(w^H\right) - u\left(w^H - \varepsilon\right)}{\varepsilon} < \frac{u\left(w^G + \frac{\varepsilon\pi}{(1-\pi)\theta}\right) - u\left(w^G\right)}{\frac{\varepsilon\pi}{(1-\pi)\theta}} \tag{5}$$

In the limit this is $u'\left(w^H\right) < u'\left(w^G\right)$, which is true by the negation assumption that $w^H > w^G$. Thus the IR (and similarly the IC) become slack, which is a contradiction to Lemma 2. Therefore it is impossible for w^H to be strictly greater than w^G in the optimal solution.

The same line of proof eliminates the possibility of $w^G > w^H$ by showing that increasing w^H by ε and decreasing w^G by $\frac{\pi\varepsilon}{(1-\pi)\theta}$ violates Lemma 2.

■

Lemma 4 *In the optimal solution the IC holds with equality.*

Proof. By Lemmata 1 and 3 $w^H > w^B$. If the IC is slack then the objective function can be improved. Decrease w^H by ε and increase w^B by $\frac{\varepsilon\pi}{(1-\pi)(1-\theta)}$. By construction this change does not affect the objective function. The IC still holds. Consider the part of the IR composed of $\pi u\left(w^H\right)+(1-\pi)(1-\theta)u\left(w^B\right)$. Those changes make the IR slack because it is a lottery with the same certainty equivalent but with less risk. Formally, the claim is that:

$$\pi u\left(w^H-\varepsilon\right)+(1-\pi)(1-\theta)u\left(w^B+\frac{\varepsilon\pi}{(1-\pi)(1-\theta)}\right)>\pi u\left(w^H\right)+(1-\pi)(1-\theta)u\left(w^B\right)$$
$$\pi\left(u\left(w^H\right)-u\left(w^H-\varepsilon\right)\right)<(1-\pi)(1-\theta)\left(u\left(w^B+\frac{\varepsilon\pi}{(1-\pi)(1-\theta)}\right)-u\left(w^B\right)\right). \tag{6}$$

Divide both sides by ε and rearrange to get:

$$\frac{u\left(w^H\right)-u\left(w^H-\varepsilon\right)}{\varepsilon}<\frac{u\left(w^B+\frac{\varepsilon\pi}{(1-\pi)(1-\theta)}\right)-u\left(w^B\right)}{\frac{\varepsilon\pi}{(1-\pi)(1-\theta)}}. \tag{7}$$

In the limit this is $u'\left(w^H\right)<u'\left(w^B\right)$, which is true because $w^B<w^H$. Now decrease w^H by δ in order to improve the objective function without damaging any of the constraints. ■

Lemma 5 *In the optimal solution $u\left(w^B\right)=U$.*

Proof. Since both the IR and the IC are tight, and since the LHS of both constraints is identical, the RHS of both constraints is equal and $u\left(w^B\right)=U$. ■

Claim 1 is then a combination of Lemmata 1, 3, and 5.

Proof of Theorem 1

Theorem 1 *The solution to Problem (2) has the following characteristics:*
(i) the optimal signal's precision (θ) increases with the task's disutility $(\bar{a})$;
(ii) the utility spread $\left(\frac{\bar{a}}{\pi+(1-\pi)\theta}\right)$ increases with the task's disutility $(\bar{a})$.

Proof. The two parts of the theorem are proved sequentially.

Proof of (i)
The proof is based on monotone comparative statics (Milgrom and Shannon, 1994). .

$$\begin{aligned}\frac{\partial w}{\partial\bar{a}} &= \left(u^{-1}\right)'\left(U+\frac{\bar{a}}{\alpha}\right)\\ \frac{\partial^2 w}{\partial\bar{a}\partial\theta} &= -\frac{(1-\pi)\bar{a}}{\alpha^2}\left(u^{-1}\right)''\left(U+\frac{\bar{a}}{\alpha}\right)<0\end{aligned} \tag{8}$$

According to the monotone comparative statics theorem, θ^* (weakly) increases with $\overline{a}$ if $\frac{\partial^2 w}{\partial \overline{a} \partial \theta} \leq 0$.[6]

Proof of (ii)

Inspection of the spread ($\frac{\overline{a}}{\alpha} = \frac{\overline{a}}{\pi+(1-\pi)\theta}$) shows that the effect of an increase in both $\overline{a}$ and θ on the spread is ambiguous. A further inspection into the FOC shows, however, that this is not the case. The FOC is:

$$c'(\theta) = \frac{\overline{a}}{\alpha}\left(u^{-1}\right)'\left(U + \frac{\overline{a}}{\alpha}\right) - \left(u^{-1}\right)\left(U + \frac{\overline{a}}{\alpha}\right) - \left(u^{-1}\right)(U) \tag{9}$$

As θ increases the LHS increases because $c(\cdot)$ is convex in θ. By differentiating the RHS w.r.t. the spread it can be verified that the RHS is increasing in the spread. Therefore an increase in θ must be accompanied by an increase in the spread. ∎

Proof of Theorem 2

Theorem 2 *The solution to Problem (2) has the following characteristics:*
(i) *the optimal signal's precision* (θ) *decreases with the success probability* (π)*;*
(ii) *the utility spread* $\left(\frac{\overline{a}}{\pi+(1-\pi)\theta}\right)$ *decreases with the success probability* (π).

Proof. The two parts of the theorem are proved sequentially.

Proof of (i)
The proof follows the same line of proof as in theorem 1.

$$\begin{aligned} \frac{\partial w}{\partial \pi} &= (1-\theta)\left(\left(u^{-1}\right)\left(U + \frac{\overline{a}}{\alpha}\right) - \frac{\overline{a}}{\alpha}\left(u^{-1}\right)'\left(U + \frac{\overline{a}}{\alpha}\right) - \left(u^{-1}\right)'(U)\right) - c(\theta) \qquad (10) \\ \frac{\partial^2 w}{\partial \pi \partial \theta} &= \frac{(1-\pi)(1-\theta)\,\overline{a}^2}{\alpha^3}\left(u^{-1}\right)''\left(U + \frac{\overline{a}}{\alpha}\right) > 0, \end{aligned}$$

where the derivation of $\frac{\partial^2 w}{\partial \pi \partial \theta}$ uses the fact that the FOC w.r.t. θ is zero at the optimum. This establishes by the monotone comparative statics theorem that as the parameter π increases, the monitoring precision θ decreases.

Proof of (ii)

Inspection of the spread's denominator $(\pi + (1-\pi)\theta)$ shows that the effect of a decrease in both π and θ on the spread is ambiguous. However, as directly given by the proof of Theorem 1

[6] In the case of a maximization problem *supermodularity* is required between $\{\overline{a}, \theta\}$. Here the sign of the cross derivative is opposite because it is a minimization problem (as $\min w = -\max -w$).

(ii) a decrease in θ must be accompanied by a decrease in the spread to keep the FOC.

■

Proof of Theorem 3

Theorem 3 *The solution to Problem (2) has the following characteristics if* $(u^{-1})'(\cdot)$ *is convex:*
(i) *the optimal signal's precision* (θ) *increases with the outside option* (U)*;*
(ii) *the utility spread* $\left(\frac{\overline{a}}{\pi+(1-\pi)\theta}\right)$ *decreases with the outside option* (U)*;*
(iii) *the cost of spreading out utility increases with the outside option* (U)*;*
(iv) *the converse version of* $(i)-(iii)$ *holds when* $(u^{-1})'(\cdot)$ *is concave.*

Proof. The three parts of the theorem are proved sequentially.

Proof of (i)

The proof follows the same line of proof as in Theorem 1 (i).

$$\begin{aligned}\frac{\partial w}{\partial U} &= \alpha\left(\left(u^{-1}\right)'\left(U+\frac{\overline{a}}{\alpha}\right)-\left(u^{-1}\right)'(U)\right)+\left(u^{-1}\right)'(U) \\ \frac{\partial^2 w}{\partial U\partial\theta} &= (1-\pi)\left(\left(u^{-1}\right)'\left(U+\frac{\overline{a}}{\alpha}\right)-\left(u^{-1}\right)'(U)\right)-\frac{\overline{a}\,(1-\pi)}{\alpha}\left(u^{-1}\right)''\left(U+\frac{\overline{a}}{\alpha}\right)\end{aligned} \tag{11}$$

According to the monotone comparative statics theorem, θ^* (weakly) increases with U if $\frac{\partial^2 w}{\partial U\partial\theta}\leq 0$. This is satisfied if $(u^{-1})'(\cdot)$ is convex. To see this, rewrite $\frac{\partial^2 w}{\partial U\partial\theta}$ as:

$$\frac{\overline{a}\,(1-\pi)}{\alpha}\left\{\frac{\left(u^{-1}\right)'\left(U+\frac{\overline{a}}{\alpha}\right)-\left(u^{-1}\right)'(U)}{\frac{\overline{a}}{\alpha}}-\left(u^{-1}\right)''\left(U+\frac{\overline{a}}{\alpha}\right)\right\}, \tag{12}$$

and notice that $\frac{\left(\left(u^{-1}\right)'\left(U+\frac{\overline{a}}{\alpha}\right)-\left(u^{-1}\right)'(U)\right)}{\frac{\overline{a}}{\alpha}}$ is the average slope of $(u^{-1})'(\cdot)$ between $\left\{U, U+\frac{\overline{a}}{\alpha}\right\}$, where $\frac{\overline{a}}{\alpha}>0$, and $(u^{-1})''\left(U+\frac{\overline{a}}{\alpha}\right)$ is the slope of $(u^{-1})'(\cdot)$ at $\left(U+\frac{\overline{a}}{\alpha}\right)$. If $(u^{-1})'(\cdot)$ is convex then the slope of $(u^{-1})'(\cdot)$ at $\left(U+\frac{\overline{a}}{\alpha}\right)$ is higher than the average slope at $\left\{U, U+\frac{\overline{a}}{\alpha}\right\}\Rightarrow$ (12) is negative $\Rightarrow \frac{\partial^2 w}{\partial U\partial\theta}\leq 0$.

Proof of (ii)

By (i), θ increases with U if $(u^{-1})'(\cdot)$ is convex. The spread is $\frac{\overline{a}}{\pi+(1-\pi)\theta}$, so the spread decreases as θ increases.

Proof of (iii)

The principal's cost in problem (2) is composed out of the cost of providing a transfer, equal to: $\alpha u^{-1}\left(U+\frac{\overline{a}}{\alpha}\right)+(1-\alpha)\,u^{-1}(U)$ and the monitoring cost $(1-\pi)\,c(\theta)$.

The first-best cost for the principal is $u^{-1}(U+\overline{a})$. Therefore, the difference between the principal's cost of providing a transfer in the first best and in (2) is a result of the requirement of spreading out utilities. Define the cost of spreading out utilities $\left\{U, U+\frac{\overline{a}}{\alpha}\right\}$ at utility U as the difference between the costs:

$$D(U)=\left\{\alpha u^{-1}\left(U+\frac{\overline{a}}{\alpha}\right)+(1-\alpha)\,u^{-1}(U)\right\}-u^{-1}(U+\overline{a}), \tag{13}$$

and notice that the curly brackets include a lottery with prizes $\left\{U+\frac{\overline{a}}{\alpha}, U\right\}$ with probabilities $\{\alpha, 1-\alpha\}$, whose expected prize is $U+\overline{a}$. This means that $D(U)$ is the difference between a lottery and a certainty equivalent $(U+\overline{a})$, valued by the function of $u^{-1}(\cdot)$. Since $u(\cdot)$ is concave $D(U)>0\;\forall u(\cdot)$.

The dependence of this cost on U is the following derivative:

$$D'(U)=\left\{\alpha\left(u^{-1}\right)'\left(U+\frac{\overline{a}}{\alpha}\right)+(1-\alpha)\left(u^{-1}\right)'(U)\right\}-\left(u^{-1}\right)'(U+\overline{a}). \tag{14}$$

Since under Condition 1 $(u^{-1})'$ is convex, Jensen's inequality implies that:

$$\begin{aligned}
\alpha\left(u^{-1}\right)'\left(U+\frac{\overline{a}}{\alpha}\right)+(1-\alpha)\left(u^{-1}\right)'(U) &> \left(u^{-1}\right)'\left\{\alpha\left(U+\frac{\overline{a}}{\alpha}\right)+(1-\alpha)(U)\right\}\\
&= \left(u^{-1}\right)'\{U+\overline{a}\}\\
&\Rightarrow D'(U)>0
\end{aligned}$$

Proof of (iv)

The proof follows the same arguments as in the proof for parts $(i)-(iv)$ above for $(u^{-1})'(\cdot)$ *concave*. ■

Proof of Proposition 1

Proposition 1 *The following conditions are equivalent:*

1: $\left(u^{-1}\right)'(\cdot)$ *is convex*

2: $\frac{u'''(w)u'(w)}{u''(w)^2}\leq 3$

3: There is a convex function $h:\Re\rightarrow\Re_+$ *such that* $\frac{1}{u'(w)}=h\left(u\left(w\right)\right)$.

4: $\frac{\left(\frac{1}{u'}\right)''}{\left(\frac{1}{u'}\right)'}\geq\frac{u''}{u'}$

Proof. The proof goes as follows: Condition 1 ⇒ Condition 3 ⇒ Condition 4 ⇒ Condition 2 ⇒ Condition 1.

Condition 1 ⇒ Condition 3

Let $h(\cdot)$ be $(u^{-1\prime}(\cdot))$. Then $h(u(x))=u^{-1\prime}(u(x))=\frac{1}{u\prime(u^{-1}(u(x)))}=\frac{1}{u'(x)}$. By Condition 1 $u^{-1}(\cdot)'$ is convex, and therefore there exists a convex function $h(\cdot)$ such that: $\frac{1}{u'(x)}=h\left(u\left(x\right)\right)$.

Condition 3 $\Rightarrow$ Condition 4

Denote: $f \equiv \frac{1}{u'(x)}, g \equiv u$. By Condition 3 $\exists h(\cdot)$ such that $f = h(g)$. Then:

$$\begin{aligned} f' &= h'(g)\, g' \\ f'' &= h''\left(g'\right)^2 + g''h' = h''g'\frac{f'}{h'} + g''\frac{f'}{g'} \\ \frac{f''}{f'} &= \frac{h''g'}{h'} + \frac{g''}{g'} \end{aligned} \tag{15}$$

u is increasing $\Rightarrow g' > 0$,
h is increasing (because $h' = \frac{f'}{g'} = \frac{-u''}{u'} > 0$) $\Rightarrow h' > 0$,
h is convex $\Rightarrow h'' > 0$,
$\Rightarrow \frac{h''g'}{h'} \geq 0$
$\Rightarrow \frac{f''}{f'} \geq \frac{g''}{g'}$.

Condition 4 $\Rightarrow$ Condition 2

Using that $\left((u')^{-1}\right)' = -(u')^{-2}u''$ and that $\left((u')^{-1}\right)'' = 2(u')^{-3}(u'')^2 - (u')^{-2}u'''$ rewrite $\frac{\left(\frac{1}{u'}\right)''}{\left(\frac{1}{u'}\right)'}$ as $\frac{u'''-2(u')^{-1}(u'')^2}{u''}$. By Condition 4 $\frac{\left(\frac{1}{u'}\right)''}{\left(\frac{1}{u'}\right)'} \geq \frac{u''}{u'}$ so we get that:

$$\begin{aligned} \frac{u''' - 2(u')^{-1}(u'')^2}{u''} &\geq \frac{u''}{u'} \\ u'''u' - 2\left(u'\right)^{-1}\left(u''\right)^2 u' &\leq \left(u''\right)^2 \\ u'''u' &\leq 3\left(u''\right)^2 \\ \frac{u'''(w)\, u'(w)}{u''(w)^2} &\leq 3 \end{aligned} \tag{16}$$

Condition 2 $\Rightarrow$ Condition 1

Using the implicit function theorem and differentiating both sides of $w = u^{-1}(u(w))$ gives the equality $(u^{-1})'(u(w)) = \frac{1}{u'}$. Differentiate $\frac{1}{u'}$ twice with respect to *utility* gives:

$$\begin{aligned} \frac{d((u')^{-1})}{dU} &= -\left(u'\right)^{-2}u''\frac{dc}{dU} = -\left(u'\right)^{-3}u'' \\ \frac{d^2((u')^{-1})}{dU^2} &= \frac{d\left(-\left(u'\right)^{-3}u''\right)}{dU} = 3\left(u'\right)^{-4}\left(u''\right)^2\frac{dc}{dU} - \left(u'\right)^{-3}u'''\frac{dc}{dU} \\ &= 3\left(u'\right)^{-5}\left(u''\right)^2 - \left(u'\right)^{-4}u''' \end{aligned} \tag{17}$$

To prove that Condition 2 ⇒ Condition 1, show that not Condition 1 ⇒ not Condition 2. Not condition 1 ⇒ $\exists w$ s.t. $\left(u^{-1}\right)'\left(u\left(w\right)\right) = \frac{1}{u'}$ is *concave*: By (17) and using that $u'(w)$ is positive:

$$\begin{aligned} 3u''\left(w\right)^2 - u'''\left(w\right)u'\left(w\right) &< 0, \\ u'''\left(w\right)u'\left(w\right) &> 3u''\left(w\right)^2 \\ \frac{u'''\left(w\right)u'\left(w\right)}{u''\left(w\right)^2} &> 3 \end{aligned}$$

■