Web Appendix to
Binary Payment Schemes: Moral Hazard and Loss Aversion

By Fabian Herweg, Daniel Müller, and Philipp Weinschenk

This appendix consists of three parts. Part B contains the proofs of Proposition 1 and the results of Section IV. In part C, we prove the validity of the first-order approach. In part D, the general case of risk and loss aversion is analyzed.

B. Additional Proofs of Propositions

PROOF OF PROPOSITION 1:

It is readily verified that Assumptions 1-3 from Grossman and Hart (1983) are satisfied. Thus, the cost-minimization problem is well defined, in the sense that for each action $a \in (0,1)$ there exists a second-best incentive scheme. Suppose the principal wants to implement action $\hat{a} \in (0,1)$ at minimum cost. Since the agent's action is not observable, the principal's problem is given by

$$\text{(MR)} \qquad \min_{(u_s)_{s=1}^S} \sum_{s=1}^S \gamma_s(\hat{a}) h(u_s)$$

subject to

$$\text{(IR}_R\text{)} \qquad \sum_{s=1}^S \gamma_s(\hat{a}) u_s - c(\hat{a}) \geq \bar{u} \, ,$$

$$\text{(IC}_R\text{)} \qquad \sum_{s=1}^S (\gamma_s^H - \gamma_s^L) u_s - c'(\hat{a}) = 0 \, .$$

where the first constraint is the individual rationality constraint and the second is the incentive compatibility constraint. Note that the first-order approach is valid, since the agent's expected utility is a strictly concave function of his effort. The Lagrangian to the resulting problem is

$$\mathcal{L} = \sum_{s=1}^S \gamma_s(a) h(u_s) - \mu_0 \left\{ \sum_{s=1}^S \gamma_s(a) u_s - c(a) - \bar{u} \right\} - \mu_1 \left\{ \sum_{s=1}^S (\gamma_s^H - \gamma_s^L) u_s - c'(a) \right\},$$

where μ_0 and μ_1 denote the Lagrange multipliers of the individual rationality constraint and the incentive compatibility constraint, respectively. Setting the partial derivative of $\mathcal{L}$ with respect to u_s equal to zero yields

$$\text{(B.1)} \qquad \frac{\partial \mathcal{L}}{\partial u_s} = 0 \iff h'(u_s) = \mu_0 + \mu_1 \frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}, \ \ \forall s \in \mathcal{S}.$$

Irrespective of the value of μ_0, if $\mu_1 > 0$, convexity of $h(\cdot)$ implies that $u_s > u_{s'}$ if and only if $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) > (\gamma_{s'}^H - \gamma_{s'}^L)/\gamma_{s'}(\hat{a})$, which in turn is equivalent to $\gamma_s^H/\gamma_s^L > \gamma_{s'}^H/\gamma_{s'}^L$. Thus it remains to show that μ_1 is strictly positive. Suppose, in contradiction, that

$\mu_1 \leq 0$. Consider the case $\mu_1 = 0$ first. From (A.1) it follows that $u_s = u^f$ for all $s \in \mathcal{S}$, where u^f satisfies $h'(u^f) = \mu_0$. This, however, violates (IC_R), a contradiction. Next, consider $\mu_1 < 0$. From (A.1) it follows that $u_s < u_{s'}$ if and only if $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) > (\gamma_{s'}^H - \gamma_{s'}^L)/\gamma_{s'}(\hat{a})$. Let $\mathcal{S}^+ \equiv \{s | \gamma_s^H - \gamma_s^L > 0\}$, $\mathcal{S}^- \equiv \{s | \gamma_s^H - \gamma_s^L < 0\}$, and $\hat{u} \equiv \min\{u_s | s \in \mathcal{S}^-\}$. Since $\hat{u} > u_s$ for all $s \in \mathcal{S}^+$, we have

$$\begin{aligned}
\sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) u_s &= \sum_{s \in \mathcal{S}^-} (\gamma_s^H - \gamma_s^L) u_s + \sum_{s \in \mathcal{S}^+} (\gamma_s^H - \gamma_s^L) u_s \\
&< \sum_{s \in \mathcal{S}^-} (\gamma_s^H - \gamma_s^L) \hat{u} + \sum_{s \in \mathcal{S}^+} (\gamma_s^H - \gamma_s^L) \hat{u} \\
&= \hat{u} \sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) \\
&= 0,
\end{aligned}$$

again a contradiction to (IC_R). Hence, $\mu_1 > 0$ and the desired result follows.

PROOF OF PROPOSITION 6:

First consider $b \geq 0$. We divide the analysis for $b \geq 0$ into three subcases.

Case 1 ($a_0 < 0$): For the effort level $\hat{a}$ to be chosen by the agent, this effort level has to satisfy the following incentive compatibility constraint:

$$\text{(IC)} \qquad \hat{a} \in \arg\max_{a \in [0,1]} \; u + \gamma(a) b - \gamma(a)(1 - \gamma(a)) b(\lambda - 1) - \frac{k}{2} a^2.$$

For $\hat{a}$ to be a zero of $dEU(a)/da$, the bonus has to be chosen according to

$$b^*(\hat{a}) = \frac{k\hat{a}}{(\gamma^H - \gamma^L)\left[2 - \lambda + 2\gamma(\hat{a})(\lambda - 1)\right]}.$$

Since $a_0 < 0$, $b^*(a)$ is a strictly increasing and strictly concave function with $b^*(0) = 0$. Hence, each $\hat{a} \in [0,1]$ can be made a zero of $dEU(a)/da$ with a nonnegative bonus. By choosing the bonus according to $b^*(\hat{a})$, $\hat{a}$ satisfies, by construction, the first-order condition. Inserting $b^*(\hat{a})$ into $d^2EU(a)/da^2$ shows that expected utility is strictly concave function if $a_0 < 0$. Hence, with the bonus set equal to $b^*(\hat{a})$, effort level $\hat{a}$ satisfies the second-order condition for optimality and therefore is incentive compatible.

Case 2 ($a_0 = 0$): Just like in the case where $a_0 < 0$, each effort level $a \in [0,1]$ turns out to be implementable with a nonnegative bonus. To see this, consider bonus

$$b_0 = \frac{k}{2(\gamma^H - \gamma^L)^2(\lambda - 1)}.$$

For $b < b_0$, $dEU(a)/da < 0$ for each $a > 0$, that is, lowering effort increases expected utility. Hence, the agent wants to choose an effort level as low as possible and therefore exerts no effort at all. If, on the other hand, $b > b_0$, then $dEU(a)/da > 0$. Now, increasing effort increases expected utility, and the agent wants to choose effort as high as possible. For $b = b_0$, expected utility is constant over all $a \in [0,1]$, that is, as long as his participation constraint is satisfied, the agent is indifferent which effort level to choose. As a tie-breaking rule we assume that, if indifferent between several effort levels, the agent chooses the effort level that the principal prefers.

Case 3 ($a_0 > 0$): If $a_0 > 0$, the agent either chooses $a = 0$ or $a = 1$. To see this, again consider bonus b_0. For $b \leq b_0$, $dEU(a)/da < 0$ for each $a > 0$. Hence, the agent wants to exert as little effort as possible and chooses $a = 0$. If, on the other hand, $b > b_0$, then $d^2EU(a)/da^2 > 0$, that is, expected utility is a strictly convex function of effort. In order to maximize expected utility, the agent will choose either $a = 0$ or $a = 1$ depending on whether $EU(0)$ exceeds $EU(1)$ or not.

Negative Bonus: $b < 0$

Let $b^- < 0$ denote the monetary punishment that the agent receives if the good signal is observed. With a negative bonus, the agent's expected utility is

$$EU(a) = u + \gamma(a)b^- + \gamma(a)(1-\gamma(a))\lambda b^- + (1-\gamma(a))\gamma(a)(-b^-) - \frac{k}{2}a^2. \tag{B.2}$$

The first derivative with respect to effort,

$$\frac{dEU(a)}{da} = \underbrace{(\gamma^H - \gamma^L)b^- \left[\lambda - 2\gamma(a)(\lambda - 1)\right]}_{MB^-(a)} - \underbrace{ka}_{MC(a)},$$

reveals that $MB^-(a)$ is a positively sloped function, which is steeper the harsher the punishment is, that is, the more negative b^- is. It is worthwhile to point out that if bonus and punishment are equal in absolute value, $|b^-| = b$, then also the slopes of $MB^-(a)$ and $MB(a)$ are identical. The intercept of $MB^-(a)$ with the horizontal axis, a_0^- again is completely determined by the model parameters:

$$a_0^- = \frac{\lambda - 2\gamma^L(\lambda - 1)}{2(\gamma^H - \gamma^L)(\lambda - 1)}.$$

Note that $a_0^- > 0$ for $\gamma^L \leq 1/2$. For $\gamma^L > 1/2$ we have $a_0^- < 0$ if and only if $\lambda > 2\gamma^L/(2\gamma^L - 1)$. Proceeding in exactly the same way as in the case of a nonnegative bonus yields a familiar results: effort level $\hat{a} \in [0, 1]$ is implementable with a strictly negative bonus if and only if $a_0^- \leq 0$. Finally, note that $a_0 < a_0^-$. Hence a negative bonus does not improve the scope for implementation.

PROOF OF PROPOSITION 7:

Throughout the analysis we restricted attention to nonnegative bonus payment. It remains to be shown that the principal cannot benefit from offering a negative bonus payment: implementing action $\hat{a}$ with a negative bonus is at least as costly as implementing action $\hat{a}$ with a positive bonus. In what follows, we make use of notation introduced in the paper as well as in the proof of Proposition 6. Let $a_0(p)$, $a_0^-(p)$, $b^*(p;\hat{a})$, and $u^*(p;\hat{a})$ denote the expressions obtained from a_0, a_0^-, $b^*(\hat{a})$, and $u^*(\hat{a})$, respectively, by replacing $\gamma(\hat{a})$, γ^L, and γ^H with $\gamma(p,\hat{a})$, $\gamma^L(p)$, and $\gamma^H(p)$. From the proof of Proposition 6 we know that (i) action $\hat{a}$ is implementable with a nonnegative bonus (negative bonus) if and only if $a_0(p) \leq 0$ ($a_0^-(p) \leq 0$), and (ii) $a_0^-(p) \leq 0$ implies $a_0(p) < 0$. We will show that, for a given value of p, if $\hat{a}$ is implementable with a negative bonus then it is less costly to implement $\hat{a}$ with a nonnegative bonus.

Consider first the case where $a_0^-(p) < 0$. The negative bonus payment satisfying incentive compatibility is given by

$$b^-(p;\hat{a}) = \frac{k\hat{a}}{(\gamma^H(p)-\gamma^L(p))\left[\lambda - 2\gamma(p,\hat{a})(\lambda-1)\right]}.$$

It is easy to verify that the required punishment to implement $\hat{a}$ is larger in absolute value than than the respective nonnegative bonus which is needed to implement $\hat{a}$, that is, $b^*(p;\hat{a}) < |b^-(p;\hat{a})|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$. When punishing the agent with a negative bonus $b^-(p;\hat{a})$, $u^-(p;\hat{a})$ will be chosen to satisfy the corresponding participation constraint with equality, that is,

$$u^-(p;\hat{a}) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^-(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda-1)\right].$$

Remember that, if $\hat{a}$ is implemented with a nonnegative bonus, we have

$$u^*(p;\hat{a}) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[2 - \lambda + \gamma(p,\hat{a})(\lambda-1)\right].$$

It follows immediately that the minimum cost of implementing $\hat{a}$ with a nonnegative bonus is lower than the minimum implementation cost with a strictly negative bonus:

$$\begin{aligned}
C^-(p;\hat{a}) &= u^-(p;\hat{a}) + \gamma(p,\hat{a})b^-(p;\hat{a}) \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^-(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda-1) - 1\right] \\
&> \bar{u} + \frac{k}{2}\hat{a}^2 + \gamma(p,\hat{a})b^*(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda-1) - 1\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[1 - \lambda + \gamma(p,\hat{a})(\lambda-1)\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[2 - \lambda + \gamma(p,\hat{a})(\lambda-1)\right] + \gamma(p,\hat{a})b^*(p;\hat{a}) \\
&= u^*(p;\hat{a}) + \gamma(p,\hat{a})b^*(p;\hat{a}) \\
&= C(p;\hat{a}).
\end{aligned}$$

The same line of argument holds when $a_0^- = 0$: the bonus which satisfies the (IC) is

$$b_0^-(p;\hat{a}) = -\frac{k}{2(\gamma^H(p)-\gamma^L(p))^2(\lambda-1)},$$

and so $b^*(p;\hat{a}) < |b_0^-(p;\hat{a})|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$.

PROOF OF COROLLARY 1:

Let $p \in (0,1)$. With $\hat{\boldsymbol{\zeta}}$ being a convex combination of $\hat{\boldsymbol{\gamma}}$ and $\mathbf{1}$ we have $(\zeta^H,\zeta^L) = p(1,1) + (1-p)(\gamma^H,\gamma^L) = (\gamma^H + p(1-\gamma^H), \gamma^L + p(1-\gamma^L))$. The desired result follows immediately from Proposition 7. Consider $\lambda > 2$. Implementation problems are less likely to be encountered under $\hat{\boldsymbol{\zeta}}$ than under $\hat{\boldsymbol{\gamma}}$. Moreover, if implementation problems are not an issue under both performance measures, then implementation of a certain action is less costly under $\hat{\boldsymbol{\zeta}}$ than under $\hat{\boldsymbol{\gamma}}$. For $\lambda = 2$ implementation problems do not arise and implementation costs are identical under both performance measures. Last, if $\lambda < 2$, implementation problems are not an issue under either performance measure, but the cost of implementation is strictly lower under $\hat{\boldsymbol{\gamma}}$ than under $\hat{\boldsymbol{\zeta}}$.

C. Validity of the First-Order Approach

LEMMA C.1: *Suppose (A1)-(A3) hold, then the incentive constraint in the principal's cost minimization problem can be represented as $EU'(\hat{a}) = 0$.*

PROOF:
Consider a contract $(u_1, (b_s)_{s=2}^S)$ with $b_s \geq 0$ for $s = 2, \ldots, S$. In what follows, we write β_s instead of $\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$ to cut back on notation. The proof proceeds in two steps. First, for a given contract with the property $b_s > 0$ only if $\beta_s > 0$, we show that all actions that satisfy the first-order condition of the agent's utility maximization problem characterize a local maximum of his utility function. Since the utility function is twice continuously differentiable and all extreme points are local maxima, if there exists some action that fulfills the first-order condition, this action corresponds to the unique maximum. In the second step we show that under the optimal contract we cannot have $b_s > 0$ if $\beta_s \leq 0$.

Step 1: The second derivative of the agent's utility with respect to a is

$$EU''(a) = -2(\lambda - 1)\sum_{s=2}^{S} b_s \sigma_s - c''(a) \ , \tag{C.1}$$

where $\sigma_s := (\sum_{i=1}^{s-1}(\gamma_i^H - \gamma_i^L))(\sum_{i=s}^{S}(\gamma_i^H - \gamma_i^L)) < 0$. Suppose action $\hat{a}$ satisfies the first-order condition. Formally

$$\sum_{s=2}^{S} b_s \beta_s = c'(\hat{a}) \iff \sum_{s=2}^{S} b_s \frac{\beta_s}{\hat{a}} = \frac{c'(\hat{a})}{\hat{a}} \ . \tag{C.2}$$

Action $\hat{a}$ locally maximizes the agent's utility if

$$-2(\lambda - 1)\sum_{s=2}^{S} b_s \sigma_s < c''(\hat{a}) \ . \tag{C.3}$$

Under Assumption (A3), we have $c''(\hat{a}) > c'(\hat{a})/\hat{a}$. Therefore, if

$$\sum_{s=2}^{S} b_s \big[-2(\lambda - 1)\sigma_s - \beta_s/\hat{a}\big] < 0 \ , \tag{C.4}$$

then (C.2) implies (C.3), and each action $\hat{a}$ satisfying the first-order condition of the agent's maximization problem is a local maximum of his expected utility. Inequality (C.4) obviously is satisfied if each element of the sum is negative. Summand s is negative if and only if

$$
\begin{aligned}
&-2(\lambda-1)\left(\sum_{i=1}^{s-1}(\gamma_i^H-\gamma_i^L)\right)\left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\hat{a}\\
&\qquad -\left(\sum_{\tau=s}^{S}(\gamma_\tau^H-\gamma_\tau^L)\right)\left[1-(\lambda-1)\left(\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right)\right]\\
&\qquad\qquad +(\lambda-1)\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right]\left(\sum_{t=1}^{s-1}(\gamma_t^H-\gamma_t^L)\right)<0\ .
\end{aligned}
$$

Rearranging the above inequality yields

$$
\begin{aligned}
\left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\left\{\lambda+2(\lambda-1)\left[\hat{a}\sum_{i=1}^{s-1}(\gamma_i^H-\gamma_i^L)-\sum_{i=1}^{s-1}\gamma_i(\hat{a})\right]\right\} &> 0\\
\iff \left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\left\{\lambda\left(1-\sum_{i=1}^{s-1}\gamma_i^L\right)+(2-\lambda)\sum_{i=1}^{s-1}\gamma_i^L\right\} &> 0. \qquad \text{(C.5)}
\end{aligned}
$$

The term in curly brackets is positive, since $\lambda \leq 2$ and $\sum_{i=1}^{s-1}\gamma_i^L < 1$. Furthermore, note that $\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L) > 0$ since $\beta_s > 0$ for all $b_s > 0$. This completes the first step of the proof.

Step 2: Consider a contract with $b_s > 0$ and $\beta_s \leq 0$ for at least one signal $s \in \{2,\ldots,S\}$ that implements $\hat{a} \in (0,1)$. Then, under this contract, (IC′) is satisfied and there exists at least one signal t with $\beta_t > 0$ and $b_t > 0$. Obviously, the principal can reduce both b_s and b_t without violating (IC′). This reasoning goes through up to the point where (IC′) is satisfied and $b_s = 0$ for all signals s with $\beta_s \leq 0$. From the first step of the proof we know that the resulting contract implements $\hat{a}$ incentive compatibly. Next, we show that reducing any spread, say b_k, always reduces the principal's cost of implementation.

$$
C(\boldsymbol{b})=\sum_{s=1}^{S}\gamma_s(\hat{a})h\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\ , \qquad \text{(C.6)}
$$

where

$$
u_1(\boldsymbol{b})=\bar{u}+c(\hat{a})-\sum_{s=2}^{S}b_s\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})-(\lambda-1)\left(\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right)\right]\ .
$$

The partial derivative of the cost function with respect to an arbitrary b_k is

$$
\begin{aligned}
\frac{\partial C(\boldsymbol{b})}{\partial b_k}=&\sum_{s=1}^{k-1}\gamma_s(\hat{a})h'\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\left[\frac{\partial u_1}{\partial b_k}\right]\\
&\qquad +\sum_{s=k}^{S}\gamma_s(\hat{a})h'\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\left[\frac{\partial u_1}{\partial b_k}+1\right]\ .
\end{aligned}
$$

Rearranging yields

$$\text{(C.7)}\quad \frac{\partial C(\boldsymbol{b})}{\partial b_k} = \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h'(u_s) \underbrace{\left[(\lambda - 1)\left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1} \gamma_t(\hat{a})\right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a})\right]}_{<0}$$
$$+ \sum_{s=k}^{S} \gamma_s(\hat{a}) h'(u_s) \underbrace{\left[(\lambda - 1)\left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1} \gamma_t(\hat{a})\right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) + 1\right]}_{>0} .$$

Note $u_s \leq u_{s+1}$ which implies that $h'(u_s) \leq h'(u_{s+1})$. Thus, the following inequality holds

$$\text{(C.8)}\quad \frac{\partial C(\boldsymbol{b})}{\partial b_k} \geq \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1)\left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1} \gamma_t(\hat{a})\right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a})\right]$$
$$+ \sum_{s=k}^{S} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1)\left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1} \gamma_t(\hat{a})\right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) + 1\right] .$$

The above inequality can be rewritten as follows

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} \geq h'(u_k) \left[(\lambda - 1)\left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1} \gamma_t(\hat{a})\right)\right] > 0 .$$

Since reducing any bonus lowers the principal's cost of implementation, it cannot be optimal to set $b_s > 0$ for $\beta_s \leq 0$. This completes the second step of the proof. In combination with Step 1, this establishes the desired result.

D. The General Case: Loss Aversion and Risk Aversion

In this part of the Web Appendix we provide a thorough discussion of the intermediate case where the agent is both risk and loss averse. The agent's intrinsic utility for money is a strictly increasing and strictly concave function, which implies that $h(\cdot)$ is strictly increasing and strictly convex. Moreover, the agent is loss averse, i.e., $\lambda > 1$. From Lemma 1, we know that the constraint set of the principal's problem is nonempty. By relabeling signals, each contract can be interpreted as a contract that offers the agent a (weakly) increasing intrinsic utility profile. This allows us to assess whether the agent perceives receiving u_s instead of u_t as a gain or a loss. As in the case of pure loss aversion, we analyze the optimal contract for a given feasible ordering of signals.

The principal's problem for a given arrangement of the signals is given by

PROGRAM MG:

$$\min_{u_1,\ldots,u_S} \sum_{s=1}^{S} \gamma_s(\hat{a})h(u_s)$$

subject to

$$(\text{IR}_G) \qquad \sum_{s=1}^{S} \gamma_s(\hat{a})u_s - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S} \gamma_s(\hat{a})\gamma_t(\hat{a})[u_t-u_s] - c(\hat{a}) = \bar{u} \ ,$$

$$(\text{IC}_G) \qquad \sum_{s=1}^{S}(\gamma_s^H-\gamma_s^L)u_s - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\left[\gamma_s(\hat{a})(\gamma_t^H-\gamma_t^L)+\gamma_t(\hat{a})(\gamma_s^H-\gamma_s^L)\right][u_t-u_s] = c'(\hat{a}) \ ,$$

$$(\text{OC}_G) \qquad u_S \geq u_{S-1} \geq \ldots \geq u_1 \ .$$

Since the objective function is strictly convex and the constraints are all linear in $\boldsymbol{u} = (u_1,\ldots,u_S)$, the Kuhn-Tucker theorem yields necessary and sufficient conditions for optimality. Put differently, if there exists a solution to the problem (MG) the solution is characterized by the partial derivatives of the Lagrangian associated with (MG) set equal to zero.

LEMMA D.1: *Suppose (A1)-(A3) hold and $h''(\cdot) > 0$, then there exists a second-best optimal incentive scheme for implementing action $\hat{a} \in (0,1)$, denoted $\boldsymbol{u^*} = (u_1^*,\ldots,u_S^*)$.*

PROOF:

We show that program (MG) has a solution, i.e., $\sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s)$ achieves its greatest lower bound. First, from Lemma 1 we know that the constraint set of program (MG) is not empty for action $\hat{a} \in (0,1)$. Next, note that from (IR_G) it follows that $\sum_{s=1}^{S}\gamma_s(\hat{a})u_s$ is bounded below. Following the reasoning in the proof of Proposition 1 of Grossman and Hart (1983), we can artificially bound the constraint set—roughly spoken because unbounded sequences in the constraint set make $\sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s)$ tend to infinity by a result from Dimitri Bertsekas (1974). Since the constraint set is closed, the existence of a minimum follows from Weierstrass' theorem.

In order to interpret the first-order conditions of the Lagrangian to problem (MG) it is necessary to know whether the Lagrangian multipliers are positive or negative.

LEMMA D.2: *The Lagrangian multipliers of program (MG) associated with the incentive compatibility constraint and the individual rationality constraint are both strictly positive, i.e., $\mu_{IC} > 0$ and $\mu_{IR} > 0$.*

PROOF:

Since (IR_G) will always be satisfied with equality due to an appropriate adjustment of the lowest intrinsic utility level offered, relaxing (IR_G) will always lead to strictly lower costs for the principal. Therefore, the shadow value of relaxing (IR_G) is strictly positive, so $\mu_{IR} > 0$.

Next, we show that relaxing (IC_G) has a positive shadow value, $\mu_{IC} > 0$. We do this by showing that a decrease in $c'(\hat{a})$ leads to a reduction in the principal's minimum cost of implementation. Let $(u_s^*)_{s\in\mathcal{S}}$ be the optimal contract under (the original) Program MG, and suppose that $c'(\hat{a})$ decreases. Now the principal can offer a new contract $(u_s^N)_{s\in\mathcal{S}}$ of the form

$$u_s^N = \alpha u_s^* + (1-\alpha)\sum_{t=1}^{S}\gamma_t(\hat{a})u_t^* \ , \tag{D.1}$$

where $\alpha \in (0,1)$, which also satisfies (IR_G), the relaxed (IC_G), and (OC_G), but yields strictly lower costs of implementation than the original contract $(u_s^*)_{s\in\mathcal{S}}$.

Clearly, for $\hat{\alpha} \in (0,1)$, $u_s^N < u_{s'}^N$ if and only if $u_s^* < u_{s'}^*$, so (OC_G) is also satisfied under contract $(u_s^N)_{s\in\mathcal{S}}$.

Next, we check that the relaxed (IC_G) holds under $(u_s^N)_{s\in\mathcal{S}}$. To see this, note that for $\alpha = 1$ we have $(u_s^N)_{s\in\mathcal{S}} \equiv (u_s^*)_{s\in\mathcal{S}}$. Thus, for $\alpha = 1$, the relaxed (IC_G) is oversatisfied under $(u_s^N)_{s\in\mathcal{S}}$. For $\alpha = 0$, on the other hand, the left-hand side of (IC_G) is equal to zero, and the relaxed (IC_G) in consequence is not satisfied. Since the left-hand side of (IC_G) is continuous in α under contract $(u_s^N)_{s\in\mathcal{S}}$, by the intermediate-value theorem there exists $\hat{\alpha} \in (0,1)$ such that the relaxed (IC_G) is satisfied with equality.

Last, consider (IR_G). The left-hand side of (IR_G) under contract $(u_s^N)_{s\in\mathcal{S}}$ with $\alpha = \hat{\alpha}$ amounts to

$$\begin{aligned}
&\sum_{s=1}^{S}\gamma_s(\hat{a})u_s^N - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^N - u_s^N\right] \\
&= \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - \hat{\alpha}(\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&> \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&= \bar{u} + c(\hat{a}) \ ,
\end{aligned} \tag{D.2}$$

where the last equality follows from the fact that $(u_s^*)_{s\in\mathcal{S}}$ fulfills the (IR_G) with equality. Thus, contract $(u_s^N)_{s\in\mathcal{S}}$ is feasible in the sense that all constraints of program (MG) are met. It remains to show that the principal's costs are reduced. Since $h(\cdot)$ is strictly convex, the principal's objective function is strictly convex in α, with a minimum at $\alpha = 0$. Hence, the principal's objective function is strictly increasing in α for $\alpha \in (0,1]$. Since $(u_s^N)_{s\in\mathcal{S}} \equiv (u_s^*)_{s\in\mathcal{S}}$ for $\alpha = 1$, for $\alpha = \hat{\alpha}$ we have

$$\sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s^*) > \sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s^N),$$

which establishes the desired result.

We now give a heuristic reasoning why pooling of information may well be optimal in this more general case. For the sake of argument, suppose there is no pooling of information in the sense that it is optimal to set distinct wages for distinct signals. In this case all order constraints are slack; formally, if $u_s \neq u_{s'}$ for all $s, s' \in \mathcal{S}$ and $s \neq s'$, then $\mu_{OC,s} = 0$ for all $s \in \{2, \dots, S\}$. In this case, the first-order condition of optimality with respect to u_s, $\partial\mathcal{L}(\boldsymbol{u})/\partial u_s = 0$, can be written as follows:

$$\text{(D.3)}\quad h'(u_s) = \underbrace{\left(\mu_{IR} + \mu_{IC}\frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}\right)}_{=:H_s}\underbrace{\left[1-(\lambda-1)\left(2\sum_{t=1}^{s-1}\gamma_t(\hat{a}) + \gamma_s(\hat{a}) - 1\right)\right]}_{=:\Gamma_s} \\ \underbrace{-\,\mu_{IC}(\lambda-1)\left[2\sum_{t=1}^{s-1}(\gamma_t^H - \gamma_t^L) + (\gamma_s^H - \gamma_s^L)\right]}_{=:\Lambda_s}.$$

For $\lambda = 1$ we have $h'(u_s) = H_s$, the standard "Holmström-formula".[1] Note that $\Gamma_s > 0$ for $\lambda \leq 2$. More importantly, irrespective of the signal ordering, we have $\Gamma_s > \Gamma_{s+1}$. The third term, Λ_s, can be either positive or negative. If the compound signal of all signals below s and the signal s itself are bad signals, then $\Lambda_s < 0$.

Since the incentive scheme is nondecreasing, when the order constraints are not binding it has to hold that $h'(u_s) \geq h'(u_{s-1})$. Thus, if $\mu_{OC,s-1} = \mu_{OC,s} = \mu_{OC,s+1} = 0$ the following inequality is satisfied:

$$\text{(D.4)}\qquad H_s \times \Gamma_s - \Lambda_s \geq H_{s-1} \times \Gamma_{s-1} - \Lambda_{s-1}.$$

Note that for the given ordering of signals, if there exists any pair of signals $s, s-1$ such that (D.4) is violated, then the optimal contract for this ordering involves pooling of wages. Even when $H_s > H_{s-1}$, as it is the case when signals are ordered according to their likelihood ratio, it is not clear that inequality (D.4) is satisfied. In particular, when s and $s-1$ are similarly informative it seems to be optimal to pay the same wage for these two signals as can easily be illustrated for the case of two good signals: If s and $s-1$ are similarly informative good signals then $H_s \approx H_{s-1} > 0$ but $\Gamma_s < \Gamma_{s-1}$ and $\Lambda_s > \Lambda_{s-1}$, thus condition (D.4) is violated. In summary, it may well be that for a given incentive-feasible ordering of signals, and thus overall as well, the order constraints are binding, i.e., it may be optimal to offer a contract which is less complex than the signal space allows for.

Application with Constant Relative Risk Aversion.—Suppose $h(u) = u^r$, with $r \geq 0$ being a measure for the agent's risk aversion. More precisely, the Arrow-Pratt measure for relative risk aversion of the agent's intrinsic utility function is $R = 1 - \frac{1}{r}$ and therefore constant. The following result states that the optimal contract is still a bonus contract when the agent is not only loss averse, but also slightly risk averse.

PROPOSITION D.1: *Suppose (A1)-(A3) hold, $h(u) = u^r$ with $r > 1$, and $\lambda > 1$. Generically, for r sufficiently small the optimal incentive scheme $(u_s^*)_{s=1}^S$ is a bonus scheme, i.e., $u_s^* = u_H^*$ for $s \in \mathcal{B}^* \subset \mathcal{S}$ and $u_s^* = u_L^*$ for $s \in \mathcal{S}\backslash\mathcal{B}^*$ where $u_L^* < u_H^*$.*

PROOF:

For the agent's intrinsic utility function being sufficiently linear, the principal's costs are approximately given by a second-order Taylor polynomial about $r = 1$, thus

$$\text{(D.5)}\qquad C(\boldsymbol{u}|r) \approx \sum_{s\in\mathcal{S}}\gamma_s(\hat{a})u_s + \Omega(\boldsymbol{u}|r)\ ,$$

[1]See Holmström (1979).

where

$$\Omega(\boldsymbol{u}|r) \equiv \sum_{s\in\mathcal{S}} \gamma_s(\hat{a})\left[(u_s \ln u_s)(r-1) + (1/2)u_s(\ln u_s)^2(r-1)^2\right] . \tag{D.6}$$

Relabeling signals such that the wage profile is increasing allows us to express the incentive scheme in terms of increases in intrinsic utility. The agent's binding participation constraint implies that

$$u_1 = \bar{u} + c(\hat{a}) - \sum_{s=2}^{S} b_s \left\{\sum_{\tau=s}^{S}\gamma_\tau(\hat{a}) - (\lambda-1)\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right]\left[\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right]\right\} \equiv u_1(\boldsymbol{b}) \tag{D.7}$$

and $u_s = u_1(\boldsymbol{b}) + \sum_{t=2}^{s} b_t \equiv u_s(\boldsymbol{b})$ for all $s = 2, \ldots, S$. Inserting the binding participation constraint into the above cost function and replacing $\Omega(\boldsymbol{u}|r)$ equivalently by $\tilde{\Omega}(\boldsymbol{b}|r) \equiv \Omega(u_1(\boldsymbol{b}), \ldots, u_S(\boldsymbol{b})|r)$ yields

$$C(\boldsymbol{b}|r) \approx \bar{u} + c(\hat{a}) + (\lambda-1)\sum_{s=2}^{S} b_s \left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right]\left[\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right] + \tilde{\Omega}(\boldsymbol{b}|r) . \tag{D.8}$$

Hence, for a given increasing wage profile the principal's cost minimization problem is:

Program ME:

$$\min_{\boldsymbol{b}\in\mathbb{R}_+^{S-1}} \ \boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) + \tilde{\Omega}(\boldsymbol{b}|r)$$

$$\text{subject to} \quad \boldsymbol{b}'\boldsymbol{\beta}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \tag{IC$'$}$$

If r is sufficiently close to 1, then the incentive scheme that solves Program ML also solves Program ME. Note that generically Program ME is solved only by bonus schemes. Put differently, even if there are multiple optimal contracts for Program ML, all these contracts are generically simple bonus contracts. Thus, from Proposition 2 it follows that generically for r close to 1 the optimal incentive scheme entails a minimum of wage differentiation. Note that for $\lambda = 1$ the principal's problem is to minimize $\tilde{\Omega}(\boldsymbol{b}|r)$ even for r sufficiently close to 1.

REFERENCES

Bertsekas, Dimitri. 1974. "Necessary and Sufficient Conditions for Existence of an Optimal Portfolio." *Journal of Economic Theory*, 8(2): 235–47.