Appendix for: Power of Incentives in Private vs. Public Organizations

by
Avinash Dixit
Princeton University

Revised October 2004

Notation

There are n principals indexed by i, and an agent. The agent's effort is a vector $x \in R^k$; the outcome is a vector $y \in R^m$; the two are linked by

$$y = F\,x + \epsilon \tag{A-1}$$

where F is an m-by-k matrix, and ϵ is an m-dimensional random error vector, normally distributed with mean 0 and variance-covariance matrix Ω.

If the i-th principal pays the agent z_i, that principal's expected utility is

$$u_i = \mathrm{E}\,[\,-\exp\{-r_i\,(b_i'\,y - z_i)\,\}\,] \tag{A-2}$$

and the agent's expected utility is

$$u_a = \mathrm{E}\,[\,-\exp\{-r_a\,(b_a'\,y + \Sigma_i\,z_i - \tfrac{1}{2}\,x'Cx)\,\}\,]\,. \tag{A-3}$$

The agent's opportunity utility level is $\underline{u}_a$; thus his participation constraint is $u_a \geq \underline{u}_a$.

Thus all players have constant absolute risk aversion, r_i for principal i and r_a for the agent. All value outputs linearly; the vectors b_i and b_a are their unit valuations. The quadratic form $\frac{1}{2}\,x'Cx$ is the agent's disutility of effort, and C is a k-by-k positive definite matrix.

The unit valuations b_i and b_a, the risk aversion parameters r_i and r_a, the agent's outside utility u_a, and the matrices F, C and Ω, are all common knowledge.

First-Best Risk Sharing

As a hypothetical ideal standard, consider the situation where the agent's effort x and the outcome y are directly observable and verifiable. Then the agent's effort can be stipulated as a part of the contract, ϵ can be inferred, and the payments z_i can be conditioned on it.

The principals' Pareto optimum will be found by maximizing

$$\begin{aligned}\sum_i \theta_i\, u_i + \lambda\, u_a \;&=\; \sum_i \theta_i \;\mathrm{E}[-\exp\{-r_i\,[b_i'\,(Fx+\epsilon) - z_i(\epsilon)]\,\}\,] \\ &\quad +\lambda\;\mathrm{E}[-\exp\{-r_a\,[b_a'(Fx+\epsilon) + \Sigma_i\, z_i(\epsilon) - \tfrac{1}{2}\,x'Cx]\,\}\,]\,.\end{aligned}$$

Here θ_i are the weights attached to the principals' utilities (which could be implicitly determined as the Lagrange multipliers on the problem of maximizing the utility of one of them subject to all the others achieving specified minimum levels), and λ is the Lagrange multiplier on the agent's participation constraint.

The first-order condition for x is

$$\Sigma_i\;\theta_i\;u_i\;r_i\,F'\,b_i + \lambda\;u_a\;r_a\,(F'b_a - Cx) = 0\,.$$

The conditions for the $z_i(\epsilon)$ are

$$\begin{aligned}&-\theta_i\;\;\exp\{-r_i\,[b_i'\,(Fx+\epsilon) - z_i(\epsilon)]\,\}\,r_i \\ &+\lambda\;\;\exp\{-r_a\,[b_a'(Fx+\epsilon) + \Sigma_i\,z_i(\epsilon) - \tfrac{1}{2}\,x'Cx]\,\}\;r_a \;\;=\;\; 0.\end{aligned}$$

Multiplying these by $F'\,b_i$, taking expectations over ϵ and adding over i, we get

$$-\Sigma_i\;\theta_i\;u_i\;r_i\,F'\,b_i + \lambda\;u_a\;r_a\,F'\,b_0 = 0\,,$$

where $b_0 \;=\; \Sigma_i\,b_i$ is the principals' aggregate unit valuation vector. Substituting into the condition for x and simplifying, we find

$$C\,x = F'\,(b_0 + b_a)\,,$$

or

$$x = C^{-1}\,F'\,(b_0 + b_a)\,. \tag{A-4}$$

The first-best optimal effort is independent of all parties' risk-aversion parameters; this is due to the assumption of constant absolute risk-aversion.

Now let $z(\epsilon) = \Sigma_i\,z_i(\epsilon)$. Taking logs in the first-order condition for $z_i(\epsilon)$, we find

$$r_i\,[b_i'\,\epsilon - z_i(\epsilon)] = r_a\,[b_a'\,\epsilon + z(\epsilon)] + T_i\,,$$

where in this context the symbol T is used as generic notation to indicate terms independent of ϵ. Dividing by r_i, adding over i, and rearranging terms, we get

$$z(\epsilon) = \frac{r_0\,b_0' - r_a\,b_a'}{r_0 + r_a}\;\epsilon + T\,,$$

where r_0 is defined by

$$1/r_0 = \Sigma_i\;(1/r_i)\,. \tag{A-5}$$

In other words, the principals act like a single entity with an aggregate risk-tolerance (reciprocal of risk-aversion) equal to the sum of their individual risk-tolerances. Then each party's exposure to risk, $[b_i'\,\epsilon - z_i(\epsilon)]$ for the principals and $[b_a'\,\epsilon + z(\epsilon)]$ for the agent, is proportional to his individual risk-tolerance. This is a standard result in the general theory of optimal risk-bearing, especially in the context of financial markets; see Huang and Litzenberger (1988, pp. 134–5).

Constrained Optimal Incentives

From now on I assume that the agent's effort x is not observable by the principals or verifiable. Unless explicitly stated, the outcome y is observable by all players and verifiable.

I restrict the principals' payments to linear functions

$$z_i = \alpha_i' \, y + \beta_i \, ; \tag{A-6}$$

leading to an aggregate payment

$$z = \alpha' \, y + \beta \, , \tag{A-7}$$

where $\alpha = \Sigma_i \; \alpha_i$ and $\beta = \Sigma_i \; \beta_i$.

With a single principal, restriction to linear schemes can be rigorously justified if the model is the reduced form of a dynamic one where the agent controls the drift of a Brownian motion; see Holmström and Milgrom (1987). However, such a formulation is used in applications without explicitly invoking the dynamic model, because the optimal linear strategies are easily computable and yield useful insights; see Holmström and Milgrom (1991) and Dixit (1996, appendix). When I consider a non-cooperative game among several principals, if in the corresponding dynamic game all the other principals are using linear strategies, then the optimal response of any one principal can be achieved by using a linear strategy. Thus there is an equilibrium in which strategies are linear, but there may be other equilibria involving complex and nonlinear history-dependent strategies. Again, the linear equilibrium is useful because of its tractability and intuitive results; see Dixit (1996, Appendix) for a similar and less general model.

Using the i-th principal's linear payment function (A-6) in his expected utility function (A-2), we find

$$\begin{aligned} u_i &= \mathrm{E}[-\exp\{-r_i\,[(b_i-\alpha_i)'\,y-\beta_i]\,\}\,] \\ &= \mathrm{E}[-\exp\{-r_i\,[(b_i-\alpha_i)'\,(Fx+\epsilon)-\beta_i]\,\}\,] \\ &= -\exp\{-r_i\,[(b_i-\alpha_i)'\,Fx-\tfrac{1}{2}\,r_i\,(b_i-\alpha_i)'\,\Omega\,(b_i-\alpha_i)-\beta_i]\,\}\,, \end{aligned}$$

where the last step uses the standard formula for the expectation of an exponential (the moment generating function) of a multidimensional normal variate; see Billingsley (1986, p. 286). Principal i's choice problem therefore simplifies to that of maximizing the "certainty-equivalent income"

$$CE_i = (b_i-\alpha_i)'\,Fx-\tfrac{1}{2}\,r_i\,(b_i-\alpha_i)'\,\Omega\,(b_i-\alpha_i)-\beta_i \tag{A-8}$$

Likewise, substituting the aggregate linear payment function (A-7) into the agent's expected utility function (A-3) and simplifying, the agent's problem reduces to maximizing his certainty-equivalent income

$$CE_a = (b_a+\alpha)'\,F\,x-\tfrac{1}{2}\,r_a\,(b_a+\alpha)'\,\Omega\,(b_a+\alpha)-\tfrac{1}{2}\,x'\,C\,x+\beta\,. \tag{A-9}$$

Given the aggregate incentive scheme (A-7), the agent chooses x to maximize his own certainty-equivalent income defined by (A-9). The first-order condition for this is

$$F'\,(b_a+\alpha)-C\,x=0\,,$$

yielding

$$x = C^{-1} F' (b_a + \alpha) . \tag{A-10}$$

Using this, we get the "indirect" forms of utility functions for the principals and the agent can. The corresponding certainty-equivalent incomes are

$$CE_i = (b_i - \alpha_i)' G (b_a + \alpha) - \tfrac{1}{2} r_i (b_i - \alpha_i)' \Omega (b_i - \alpha_i) - \beta_i , \tag{A-11}$$

$$CE_a = \tfrac{1}{2} (b_a + \alpha)' G (b_a + \alpha) - \tfrac{1}{2} r_a (b_a + \alpha)' \Omega (b_a + \alpha) + \beta , \tag{A-12}$$

where

$$G = F C^{-1} F' \tag{A-13}$$

is an m-by-m positive definite matrix.

Colluding Principals

Suppose the principals collusively choose the parameters α_i, β_i of their linear schemes, to achieve a Pareto optimum subject to the agent's participation constraint.

The principals can change the β_i by offsetting amounts to achieve lump-sum transfers among themselves without affecting the agent's participation or incentives. Similarly, the aggregate β serves to achieve transfers between the principals and the agent, and thereby to meet the agent's participation constraint. Therefore the optimum should maximize the sum of the certainty-equivalent incomes of all players, and the aggregate α is the crucial choice variable that controls the agent's incentive to make effort. Thus the principals' problem becomes to maximize

$$(b_0 - \alpha)' G (b_a + \alpha) + \tfrac{1}{2} (b_a + \alpha)' G (b_a + \alpha) - \tfrac{1}{2} \sum_i r_i (b_i - \alpha_i)' \Omega (b_i - \alpha_i) - \tfrac{1}{2} r_a (b_a + \alpha)' \Omega (b_a + \alpha) .$$

The first-order condition for α_i is

$$\begin{aligned} 0 &= G (b_0 - \alpha) - G (b_a + \alpha) + G (b_a + \alpha) + r_i \Omega (b_i - \alpha_i) - r_a \Omega (b_a + \alpha) \\ &= G (b_0 - \alpha) + r_i \Omega (b_i - \alpha_i) - r_a \Omega (b_a + \alpha) . \end{aligned}$$

Divide this by r_i and add over i. Using the definition (A-5) of the principals' aggregate risk-aversion r_0, we have

$$(1/r_0) G (b_0 - \alpha) + \Omega (b_0 - \alpha) - (r_a/r_0) \Omega (b_a + \alpha) = 0.$$

Finally, writing a^j for the incentive payment vector in the principals' joint optimum, we have

$$[G + (r_0 + r_a) \Omega] (b_a + \alpha^j) = [G + r_0 \Omega] (b_0 + b_a) . \tag{A-14}$$

This is equation (4) in the text.

The α_i of the individual principals can, if desired, be retrieved by using this in the separate first-order conditions for each i.

The equation can be solved explicitly for α, but the form above is more informative. The vector on the right-hand side, $(b_0 + b_a)$, is the sum of everyone's unit valuations, while the vector on the left-hand side is the agent's overall benefit from the marginal output – his own valuation vector b_a plus the payment vector α offered by the principals – and it determines his effort. Thus we see how the overall valuation affects the effort and the outcomes. We have

$$x = C^{-1} F \, [G + (r_0 + r_a)\, \Omega]^{-1} \, [G + r_0 \, \Omega] \, (b_0 + b_a) \, ,$$

and the resulting expected output is

$$\mathrm{E}[y] = G \, [G + (r_0 + r_a)\, \Omega]^{-1} \, [G + r_0 \, \Omega] \, (b_0 + b_a) \, .$$

The economic interpretation and some implications of (A-14) are discussed in the text. Here I briefly note some further points: [1] The principals pool their risks and have a large risk-tolerance (low risk-aversion). [2] The principals recognize the agent's own concern for output, and partially offset their incentive payments.

Competing Principals

Next let principals act non-cooperatively. Consider the Nash equilibrium of their choices of α_i and β_i, taking into account the agent's optimum response, given by (A-10) to the aggregate incentives he faces.

First we determine one principal's best response to given strategies of others. We ask what would happen if this principal were not to offer any incentives, compare that to what happens when he does, and thereby calculate the surplus that is due to the bilateral interaction between him and the agent. Then his α_1 is chosen to maximize this surplus, and β_1 to divide it between him and the agent so as to meet the latter's participation constraint. This procedure follows Holmström and Milgrom (1991).

Fix on say principal 1, and denote the parameters of the aggregate schemes of the rest by

$$A_1 = \sum_{i \neq 1} \alpha_i, \qquad B_1 = \sum_{i \neq 1} \beta_i \, .$$

If principal 1 were to offer nothing, the agent would choose

$$x = C^{-1} F' \, (b_a + A_1) \, .$$

Using this in the formulas (A-11) and (A-12) above, principal 1's certainty-equivalent income would be

$$b_1 \, G \, (b_a + A_1) - \tfrac{1}{2} \, r_1 \, b_1' \, \Omega \, b_1 \, ,$$

and the agent's,

$$\tfrac{1}{2} \, (b_a + A_1)' \, G \, (b_a + A_1) - \tfrac{1}{2} \, r_a \, (b_a + A_1)' \, \Omega \, (b_a + A_1) + B_1 \, .$$

When principal 1 does offer a payment scheme $z_1 = \alpha_1 \, y + \beta_1$, the expressions (A-11) and (A-12) apply, with the aggregate parameters $\alpha = \alpha_1 + A_1$ and $\beta = \beta_1 + B_1$.

Subtracting, we find the expression for the bilateral surplus,

$$b_1'\, G\, \alpha_1 - \tfrac{1}{2}\, \alpha_1'\, G\, \alpha_1 + r_1\, b_1'\, \Omega\, \alpha_1 - \tfrac{1}{2}\, r_1\, \alpha_1'\, \Omega\, \alpha_1 - r_a\, (b_a + A_1)'\, \Omega\, \alpha_1 - \tfrac{1}{2}\, r_a\, \alpha_1'\, \Omega\, \alpha_1\, .$$

To maximize this, the first-order condition for α_1 is

$$G\, b_1 - G\, \alpha_1 + r_1\, \Omega\, b_1 - r_1\, \Omega\, \alpha_1 - r_a\, \Omega\, (b_a + A_1) - r_a\, \Omega\, \alpha_1 = 0\, .$$

Using $A_1 = \alpha - \alpha_1$, this can be written

$$[G + r_1\, \Omega]\, \alpha_1 = [G + r_1\, \Omega]\, b_1 - r_a\, \Omega\, (b_a + \alpha)\, .$$

There are similar equations for each principal. Multiply the i-th by $[G + r_i\, \Omega]^{-1}$, and add across i, to get

$$\sum_i \alpha_i = \sum_i\; b_i - \left\{ \sum_i\; [G + r + i\, \Omega]^{-1} \right\}\; r_a\, \Omega\, (b_a + \alpha)\, .$$

Finally, denote the aggregate incentive parameter α in this Nash equilibrium where the principals act separately by α^s, to distinguish it from the α^j above where the principals act jointly. Then

$$\left\{\, I + r_a\; \sum_i\; [G + r_i\, \Omega]^{-1}\; \Omega \right\}\; (b_a + \alpha^s) = b_0 + b_a\, . \tag{A-15}$$

If all principals have the equal risk-aversion, this takes the special form

$$[\, G + n\, (r_0 + r_a)\, \Omega\,]\; (b_a + \alpha^s) = [\, G + n\, r_0\, \Omega\,]\; (b_0 + b_a)\, , \tag{A-16}$$

which is equation (5) in the text. The economic implications are discussed in the text.

Most importantly, if we let n go to infinity holding risk-aversions of all individual players fixed, $n\, r_0$ has a positive finite limit while $n\, r_a$ goes to ∞. Then (A-15) or (A-16) reduce to $b_a + \alpha^s = 0$.

The annihilation of incentives with a large number of principals arises because of their interaction. This is most easily seen in the case where the number of outputs m equals the number of principals n, and each principal cares for only one dimension of the output, so only the i-th component of b_i is positive and all others are zero. But the α_i have negative components for $j \neq i$. Thus principal i pays the agent more when output of j is less; in other words, the principal offers insurance to the agent for risks associated with outputs $j \neq i$. Since principal i does not care about these outputs, he is not concerned about the resulting moral hazard, namely the agent's diminution of effort in those dimensions. The other principals bear this cost. This negative externality is as usual oversupplied in the non-cooperative Nash equilibrium of the principals. The existence of these negative incentives for $j \neq i$ also reduces the benefit that principals j stand to get by offering positive incentives for the dimensions j that concern them. The result is that incentives are weak all round.

Restricted Contracts

The above discussion suggests that power of incentives might be restored if each principal were restricted to schemes that depended only on the dimension of output that is of most concern to him. This might be done by restricting observability (each principal is prohibited from seeing the outputs that concern the other principals), or by not allowing them to make use of these observations (constraining contracts).

Continue the assumption at the end of the last section, namely that each principal is concerned with only one dimension of the outcome: $b_{ij} = 0$ for $i, j = 1, 2, \ldots n, i \neq j$. Add the restriction that each α_i can have only the i-th component α_{ii} non-zero, the bilateral surplus between principal 1 and the agent becomes[1]

$$b_{11}\, G_{11}\, \alpha_{11} - \tfrac{1}{2}\, G_{11}\, (\alpha_{11})^2 + r_1\, b_{11}\, \Omega_{11}\, \alpha_{11} - \tfrac{1}{2}\, (r_1 + r_a)\, \Omega_{11}\, (\alpha_{11})^2$$
$$- r_a \sum_{j=1}^{n} b_{aj}\, \Omega_{j1}\, \alpha_{11} - r_a \sum_{j=2}^{n} A_{1j}\, \Omega_{j1}\, \alpha_{11}\,,$$

leading to the first-order condition

$$[G_{11} + (r_1 + r_a)\, \Omega_{11}]\; \alpha_{11} = [G_{11} + r_1\, \Omega_{11}]\, b_{11} - r_a \sum_{j=1}^{n} b_{aj}\, \Omega_{j1} - r_a \sum_{j=2}^{n} A_{1j}\, \Omega_{j1}\,.$$

The most important new feature is the absence of the factor n multiplying the risk-aversion parameters. This is most easily seen if the error variance-covariance matrix Ω is diagonal, when the first-order condition immediately yields a solution for α_{11}:

$$[G_{11} + (r_1 + r_a)\, \Omega_{11}]\; \alpha_{11} = [G_{11} + r_1\, \Omega_{11}]\, b_{11} - r_a\, \Omega_{11}\, b_{a1}\,.$$

Now, if the different dimensions of effort are close substitutes in the agent's utility function, the matrix C is nearly singular and G_{11} is large. The solution is $\alpha_{11} \approx b_{11}$: the principal offers the agent full-powered incentives, with a bonus coefficient almost equal to the principal's own marginal valuation of the outcome. If the agent does not value outcomes directly, this approximates the first-best.

This is another instance of the general theory of the second best: the externality arising from the principals' non-cooperation may be partially offset by worsening another problem, namely that of asymmetric observability of effort. Of course the risk-sharing is not first-best, so the full optimum is not replicated.

[1] I thank Mr. Xinshuai Guo of the University of Science and Technology of China for pointing out an error in the original version of this formula.

References

Billingsley, Patrick. 1986. *Probability and Measure.* Second edition. New York: John Wiley.

Dixit, Avinash. 1996. *The Making of Economic Policy: A Transaction-Cost Politics Perspective.* Cambridge, MA: MIT Press.

Holmström, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55, no. 2: 303–328.

Holmström, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization,* vol. 7, Special Issue, pp. 24–51.

Hunag, Chi-fu and Robert H. Litzenberger. 1988. *Foundations for Financial Economics.* Amsterdam: North-Holland.