

Moral hazard in teams

KC Border

November 2004

These notes are based on the first part of “Moral hazard in teams” by Bengt Holmstrom [1], and fills in the gaps in the proof in his appendix. The problem is to

find a scheme to compensate team members when individuals cannot observe the effort level of others, only the total output of the team.

Technology

There are $n > 1$ **agents**. Each agent i chooses an **action** or **effort level** $a_i \in \mathbf{R}_+$. The team’s monetary reward x depends on the effort of each agent. That is,

$$x: \mathbf{R}_+^n \rightarrow \mathbf{R}_+.$$

We assume that x is continuous, strictly increasing, concave, and differentiable.

Tastes

Each agent cares only about his effort level and his monetary compensation. We will consider the special case of quasi-linear utility

$$u_i(m, a) = m - v_i(a),$$

where m is monetary compensation and a is effort. The value $v_i(a)$ gives the minimum compensation need to induce agent i to exert effort level a .

We will assume that each v_i is continuous, strictly increasing, convex, and differentiable.

Allocations

An **allocation** is an ordered list

$$(m, a) = ((m_1, \dots, m_n), (a_1, \dots, a_n))$$

satisfying

$$\sum_{i=1}^n m_i \leq x(a_1, \dots, a_n). \tag{A}$$

The interpretation is that m_i is agent i ’s monetary compensation, and a_i is his effort. The condition (A) requires that the compensation be derived from the team’s output.

Efficiency

An allocation is **efficient** if no other allocation gives every agent greater utility.

1 Proposition *Due to the quasi-linearity of utility, the allocation (m^*, a^*) is efficient if and only $a^* = (a_1^*, \dots, a_n^*)$ maximizes the **total surplus***

$$S(a) = x(a) - \sum_{i=1}^n v_i(a_i), \quad (\text{S})$$

and the total reward is fully distributed:

$$\sum_{i=1}^n m_i^* = x(a^*). \quad (\text{D})$$

Proof: (\Leftarrow) Assume that (m^*, a^*) maximizes (S) and satisfies (D). The for any other allocation (m', a') we have

$$\sum_{i=1}^n m'_i - v_i(a'_i) \leq x(a') - \sum_{i=1}^n v_i(a'_i) \leq x(a^*) - \sum_{i=1}^n v_i(a_i^*) = \sum_{i=1}^n m_i^* - v_i(a_i^*),$$

which implies that we cannot have $u_i(m'_i, a'_i) = m'_i - v_i(a'_i) > m_i^* - v_i(a_i^*) = u_i(m_i^*, a_i^*)$ for each i . In other words, (m^*, a^*) is efficient.

(\Rightarrow) Assume by way of contraposition that either a^* does not maximize (S) or violates (D). That is, there is some a' (possibly $a' = a^*$) satisfying

$$x(a') - \sum_{i=1}^n v_i(a'_i) > \sum_{i=1}^n m_i^* - \sum_{i=1}^n v_i(a_i^*),$$

so define

$$c = x(a') - \sum_{i=1}^n (m_i^* - v_i(a_i^*) + v_i(a'_i)) > 0.$$

Setting $m'_i = m_i^* - v_i(a_i^*) + v_i(a'_i) + c/n$ gives $u_i(m'_i, a'_i) = m'_i - v_i(a'_i) > m_i^* - v_i(a_i^*) = u_i(m_i^*, a_i^*)$ for each i , and $\sum_{i=1}^n m'_i = x(a')$. That is, (m^*, a^*) is not efficient. \blacksquare

Nontriviality

We now add another key assumption.

2 Assumption (Nontriviality) *Assume that a surplus maximizer $a^* \gg 0$ exists, and that the surplus is positive:*

$$x(a^*) - \sum_{i=1}^n v_i(a_i^*) > 0.$$

This is the simplest assumption we could make, but it is not a primitive assumption. The following additional assumptions imply that if a maximizer a^* exists, then it must have $a_i^* > 0$: $x(0) = 0$, $v_i(0) = 0$ and $v'_i(0) = 0$ for all i . (I'll leave the proof as an exercise.) In order to guarantee that a maximizer exists, sufficient conditions are that for all i , $v'_i(a) \rightarrow \infty$ as $a \rightarrow \infty$, or for all i , $D_i(a) \rightarrow 0$ as $a_i \rightarrow \infty$. (Again, I'll leave it to you to figure out why.)

Incentives

Since only the output of the team is observable, compensation can depend only on the total output. A **sharing rule** is a function

$$s: \mathbf{R}_+ \rightarrow \mathbf{R}^n$$

satisfying the budget balance condition

$$\sum_{i=1}^n s_i(x) = x \quad \text{for all } x \in \mathbf{R}_+. \quad (\text{B})$$

The interpretation is that $s_i(x)$ is agent i 's compensation when the total output is x . Condition (B) just asserts that the output is divided among the team members. Note that we allow $s_i(x)$ to be negative, which amounts to fining agent i .

A sharing rule defines a **game** among the agents, where the **strategies** are effort levels and the **payoff functions** are given by

$$\pi_i(a_1, \dots, a_n) = s_i(x(a_1, \dots, a_n)) - v_i(a_i).$$

An ordered strategy list $\bar{a} = (\bar{a}_1, \dots, \bar{a}_n)$ is a **Cournot–Nash equilibrium** if for every agent i and every effort level a_i ,

$$s_i(x(\bar{a}_{-i}, a_i)) - v_i(a_i) \leq s_i(x(\bar{a})) - v_i(\bar{a}_i),$$

where $(\bar{a}_{-i}, a_i) = (\bar{a}_1, \dots, \bar{a}_{i-1}, a_i, \bar{a}_{i+1}, \dots, \bar{a}_n)$. That is, no agent can unilaterally deviate from \bar{a} and get higher payoff.

In this case we also say that the sharing rule s **implements** the effort vector \bar{a} .

Incentives vs. efficiency

3 Theorem *Under the assumptions we have made, if $a^* \gg 0$ maximizes surplus (S), then there is no sharing rule satisfying budget balance (B), for which a^* is an equilibrium.*

Fake Proof: The first order conditions for surplus maximization are

$$D_i x(a^*) - v'_i(a_i^*) = 0, \quad i = 1, \dots, n. \quad (*)$$

If a^* is also an equilibrium, it satisfies the first order conditions

$$s'_i(x(a^*)) D_i x(a^*) - v'_i(a_i^*) = 0, \quad i = 1, \dots, n.$$

Together these imply $s'_i(x(a^*)) = 1$ for each i . But by budget balance $\sum_{i=1}^n s'_i(x) = 1$ for every x . But this contradicts $n > 1$. ■

Why is this not a proof? What we have proved is that no differentiable sharing rule can implement a^* . But we are allowed to pick rules that are not differentiable, and it may be that we will want to.

Better proof: We shall assume that s is a sharing rule for which a^* is an equilibrium, and derive a contradiction.

The idea is this. If s implements a^* , then any agent i must be deterred from cutting his effort from a_i^* by a loss in compensation. So consider a reduction in agent i 's effort by Δa_i that reduces output by ε . His compensation must fall more than his utility gain $v_i(a_i^*) - v_i(a_i^* - \Delta a_i)$. But here is the key: since effort is not observable, we don't know who shirked, so *everyone's* compensation must be cut, and by enough to deter each and everyone of them. That is, total compensation must fall by at least $\sum_{j=1}^n v_j(a_j^*) - v_j(a_j^* - \Delta a_i)$. Now we ask, how much does output fall when only person i shirks? The answer is approximately $D_i x(a^*) \Delta a_i$. How much must compensation fall? By approximately $\sum_{j=1}^n v'_j(a_j^*) \Delta a_i$. But here's the rub, since a^* maximizes surplus, the first order conditions imply $D_j x(a^*) = v'_j(a_j^*)$. Now think of the person i who gains least by deviating by Δa_i . The total compensation must fall at least n times as much as for that individual. But now the budget balance condition kicks in. Total compensation falls only by the reduction in total reward, which is approximately proportional to the agent's individual utility gain, not n times as much. Here are the details:

For convenience, set $x^* = x(a^*)$. Since x is continuous and strictly increasing, and $a^* \gg 0$, for every $\varepsilon > 0$ small enough,¹ by the Intermediate Value Theorem, for each j there is some $a_j(\varepsilon)$ satisfying

$$0 < a_j(\varepsilon) < a_j^*$$

and

$$x(a_{-j}^*, a_j(\varepsilon)) = x^* - \varepsilon.$$

As we shall see, it is important that the right-hand side is independent of j .

If a^* is an equilibrium, then by definition, for each j , it cannot pay agent j to switch from a_j^* to $a_j(\varepsilon)$, so

$$s_j(x(a^*)) - v_j(a_j^*) \geq s_j(x(a_{-j}^*, a_j(\varepsilon))) - v_j(a_j(\varepsilon)),$$

or

$$s_j(x(a^*)) - s_j(x(a_{-j}^*, a_j(\varepsilon))) \geq v_j(a_j^*) - v_j(a_j(\varepsilon)),$$

or, using the fact that $x(a_{-j}^*, a_j(\varepsilon)) = x^* - \varepsilon$, independent of j , we have

$$s_j(x^*) - s_j(x^* - \varepsilon) \geq v_j(a_j^*) - v_j(a_j(\varepsilon)). \quad (1)$$

Summing over j gives

$$\sum_{j=1}^n s_j(x^*) - \sum_{j=1}^n s_j(x^* - \varepsilon) \geq \sum_{j=1}^n v_j(a_j^*) - v_j(a_j(\varepsilon)),$$

so by budget balance

$$x^* - (x^* - \varepsilon) \geq \sum_{j=1}^n v_j(a_j^*) - v_j(a_j(\varepsilon)).$$

Now let i be an/the agent for whom

$$v_i(a_i^*) - v_i(a_i(\varepsilon)) \leq v_j(a_j^*) - v_j(a_j(\varepsilon)) \quad \text{for all } j.$$

¹For each j , we know that $x(a^*) - x(a_{-j}^*, 0) > 0$, so setting $m = \min_j x(a^*) - x(a_{-j}^*, 0)$, we have $m > 0$. Any ε satisfying $0 < \varepsilon < m$ is small enough.

Then

$$x^* - (x^* - \varepsilon) \geq n [v_i(a_i^*) - v_i(a_i(\varepsilon))] > 0. \quad (2)$$

(Note that i may depend on ε , but our notation does not reflect that.)

By the definition of derivative or Taylor's Theorem (take your pick),

$$v_i(a_i^*) - v_i(a_i(\varepsilon)) = v_i'(a_i^*)(a_i^* - a_i(\varepsilon)) - \hat{r}_i(\varepsilon), \quad \text{where } \hat{r}_i(\varepsilon)/(a_i^* - a_i(\varepsilon)) \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \quad (3)$$

Likewise,

$$x(a^*) - x(a_{-j}^*, a_i(\varepsilon)) = D_i x(a^*)(a_i^* - a_i(\varepsilon)) - \tilde{r}_i(\varepsilon), \quad \text{where } \tilde{r}_i(\varepsilon)/(a_i^* - a_i(\varepsilon)) \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \quad (4)$$

Now since a^* maximizes surplus, it satisfies the first order conditions (*), so

$$v_i'(a_i^*) = D_i x(a^*).$$

Thus (3) becomes

$$v_i(a_i^*) - v_i(a_i(\varepsilon)) = D_i x(a^*)(a_i^* - a_i(\varepsilon)) - \hat{r}_i(\varepsilon)$$

Now rewrite (2) as

$$\begin{aligned} D_i x(a^*)(a_i^* - a_i(\varepsilon)) - \tilde{r}_i(\varepsilon) &= x^* - (x^* - \varepsilon) \\ &\geq n [v_i(a_i^*) - v_i(a_i(\varepsilon))] \\ &= n [D_i x(a^*)(a_i^* - a_i(\varepsilon)) - \hat{r}_i(\varepsilon)]. \end{aligned}$$

Divide by $a_i^* - a_i(\varepsilon)$ and gather remainders to get

$$D_i x(a^*) = n D_i x(a^*) + \frac{\tilde{r}_i(\varepsilon) - n \hat{r}_i(\varepsilon)}{a_i^* - a_i(\varepsilon)}.$$

Letting $\varepsilon \rightarrow 0$ gives

$$D_i x(a^*) = n D_i x(a^*). \quad (5)$$

Now there is a caveat here: Remember I told you that i in (2) depends on ε , so as $\varepsilon \rightarrow 0$, the index i may change. Nevertheless, since there are only n choices for i , some index i must occur infinitely often, so for such an index (5) must hold.

But (5) can only hold if either $n = 1$ or $D_i x(a^*) = 0$. By hypothesis $n > 1$, and also by hypothesis x is concave and strictly increasing, so $D_i x(a^*)$ cannot be zero.² Thus (5) cannot hold, a contradiction.

This contradiction proves that no sharing rule (even a non-differentiable rule) can implement the efficient effort vector a^* . ■

²To see this, define $g(a) = x(a_{-i}^*, a)$. Then g is also concave and strictly increasing, but if $D_i x(a^*) = 0$, then $g'(a_i^*) = 0$, so g has a global maximum there, contradicting the fact that it is strictly increasing.

Breaking the budget

However, if we are willing to work with the weaker budget balance condition

$$\sum_{i=1}^n s_i(x) \leq x \quad \text{for all } x \in \mathbf{R}_+, \quad (\text{B}')$$

that is, if we are willing to throw away money, there are many sharing rules that implement a^* . Here is a typical example—note that it is not even continuous, let alone differentiable everywhere. Choose b_i , $i = 1, \dots, n$, to satisfy

$$b_i > v_i(a_i^*)$$

and

$$\sum_{i=1}^n b_i < x(a^*),$$

(this can certainly be done under Assumption 2) and define

$$s_i(x) = \begin{cases} b_i & \text{if } x \geq x(a^*) \\ 0 & \text{otherwise.} \end{cases}$$

I leave it to you to verify that this s implements a^* . The curious feature is that unless $a = 0$, the team always produces more than its members receive in compensation.

References

- [1] B. Holmstrom. 1982. Moral hazard in teams. *Bell Journal of Economics* 13:324–40.