Principal Agent Problems*

The examples we have used for mechanisms have primarily been drawn from the auction literature. Another standard problem that we can analyze using this framework is that of the problem between the principal of a firm (its owner(s)) and the agent (manager). Typically, it is impossible or extremely costly for principals to monitor the effort level of the agents. Thus, there is a fear that agents might provide less than their highest effort. This fear is why some contracts have incentive clauses in them – consider profit sharing between the agent and the firm, or sports contracts with performance goals. The idea is to tie the agent's payment to the outcome with the hope that this incentive will induce the agent to give high effort. Some of these contracts can be analyzed as Bayes-Nash equilibria, while others might require the Perfect Bayesian equilibrium solution concept (next set of notes).

1 Basic examples

We begin with a discussion of games where the range of effort is discrete (the agent can exert high effort or low effort) and the revenue to the principal is also discrete (there is a high revenue or low revenue) to motivate the problem and develop some general principles.

1.1 Observable effort

We begin the discussion of the principal-agent problem by assuming that the agent's effort is observable. The game begins with the principal deciding whether or not to make a contract offer to the agent. If the principal does not make a contract offer then the principal receives 0 and the agent receives some reservation utility U. If the principal makes a contract offer then it is the agent's move. The contract specifies that the agent will receive W_H if he exerts high effort and W_L if he exerts low effort. The agent has 3 choices – he can choose to accept the contract and exert high effort, accept the contract and exert low effort, or reject the contract. If he rejects the contract then the principal receives 0 and the agent receives U. If he accepts the contract and exerts low effort then the agent receives $W_L - e_L$, where e_L is the effort cost of exerting low effort and the principal receives $R_L - W_L$, where R_L is the revenue the principal receives if the agent exerts low effort. If the agent accepts the contract and exerts high effort then he receives $W_H - e_H$, where e_H is the effort cost of exerting high effort and the principal receives $R_H - W_H$, where R_H is the revenue the principal receives if the agent exerts high effort. Assume that $e_H > e_L$. Again, note that the agent's effort level is perfectly observable in this game. The game tree is:

What restrictions are necessary on the parameters to have an SPNE (there are no information sets that contain multiple nodes, so we can use SPNE) of the game be that the principal offers the contract and the agent accepts and exerts high effort? From the principal's point of view, we need that $R_H - W_H > 0$ because if it is not then the principal could be better off choosing to not offer the contract. It would also be helpful if $R_H - W_H > R_L - W_L$ because then the principal would prefer the agent to exert high effort. If the principal prefers that the agent exert low effort then it is fairly easy to ensure this by simply setting $W_H = W_L$ so that the agent receives the same payment regardless of which effort level is chosen. From the agent's point of view, two things need to happen. One is that $W_H - e_H > W_L - e_L$, so that the agent finds it more profitable to exert high effort rather than low effort. It also needs to be the case that $W_H - e_H > U$ so that the agent chooses to accept the contract rather than reject the contract. It is possible that $W_H - e_H > W_L - e_L$ but that exerting effort for this agent is too costly relative to his opportunity cost (U) so the agent would simply

*These notes generally correspond to Chapter 14 of MWG.

choose to reject the contract. As we have already discussed with mechanisms, these two conditions are just the incentive compatibility constraint and the participation constraint.

Incentive compatibility constraint: The principal must structure the contract such that it gives the agent the incentive to act in the principal's best interest. In this example, choosing high effort over low effort would mean $W_H - e_H > W_L - e_L$.

Participation constraint: The principal must structure the contract such that participation by the agent is better than non-participation. In this example, $W_H - e_H > U$.

Now, what should the principal set W_H to be in this example? Assuming that $R_H - W_H > R_L - W_L$, the principal wants the agent to exert high effort. The principal needs $W_H - e_H > W_L - e_L$ and $W_H - e_H > U$. The principal can guarantee that the agent will exert high effort if he accepts the contract by setting W_L such that $U > W_L - e_L$. With that choice for W_L, the principal will satisfy the agent's incentive compatibility constraint if he satisfies the participation constraint because now:

$$W_H - e_H > U > W_L - e_L.$$

Thus the principal should set $W_H > U + e_H$. But for the principal to maximize profits the wage should be set as close to the boundary condition as possible, so $W_H = U + e_H + \varepsilon$, for some small $\varepsilon > 0$. If we set up the constraint as $W_H - e_H \geq U$ then we would get that the wage must equal the reservation utility plus the effort cost. We will show this result in a more general manner below.

1.2 Unobservable effort

Suppose now that the principal cannot observe effort and can only observe outcome (either R_H or R_L). Moreover, there is a possibility that the principal receives R_L even if the agent exerts high effort e_H and a chance that the principal receives R_H even if the principal exerts low effort e_L. Thus, the outcome does not perfectly represent the agent's effort. Because the principal can only observe outcome he bases the contract on the observed outcome – if he observes R_H then the agent receives W_H and if he observes R_L then the agent receives W_L. The game is as follows:

Again, we can work through the participation and incentive compatibility constraints to determine what the parameter restrictions need to be to ensure a particular equilibrium. Suppose we want the principal to offer a contract and the agent to accept the contract and put forth high effort. Assuming the agent is risk neutral, and that the agent's effort does not affect the probability of the good and bad states,[1] the agent's incentive compatibility constraint is:

$$\Pr(Good) * (W_H - e_H) + \Pr(Bad) * (W_L - e_H) \geq \Pr(Good) * (W_H - e_L) + \Pr(Bad) * (W_L - e_L)$$

and the agent's participation constraint is:

$$\Pr(Good) * (W_H - e_H) + \Pr(Bad) * (W_L - e_H) \geq U.$$

Rewriting the IC constraint:

$$\begin{aligned}
\Pr(Good) * (W_H - e_H) + (1 - \Pr(Good)) * (W_L - e_H) &\geq \Pr(Good) * (W_H - e_L) + (1 - \Pr(Good)) * (W_L - e_L) \\
-e_H \Pr(Good) - e_H + e_H \Pr(Good) &\geq -e_L \Pr(Good) - e_L + e_L \Pr(Good) \\
-e_H &\geq -e_L \\
e_L &\geq e_H
\end{aligned}$$

Thus, because the agent's effort does not affect the state of the world the principal will be unable to offer the agent a contract that is incentive compatible and induces high effort. This result should be intuitive – if the agent cannot influence the outcome through effort, why would the agent exert high effort?

Now supposing that the agent's effort affects the probability of the good and bad states, the agent's incentive compatibility constraint is:

$$\Pr(Good|e_H) * (W_H - e_H) + (1 - \Pr(Good|e_H)) * (W_L - e_H) \geq \Pr(Good|e_L) * (W_H - e_L) + (1 - \Pr(Good|e_L)) * (W_L - e_L).$$

[1] This assumption is changed shortly.

where $\Pr(Good|e_H)$ is the probability of the good state when high effort is chosen and $\Pr(Good|e_L)$ is the probability of the good state when low effort is chosen. The agent's participation constraint for exerting high effort is now:

$$\Pr(Good|e_H) * (W_H - e_H) + \Pr(Bad|e_H) * (W_L - e_H) \geq U.$$

The principal also has a constraint that must be met:

$$\Pr(Good|e_H)*(R_H - W_H)+\Pr(Bad|e_H)*(R_L - W_L) \geq \Pr(Good|e_L)*(R_H - W_H)+\Pr(Bad|e_L)*(R_L - W_L).$$

Thus, if these conditions are satisfied then the principal will offer the agent a contract and the agent will accept the offer and exert high effort. A more formal discussion, with profit levels drawn from a continuum, is provided below.

2 More formal description

Consider a similar structure as the prior examples with a principal and an agent. Now, however, the principal's profit is continuous over the range $[\underline{\pi}, \overline{\pi}]$. The agent's effort $e \in E$ can be any effort level from the set of effort levels. The agent's effort choice affects the profit stochastically – thus any profit can occur with any effort choice. The conditional density function $f(\pi|e)$ describes the relationship between profit and effort, and $f(\pi|e) > 0$ for all $e \in E$ and all $\pi \in [\underline{\pi}, \overline{\pi}]$. We again restrict the agent's effort choice to $\{e_H, e_L\}$ but assume that the distribution of π conditional on e_H first-order stochastically dominates the distribution conditional on e_L. Thus, the expected level of profits when the agent chooses e_H is greater than when he chooses e_L:

$$\int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e_H)\, d\pi > \int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e_L)\, d\pi.$$

The agent has utility over wage and effort, $u(w, e)$. We focus on a slightly more general case than above, with $u(w, e) = v(w) - g(e)$. Assume that $v'(w) > 0$, $v''(w) < 0$, and $g(e_H) > g(e_L)$. The principal receives the profit realization π but must pay the wage from that profit, so ultimately the principal receives $\pi - w$.

2.1 Observable effort

Again, consider the case of observable effort. The principal offers a contract that the agent can accept or reject. The contract specifies the effort level $e \in \{e_L, e_H\}$ and the wage as a function of observed profit, $w(\pi)$. The agent's reservation utility of accepting the contract is $\overline{u}$. If the agent rejects the contract the principal receives a payoff of zero.

Assume that the principal will want to offer the agent a contract such that the agent will accept (expected payoff is greater than $\overline{u}$). The principal's problem then is:

$$\max_{e \in \{e_L, e_H\},\ w(\pi)} \int_{\underline{\pi}}^{\overline{\pi}} (\pi - w(\pi)) f(\pi|e)\, d\pi$$

$$\text{s.t.} \int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|e)\, d\pi - g(e) \geq \overline{u}$$

Given that effort is observable, what is the best choice of $w(\pi)$ for each choice of e? Once that is known, what is the best choice of e?

If we split the first term into two pieces, $\int_{\underline{\pi}}^{\overline{\pi}} (\pi - w(\pi)) f(\pi|e)\, d\pi = \int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e)\, d\pi - \int_{\underline{\pi}}^{\overline{\pi}} w(\pi) f(\pi|e)\, d\pi$, we can see that choosing $w(\pi)$ to maximize $\int_{\underline{\pi}}^{\overline{\pi}} (\pi - w(\pi)) f(\pi|e)\, d\pi$ is the same as choosing $w(\pi)$ to minimize $\int_{\underline{\pi}}^{\overline{\pi}} w(\pi) f(\pi|e)\, d\pi$. Thus, we have:

$$\min_{w(\pi)} \int_{\underline{\pi}}^{\overline{\pi}} w(\pi) f(\pi|e)\, d\pi$$

$$\text{s.t.} \int_{\underline{\pi}}^{\overline{\pi}} v\left(w\left(\pi\right)\right) f\left(\pi|e\right) d\pi - g\left(e\right) \geq \overline{u}$$

We can argue that the constraint always binds because why would the principal give the agent more than is needed to accept the contract? There may be reasons why that occurs in other situations, but in this specific model there is no reason for the principal to give the agent more than is required. The problem is a constrained optimization problem. Let γ be the multiplier on the constraint, so the agent's wage $w\left(\pi\right)$ for each $\pi \in [\underline{\pi}, \overline{\pi}]$ must satisfy:

$$\begin{aligned} -f\left(\pi|e\right) + \lambda v'\left(w\left(\pi\right)\right) f\left(\pi|e\right) &= 0 \\ \frac{1}{v'\left(w\left(\pi\right)\right)} &= \gamma \end{aligned}$$

What happens if the agent is strictly risk averse? Then the risk-neutral principal just offers a fixed payment to the risk-averse agent based on the observable effort, and the principal fully insures the agent. Thus, the principal offers some fixed wage w_e^* such that $v\left(w_e^*\right) - g\left(e\right) = \overline{u}$. The wage w_e^* will depend on the effort level that is provided, with $w_{e_H}^* > w_{e_L}^*$ because it is more costly for the agent to exert high effort.[2]

What is the optimal choice of e? Recall that effort is observable, so the principal can specify e. The effort level is the one that maximizes expected profit minus wages:

$$\int_{\underline{\pi}}^{\overline{\pi}} \pi f\left(\pi|e\right) d\pi - v^{-1}\left(\overline{u} + g\left(e\right)\right).$$

The specific choice of e_H or e_L depends on both $f\left(\pi|e\right)$ and $g\left(e\right)$.

Proposition 1 *In the principal-agent model with observable effort, an optimal contract specifies that the agent choose the effort* e^* *that maximizes* $\left[\int_{\underline{\pi}}^{\overline{\pi}} \pi f\left(\pi|e\right) d\pi - v^{-1}\left(\overline{u} + g\left(e\right)\right)\right]$ *and pays the agent a fixed wage* $w^* = v^{-1}\left(\overline{u} + g\left(e\right)\right)$. *This contract is the uniquely optimal contract if* $v''\left(w\right) < 0$ *at all* w.

2.2 Unobservable effort

The optimal contract when effort is observable specifies an efficient effort choice and fully insures the agent against risk. When effort is not observable, these goals are in conflict because the principal pays a wage based on realized profit, not effort. Thus, the agent could exert high effort, get a bad profit draw, and be paid below his effort cost. When the agent is risk-neutral the principal can still achieve the same expected payoff as when effort is observable.

2.2.1 Risk-neutral agent

Let $v\left(w\right) = w$. The optimal effort level e^* when effort is observable solves:

$$\max_{e \in \{e_L, e_H\}} \int_{\underline{\pi}}^{\overline{\pi}} \pi f\left(\pi|e\right) d\pi - g\left(e\right) - \overline{u}.$$

Now consider the case when effort is not observable. First, note that the principal can never do better when effort is unobservable (and unable to be specified by the principal) than when effort is observable (and able to be specified by the principal). If the principal could do better with unobservable effort, the principal could just specify the unobservable effort contract when effort is observable and allow the agent to choose effort.

Now let the principal specify $w\left(\pi\right) = \pi - \alpha$, where α is a fixed payment. Thus, the principal has essentially sold the project to the agent for α. Suppose the agent accepts. The agent then chooses e to maximize (remember the agent is now the owner of the project, so there is no issue with aligning incentives):

$$\int_{\underline{\pi}}^{\overline{\pi}} w\left(\pi\right) f\left(\pi|e\right) d\pi - g\left(e\right) = \int_{\underline{\pi}}^{\overline{\pi}} \pi f\left(\pi|e\right) d\pi - \alpha - g\left(e\right).$$

[2] If the agent is risk-neutral then any compensation scheme, including the fixed payment, is optimal as long as the expected wage payment is equal to $\overline{u} + g\left(e\right)$.

Note that e^* maximizes both $\int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e)\, d\pi - \alpha - g(e)$ (the unobservable payoff) and $\int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e)\, d\pi - g(e) - \overline{u}$ (the observable payoff) because the only difference between the two is the constants α and $\overline{u}$. So we have shown that the effort level is the same in the two problems.

Now, the agent will accept the contract $w(\pi) = \pi - \alpha$ as long as the agent's utility is at least $\overline{u}$, so:

$$\int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e^*)\, d\pi - \alpha - g(e^*) \geq \overline{u}.$$

Let α^* be the value of the fixed payment such that $\int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e^*)\, d\pi - \alpha^* - g(e^*) = \overline{u}$. Now, $\alpha^* = \int_{\underline{\pi}}^{\overline{\pi}} \pi f(\pi|e^*)\, d\pi - g(e^*) - \overline{u}$, which is the exact same expected payoff the principal had when effort was observable. There are no risk sharing problems when the agent is risk-neutral.

2.2.2 Risk-averse agent

Now consider the case of a risk-averse agent. With risk-neutrality and unobservable effort we did not really utilize an incentive compatibility constraint because when the principal sold the project to the agent there was no need to align incentives. Formally:

$$\min_{w(\pi)} \int_{\underline{\pi}}^{\overline{\pi}} w(\pi) f(\pi|e)\, d\pi$$

$$\text{s.t.} \quad \int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|e)\, d\pi - g(e) \geq \overline{u}$$

$$e \text{ solves } \max_{\widetilde{e}} \int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|\widetilde{e})\, d\pi - g(\widetilde{e})$$

So the principal must now specify the contract in order to elicit the "correct" effort amount from the agent.

What wage function should the principal specify if he wants the agent to exert low effort e_L? He just offers $w_e^* = v^{-1}(\overline{u} + g(e_L))$. In order to induce the low effort level the principal simply specifies the fixed payment *as if effort is observable*. In this case, while effort is not observable, the principal knows that with this contract the agent will not choose high effort (unless the agent makes a mistake), so effectively the principal knows the effort level choice of the agent. This wage contract yields the same payoff to the principal in the unobservable effort case as when effort is observable, and we know that the principal can never earn more when effort is unobservable, so this contract must be a solution.

The contract to induce the agent to exert high effort is more interesting. As in the discrete payoff case above, the incentive compatibility constraint can be written as a comparison between exerting high effort and low effort, where the payoff to the agent exerting high effort must be greater than or equal to the payoff to the agent exerting low effort:

$$\int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|e_H)\, d\pi - g(e_H) \geq \int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|e_L)\, d\pi - g(e_L)$$

Now we have a constrained optimization problem with two inequality constraints. Let γ be the multiplier on the participation constraint and μ be the multiplier on the incentive compatibility constraint and we have:[3]

$$\begin{aligned}
&= \int_{\underline{\pi}}^{\overline{\pi}} w(\pi) f(\pi|e_H)\,d\pi + \gamma\left(\overline{u} - \int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|e_H)\,d\pi\right) + \\
&\quad \mu\left(\int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|e_L)\,d\pi - g(e_L) - \int_{\underline{\pi}}^{\overline{\pi}} v(w(\pi)) f(\pi|e_L)\,d\pi + g(e_H)\right) \\
0 &= f(\pi|e_H) - \gamma v'(w(\pi)) f(\pi|e_H) + \mu\left(f(\pi|e_L) - f(\pi|e_H)\right) v'(w(\pi)) \\
f(\pi|e_H) &= \gamma v'(w(\pi)) f(\pi|e_H) - \mu\left(f(\pi|e_L) - f(\pi|e_H)\right) v'(w(\pi)) \\
f(\pi|e_H) &= v'(w(\pi))\left[\gamma f(\pi|e_H) - \mu\left(f(\pi|e_L) - f(\pi|e_H)\right)\right] \\
\frac{1}{v'(w(\pi))} &= \frac{\gamma f(\pi|e_H) - \mu\left(f(\pi|e_L) - f(\pi|e_H)\right)}{f(\pi|e_H)} \\
\frac{1}{v'(w(\pi))} &= \frac{\gamma f(\pi|e_H) + \mu\left(f(\pi|e_H) - f(\pi|e_L)\right)}{f(\pi|e_H)} \\
\frac{1}{v'(w(\pi))} &= \gamma + \frac{\mu\left(f(\pi|e_H) - f(\pi|e_L)\right)}{f(\pi|e_H)} \\
\frac{1}{v'(w(\pi))} &= \gamma + \mu\left[1 - \frac{f(\pi|e_L)}{f(\pi|e_H)}\right]
\end{aligned}$$

Note that this result looks fairly similar to the observable effort case – in fact, it is the same, except for the $\mu\left[1 - \frac{f(\pi|e_L)}{f(\pi|e_H)}\right]$ term, which arises because of the presence of the incentive compatibility constraint. We know both γ and μ are positive, so both constraints bind. How? Consider $\mu = 0$. If that is true, then the payment is fixed at γ. For a fixed payment the agent will choose e_L because the disutility of e_H is greater than the disutility of e_L. For $\gamma > 0$, suppose that the constraint is not binding, so $\gamma = 0$. Because $F(\pi|e_H)$ first-order stochastically dominates $F(\pi|e_L)$, there must exist an open set of profit levels $\Pi \subset [\underline{\pi}, \overline{\pi}]$ such that $\left[\frac{f(\pi|e_L)}{f(\pi|e_H)}\right] > 1$. If that is true, then $v'(w(\pi)) < 0$, which is a contradiction.

That γ and μ are positive leads to some interesting results. Suppose that $\widehat{w}$ is the fixed wage payment that leads to $\frac{1}{v'(\widehat{w})} = \gamma$. Then:

$$\begin{aligned}
w(\pi) &> \widehat{w} \text{ if } \frac{f(\pi|e_L)}{f(\pi|e_H)} < 1 \\
w(\pi) &< \widehat{w} \text{ if } \frac{f(\pi|e_L)}{f(\pi|e_H)} > 1
\end{aligned}$$

Intuitively, this result suggests that the principal is paying the agent more for outcomes that are statistically more likely to happen under e_H than under e_L.

What is interesting is that there are cases in which higher profits do not necessarily lead to higher wages (the wage structure is not monotonically increasing). If that is true, then $\frac{f(\pi|e_L)}{f(\pi|e_H)}$ must be decreasing in π. That means that as π increases, the likelihood of obtaining profit level π if effort is e_H relative to the likelihood if effort level is e_L must increase. While that seems logical, *first-order stochastic dominance does not guarantee it*. The text has an example of a distribution that first-order stochastically dominates another distribution, yet the optimal wage scheme is nowhere close to monotonic (see page 486). One final note is that if the principal wants to induce high effort in the unobservable case then he will have to pay the agent a higher wage than in the observable case. The higher wage compensates the agent for the risk that must be borne.

Note the effect that the unobservable case has on welfare. First, consider the case when the principal wants to induce effort e_L by the agent. We have seen that the optimal wage scheme under both is the same, so there is no welfare loss. Now, consider the principal inducing high effort. If high effort is optimal when observable, then either (1) the principal must compensate the agent more in order to induce the agent to exert high effort when effort is unobservable or (2) the principal may not find it optimal to induce high effort when effort is unobservable, thus leading the principal to induce low effort. Both lead to a welfare loss for the principal.

[3] Clearly there is a set of first-order conditions. As in MWG, we focus on one first-order condition.