
Principal-Agent Models

Things are gettin' better
 It's people that are gettin' worse
 Mose Allison

In the *principal-agent model*, the payoff to the *principal* depends on an action taken by the *agent*. The principal cannot contract for the action, but can compensate the agent based on some observable signal that is correlated with the action. The principal is first mover, and chooses an incentive scheme for paying the agent that depends on the observed signal. The agent then determines the optimal action to take, given the incentives, then decides whether to accept the principal's offer, based on the expected payment and the subjective cost of performing the action. Upon accepting, the agent chooses an action that maximizes his payoff, and the principal observes the signal correlated with the action, pays the agent according to the incentive scheme, and receives a payoff dependent upon the signal. The incentive scheme is a precommitment by the principal, even though the agent's action is not.

7.1 Gift Exchange

In a famous paper, "Labor Contracts as Partial Gift Exchange," George Akerlof (1982) suggested that sometimes employers pay employees more than they must to attract the labor they need, and employees often reciprocate by working harder or more carefully than they otherwise would. He called this *gift exchange*. This section analyzes a simple model of gift exchange in labor markets. A firm hires N identical employees, each of whom supplies effort level $e(w - z)$, where $e(\cdot)$ is increasing and concave, w is the wage, and z is a benchmark wage such that $w > z$ indicates that the employer is being generous, and conversely $w < z$ indicates that the boss is ungenerous. The firm's revenue is an increasing and concave function $f(eN)$ of total amount of effort supplied by the N employees, so the firm's net profit is given by

$$\pi(w, N) = f(e(w - z)N) - wN.$$

Suppose that the firm chooses w and N to maximize profits. Show that the *Solow condition*

$$\frac{de}{dw} = \frac{e}{w} \tag{7.1}$$

holds (Solow 1979) and that the second-order condition for a profit maximum is satisfied. Then, writing the equilibrium wage w^* , equilibrium effort e^* , and equilibrium profits π^* as a function of the benchmark wage z , show that

$$\frac{de^*}{dz} > 0; \quad \frac{d\pi^*}{dz} < 0; \quad \frac{dw^*}{dz} > 1.$$

7.2 Contract Monitoring

An employer hires supervisors to oversee his employees, docking the pay of any employee who is caught shirking. Employee effort consists of working a fraction e of the time, so if there are N employees, and each employee works at effort level e for one hour, then total labor supplied is eN . The employer's revenue in this case is $q(eN)$, where $q(\cdot)$ is an increasing function. All employees have the same utility function $u(w, e) = (1 - e)w$, where w is the wage rate and e is the effort level. An employee, who is normally paid w , is paid $z < w$ if caught shirking.

Suppose that an employee who chooses effort level e is caught shirking with probability $p(e) = 1 - e$, so the harder the employee works, the lower the probability of being caught shirking. The game tree for this problem is depicted in figure 7.1.

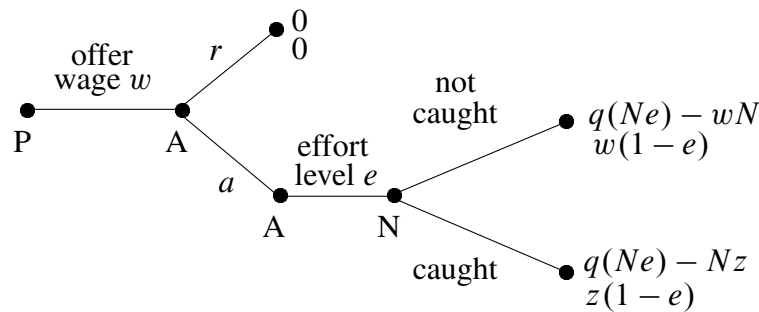


Figure 7.1. Labor discipline with monitoring

- a. Show that $w(1 - e)e + z(1 - e)^2$ is the payoff to an employee who chooses effort level e .

- b. Show that if the employer offers wage $w > 2z$, the employee's best response is to choose

$$e(w) = \frac{w - 2z}{2(w - z)}.$$

Show that this employee's *best-response schedule* is increasing and concave, as depicted in figure 7.2.

- c. If the employer chooses w and N to maximize profits, show that the choice of w in fact maximizes $e(w)/w$, the amount of effort per dollar of wages, which is the slope of the employer iso-cost line in figure 7.2.
- d. Show that Nash equilibrium (w^*, e^*) satisfies the *Solow condition* (Solow 1979),

$$e'(w^*) = \frac{e(w^*)}{w^*}.$$

This is where the employer iso-cost line is tangent to the employee's best-response schedule at (w^*, e^*) in figure 7.2.

- e. Show that

$$w^* = (2 + \sqrt{2})z \approx 3.41z, \quad e^* = \frac{1}{\sqrt{2}(1 + \sqrt{2})} \approx 0.29.$$

- f. Suppose the employee's reservation utility is $z_0 > 0$, so the employee must be offered expected utility z_0 to agree to come to work. Show that the employer will set $z = 2z_0/(1 + \sqrt{2}) \approx 0.83z_0$.

7.3 Profit Signaling

An employer hires an employee to do a job. There are two possible levels of profits for the employer, high (π_H) and low ($\pi_L < \pi_H$). The employee can affect the probability of high profits by choosing to work with either high or low effort. With high effort the probability of high profits is p_h , and with low effort the probability of high profits is p_l , where $0 < p_l < p_h < 1$.

If the employer could see the employee's choice of effort, he could simply write a contract for high effort, but he cannot. The only way he can induce A to work hard is to offer the proper *incentive contract*: pay a wage w_H if profits are high and $w_L < w_H$ if profits are low.

How should the employer choose the incentives w_L and w_H to maximize expected profits? The game tree for this situation is shown in figure 7.3,

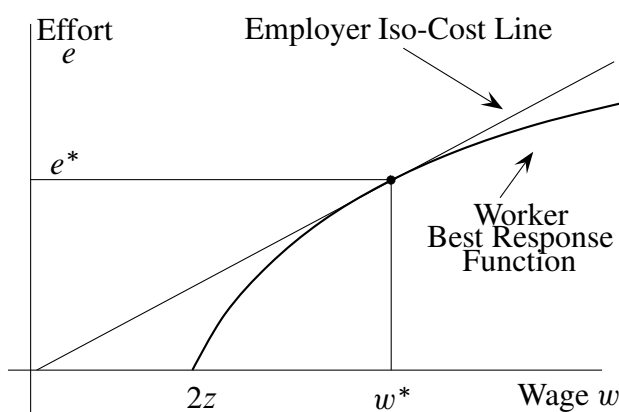


Figure 7.2. Equilibrium in the labor discipline model

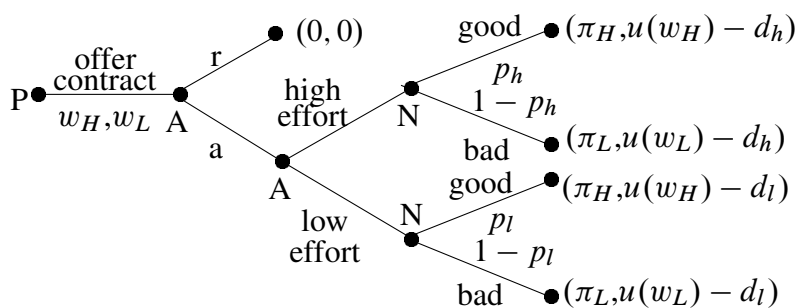


Figure 7.3. Labor incentives

where we assume the utility of the wage is $u(w)$, the cost of high effort to the employee is d_h and the cost of low effort is $d_l < d_h$. By working hard, the employee faces a lottery with payoffs $u(w_H) - d_h, u(w_L) - d_h$ with probabilities $(p_h, 1 - p_h)$, the expected value of which is

$$\begin{aligned}
 & p_h(u(w_H) - d_h) + (1 - p_h)(u(w_L) - d_h) \\
 & = p_h u(w_H) + (1 - p_h)u(w_L) - d_h.
 \end{aligned}$$

With low effort, the corresponding expression is $p_l u(w_H) + (1 - p_l)u(w_L) - d_l$. Thus, the employee will choose high effort over low effort only if the first of these expressions is at least as great as the second, which gives

$$(p_h - p_l)(u(w_H) - u(w_L)) \geq d_h - d_l. \tag{7.2}$$

This is called the *incentive compatibility constraint* for eliciting high effort.

Now suppose the employee's next-best job prospect has expected value z . Then to get the employee to take the job, the employer must offer the employee at least z . This gives the *participation constraint*:

$$p_h u(w_H) + (1 - p_h)u(w_L) - d_h \geq z, \quad (7.3)$$

if we assume that the principal wants the employee to work hard.

The expected profit of the employer, if we assume that the employee works hard, is given by

$$p_h(\pi_H - w_H) + (1 - p_h)(\pi_L - w_L). \quad (7.4)$$

It is clear that, to minimize the expected wage bill, the employer should choose w_H and w_L so that equation (7.3) is satisfied as an equality. Also, the employee should choose w_H and w_L so that equations (7.2) and (7.3) are satisfied as equalities. This is illustrated in figure 7.4.

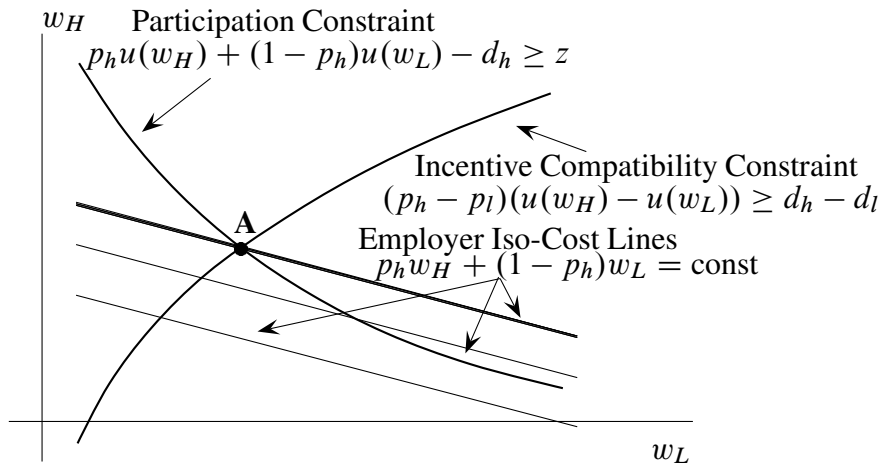


Figure 7.4. Minimizing the cost of inducing an action, given participation and incentive compatibility constraints

Using this figure, we note that the employer's iso-cost lines are of the form $p_h w_H + (1 - p_h)w_L = \text{const.}$, and we show that the participation constraint is decreasing and convex. We treat w_H as a function of w_L and differentiate the participation constraint, getting

$$p_h u'(w_H) \frac{dw_H}{dw_L} + (1 - p_h)u'(w_L) = 0.$$

Thus,

$$\frac{dw_H}{dw_L} = -\frac{1-p_h}{p_h} \frac{u'(w_L)}{u'(w_H)} < -\frac{1-p_h}{p_h} < 0. \quad (7.5)$$

The second inequality (which we use later) holds because $w_L < w_H$, so if the agent is strictly risk averse, u' is decreasing. The participation constraint is thus decreasing. Now take the derivative of equation (7.5), getting

$$\frac{d^2w_H}{dw_L^2} = -\frac{1-p_h}{p_h} \left[\frac{u''(w_H)}{u'(w_H)} - \frac{u'(w_L)u''(w_H)}{u'(w_H)^2} \frac{dw_H}{dw_L} \right] > 0.$$

Thus, the participation constraint is convex.

The incentive compatibility constraint is increasing and cuts the w_L -axis for some $w_L > 0$. If the agent is weakly decreasingly risk averse (that is, if $u''' > 0$), then the incentive compatibility constraint is concave. To see this, we differentiate the incentive compatibility constraint $u(w_H) = u(w_L) + \text{constant}$, getting

$$u'(w_H) \frac{dw_H}{dw_L} = u'(w_L),$$

so $dw_H/dw_L > 1 > 0$, and the incentive compatibility constraint is increasing. Differentiate again, getting

$$u''(w_H) \frac{dw_H}{dw_L} + u'(w_H) \frac{d^2w_H}{dw_L^2} = u''(w_L).$$

Thus

$$u'(w_H) \frac{d^2w_H}{dw_L^2} = u''(w_L) - u''(w_H) \frac{dw_H}{dw_L} < u''(w_L) - u''(w_H) < 0,$$

and the constraint is concave.

If the agent is strictly risk averse (§2.4), the slope of the iso-cost lines is less than the slope of the participation constraint at its intersection A with the incentive compatibility constraint. To see this, note that the slope of the iso-cost line is $|dw_H/dw_L| = (1-p_h)/p_h$, which is less than the slope of the participation constraint, which is

$$|(1-p_h)u'(w_L)/p_h u'(w_H)|,$$

by equation (7.5).

It follows that the solution is at **A** in figure 7.4.

7.4 Properties of the Employment Relationship

The unique Nash equilibrium in the labor discipline model of the previous section is the solution to the two equations $p_h u(w_H) + (1 - p_h)u(w_L) - d_h = z$ and $(p_h - p_l)(u(w_H) - u(w_L)) = d_h - d_l$. Solving simultaneously, we get

$$u(w_L) = z + \frac{p_h d_l - p_l d_h}{p_h - p_l}, \quad u(w_H) = u(w_L) + \frac{d_h - d_l}{p_h - p_l}. \quad (7.6)$$

Note that the employee exactly achieves his reservation position. As we might expect, if z rises, so do the two wage rates w_L and w_H . If d_h rises, you can check that w_H rises and w_L falls. Similar results hold when p_h and p_l vary.

Now that we know the cost to the principal of inducing the agent to take each of the two actions, we can determine which action the principal should ask the agent to choose. If H and L are the expected profits in the good and bad states, respectively, then the return $\pi(a)$ for inducing the agent to take action $a = h, l$ is given by

$$\pi(h) = Hp_h + L(1 - p_h) - \mathbf{E}_h w, \quad \pi(l) = Hp_l + L(1 - p_l) - \mathbf{E}_l w, \quad (7.7)$$

where $\mathbf{E}_h w$ and $\mathbf{E}_l w$ are the expected wage payments if the agent takes actions h and l , respectively; that is, $\mathbf{E}_h w = p_h w_H + (1 - p_h)w_L$ and $\mathbf{E}_l w = p_l w_H + (1 - p_l)w_L$.

Is it worth inducing the employee to choose high effort? For low effort, only the participation constraint $u(w_l) = d_l + z$ must hold, where w_l is the wage paid independent of whether profits are H or L , with expected profit $p_l H + (1 - p_l)L - w_l$. Choose the incentive wage if and only if $p_h(H - w_H) + (1 - p_h)(L - w_L) \geq p_l H + (1 - p_l)L - w_l$. This can be written

$$(p_h - p_l)(H - L) \geq p_h w_H + (1 - p_h)w_L - w_l. \quad (7.8)$$

We will see that, in general, if the employee is risk neutral and it is worth exerting high effort, then the optimum is to make the principal the fixed claimant and the agent the residual claimant (§7.7). To see this for the current example, we can let $u(w) = w$. The participation constraint is then $p_h u(w_H) + (1 - p_h)u(w_L) = p_h w_H + (1 - p_h)w_L = z + d_h$, and the employer's profit is then $A = p_h H + (1 - p_h)L - (z + d_h)$. Suppose we give

A to the employer as a fixed payment and let $w_H = H - A$, $w_L = L - A$. Then the participation constraint holds, because

$$p_h w_H + (1 - p_h) w_L = p_h H + (1 - p_h) L - A = z + d_h.$$

Because high effort is superior to low effort for the employer, 7.8) must hold, giving

$$\begin{aligned} (p_h - p_l)(H - L) &\geq p_h w_H + (1 - p_h) w_L - (z + d_l) \\ &= z + d_h - (z + d_l) = d_h - d_l. \end{aligned}$$

But then,

$$w_H - w_L = H - L \geq \frac{d_h - d_l}{p_h - p_l},$$

which says that the incentive compatibility constraint is satisfied.

Figure 7.5 is a graphical representation of the principal's problem. Note that in this case there are many profit-maximizing contracts. Indeed, any point on the heavy solid line in the figure maximizes profits.

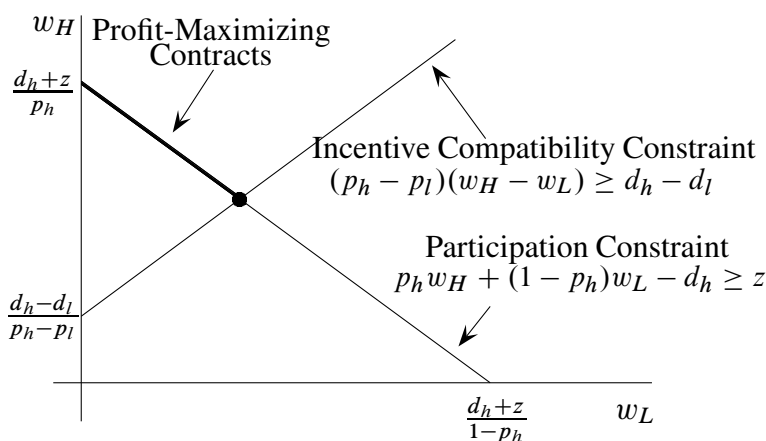


Figure 7.5. The principal's problem when the agent is risk neutral

7.5 Peasant and Landlord

A landlord hires a peasant to tend a cornfield. The landlord's profit is H if the crop is good and $L < H$ if the crop is poor. The peasant can work at either high effort h or low effort l , and the probability p_h of a good crop

when he exerts high effort is greater than the probability p_l of a good crop when he expends low effort, with $0 < p_l < p_h < 1$. The landowner cannot observe the peasant's effort.

Suppose the peasant's utility function when the wage is w is given by $u(w) - d_h$ with high effort, and $u(w) - d_l$ with low effort. We assume $d_h > d_l$, so unless given some inducement, the peasant will not work hard and $u' > 0, u'' < 0$, so the peasant has diminishing marginal utility of the wage. The peasant's fallback utility is z .

To induce the peasant to work hard, the landlord chooses a pair of wages w_H and w_L , and pays the peasant w_H if profit is H , and w_L if profit is L . This is called an *incentive wage*.

What should the landlord pay the peasant if he wants to minimize the expected wage $Ew = p_h w_H + (1 - p_h)w_L$, subject to eliciting high effort? First, w_H and w_L must satisfy a *participation constraint*: w_H and w_L must be sufficiently large that the peasant is willing to work at all. Suppose the peasant's next-best alternative gives utility z . Then the landowner must choose w_H and w_L so that the peasant's expected utility is at least z :

$$p_h u(w_H) + (1 - p_h)u(w_L) - d_h \geq z. \quad (\text{PC})$$

Second, w_H and w_L must satisfy an *incentive compatibility constraint*: the payoff (that is, the expected return) to the peasant for working hard must be at least as great as the payoff to not working hard. Thus, we must have

$$p_h u(w_H) + (1 - p_h)u(w_L) - d_h \geq p_l u(w_H) + (1 - p_l)u(w_L) - d_l.$$

We can rewrite this second condition as

$$[u(w_H) - u(w_L)](p_h - p_l) \geq d_h - d_l. \quad (\text{ICC})$$

We now prove that both the PC and the ICC must hold as equalities. The problem is to minimize $p_h w_H + (1 - p_h)w_L$ subject to PC and ICC. This is the same as maximizing $-p_h w_H - (1 - p_h)w_L$ subject to the same constraints, so we form the Lagrangian

$$\begin{aligned} \mathcal{L}(w_H, w_L, \lambda, \mu) = & -p_h w_H - (1 - p_h)w_L \\ & + \lambda[p_h u(w_H) + (1 - p_h)u(w_L) - d_h - z] \\ & + \mu[(u(w_H) - u(w_L))(p_h - p_l) - (d_h - d_l)]. \end{aligned}$$

The first-order conditions can be written:

$$\mathcal{L}_H = 0, \mathcal{L}_L = 0, \quad \lambda, \mu \geq 0;$$

if $\lambda > 0$, then the PC holds with equality;

if $\mu > 0$, then the ICC holds with equality.

But we have

$$\mathcal{L}_H = -p_h + \lambda p_h u'(w_H) + \mu u'(w_H)(p_h - p_l) = 0,$$

$$\mathcal{L}_L = -1 + p_h + \lambda(1 - p_h)u'(w_L) - \mu u'(w_L)(p_h - p_l) = 0.$$

Suppose $\lambda = 0$. Then, by adding the two first-order conditions, we get

$$\mu(u'(w_H) - u'(w_L))(p_h - p_l) = 1,$$

which implies $u'(w_H) > u'(w_L)$, so $w_H < w_L$ (by declining marginal utility of income). This, of course, is not incentive compatible, because ICC implies $u(w_H) > u(w_L)$, so $w_H > w_L$. It follows that our assumption that $\lambda = 0$ is contradictory and hence $\lambda > 0$, from which it follows that the participation constraint holds as an equality.

Now suppose $\mu = 0$. Then the first-order conditions $\mathcal{L}_H = 0$ and $\mathcal{L}_L = 0$ imply $u'(w_H) = 1/\lambda$ and $u'(w_L) = 1/\lambda$. Because $u'(w_H) = u'(w_L) = 1/\lambda$, $w_H = w_L$ (because u' is strictly decreasing). This also is impossible by the ICC. Hence $\mu > 0$, and the ICC holds as an equality.

The optimal incentive wage for the landlord is then given by

$$u(w_L) = d_h - p_h(d_h - d_l)/(p_h - p_l) + z$$

$$u(w_H) = d_h + (1 - p_h)(d_h - d_l)/(p_h - p_l) + z.$$

To see this, suppose the landlord has concave utility function v , with $v' > 0$ and $v'' < 0$. The peasant is risk neutral, so we can assume her utility function is $u(w, d) = w - d$, where w is income and d is effort. The assumption that high effort produces a surplus means that the following social optimality (SO) condition holds:

$$p_h H + (1 - p_h)L - d_h > p_l H + (1 - p_l)L - d_l,$$

or

$$(p_h - p_l)(H - L) > d_h - d_l. \quad (\text{SO})$$