This is designed for one 75-minute lecture using *Games and Information*.

October 3, 2006

7 Moral Hazard: Hidden Actions

PRINCIPAL-AGENT MODELS

The **principal** *(or* **uninformed player***) is the player who has the coarser information partition.*

The **agent** *(or* **informed player***) is the player who has the finer information partition.*

Production Game I: Full Information

Players
The principal and the agent.

The order of play
1 The principal offers the agent a wage w.
2 The agent decides whether to accept or reject the contract.
3 If the agent accepts, he exerts effort e.
4 Output equals $q(e)$, where $q' > 0$.

Payoffs
If the agent rejects the contract, then $\pi_{agent} = \bar{U}$ and $\pi_{principal} = 0$.
If the agent accepts the contract, then $\pi_{agent} = U(e, w), U_e < 0, U_w > 0$ and $\pi_{principal} = V(q - w), V' > 0$.

reservation utility, $\bar{U}$

Production Game I: Full Information

In the first version of the game, every move is common knowledge and the contract is a function $w(e)$.

The agent must be paid some amount $\tilde{w}(e)$ to exert effort e, where $\tilde{w}(e)$ is the function that makes him just willing to accept the contract, so

$$U(e, w(e)) = \overline{U}. \tag{1}$$

Thus, the principal's problem is

$$\underset{e}{Maximize} \;\; V(q(e) - \tilde{w}(e)) \tag{2}$$

The first-order condition for this problem is

$$V'(q(e) - \tilde{w}(e)) \left(\frac{\partial q}{\partial e} - \frac{\partial \tilde{w}}{\partial e} \right) = 0, \tag{3}$$

which implies that

$$\frac{\partial q}{\partial e} = \frac{\partial \tilde{w}}{\partial e}. \tag{4}$$

From condition (1), using the implicit function theorem (see section 13.4), we get

$$\frac{\partial \tilde{w}}{\partial e} = - \left(\frac{\frac{\partial U}{\partial e}}{\frac{\partial U}{\partial \tilde{w}}} \right). \tag{5}$$

Combining equations (4) and (5) yields

$$\left(\frac{\partial U}{\partial \tilde{w}} \right) \left(\frac{\partial q}{\partial e} \right) = - \left(\frac{\partial U}{\partial e} \right). \tag{6}$$

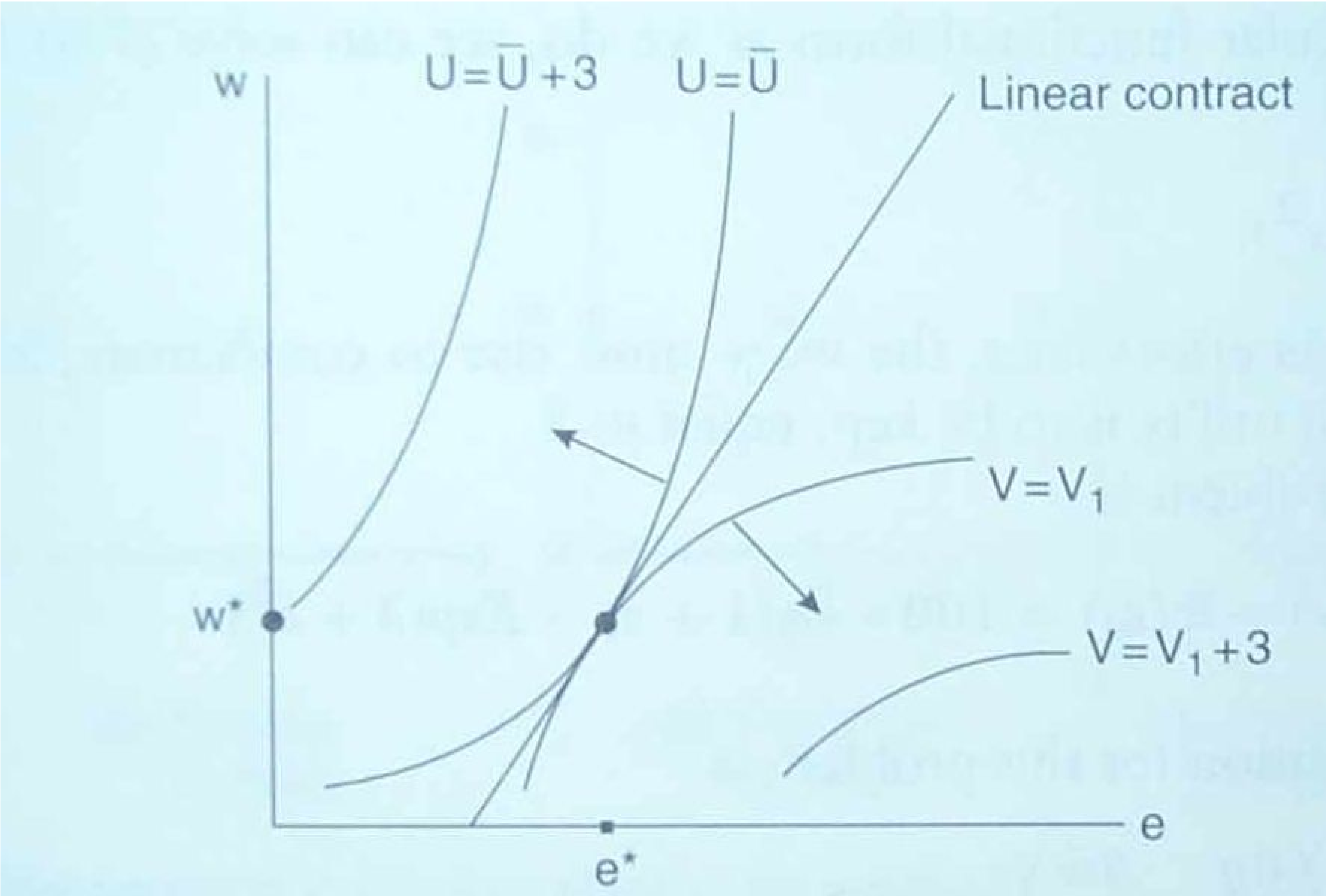
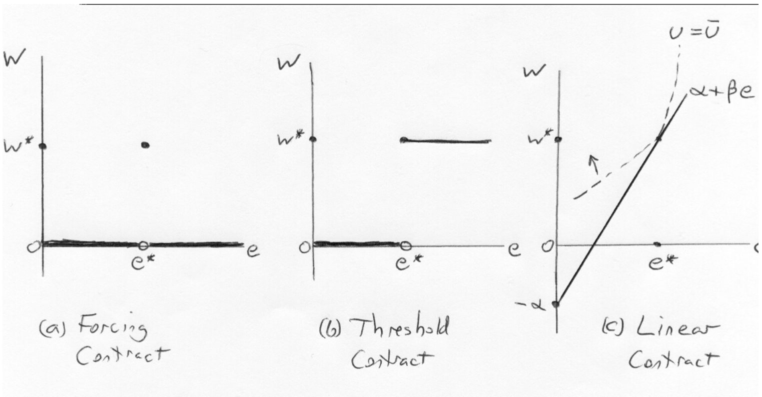


Figure 2: The Efficient Effort Level in Production Game I

Under perfect competition among the principals the profits are zero, so the reservation utility, $\overline{U}$, will be at the level such that at the profit-maximizing effort e^*, $\tilde{w}(e^*) = q(e^*)$, or

$$U(e^*, q(e^*)) = \overline{U}. \tag{7}$$

The principal selects the point on the $U = \overline{U}$ indifference curve that maximizes his profits, at effort e^* and wage w^*.



The principal must then design a contract that will induce the agent to choose this effort level. The following three contracts are equally effective under full information.

1 The **forcing contract** sets $w(e^*) = w^*$ and $w(e \neq e^*) = 0$. This is certainly a strong incentive for the agent to choose exactly $e = e^*$.

2 The **threshold contract** sets $w(e \geq e^*) = w^*$ and $w(e < e^*) = 0$. This can be viewed as a flat wage for low effort levels, equal to 0 in this contract, plus a bonus if effort reaches e^*. Since the agent dislikes effort, the agent will choose exactly $e = e^*$.

3 The **linear contract** sets $w(e) = \alpha + \beta e$, where α and β are chosen so that $w^* = \alpha + \beta e^*$ and the contract line is tangent to the indifference curve $U = \bar{U}$ at e^*.

Let's now fit out Production Game I with specific functional forms. Suppose the agent exerts effort $e \in [0, \infty]$, and output equals

$$q(e) = 100 * log(1 + e), \tag{8}$$

so $q' = \frac{100}{1+e} > 0$ and $q'' = \frac{-100}{(1+e)^2} < 0$. If the agent rejects the contract, let $\pi_{agent} = \bar{U} = 3$ and $\pi_{principal} = 0$, whereas if the agent accepts the contract, let $\pi_{agent} = U(e, w) = log(w) - e^2$ and $\pi_{principal} = q(e) - w(e)$.

The agent must be paid some amount $\tilde{w}(e)$ to exert effort e, where $\tilde{w}(e)$ is defined to be the wage that makes the agent willing to participate, i.e., as in equation (1),

$$U(e, w(e)) = \overline{U}, \quad so \;\; log(\tilde{w}(e)) - e^2 = 3. \tag{9}$$

Knowing the particular functional form as we do, we can solve (9) for the wage function:

$$\tilde{w}(e) = Exp(3 + e^2), \tag{10}$$

where we use $Exp(x)$ to mean Euler's constant (about 2.718) to the power x, since the conventional notation of e^x would be confused with e as effort.

Equation (10) makes sense. As effort rises, the wage must rise to compensate, and rise more than exponentially if utility is to be kept equal to 3.

Now that we have a necessary-wage function $\tilde{w}(e)$, we can attack the principal's problem, which is

$$\underset{e}{Maximize} \;\; V(q(e) - \tilde{w}(e)) = 100 * log(1+e) - Exp(3+e^2) \tag{11}$$

The first-order condition for this problem is

$$V'(q(e) - \tilde{w}(e)) \left(\frac{\partial q}{\partial e} - \frac{\partial \tilde{w}}{\partial e} \right) = 0, \tag{12}$$

so for our problem,

$$\left(\frac{100}{1+e} \right) - 2e(Exp(3+e^2)) = 0, \tag{13}$$

which cannot be solved analytically. Using the computer program Mathematica, I found that $e^* \approx 0.77$, from which, using the formulas above, we get $q^* \approx 57$ and $w^* \approx 37$. The payoffs are $\pi_{agent} = 3$ and $\pi_{principal} \approx 20$.

If $\overline{U}$ were high enough, the principal's payoff would be zero. If the market for agents were competitive, this is what would happen, since the agent's reservation payoff would be the utility of working for another principal instead of $\overline{U} = 3$.

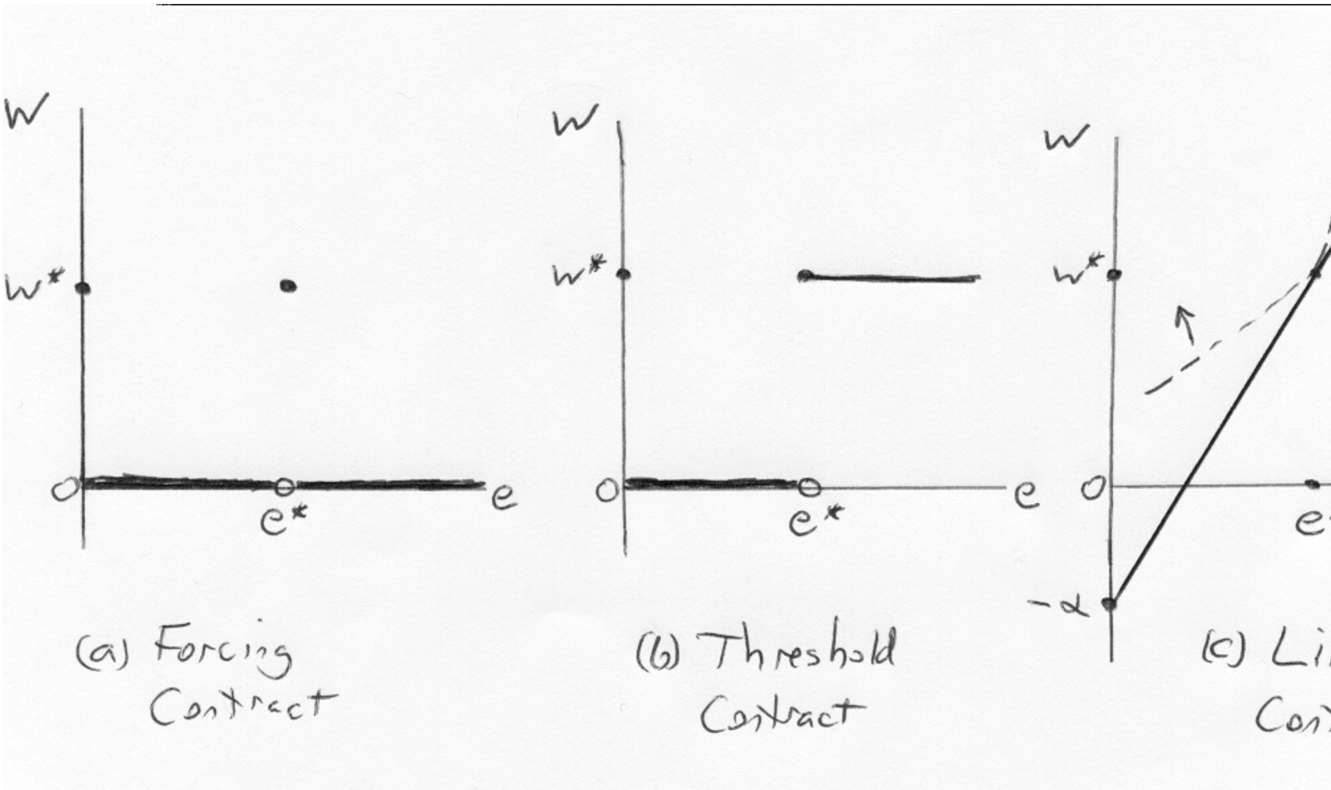


Figure 3: Three Contracts that Induce Effort e^* for Wage w^*

To obtain $e^* = 0.77$, a number of styles of contract could be used.

1 The **forcing contract** sets $w(e^*) = w^*$ and $w(e \neq 0.77) = 0$. Here, $w(0.77) = 37$ (rounding up) and $w(e \neq e^*) = 0$.

2 The **threshold contract** sets $w(e \geq e^*) = w^*$ and $w(e < e^*) = 0$. Here, $w(e \geq 0.77) = 37$ and $w(e < 0.77) = 0$.

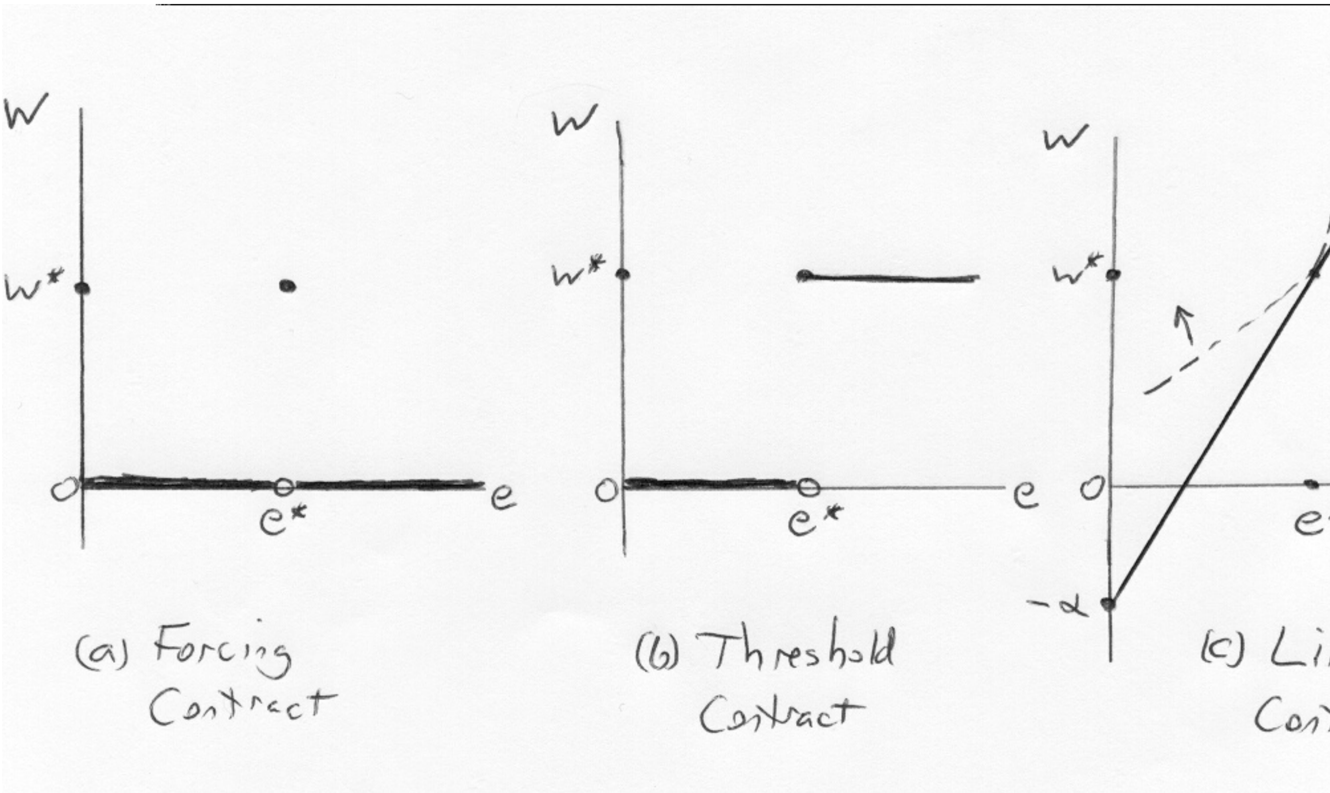


Figure 3: Three Contracts that Induce Effort e^* for Wage w^*

3 The **linear contract** sets $w(e) = \alpha + \beta e$, where α and β are chosen so that $w^* = \alpha + \beta e^*$ and the contract line is tangent to the indifference curve $U = \bar{U}$ at e^*. The slope of that indifference curve is the derivative of the $\tilde{w}(e)$ function, which is

$$\frac{\partial \tilde{w}(e)}{\partial e} = 2e * Exp(3 + e^2). \qquad (14)$$

At $e^* = 0.77$, this takes the value 56 (which only coincidentally is near the value of $q^* = 57$). That is the β for the linear contract. The α must solve $w(e^*) = 37 = \alpha + 56(0.77)$, so $\alpha = -7$.

We ought to be a little concerned as to whether the agent will choose the effort we hope for if he is given the linear contract. We constructed it so that he would be willing to accept the contract, because if he chooses $e = 0.77$, his utility will be 3. But might he prefer to choose some larger or smaller e and get even more utility? No, because his utility is concave. That makes the indifference curve convex, so its slope is always increasing and no preferable indifference curve touches the equilibrium contract line.

Quasilinearity and Alternative Functional Forms for the Production Game

Consider the following three functional forms for utility:

$$
\begin{aligned}
U(e,w) &= log(w) - e^2 && (a) \\
U(e,w) &= w - e^2 && (b) \qquad (15) \\
U(e,w) &= log(w - e^2) && (c)
\end{aligned}
$$

Utility function (a) is what we just used in Production Game I. Utility function (b) is an example of **quasilinear preferences**, because utility is separable in one good— money, here— and linear in that good. This kind of utility function is commonly used to avoid wealth effects that would otherwise occur in the interactions among the various goods in the utility function. Separability means that giving an agent a higher wage does not, for example, increase his marginal disutility of effort. Linearity means furthermore that giving an agent a higher wage does not change his tradeoff between money and effort, his marginal rate of substitution, as it would in function (a), where a richer agent is less willing to accept money for higher effort. In effort-wage diagrams, quasilinearity implies that the indifference curves are parallel along the effort axis (which they are *not* in Figure 2).

If utility is quasilinear, the efficient effort level is independent of which side has the bargaining power because the gains from efficient production are independent of how those gains are distributed so long as each party has no incentive to abandon the relationship. This as the same lesson as the Coase Theorem's:, under general conditions the activities undertaken will be efficient and independent of the distribution of property rights (Coase [1960]). This property of the efficient-effort level means that the modeller is free to make the assumptions on bargaining power that help to focus attention on the information problems he is studying.

There are thus three reasons why modellers so often use take-it-or-leave-it offers. The first two reasons were discussed earlier in the context of Production Game I: (1) such offers are a good way to model competitive markets, and (2) if the reservation payoff of the player without the bargaining power is set high enough, such offers lead to the same outcome as would be reached if that player had more bargaining power. Quasi-linear utility provides a third reason: (3) if utility is quasi-linear, the optimal effort level does not depend on who has the bargaining power, so the modeller is justified in choosing the simplest model of bargaining.

$$
\begin{aligned}
U(e,w) &= log(w) - e^2 && (a) \\
U(e,w) &= w - e^2 && (b) \qquad (16) \\
U(e,w) &= log(w - e^2) && (c)
\end{aligned}
$$

Quasilinear utility functions most often are chosen to look like (b), but what quasilinearity really requires is just linearity in the special good (w here) for some monotonic transformation of the utility function.

Utility function (c) is a logarithmic transformation of (b), which is a monotonic transformation, so it too is quasilinear. That is because marginal rates of substitution, which is what matter here, are a feature of general utility functions, not the Von Neumann-Morgenstern functions we typically use.

Thus, utility function (c) is also a quasi-linear function, because it is just a monotonic function of (b).

$$
\begin{aligned}
U(e,w) &= log(w) - e^2 && (a) \\
U(e,w) &= w - e^2 && (b) \qquad (17) \\
U(e,w) &= log(w - e^2) && (c)
\end{aligned}
$$

Returning to the solution of Production Game I, let us now use a different approach to get to the same answer as we did using the principal's maximization problem (11). Instead, we will return to the general optimality condition(6), here repeated.

$$\left(\frac{\partial U}{\partial \tilde{w}}\right)\left(\frac{\partial q}{\partial e}\right) = -\frac{\partial U}{\partial e} \qquad (6)$$

For any of our three utility functions we will continue using the same output function $q(e) = 100 * log(1+e)$ from (8), which has the first derivative $q' = \frac{100}{1+e}$.

Using utility function (a), $\frac{\partial U}{\partial \tilde{w}} = 1/w$. and $\frac{\partial U}{\partial e} = -2e$, so equation (6) becomes

$$\left(\frac{1}{w}\right)\left(\frac{100}{1+e}\right) = -(-2e). \qquad (18)$$

If we substitute for w using the function $\tilde{w}(e) = Exp(3+e^2)$ that we found in equation (10), we get essentially the same equation as (13), and so outcomes are the same— $e^* \approx 0.77$, $q^* \approx 57$, and $w^* \approx 37$, $\pi_{agent} = 3$, and $\pi_{principal} \approx 20$.

$$
\begin{aligned}
U(e,w) &= log(w) - e^2 && (a) \\
U(e,w) &= w - e^2 && (b) \qquad (19) \\
U(e,w) &= log(w - e^2) && (c)
\end{aligned}
$$

Returning to the solution of Production Game I, let us now use a different approach to get to the same answer as we did using the principal's maximization problem (11). Instead, we will return to the general optimality condition(6), here repeated.

$$\left(\frac{\partial U}{\partial \tilde{w}}\right)\left(\frac{\partial q}{\partial e}\right) = -\frac{\partial U}{\partial e} \qquad (6)$$

For any of our three utility functions we will continue using the same output function $q(e) = 100 * log(1+e)$ from (8), which has the first derivative $q' = \frac{100}{1+e}$.

Using utility function (b), $\frac{\partial U}{\partial \tilde{w}} = 1$ and $\frac{\partial U}{\partial e} = -2e$, so equation (6) becomes

$$(1)\left(\frac{100}{1+e}\right) = -\left(-2e\right) \qquad (20)$$

Notice that w has disappeared. The optimal effort no longer depends on the agent's wealth. Thus, we don't need to use the wage function to solve for the optimal effort. Solving directly, we get $e^* \approx 6.59$ and $q^* \approx 203$. The wage function will be different now, solving

$w - e^2 = 3$, so $w^* \approx 43$, $\pi_{agent} = 3$, and $\pi_{principal} \approx 160$. (These numbers are not really comparable to when we used utility function (a), but they will be useful in Production Game II.)

$$
\begin{aligned}
U(e,w) &= log(w) - e^2 && (a) \\
U(e,w) &= w - e^2 && (b) \\
U(e,w) &= log(w - e^2) && (c)
\end{aligned} \tag{21}
$$

Returning to the solution of Production Game I, let us now use a different approach to get to the same answer as we did using the principal's maximization problem (11). Instead, we will return to the general optimality condition(6), here repeated.

$$
\left(\frac{\partial U}{\partial \tilde{w}}\right)\left(\frac{\partial q}{\partial e}\right) = -\frac{\partial U}{\partial e} \tag{6}
$$

For any of our three utility functions we will continue using the same output function $q(e) = 100 * log(1+e)$ from (8), which has the first derivative $q' = \frac{100}{1+e}$.

Using utility function (c), $\frac{\partial U}{\partial \tilde{w}} = 1/(w - e^2)$ and $\frac{\partial U}{\partial e} = -2e/(w - e^2)$, so equation (6) becomes

$$
\left(\frac{1}{w - e^2}\right)\left(\frac{100}{1+e}\right) = -\left(\frac{-2e}{w - e^2}\right) \tag{22}
$$

and with a little simplification,

$$
\frac{100}{1+e} = 2e. \tag{23}
$$

The variable w has again disappeared, so as with utility function (b) the optimal effort does not depend on the agent's wealth. Solving for the optimal effort yields $e^* \approx 6.59$ and $q^* \approx 203$, the same as with utility function (b). The wage function is different, however. Now it solves $log(w - e^2) = 3$, so $w = e^2 + exp(3)$ and $w^* \approx 63$, $\pi_{agent} = 3$, and $\pi_{principal} \approx 140$.

Production Game III: A Flat Wage Under Certainty

In this version of the game, the principal can condition the wage neither on effort nor on output. This is modelled as a principal who observes neither effort nor output, so information is asymmetric. In Production Game III, we have finally reached "moral hazard".

Production Game IV: An Output-Based Wage under Certainty

In this version, the principal cannot observe effort but he can observe output and specify the contract to be $w(q)$.

Unlike in Production Game III, the principal now picks not a number w but a function $w(q)$.

The forcing contract, for example, would be any wage function such that $U(e^*, w(q^*)) = \bar{U}$ and $U(e, w(q)) < \bar{U}$ for $e \neq e^*$.

Production Game V: An Output-Based Wage under Uncertainty.

In this version, the principal cannot observe effort but can observe output and specify the contract to be $w(q)$. Output, however, is a function $q(e, \theta)$ both of effort and the state of the world $\theta \in \mathbf{R}$, which is chosen by Nature according to the probability density $f(\theta)$.

Because of the uncertainty about the state of the world, effort does not map cleanly onto observed output in Production Game V. A given output might have been produced by any of several different effort levels, so a forcing contract based on output will not necessarily achieve the desired effort. Unlike in Production Game IV, here the principal cannot deduce $e \neq e^*$ from $q \neq q^*$. Moreover, even if the contract does induce the agent to choose e^*, if it does so by penalizing him heavily when $q \neq q^*$ it will be expensive for the principal. The agent's expected utility must be kept equal to $\bar{U}$ so he will accept the contract, and if he is sometimes paid a low wage because output happens not to equal q^* despite his correct effort, he must be paid more when output does equal q^* to make up for it. If the agent is risk averse, this variability in his wage requires that his expected wage be higher than the w^* found earlier, because he must be compensated for the extra risk. There is a tradeoff between incentives and insurance against risk.

Put more technically, moral hazard is a problem when $q(e)$ is not a one-to- one function and a single value of e might result in any of a number of values of q, depending on the value of θ. In this case the output function is not invertible; knowing q, the principal cannot deduce the value of e perfectly without assuming equilibrium behavior on the part of the agent.

A **first-best contract** *achieves the same allocation as the contract that is optimal when the principal and the agent have the same information set and all variables are contractible.*

A **second-best contract** *is Pareto optimal given information asymmetry and constraints on writing contracts.*

7.3 The Incentive Compatibility and Participation Constraints

The principal's objective in Production Game V is to maximize his utility knowing that the agent is free to reject the contract entirely and that the contract must give the agent an incentive to choose the desired effort. These two constraints arise in every moral hazard problem, and they are named the **participation constraint** and the **incentive compatibility contraint**. Mathematically, the principal's problem is

$$\underset{w(\cdot)}{Maximize} \;\; EV(q(\tilde{e},\theta) - w(q(\tilde{e},\theta))) \tag{24}$$

subject to

$$\tilde{e} = \overset{argmax}{e} \;\; EU(e, w(q(e,\theta))) \;\; \text{(incentive compatibility constra}$$

$$EU(\tilde{e}, w(q(\tilde{e},\theta))) \geq \bar{U} \qquad \text{(participation constraint) (24b}$$

To support the effort level e, the wage contract $w(q)$ must satisfy the incentive compatibility and participation constraints. Mathematically, the problem of finding the least cost $C(\tilde{e})$ of supporting the effort level $\tilde{e}$ combines steps one and two.

$$C(\tilde{e}) = \underset{w(\cdot)}{Minimum} \;\; Ew(q(\tilde{e},\theta)) \tag{25}$$

subject to constraints (24a) and (24b).

Step three takes the principal's problem of maximizing his payoff, expression (24), and restates it as

$$\underset{\tilde{e}}{Maximize}\ \ EV(q(\tilde{e},\theta) - C(\tilde{e})). \tag{26}$$

After finding which contract most cheaply induces each effort, the principal discovers the optimal effort by solving problem (26).

7.1 Categories of Asymmetric Information Models

It used to be that the economist's first response to peculiar behavior which seemed to contradict basic price theory was "It must be some kind of price discrimination." Today, we have a new answer: "It must be some kind of asymmetric information." In a game of asymmetric information, player Smith knows something that player Jones does not. This covers a broad range of models (including price discrimination itself), so it is not surprising that so many situations come under its rubric. We will look at them in five chapters.

Moral hazard with hidden actions (Chapters 7 and 8)
Smith and Jones begin with symmetric information and agree to a contract, but then Smith takes an action unobserved by Jones. Information is complete.

Adverse selection (Chapter 9)
Nature begins the game by choosing Smith's type, unobserved by Jones. Smith and Jones then agree to a contract. Information is incomplete.

Mechanism design in adverse selection and **post-contractual hidden knowledge**) (Chapter 10)
Jones is designing a contract for Smith designed to elicit Smith's private information. This may happen under

adverse selection— in which case Smith knows the information prior to contracting— or post-contractual hidden knowledge (also called moral hazard with hidden information)—in which case Smith will learn it after contracting.

Signalling and **Screening** (Chapter 11)

Nature begins the game by choosing Smith's type, unobserved by Jones. To demonstrate his type, Smith takes actions that Jones can observe. If Smith takes the action before they agree to a contract, he is signalling. If he takes it afterwards, he is being screened. Information is incomplete.

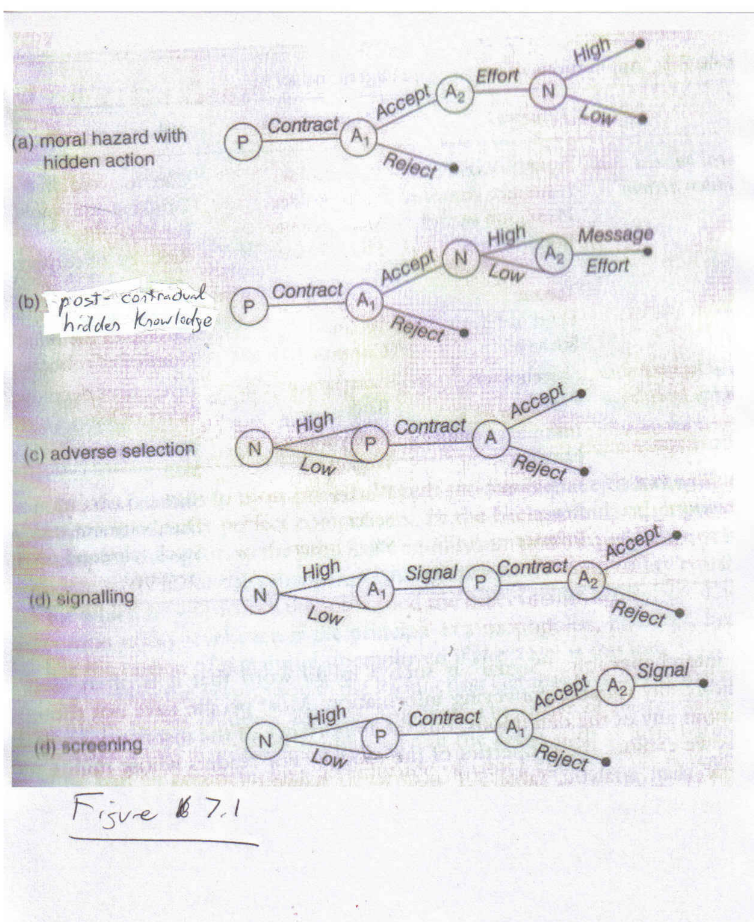


Figure 1: Categories of Asymmetric Information Models

Production Game II: Full Information. Agent Moves First.

In this version, every move is common knowledge and the contract is a function $w(e)$. The order of play, however, is now as follows

The Order of Play
1 The agent offers the principal a contract $w(e)$.
2 The principal decides whether to accept or reject the contract.
3 If the principal accepts, the agent exerts effort e.
4 Output equals $q(e)$, where $q' > 0$.

Now the agent has all the bargaining power, not the principal. Thus, instead of requiring that the contract be at least barely acceptable to the agent, our concern is that the contract be at least barely acceptable to the principal, who must earn zero profits so $q(e) - w(e) \geq 0$.

The agent will maximize his own payoff by driving the principal to exactly zero profits, so $w(e) = q(e)$. Substituting $q(e)$ for $w(e)$ to account for this constraint, the maximization problem for the agent in proposing an effort level e at a wage $w(e)$ can therefore be written as

$$\underset{e}{Maximize} \quad U(e, q(e)) \tag{27}$$

The first-order condition is

$$\frac{\partial U}{\partial e} + \left(\frac{\partial U}{\partial q}\right)\left(\frac{\partial q}{\partial e}\right) = 0. \tag{28}$$

Since $\frac{\partial U}{\partial q} = \frac{\partial U}{\partial w}$ when the wages equals output, equation (28) implies that

$$\left(\frac{\partial U}{\partial w}\right)\left(\frac{\partial q}{\partial e}\right) = -\left(\frac{\partial U}{\partial e}\right). \tag{29}$$

We can see the wealth effect by solving out optimality equation (29) for the specific functional forms of Production Game I from expression (21).

Using utility function (a) from expression (21)

$$\left(\frac{1}{w}\right)\left(\frac{100}{1+e}\right) = -(-2e). \tag{30}$$

That is the same as in Production Game I, equation (18), but now w is different. It is not found by driving the agent to his reservation payoff, but by driving the principal to zero profits: $w = q$. Since $q = 100 * log(1+e)$, we can substitute that in for w to get

$$\left(\frac{1}{100 * log(1+e)}\right)\left(\frac{100}{1+e}\right) = 2e. \tag{31}$$

When solved numerically, this yields $e^* \approx 0.63$, and thus $q = w \approx 49$, and $\pi_{principal} = 0$ and $\pi_{agent} \approx 3.49$. In Production Game I, the optimal effort using this utility function was 0.77 and the agent's payoff was 3. The difference arises because there the agent's wealth was lower because the principal had the bargaining power. In Production Game II the agent is, in effect, wealthier, and since his marginal utility of money is lower, he chooses to convert some (but not all) of that extra wealth into what we might call leisure— working less hard.

Using the quasilinear utility functions (b) and (c) from expression (21), recall that both have the same optimality condition, the one we found in equations (20) and (23):

$$\frac{100}{1+e} = 2e \qquad (23)$$

As we observed before, w does not appear in equation (23), so the wage equation does not matter to e^*. But that means that in Production Game II, $e^* \approx 6.59$ and $q^* \approx 203$, just as in Production Game I. With quasilinear utility, the efficient action does not depend on bargaining power. Of course, the wage and payoffs do depend on who has the bargaining power. In Production Game II, $w^* = q^* \approx 203$, and $\pi_{principal} = 0$. The agent's payoff is higher than in Production Game I, but it differs, of course, depending on the payoff function. For utility function (b) it is $\pi_{agent} \approx 160$ and for utility function (c) it is $\pi_{agent} \approx 5.08$.