Moral Hazard Example

Well, then says I, what’s the use you learning to do right when it’s troublesome to do right and ain’t no trouble to do wrong, and the wages is just the same? I was stuck. I couldn’t answer that. So I reckoned I wouldn’t bother no more about it, but afterwards always do whichever come handiest at the time. (Quoted by Mr. Salanié from *The Adventures of Huckleberry Finn* by Mark Twain.)

Consider a simple 2 action, 2 outcome model. The agent can choose either effort $e_0$ (shirk) or $e_1$ (work). The cost of either action $e$ is normalized so that it has utility cost $e$. Hence the Agent’s utility if she gets wage $w$ and chooses action $e$ is

$$u(w, e) = v(w) - e,$$

where $v'(w) > 0$ and either $v''(w) < 0$ (risk averse case) or $v''(w) = 0$ (risk neutral case).

There are 2 possible outcomes of the agent’s effort, she can either succeed or fail in a project. Higher effort increases the probability of success, which equals $P$ if she chooses $e_1$ and $p$ if she chooses $e_0$, where $1 > P > p > 0$. Thus the agent may fail even if she chooses to work, and she may succeed even if she shirks.

The Principal is risk neutral. His revenue in case of failure is $y_1 \equiv y_F$, and is $y_2 \equiv y_S$ in case of success, where $0 \leq y_F < y_S$. The Principal only observes the outcome $y$, but not the effort $e$. Hence he can only condition the contract he offers to A on whether the project succeeds or fails. An incentive contract is $C = (w_1, w_2) \equiv (w_F, w_S)$, offering a wage of $w_F$ if the project fails and $w_S$ if it succeeds. If $w_F \neq w_S$, the agent is bearing some risk. A special case is a straight wage contract $C = (w, w)$ that offers the same wage $w$ regardless of the project’s outcome.

1. The Agent’s Problem

Let $\phi^e \equiv (\phi^e_1, \phi^e_2)$ denote the probability distribution over the 2 states if the Agent chooses effort $e$. So, $\phi^e_2 = p$ if $e = e_0$ and $\phi^e_2 = P$ if $e = e_1$. Let $Eu(C, e) \equiv \phi^e_1 v(w_1) + \phi^e_2 v(w_2) - e$ denote the Agent’s expected utility from the contract $C = (w_1, w_2)$ if she chooses $e$.

Given any contract offer $C = (w_1, w_2)$ from the Principal, the Agent must first decide whether to accept the contract or not. This leads to the Agent’s participation or individual rationality constraint, that the Agent finds it more worthwhile to accept the contract $C$ than to refuse it and get her outside option utility level $U$. That is, when offered $C$, either she finds working more worthwhile than refusing the contract:

$$Eu(C, e_1) \geq U \quad IR(e_1)$$

or she finds shirking more worthwhile than refusing the contract:

$$Eu(C, e_0) \geq U \quad IR(e_0)$$

If she accepts $C$, she must then decide whether to work or shirk. This leads to her incentive compatibility constraint. The interesting case is when the Principal wants the Agent to work. Then he must offer her an incentive contract $C$ satisfying:

$$Eu(C, e_1) \geq Eu(C, e_0) \quad IC(e_1)$$

Alternatively, the Agent will shirk rather than work if:

$$Eu(C, e_0) \geq Eu(C, e_1) \quad IC(e_0)$$
Summarizing, the agent accepts the contract if either $IR(e_0)$ or $IR(e_1)$ is satisfied. Given she accepts, she will shirk if $IC(e_0)$ is satisfied and work if $IC(e_1)$ is satisfied.

2. The Principal’s Problem

Let $Ey(e) \equiv \phi_1^e y_1 + \phi_2^e y_2$ denote the Principal’s expected revenue if the Agent chooses $e$. Similarly, let $Ew(e) \equiv \phi_1^e w_1 + \phi_2^e w_2$ denote the Principal’s expected cost if he offers the contract $C = (w_1, w_2)$ and the Agent chooses $e$.

Folding back the game tree, the Principal will choose an effort level $e \in \{e_0, e_1\}$ and a contract $C = (w_1, w_2)$ that maximizes his expected profit subject to $IR(e)$ (the agent accepting) and $IC(e)$ (the agent choosing the effort level the Principal wants to implement). Hence the Principal’s optimization problem may be written:

$$\max_{C=(w_S,w_F), e \in \{e_0,e_1\}} Ey(e) - Ew(e) \quad \text{subject to:}$$

$$Eu(C,e) \geq U \quad IR(e)$$

$$Eu(C,e) \geq Eu(C,e'), \quad IC(e)$$

where it is understood that $e' \neq e$.

This problem can be divided into 2 smaller steps. First, for any effort level $e$, let the Principal find the least cost way of implementing $e$. The program for implementing $e$ is:

$$\min_{C=(w_S,w_F)} Ew(e) \quad \text{subject to:}$$

$$Eu(C,e) \geq U \quad IR(e)$$

$$Eu(C,e) \geq Eu(C,e'), \quad IC(e)$$

Let the contract $C^*(e) \equiv (w_1^*(e), w_2^*(e))$ solve this implementation problem; hence $c(e) \equiv Ew^*(e)$ is the Principal’s least cost way of implementing $e$. Second, using the cost function $c$, let the Principal choose the effort level that maximizes his expected profit:

$$\max_{e \in \{e_0,e_1\}} Ey(e) - c(e).$$

The above is the Grossman-Hart 2-step procedure for solving the Principal’s optimization problem. It is analogous to the 2-step procedure for solving a firm’s profit-maximization problem: Instead of

$$\max_{q,L} pq - wL \quad \text{s.t.} \quad q = f(L),$$

it is often more convenient to first solve the problem of finding the least-cost way of producing $q$ units given the production function $f$:

$$\min_{L} wL \quad \text{s.t.} \quad f(L) \geq q.$$

This yields a cost function $c(q)$. Then use this cost function to maximize profits by solving:

$$\max_{q} pq - c(q).$$

Below we will sometimes write $C_0^*$ for $C^*(e_0)$ and $C_1^*$ for $C^*(e_1)$.

3. Implementing low effort

To make our moral hazard example interesting, we will assume that high effort is socially efficient. To formalize, define
the straight reservation wage for effort level \( e \), \( w(e) \), by the implicit equation \( v(w(e)) - e = U \). That is, \( w(e) \) is the straight wage that leaves the Agent indifferent between choosing action \( e \) at the Principal’s firm and choosing her outside option. Define the expected gains from trade from shirking or working, respectively, as:

\[
GFT_0 \equiv Ey(e_0) - w(e_0) \\
GFT_1 \equiv Ey(e_1) - w(e_1).
\]

Notice that the Agent receives a straight wage in the above calculations, hence there is efficient risk sharing. We will assume that \( GFT_1 > GFT_0 \), so working is more efficient than shirking. But will the Principal decide to implement \( e_1 \) or \( e_0 \)?

Implementing \( e_0 \) is easy. The Principal offers the Agent a straight wage of \( w(e_0) \). The straight wage contract \( C_0^* \equiv (w(e_0), w(e_0)) \) satisfies both \( IR(e_0) \) (this follows immediately from the construction of \( w(e_0) \)), and it also satisfies \( IC(e_0) \) (which follows from the fact that, given any straight wage, the Agent will choose her least costly action since she has no incentive to work harder; recall the quote from Mark Twain at the beginning). Thus the Principal’s expected profit from implementing \( e_0 \) is

\[
\Pi_0^* \equiv Ey(e_0) - w(e_0) \equiv GFT_0.
\]

That is, the Principal can fully appropriate the potential GFT from low effort.

### 4. Implementing high effort

Implementing high effort is not so easy. The participation constraint requires:

\[
Pv(w_S) + (1 - P)v(w_F) - e_1 \geq U \quad IR(e_1)
\]

The IC constraint for implementing \( e_1 \) requires:

\[
Pv(w_S) + (1 - P)v(w_F) - e_1 \geq pv(w_S) + (1 - p)v(w_F) - e_0.
\]

Henceforth we will assume \( e_1 - e_0 = 1 \). Thus this IC constraint can be re-written:

\[
(P - p)[v(w_S) - v(w_F)] \geq 1. \quad IC(e_1)
\]

Since \( P - p > 0 \), this implies \( v(w_S) - v(w_F) \) must be strictly positive; i.e., to implement high effort the Principal must offer the Agent a strictly higher wage when the project succeeds than when it fails — to give her an incentive to choose the more costly action \( e_1 \). That is, the spread between \( w_S \) and \( w_F \) must be positive (and sufficiently large) for the Agent to want to choose \( e_1 \); again recall the Mark Twain quote.

Let \( v_F \equiv v(w_F) \) and \( v_S \equiv v(w_S) \) denote, respectively, the agent’s utility wages if the project fails or succeeds. Notice both the constraints are linear in utility wages. In particular, with a bit of algebra, they can be re-written:

\[
v_S \geq \frac{U + e_1}{P} - \frac{1 - P}{P}v_F \quad IR(e_1)
\]

\[
v_S \geq \frac{1}{P - p} + v_F. \quad IC(e_1)
\]

These constraints are graphed below. The IC region is shaded in the Figure, while the IR region is marked by hatch marks.
The Principal’s least-cost way of implementing $e_1$ will be somewhere in the intersection of these 2 regions. We will show that if the Agent is risk averse, both constraints will bind in $C_1^*$. 

4.1 The risk averse case

$IR(e_1)$ will always bind—whether the Agent is risk averse or risk neutral. Otherwise the Principal could decrease $w_F$ and hence $v_F$ while keeping $w_S$ and hence $v_S$ unchanged. This tightens IR, and at the same time it loosens IC since the spread between $w_S$ and $w_F$ increases. See the movement from $V$ to $V'$ in the Figure.

Moreover, if the agent is risk averse, $IC(e_1)$ will also bind: By moving southeast on the IR-binding line, the Principal is moving closer to a straight wage contract, decreasing the risk the Agent must bear. Hence the Principal’s expected wage bill will decrease because he has to pay the Agent a smaller risk premium. See the movement from $V'$ to $V^*$ in the Figure.

Combining we see that if the Agent is risk averse, the cost-minimizing way to implement $e_1$ is using the utility-wage contract $V^*$ where both constraints bind. Solving the 2 equations in 2 unknowns yields:

$$v_F^* = \frac{U + e_1 - P}{P - p}$$

$$v_S^* = \frac{U + e_1 + 1 - P}{P - p}.$$ 

These cost-minimizing utility wages correspond to money wages $w_F^* = v^{-1}(v_F^*)$ and $w_S^* = v^{-1}(v_S^*)$. Hence, to implement high effort, the Principal will offer the incentive contract $C_1^* = (w_F^*, w_S^*)$. This will result in an expected profit of $\Pi_1^* = Ey(e_1) - Ew^*(e_1) = Ey(e_1) - c(e_1)$.

While $IR(e_1)$ is binding under $C_1^*$, $\Pi_1^* < GFT_1$ because the Principal must offer the Agent an incentive contract to implement $e_1$, which implies that $c(e_1) - w(e_1) \equiv RP(e_1) > 0$. That is, the Principal must offer the Agent a risk premium above her straight reservation wage in order to induce her to accept the incentive contract $C_1^*$. Hence, when there is asymmetric information, the Principal cannot fully appropriate the potential gains from trade from implementing high effort:

$$\Pi_1^* = GFT_1 - RP(e_1) < GFT_1.$$
Recall by way of contrast,
\[ \Pi_0^* = GFT_0. \]
See Figure 2.

![Diagram](image)

**Figure 2: Implement \( e_0 \) or \( e_1 \)?**

Which effort level will the Principal implement if the Agent is risk averse? Since \( GFT_1 > GFT_0 \), the answer rests on the size of the risk premium. If the agent is very risk averse, \( RP(e_1) \) will be very high, hence \( \Pi_1^* \) will be smaller than \( \Pi_0^* \), and the Principal will accept shirking: he will offer the Agent a straight wage. But if the Agent is not very risk averse, \( RP(e_1) \) will be small, hence he will implement high effort using the incentive contract \( C_1^* \). The Figure illustrates a case when \( \Pi_1^* > \Pi_0^* \).

Notice if the Agent is risk averse, the outcome is never socially efficient: If the Principal implements \( e_0 \), there is efficient risk sharing since the Agent receives a straight wage, but the Agent is not working enough since \( GFT_1 > GFT_0 \). On the other hand, if the Principal implements \( e_1 \), the Agent is given good incentives for working, but the contract that induces \( e_1 \) entails inefficient risk sharing since the Agent must bear some risk. Sometimes this strong inefficiency result is referred to as the horns of a dilemma: if the Agent is risk averse, there can be efficient risk sharing or efficient incentives, but not both. We will see that, if the Agent is risk neutral, this dilemma does not arise because the Agent does not mind bearing risk.

### 4.2 The risk neutral case

If the agent is risk neutral, we can assume without loss of generality that her utility function over income is \( v(w) = w \), hence there is no distinction between her utility wage and her money wage: \( v_F = w_F \) and \( v_S = w_S \). Recall that \( (1 - P)/P \) — the ratio of the probabilities of the 2 states when \( e = e_1 \) — equals the slope of the IR-binding line when \( e = e_1 \). But this also equals the slope of the Principal’s isocost lines \( (1 - P)w_F + Pw_S = a \) constant. Hence, it should be clear from Figure 1 that there are a continuum of contracts that minimize the Principal’s cost of implementing high effort when the Agent is risk neutral: any contract that is on the IR-binding line and that satisfies the IC constraint does the trick.

The most interesting contract in this family is the **residual claimant contract**:

\[ \tilde{C}_F = y_F - GFT_1 \quad \text{and} \quad \tilde{C}_S = y_S - GFT_1. \]

This contract is equivalent to the Principal selling the project to the Agent for \( GFT_1 \) (hence \( IR(e_1) \) is binding) and then letting the Agent do as she pleases (hence \( IC(e_1) \) is loose.
because the Agent will strictly prefer working hard as the residual claimant). Under this contract, both parties may be viewed as fully appropriating: The Principal fully appropriates the maximum potential gains from trade by selling the project for $GFT_1$. And, after buying the project, the Agent also fully appropriates because she is the residual claimant of the random variable $y(e)$, which depends on her effort choice. Once the project is sold, she can no longer impose a delivery externality on the Principal.

Since in this case $\Pi_1^* = GFT_1$, it follows from our assumption that $GFT_1 > GFT_0$ that the Principal will implement $e_1$. Hence the outcome is socially efficient. Unlike the risk averse case, now offering the Agent an incentive contract does not entail the expense of a costly risk premium; one of the horns of the dilemma has been eliminated.

4.3 A note on liquidity constraints, corner solutions, and the possibility of rationing in equilibrium

So far we have not imposed the non-negativity constraints

$$w_F \geq 0 \quad \text{and} \quad w_S \geq 0.$$  

Hence, for some parameter values, the cost-minimizing contract for implementing high effort will call for $w_F < 0$. Figure 3 illustrates for the case of a risk neutral agent. Notice for example, under the residual claimant contract $\hat{C}$ in the Figure, $GFT_1$ is sufficiently large that when the project fails, the agent loses money.

If the Agent has insufficient outside income to cover any negative wage outcome, we would impose on the Principal’s problem the additional non-negativity constraints above. In this even, for some parameter values, the cost-minimizing contract for implementing $e_1$ will satisfy “IC is binding” but “IR is loose.” To illustrate, see the contract $C^{**} = (0, w_S^{**})$ in the Figure.

This leads to the possibility of equilibria with “good jobs” that have an excess supply of applicants and “bad jobs” that have no excess supply. To illustrate suppose because of liquidity constraints $w_F$ must be non-negative, so $C^{**}$ is the least-cost way for the Principal to implement $e_1$. Also suppose the Principal pulls his manager/Agent out of a pool of identical potential managers all with the same outside option $U$. Then there will be rationing in equilibrium: All the potential managers in the pool will want the “good job” of being the
manager of the Principal’s project. But, unlike traditional neoclassical trading (without delivery problems), competition among the potential managers will not lead to a lower wage that eliminates the excess supply of labor. The Principal will be unwilling to pay any manager less than $w_s^{**}$ if the project succeeds because he realizes that any lower wage will lead to shirking (a failure of IC).