This lecture examines how ideas from agency theory help shape our understanding of key issues in labor economics. We’ll be using three basic models of imperfect or incomplete information:

▶ Moral Hazard. The uninformed party has imperfect information about actions taken by a second part, and the uninformed party moves first.
▶ Adverse Selection. The uniformed party has incomplete information about the characteristics or traits of the informed party. The uniformed party moves first.
▶ Signaling. The uniformed party has incomplete information about the informed party, but the informed party moves first.

In the cases we examine—one party makes a take-it-or-leave-it offer and the other party responds in her own interest.
We'll spend most of our time with the moral hazard. Our uniformed party is a principal and the informed party is an agent.

In the follow-up lecture we'll continue down this same path, but take a more behavioral approach. In particular, we'll ask what happens to key ideas from agency theory when people have nonstandard preferences (e.g., social preferences).
2.1. The Holmstrom Model of Moral Hazard

Suppose an agent takes a hidden action that affects a principal’s payoff. A key paper is Holmstrom (1979).

- The agent’s utility is
  \[ U(w, e) = u(w) - c(e), \]
  where \( w \) is wage, \( e \geq 0 \) is “effort,” \( u(\ ) \) is concave, and \( c(\ ) \) is convex. The agent has an outside option that gives utility \( \hat{U} \).

- Output \( y \) is determined by effort (with \( \frac{\partial y}{\partial e} > 0 \)) and by “nature.” We’ll work with the cumulative distribution of \( y \) conditional on \( e \), \( F(y|e) \).

- Value created for the principal is
  \[ v(y - w), \]
  where \( v(\ ) \) is concave.

A contract \( w(\ ) \) specifies payment from the principal to the agent, given outcomes that are observable and contractible; for now assume we observe only \( y \).

Now we have a game with the following timing:

- The principal offers the contract.
- The agent accepts the offer if it meets the participation constraint, and, if so, chooses effort \( e \).
- Nature draws output \( y(e) \), using \( F(y|e) \).
- The agent is paid the agree-upon wage, \( w(y) \).

Conceptually the solution is pretty simple; we solve the problem using backward induction.

1. Figure out the agent’s best response \( (e^*) \) for an offered contract (while making sure the agent accepts the job).
2. Given the best response, find the contract that maximizes the principal’s objective.
Actually, the problem can be a bit of a mess. The principal wants to maximize
\[ \int v(y - w(y))dF(y|e) \]
subject to the constraints:
\[ \int [u(w(y)) - c(e)]dF(y|e) \geq \hat{U}, \quad \text{and} \]
\[ e \in \arg \max_{\hat{e}} \int u(w(y)) - c(\hat{e})dF(y|\hat{e}). \]

- Here’s a concern: If \( w(\cdot) \) is non-convex, the agent’s best response problem is non-concave.
- If we assume away the first point, and suppose further that \( F \) is differentiable (\( f \) is the p.d.f.), then the agent’s first order condition is
\[ \int u(w(y))f_e(y|e)dy = c'(e). \]

I’ll try to persuade you that interpretation of (1) isn’t too hard!
- In the absence of the moral hazard problem, this is just a risk-sharing problem.
- When we do have moral hazard, the principal will engage in risk sharing and will typically reward better outcomes.
Intuitively, we expect that the reward structure \( w(y) \) should have higher pay for better outcomes \( y \).

- This needn't always happen. Lecture notes by Autor and Acemoglu provide this example: Suppose the agent can pick effort \( e_L \) or \( e_H \). If the low effort is chosen, output is

\[
y|e_L = \begin{cases} 
3 & \text{with probability } \frac{1}{2} \\
1 & \text{with probability } \frac{1}{2}
\end{cases}
\]

while if the high effort level is chosen

\[
y|e_H = \begin{cases} 
4 & \text{with probability } \frac{1}{2} \\
2 & \text{with probability } \frac{1}{2}
\end{cases}
\]

- Notice that the following contract perfectly insures a worker and gives the right incentive: pay \( w(2) = w(4) = \hat{U} + c(e_H) \) and \( w(1) = w(3) = -M \), making \( M \) sufficiently large.

- That contract is odd in that pay is higher when \( y \) is 2 than when it is 3; pay is not generally increasing in output.
Milgrom (1981) gives sufficient conditions to rule out weird cases like that one on the previous slide:

- $w(y)$ will be increasing in $y$, i.e., high values of $y$ will be good news if
  \[ \frac{f_a(y|e)}{f(y|e)} \]
  is increasing in $y$.

- Equivalently,
  \[ \frac{f(y|e_1)}{f(y|e_2)} \]
  is increasing in $y$ for $e_1 > e_2$.

This is known as the monotone likelihood ratio property (MLRP).
A final point from Holmstrom (1979):

- Suppose the principal observes $y$ and in addition observes $z$, a random variable that might be related to $e$. When should $w(\cdot)$ include $z$?
- The answer has to do with *sufficient statistics*. Suppose, using Bayes rule,
  \[ f(e|y, z) = f(e|y). \]
  Then $z$ adds no new information at the margin; $y$ is a sufficient statistic and the firm will *not* use $z$ as part of the contract.

To summarize, our principal agent problem leads to these insights:

- The solution entails risk sharing.
- Under a reasonable regularity condition (MLRP), pay will be increasing in observed performance.
- Pay will be based only on measures of performance that (on the margin) reveals information about the agent’s effort.

To make additional headway, we need more structure. For example

- Many papers study the case in which the agent has constant absolute risk aversion, $u(w) = -e^{-\alpha w}$, and assume normality in the stochastic structure. See, e.g., Holmstrom and Milgrom (1987) and Prendergast (1999).

There are many elaborations, including models that are based on both *objective* and *subjective* evaluation (e.g., Holmstrom and Milgrom, 1991, and Prendergast, 1999).
2.2. A Simple Building-Block Model

Our approach here is to simplify greatly—taking out the agent’s risk aversion, focusing only on the incentive part of pay, and doing so in a model that uses subjective evaluation.

- The principal has a payoff is given by
  \[ g(e) - w, \]
  where \( e \) is the agent’s effort and \( w \) is the wage.

- An agent has utility given by
  \[ w - e, \]
  and is willing to work for the principal if \( w - e \geq \hat{U} \).

For future reference, note that if the agent worked for herself, she would provide effort such that
\[ g(e^*) - 1 = 0. \]
To make the problem interesting, we assume:

- The principal has only a noisy signal of effort,
  \[ x = e + \epsilon, \]
  where \( \epsilon \) is drawn from density \( f \) (with c.d.f. given by \( F \)).
- Pay is based only on \( x \).

Principal-agent interaction then is as follows:

- The principal announces the policy of paying \( w_0 \) if \( x \) is less than some cut-off, \( \bar{x} \), and \( w_1 > w_0 \) if \( x \) is greater than \( \bar{x} \).
- The agent decides whether to accept the job, and if she does, takes hidden action \( e \).
- Nature plays \( x \) (observed performance).
- Given \( x \), the firm pays the agreed-upon wage.

To find an optimal wage policy, \((w_0^\ast, w_1^\ast)\), we start with the agent’s best response.

- At effort level \( e \), the probability of earning \( w_0 \) is \( F(\bar{x} - e) \) and the probability of earning \( w_1 \) is \( 1 - F(\bar{x} - e) \).
- So the agent wants to maximize \( w_0 F(\bar{x} - e) + w_1 [1 - F(\bar{x} - e)] - e \).
- The best response, \( \hat{e} \), then solves \( bf(\bar{x} - \hat{e}) - 1 = 0 \),

where \( b \equiv w_1 - w_0 \) is the “bonus” that the agent gets with the high-performance outcome. Notice that effort is a function of that bonus, \( \hat{e}(b) \), with

\[ \hat{e}'(b) = f(\bar{x} - \hat{e})/[bf'(\bar{x} - \hat{e})] > 0 \]

under the assumption that the second order condition holds.
Finally, to pin down wages we use the participation constraint,
\[ w_0 F(\bar{x} - \hat{e}) + w_1 [1 - F(\bar{x} - \hat{e})] \geq \hat{U} + \hat{e}. \]

Let’s now look at the principal’s problem.

- The principal maximizes profit, given the best response and given the participation constraint, i.e., chooses \( b \) to maximize 
  \[ g(\hat{e}(b)) - [\hat{U} + \hat{e}(b)]. \]
- The first order condition is
  \[ \left[ g'(\hat{e}(b^*)) - 1 \right] \hat{e}'(b^*) = 0. \]
- Since \( \hat{e}'(b) > 0 \) for any best response, \( e^* \equiv \hat{e}(b^*) \),
  \[ g'(e^*) - 1 = 0. \]

Compare with the case with no agency problem (Slide 13)!

To summarize, the agency problem is solved as follows:

- The principal sets a base wage \( w_0 \).
- The incentive is simple: a higher wage \( w_1 > w_0 \) is given when performance \( x \) exceeds some threshold \( \bar{x} \).
- The gap between the two wages (i.e., the bonus \( b \equiv w_1 - w_0 \)) must be set just right.
- Given the incentive structure, \( w_0 \) is simply set to just meet the participation constraint.

The solution leads to the efficient level of effort, with \( g'(e^*) = 1 \).

A potential complication:

- If \( b \) must be very large, \( w_0 \) might be negative.
- It might not be possible to enforce a payment from the agent to the principal; we might have limited liability constraint. Then the participation constraint does not bind, and the agent earns information rents.
One Principal and One Agent: CEO Compensation

- The CEO make a big contribution to firm profitability. Our agency model tells us that high “effort” is important here, so the reward to success should be large.
- Limited liability likely applies, which leads to higher pay yet (and possibly rents).
- This simple agency model seems to explain broad patterns in CEO compensation (e.g., Gayle and Miller, 2009).
- There may, however, be anomalies and dysfunction in CEO pay (Hall and Liebman, 1998, Bertrand and Mullainathan, 2001, Heron and Lie, 2009, etc.). Behavioral approaches might help understand these.

One Principal and Many Agents: Personnel Policies

When there are many workers, the principal (the firm) might use a “tournament” of the following form:

- All workers earn a base wage, say $w_0$.
- Workers are ordered by performance.
- A pre-announced proportion of top performers are “promoted,” earning $w_1 > w_0$.

The structure above is similar to Malcomson (1984) and to many other tournament models.
We modify our building-block model to study a tournament structure.

- A profit-maximizing employer has \( n \) workers produce output, \( Y = G(e^1, e^2, \ldots, e^n) \) per period, where \( e^i \) is worker \( i \)'s effort.
- We treat the firm’s agency problem with a given worker in terms of the function \( g(e^i) = G(e^i, e^{(-i)}) \), where \( e^{(-i)} \) is effort levels of workers other than \( i \).
- Workers chooses \( e \) and receives utility \( w - e \).

The firm does not condition compensation on \( Y \); it uses subjective performance measures \( x \) for each worker.

- Typically \( x \) cannot readily be used as the basis for forming contracts that can be enforced by an outside court.
- Workers might worry about that the firm will renege on an agreement to directly condition pay on \( x \).

How might the firm proceed? A “tournament” might help:

- Workers are ordered on the basis of \( x \), and then
- the fraction \( P \) who are lowest-performing are paid \( w_0 \), while the remaining high-performing workers are paid \( w_1 > w_0 \).

The idea is that everyone observes the agreed-upon reward structure, and can see if the firm honors that agreement. What is the equilibrium?

- Suppose worker \( i \) believes other workers play \( \hat{e}^{(-i)} \). Then using her knowledge of \( P \), she can deduce the cut-off performance value \( \hat{x} \) that will result in promotion.
- Given \( \hat{x} \), chooses effort \( e^i \) to maximize
  \[
  w_0 F(\hat{x} - e^i) + w_1 [1 - F(\hat{x} - e^i)] - e,
  \]
  so here best response is exactly as in our building-block model (see Slide 15):
  \[
  (w_1 - w_0)f(\hat{x} - \hat{e}) - 1 = 0.
  \]
So the firm has a workable plan here:

- The firm starts by setting the “tournament prizes” \( (w_0^*, w_1^*) \).
- It chooses the fraction \( P^* \) and wages so as to just satisfy the participation constraint.
- If worker \( i \) believes other workers are choosing effort level \( \tilde{e}(-i) = e^* \), she responds by also choosing \( e^* \). All workers behave the same in equilibrium, and effort is efficient.

The logic can be used to understand why, in many organizations:

- workers fall into distinct pay grades,
- workers in high-paid positions are promoted from within,
- wages typically rise with seniority (perhaps by more than productivity),
- and the variance of wages increases with seniority.

For more on this topic, see Lazear (1998), Prendergast (1999), Malcomson (1999), and Gibbons (1998).

---

Many studies show *incentives work*, pretty much as predicted by agency theory.

- See Lazear (2000), Kahn and Sherer (1990), Jacob (2003), work by Groves and McMillan and co-authors on incentives and productivity in Chinese industry, etc.

But many studies that show *incentives work* do so in cases in which incentives are poorly designed.

- See Oyer (1998), Courty and Marschke (2004), Gaynor, Rebitzer and Taylor (2004), Levitt and Jacob (2003), etc.

Here is an additional point to ponder: Our approach to agency posits that ex-post inequality generated by the pay system is irrelevant. Agents are fine with the fact that rewards are often as much “pay for luck” as “pay for performance.”
Many Principals and Agents: Unemployment and Labor Market Segmentation

Take the baseline agency model, and now add these elements:
- The game is repeated (indefinitely), with a discount rate $\rho$.
- Limited liability is invoked via an assumption that the only penalty the principal can impose is to dismiss the worker.

Now in our model, in each round:
- The principal offers wage $w$.
- The agent chooses effort $e$.
- Nature plays $x$ according to the distribution $f(\cdot)$.
- The firm pays $w$, and continues the game if $x > \bar{x}$.

This last step is new. The “reward” for good performance is to continue the game (else you go to a state with utility $= 0$).

Ritter and Taylor (2011) solve this problem. I’ll sketch out the basic idea.

First we need the worker’s best response $\hat{e}(w, \bar{x})$ in any period.
- A worker who chooses $e$ in the current period, and then reverts to $\hat{e}$ in future periods, has lifetime utility,
  \[ V(e) = w - e + \frac{[F(\bar{x} - e)V^u + (1 - F(\bar{x} - e))V(\hat{e})]}{1 + \rho}, \]
  where $V^u$ is the utility of losing one’s job.
- The worker maximizes $V(e)$, choosing $\hat{e}$ to solve
  \[ \frac{V(\hat{e}) - V^u}{1 + \rho} f(\bar{x} - \hat{e}) - 1 = 0. \]
  This is a familiar expression (see Slide 15).
- Solve for $V(\hat{e})$ and substitute into the second equation:
  \[ w = \hat{e} + \frac{\rho V^u}{1 + \rho} + \frac{\rho + F(\bar{x} - \hat{e})}{f(\bar{x} - \hat{e})}. \]
  This last expression implicitly defines the worker’s best response, $\hat{e}(w, \bar{x})$.
Now we turn to the firm’s objective.

- The firm maximizes profits given the worker’s best response, i.e., maximizes $$g(\hat{e}(w, \bar{x})) - w$$.
- It is easy to show that $$g'(e^*) = 1$$.

In Ritter and Taylor (2011) we show:

- At the optimal $$w^*$$ and $$\bar{x}^*$$, the probability of retention, $$F(\bar{x}^* - \hat{e}^*)$$, is unaffected by the level of $$\sigma^2$$, the variance of the density $$f(\cdot)$$.
- The wage is
  $$w^* = e^* + A + B\sigma,$$
  where $$z^* = (\bar{x}^* - e^*)/\sigma$$, and $$\phi(\cdot)$$ is the p.d.f. of $$\xi$$.

We conclude:

- Workers earn an “information rent.”
- The more intractable the agency problem—the greater the value of $$\sigma$$—the higher is the wage required to achieve efficient effort and so the greater the rent accruing to workers.

This is an efficiency wage model, in which the firm is using a carrot and a stick.

- The carrot is the promise of continued employment in a valuable job.
- The stick is the threat is dismissal for substandard effort.
2. The Principal-Agent Problem
3. Agency Matters
4. Signaling and Adverse Selection

Suppose we have an economy in which all firms pay efficiency wages, as described above. It is pretty easy to show that in equilibrium we have unemployment.

- First notice that $V^u$ must depend on the unemployment rate $u$ in a mechanical way.
- After a bit of algebra we find

$$w = e^* + \frac{1}{\phi(z^*)} \left( \rho + \frac{F^*}{u} \right) \sigma. \quad (2)$$

where $F^* \equiv F(\bar{x}^*, e^*)$. This expression shows potential equilibrium wage and unemployment levels for the labor market.

The Figure on the next slide shows the market equilibrium when long-run labor demand is perfectly elastic, at $w^E$.

- Equilibrium unemployment, $u^E$, solves

$$w^E = e^* + \frac{1}{\phi(z^*)} \left( \rho + \frac{F^*}{u^E} \right) \sigma.$$

Clearly $u^E > 0$.

Notice that an increase in $\sigma$ increases unemployment. This outcome is intuitive.

- The weaker the link between the dismissal threat and employee behaviors, the stronger are the incentives required to elicit the desired effort level.
So our model of incentives can lead to

- unemployment, as set out in Shapiro and Stiglitz (1984), and
- “labor market segmentation,” as discussed in Bulow and Summers (1986), if “information rents” vary by firm and industry.

Evidence is found in Krueger and Summers (1988) and many subsequent papers on inter-industry wage differences. Also Cappelli and Chauvin (1991), Nagin, Rebitzer, Sanders, and Taylor (2002), etc.

In Ritter and Taylor (2011) we suggest that the model might help understand racial differences in unemployment:

- Let $B$ workers and $W$ workers be equally productive, and $\sigma_B > \sigma_W$. The wages will be the same for the two groups of workers, but unemployment will be higher for $B$ than for $W$. 

Compensation policies are often asked to do “double duty.” In Lazear’s (2000) study of compensation practices at Safelite, a windshield installation company, an explicit piece rate system provided an \textit{incentive} to work harder, and had a \textit{selection} effect, drawing workers who liked the income-effort tradeoffs inherent in the piece rate system.

There are many examples. We’ll look at two examples—one entailing \textit{signaling} and one with \textit{adverse selection}. Let’s review the two models first. Recall:

- Adverse Selection. The uniformed party has incomplete information about the characteristics or traits of the informed party. The uniformed party moves first.
- Signaling. The uniformed party has incomplete information about the informed party, but the informed party moves first.

\section*{Signaling}

In a very famous game, formulated by Spence (1973), education is taken to be a signal of ability (which is otherwise hidden). Let’s briefly review the idea.

- Workers are born with an “ability type” which determines productivity.
- Workers know their own type, but that is information is hidden from employers.
- Workers’ acquisition of education serves to \textit{signal} employers about their type.
- For the signal to be effective it must be “costly to fake.”
Slightly more formally, here the game:

- Nature determines workers' productivity \( V \) to be either \( V_H \) or \( V_L < V_H \), drawing type L with probability \( q \).
- Workers choose education \( E \). Utility is
  \[
  u(w) - C(E, V),
  \]
  where \( u' > 0 \) and \( u'' < 0 \). Also, we'll assume
  \[
  \frac{\partial C}{\partial E} > 0, \quad \frac{\partial^2 C}{\partial E^2} > 0, \quad \frac{\partial^2 C}{\partial E \partial V} < 0.
  \]
  This last assumption, the "Mirrlees-Spence condition," is crucial.
- Firms in a competitive labor market pay wage, \( w(E) \), equal to expected worker value,
  \[
  w(E) = \mu(E)V_L + (1 - \mu(E))V_H,
  \]
  where \( \mu(E) \) gives the employers’ beliefs about the probability that the worker is type L, given \( E \).
A perfect Bayesian equilibrium is a set of strategies, 
\((E^*_L, E^*_H, w^*(E))\) and beliefs \(\mu^*(E)\), in which:

- A type-L worker is choosing her optimal level of education, 
given expectations about the wage function, i.e., solving 
\[E^*_L \in \arg\max_E \{u(w^*(E)) - C(E, V_L)\}.\]

- A type-H worker is choosing her optimal level of education, 
given expectations about the wage function, i.e., solving 
\[E^*_H \in \arg\max_E \{u(w^*(E)) - C(E, V_H)\}.\]

- Employers are paying 
\[w^*(E) = \mu^*(E)V_L + (1 - \mu^*(E))V_H.\]

- Beliefs \(\mu^*(E)\) are consistent with strategies. So if \(E^*_L \neq E^*_H\):
  - if \(E = E^*_L\) then \(\mu^*(E) = 1\), and if \(E = E^*_H\) then \(\mu^*(E) = 0\).
  - But if \(E^*_L = E^*_H = E\), then \(\mu^*(E) = q\).

There are many equilibria, but one we want to emphasize:

- There is a pure strategy separating equilibria, in which \(E^*_L = 0\) and \(E^*_H > E^*_L\).

- In that equilibrium we must have 
\[u(V_L) - c(0, V_L) \geq u(V_H) - C(E^*_H, V_L),\]
and 
\[u(V_H) - c(E^*_H, V_H) \geq u(V_L) - C(0, V_H).\]

There are many separating equilibria like the one we’ve described.

- The reason is that the concept of Bayesian equilibrium places no restrictions on out-of-equilibrium beliefs.
- Cho and Kreps (1987) provides the intuitive criterion for ruling out all but one of these equilibria (the least costly one).
Adverse selection is the mirror image of signaling. The key
difference is that the uninformed party moves first.

▶ The classic example is health insurance: A firm that offers a
contract specifying a set of benefits and prices can
disproportionately attract consumers who know themselves to
be at risk of health problems.

▶ There are many applications in labor markets.
2. The Principal-Agent Problem
3. Agency Matters
4. Signaling and Adverse Selection

3. Agency Matters

4. Signaling and Adverse Selection

4.1. High Wages as a Signal of Firm Fitness

In labor market examples, the idea typically goes like this:

▶ There are two or more types of workers who vary in terms of preferences, productivity, etc. (Types are hidden.)
▶ Employers move first by presenting potential employees with a menu of options specifying wage, bonus structure, benefits, work hours, etc.
▶ The hope is that each type of worker selects into the contract designed specifically for him or her.

Just as in the corresponding signaling model, we need a Mirrlees-Spence condition to generate a separating equilibrium.

4.2. Adverse Selection: The Rat Race

In finance and IO there are important applications in which firms use signals:

▶ By taking an equity stake in a project, an entrepreneur might credibly signal a high-quality project (Leland and Pyle, 1977).
▶ Aggressive introductory pricing and advertising can signal that a newly-introduced good has high quality (Milgrom and Roberts, 1986).
A comparable application to labor markets is found in Ritter and Taylor (1994): Firms have an unobservable characteristic, firm probability of survival.

- Worker motivation depends on the job having value—by a worker-posted bond, efficiency wages, or a combination.
- A separating equilibrium exists in which high-fitness firms signal their fitness by paying wages that provide rents.
- Efficiency wages arise without resorting to an assumption of limited liability.

In Akerlof’s (1976) “rat race” workers have a hidden characteristic, an “inclination to work hard.”

- Workers differ in productive and in their inclination to work hard.
- Workers who like to work hard are also more productive (the Spence-Mirrlees condition).
- An adverse selection equilibrium emerges in which the high-productivity workers over-work as a means of credibly establishing a hard-to-observe trait.
In Landers, Rebitzer and Taylor (1996) set the rat race in a simple dynamic (two period) model of law firms:

- All lawyers are equally productive, but workers vary in their inclination to work long hours:
  \[ u_1(w, h) = w - b_1 h^2, \]

  and

  \[ u_2(w, h) = w - b_2 h^2, \]

  where \( b_1 > b_2 \) (so type-1 lawyers find long hours more distasteful than type-2 lawyers).

- Lawyers hired in period 1 work as “associates.” In period 2 they become equity partners, at which point they tend to “free ride.”

- Type-2 lawyers do less free riding; this is where the Spence-Mirrlees comes in.

- In equilibrium, lawyers run a rat race—they over-work!

Many of the ideas used to set up the building block model stem from my work with Joseph Ritter:


Almost all of the references in the slides can be found in:

Some additional references: