Principal-Agent Theory Notes

May 22, 2005

This is a rough guide to these issues I wrote up for myself. Use with caution. I probably made mistakes (please tell me) and I did not explain a lot of things fully.

1 Basic features of principal agent problems

The key unifying features of principal-agent problems are that

1. the principal knows less than the agent about something important, and
2. their interests conflict in some way.

There seem to be two types of problems:

- Problems where agents can do some costly action to improve outcomes for the principal but the principal can't observe the action. These are known as **effort aversion/moral hazard** problems.
- Problems where there are different types of agents and principals can't tell the difference among them. These are known as **adverse selection** when the types are fixed and the question is which agents will participate.

Nolan Miller's notes identify a third type, which he refers to as **hidden information** models, but this category does not seem very well-defined. One of the examples he gives, of managers in firms who take actions that advance their own careers but hurt shareholders, sounds like an effort aversion problem with a different definition of effort. Rather than encouraging the agent to undertake a certain action that will more likely lead to good outcomes for the principal, the principal instead wants to discourage the agent from taking certain actions that will more likely lead to bad outcomes for the principal. Subtract the extent of those nefarious actions A from effort e and you've got an effort aversion model.

The other cases listed under hidden information have in common that agents act differently when they are insured than they would otherwise – insured patients make more trips to the doctor than they would without insurance, banks make more risky loans than they would if there were no bank bailouts. There is an information problem here, in the sense that the doctors would like to turn away frivolous patients and governments would like to let overly risky banks flounder, but the principals here can't tell the difference between legitimate and extraneous calls for help. (This is because all patients need medical attention sometimes, and even well-run banks fail.) But it may be more appropriate to think of these problems as externality problems, in the sense that the "social" cost of the insured agent's action (going to the doctor, taking on bad loans) is much higher than the private cost. One of the purposes of insurance is to reduce the private cost of going for help in event of mishap, but it could be argued that when there is overuse of this kind the private cost should be increased, for example by increasing the co-payment for doctor visits or completely replacing bank leadership in the event of bank failure. At any rate, these examples are traditionally referred to as "moral hazard" but Nolan Miller's notes reserve that term for effort aversion problems in which unobserved but costly efforts by the agent makes are useful to the principal. I turn to those problems now.

2 Moral Hazard/Effort Aversion

It is useful to think about contracts between principals and agents in two parts:

- **Risk-sharing**: In all of these problems the outcome and effort level are stochastically related. (If there was a deterministic relationship, it wouldn't matter that the effort was not observed.) The risk aversion of the players therefore is a factor. The contract will feature payoffs as a function of stochastic outcomes, and the agent will only sign a contract that provides him enough expected utility to overcome his reservation utility. The agent's expected utility will depend on the probabilities of various outcomes, the payoff that the he receives for those outcomes, and the utility that he receives as a function of those payoffs. With fixed probabilities and payoffs, the agent's expected utility will be a decreasing function of his risk aversion. To convince the agent to sign a contract, therefore, the principal must offer payoffs that are either more generous or more equal as the agent becomes more risk averse.

- **Incentives**: By specifying different payments for different outcomes, the contract sets up incentives for the agent as he chooses an effort level. If under a given contract payments are larger for higher outcomes, and higher effort is more likely to produce higher outcomes (as is usually the case in these models), then the agent may be motivated to exert higher effort. If his expected utility from signing the contract and exerting some level of effort is larger than his reservation utility, he will sign the contract (this is the risk issue); if his expected utility from exerting high effort is higher than his expected utility from exerting any other level of effort then he will exert high effort (this is the incentive issue).

2.1 Contracting for effort: first best

If effort is observable, we can write contracts that specify a level of effort and are thereafter enforced in courts. The task facing the principal is thus twofold:

- For a given level of effort specified in a contract e^*, the principal must choose the payoff vector $\boldsymbol{w} = (w_1, w_2, ...w_N)$ (a wage for each of N outcomes) that maximizes her own utility while convincing the agent to participate (providing the agent with an expected utility greater than or equal to his reservation utility), and

- Choose a level of effort that maximizes her expected utility.

The optimal contract at a given level of effort will provide the agent *exactly* his reservation utility (in expectation). Since the outcomes are distributed stochastically, there are a lot of ways to do this: at one extreme, the contract could be set up like a lottery, where a huge payment is given for an extremely rare outcome and no payment is given otherwise; at the other extreme, the contract could give the agent the same wage regardless of the outcome. Which contract the principal chooses will depend on the risk aversions of the principal and agent.

This risk sharing problem with observable effort (the "first-best optimal risk sharing problem") can be written as follows:

$$
\begin{aligned}
& max_{w_n \geq 0} \sum_{n=1}^{N} p_n(e^*) B(x_n - w_n) \\
& \text{s.t.} \sum_{n=1}^{N} p_n(e^*) v(w_n) - g(e^*) \geq u_0 \qquad \text{(PC)}
\end{aligned}
$$

The Lagrangian solution to this problem is:

$$
\frac{B'(x_n - w_n^*)}{v'(w_n^*)} = \lambda^*, \qquad \forall n.
$$

If the agent is risk averse and the principal is risk neutral, the best contract for the principal for a given level of effort provides the same wage regardless of the outcome (ie the agent is fully insured). If the agent is risk neutral and the principal is risk averse, the situation is reversed, and the optimal contract gives the principal the same net payoff $(x_n - w_n^*)$ regardless of the outcome, which means that the principal is fully insured. In other words $x_i - w_i^* = x_j - w_j^* = F \forall i, j$, which means that the agent pays the principal the fixed amount F for the business, and receives a net payoff of $x_n - F \forall n$. If both are risk averse to some extent, the optimal contract will share the risk. The principal will still be choosing the payoffs $\boldsymbol{w}^*$ that are most beneficial to her while (barely) convincing the agent to sign the contract.

I like to think of the Lagrangian multiplier λ as a shadow price that sets the value of $\boldsymbol{w}^*$ so that the budget constraint binds. Think of a different plot of $\frac{B'(x_n - w_n^*)}{v'(w_n^*)}$ against w_n^* for each n. In the case where both the principal and the agent are risk averse, this will start at the origin, sloping upwards, and approach a vertical asymptote where $w_n^* = x_n$. λ can be thought of as the price in this plot, which dictates an optimal wage $\boldsymbol{w}^*$ such that the constraint binds (the agent barely participates). It is therefore the "bang for the buck" we talk about so often in this kind of situation – the marginal benefit to the principal per marginal benefit to the agent. Dollars devoted to different payoffs w_n^* should have the same "bang for the buck," and this value should be such that you just barely convince the agent to participate.

To sum up, when effort is contractible the principal-agent problem reduces to one in which the principal chooses a level of agent effort that gives her the highest expected utility. She figures out what the optimal risk-sharing contract is for various levels of effort and then chooses the contract (and therefore level of effort) that is best for her.

2.2 Contracting for outcomes: second-best

When effort is not contractible (which is usually the case in these problems) the contract must specify outcomes, not effort levels. This development adds a level of complexity in that now we want to write a contract that leads the agent to choose the right amount of effort. Assume that higher effort makes high outcomes more likely. What kind of contract would a) lead the agent to sign, and b) lead the agent to choose high effort? Again, extreme solutions can be illuminating:

- The contract could be like a lottery in which all the chips go to the high outcome. If you make the prize large enough to encourage participation, then the agent will choose high effort. But this contract may not be optimal because the principal might be able to induce high effort at a lower expected payoff by moving some of the lottery prize to lower outcomes. A risk-averse agent may will be willing to participate at lower expected wage but still choose high effort.

- The contract could fully insure the agent against risk, ie the agent is paid a fixed wage regardless of outcome. But in this case the agent would never choose high effort.

You can think of this as a tradeoff between risks and incentives. The principal wants to impose some risk on the agent in order to motivate him to choose higher effort (and ultimately produce higher outcomes, which is what the principal really cares about). But imposing higher risk means providing a higher expected wage (because the risk-averse agent needs compensation to participate).

The only difference between the first-best problem (with effort observed) and the second-best problem (outcomes, not effort, observed) is the addition of the "incentive compatibility constraint":

$$\sum_{n=1}^{N} p_n(e^*)v(w_n) - g(e^*) \geq \sum_{n=1}^{N} p_n(e)v(w_n) - g(e) \qquad \forall e \qquad \text{(ICC)}$$

This says that e^* has to give the agent the highest level of expected utility of all levels of effort.

Both constraints bind. Why? For fixed probabilities $p()$, effort costs $g()$, utility functions $v()$, and reservation utilities u_0, the PC depends on the level of the payoffs for $e*$ (ie the expected utility given e^*) while the ICC depends on the relative values of the payoffs w_n across outcomes. If the PC did not bind (ie the agent is getting a higher expected utility than his reservation utility), then the principal could benefit herself by revising the payoff vector $\mathbf{w}$ such that e^* still provides the same utility to the agent as his next best choice of e (ie ICC still binds) but the absolute level of utility provided is lower. In other words, if the principal makes all of the payoffs slightly lower, then the agent will pick the same level of effort but at a lower cost to the principal. If the ICC does not bind (ie the agent is picking the right level of effort and the resulting expected utility is strictly greater than that for any other effort level), then the principal could benefit herself by revising the payoff vector $\mathbf{w}$ such that the agent is burdened with less risk. If there is less risk on the agent, the expected payoff can be made lower while providing the same expected utility for that level of effort. In other words, the original situation had incentives that were too supercharged (like the lottery scenario above) and, by turning down the volume on those incentives, the principal can pay less in expectation.

For the simplest case, with two outcomes, two effort levels, and a risk-neutral principal, the ICC and PC together dictate the wages paid for each outcome level in order to induce a particular effort level at lowest cost. Since there are only two effort levels, the ICC must bind with equality; with more effort levels, I think ICC binds with equality for only one pair of effort levels. The approach to solving the problem is thus to find the payoffs $\mathbf{w}$ that would cause the agent to choose each effort level and then evaluate which effort level is better for the principal. Inducing low effort when the principal is risk-neutral always calls for a constant wage that produces the agent's reservation utility. (Fully insuring the agent is the cheapest way to encourage participation.) The low-effort contract will be better for the principal when the agent is very risk averse; inducing participation and high effort at the same time will require a high expected wage to compensate such an agent for taking on risk.

I have tried to think about the case of more possible effort levels a bit and have not gotten far. The question is, "How do you know which incentive compatibility constraint should bind?" Which of the two effort levels should produce equal utility for the agent? First of all it may be important to recognize that the principal doesn't really care about effort levels, but rather about outcome distributions. All else equal, the principal wants the agent to do anything that produces a better outcome distribution. With that in mind, it's tempting to say that we should use FOSD and say that the ICC should bind between the two effort levels that produce outcome distributions that FOSD all others. There is not necessarily going to be such a pair. Another thought is, "Is there a $\mathbf{w}_e$ for each e that makes that e optimal for the agent?" It seems like there might be. So one point is that, for each e^*, the candidates for a binding ICC would be limited to those e's that come in second for values of $\mathbf{w}^*$ that make e^* optimal for the agent. Still, this is computationally a pretty hard problem. It may also help to make e continuous and x some distribution using e as one of its arguments.

2.3 Comparing the first- and second-best solutions

Since the second-best problem is the first-best problem with an additional constraint, the first-best solution must weakly Pareto-dominate the second-best solution: The principal would do weakly better if she could contract for effort and not output, and the agent does the same (u_0) either way. That means that the principal may be willing to devote resources to making effort contractible, and the agent wouldn't stop it.

In the simplest (two effort levels) case, if the principal chooses to implement high effort in the second-best scenario, then she would choose to do the same in the first-best. But there may be cases where she would implement high effort in the first-best but not the second-best scenario – those cases where the compensation she must give the agent to bear the incentive-producing risk is larger than the expected increase in output from the higher level of effort.

3 Bibliographic note

These notes are based largely on Nolan Miller's "Notes on the Principal-Agent Problem," revised 3/25/2005.