1. The Principal-Agent Problem

In the basic Principal-Agent problem, the Agent chooses $e$ (effort) to maximize:

$$W = R(q(e)) - V(e)$$

(1)

Where $V(e)$ is disutility of effort, $q$ is the agent’s output, and $R$ is the agent’s monetary reward. Note two important things: First, the utility function is linear in income, $R$. This means there are no income effects on effort supply. The usual justification for this is that the rewards we are thinking about are small relative to lifetime income. Second, for simplicity (1) ignores uncertainty in the relation between effort and results but the vast majority of actual treatments include it, for example in the form $q = Q(e, \varepsilon)$, where $\varepsilon$ is a random variable outside the agent’s and principal’s control. In that case, risk aversion matters. When risk aversion is considered in these models, it is pretty much always modelled by letting the utility function be $U[R(q(e))] - V(e)$, where $U' > 0$, $U'' < 0$. This is still considerably simpler and more restrictive than the general case $U(R, e)$, because $U_{CL}$ equals zero.

The Principal, in turn, chooses $R(q)$ (note this is a function) to maximize expected profits:

$$q - R(q)$$

(2)

subject to two constraints:

1. The agent’s maximization as in (1) (“incentive compatibility”). Sometimes this is characterized by a first-order condition, but care must be taken if you do that…..

2. The agent’s participation constraint, $W \geq \bar{W}$, where $\bar{W}$ is the agent’s alternative (or target) utility level.

Thus, note that this is a “Stackelberg” type problem: The principal moves first, but chooses her actions anticipating how the agent will respond to $R(q)$. To solve these problems, the economist proceeds backwards in time, solving (1) first, then using that information to solve (2).

Also, note that the participation constraint ensures that the solution is (constrained) Pareto-Optimal, i.e. it maximizes one party’s utility (the principal’s) subject to a minimum level of the agent’s. Thus, by varying $\bar{W}$, we can find not just the best solution for the principal that prevents the agent from quitting, but any Pareto-optimal reward schedule that satisfies the incentive-compatibility constraints.
2. Theoretical Predictions

a) The most basic result in P/A theory is that, when there is no uncertainty (or, more importantly, when the agent is risk neutral), the optimal reward schedule takes the form:

\[ R(q) = q + k \]

Note that this is linear and the coefficient on \( q \) is 1. In other words, this is a “100% piece rate”, where the agent receives every dollar of (net) revenue he generates for the principal. Note that, for the principal to make any profit under this arrangement, \( k \) must be negative. Essentially, the principal “sells the job” to the agent.

b) If the agent is risk averse and nothing else changes, then the optimal reward schedule, \( R(q) \), balances incentives and risk sharing. If the contract cannot be made state-contingent, the problem is now identical to the standard moral hazard problem in insurance, where \( e \) is the agent’s behavior (care). Here, the main prediction is that increases in production risk (\( \sigma \)) reduce the optimal degree of incentivization (Higgs 1973).

c) If there is more than one agent and the outputs of individual agents are observed by the principal, the optimal \( R(q) \) may involve payment by relative output (“tournaments”). Key theoretical predictions include:

- with risk-neutral agents, tournaments can replicate any piece rate, using less information
- with risk-averse agents and common shocks, tournaments can dominate piece rates (Lazear and Rosen; Knoeber)
- tournaments discourage helping, information sharing; encourage sabotage (Carpenter)
- when contestants are few, tournaments are vulnerable to collusion (Bandiera et al.)
- if agents can influence \( \sigma \), tournaments can encourage excessive risk-taking, especially for low ability agents, or agents who are behind (Brown, Harlow and Starks)
- tournaments encounter problems when agents abilities differ too much (Jennifer Brown)
- Handicaps/leagues can fix this (Calsamiglia et al.)
- self-selection into tournaments can be suboptimal, for both ‘classical’ and ‘behavioral’ reasons (Barber and Odean, Niederle and Vesterlund)

d) If there is more than one agent and the outputs of individual agents can’t be observed, we are in a team production situation. This raises very general incentives to free ride: no budget-balancing sharing rule generates optimal effort (Holmstrom).

- theoretically optimal solutions to team incentives involve extreme incentives and budget-breaking out of equilibrium (Eswaran and Kotwal).
- empirically, people free ride much less than ‘classical’ theory predicts (though you can get them to free ride if you structure an experiment just right).
- one reason people don’t free ride is altruistic punishment (Fehr et al.) or peer pressure (Kandel and Lazear).
- team-based pay incentivizes workers to cooperate, share information and monitor each other.
- in some cases (Hamilton et al.) these incentives can be strong enough to outweigh any free rider effects, and to make teams more productive than piece rates even when individual output is observed.
- when the production function exhibits complementarities (e.g. weakest-link production), there can be multiple equilibria in agents’ effort choices.
- sometimes extreme complementarities (whether inherent in production or created by the principal) can actually improve team performance (Knez and Simester)

e) If the agent performs more than one task, incentivizing one task may cause the agent to neglect the other. This is the multi-task PA problem. A number of authors have shown that, in such situations, zero incentives ($R(q) = K$) may be the Pareto-optimal policy (Farrell and Shapiro; Fehr-Schmidt lab experiment).

f) A number of interesting issues arise when we consider multi-period PA interactions:

- if firms use dismissal threats as incentives, their power can be increased by deferring compensation (Lazear 1979, 1981), but this raises incentive for firms to renege on long-term contracts.
- if either the agent’s type or the “difficulty of the job” is unknown to the firm, agents have an incentive to hide this information in early periods. This ‘ratchet effect’ problem leads to suboptimal effort levels (Charness, Kuhn and Villeval).
- if the agent’s type is unknown to firms but their job performance is public information, high-ability agents will want to signal their type in early periods by working hard. This ‘career concerns’ problem can lead to socially excessive efforts in early periods. Some have argued that within-firm incentive pay (i.e. $b>0$) may be unnecessary in these environments.
3. Evidence

There is a rich and rapidly growing body of evidence that examines the above predictions of principal-agent models, based on a combination of lab experiments, field experiments, and “insider econometrics” studies of firms’ compensation/HRM policies and productivity outcomes. For a recent survey of the lab evidence (with lots of references to field and econometric studies as well) see Charness and Kuhn (2010), available on my website. For a survey of the ‘insider econometrics’ literature, see Ichniowski and Shaw (2009).

The literature on lab and field experiments has confirmed a number of the predictions of ‘classical’ principal-agent theory, including the following:

- Suppose the reward schedule is linear, i.e. \( R(q) = a + bq \). Then, for the most part, higher piece rates (i.e. higher values of \( b \)) induce higher effort.
- Properly-designed tournaments can mimic piece rates fairly well.
- Higher incentives in tournaments can raise sabotage as well as effort.
- There is a lot of free riding under team compensation (VCMs).

At the same time, a number of highly robust ‘anomalies’ that are inconsistent with the ‘classic’ model (which assumes rational, egoistic behavior) have been identified, including:

- A zero piece rate (\( b=0 \)) can yield higher effort than a positive one (Gneezy and Rustichini 2000; Irlenbusch and Sliwka 2005).
- Restrictions on agents’ choice sets can reduce the efforts of agents on whom they are not binding (Falk and Kosfeld 2006).
- A principal’s decision to monitor the agent can also reduce agents’ efforts. (Dickinson and Villeval 2008).
- Unenforceable promises to pay bonuses for ‘satisfactory’ worker performance can elicit surprising amounts of effort, and can outperform more objective mechanisms such as random monitoring combined with punishment (Fehr, Klein and Schmidt 2007).
- Even when there is no strategic independence between workers, workers’ efforts may depend on their co-workers’ efforts (Falk and Ichino 2006; Mas and Moretti 2009).
- Teams can perform very well if members are allowed to punish each other. (Fehr and Gachter, Nature 2002).
- Mutual monitoring and information exchange can make team-based rewards more effective than individual rewards. (Hamilton, Nickerson, and Owan, JPE 2003).
One ‘anomaly’ for classical principal-agent theory that has been identified in empirical studies is the fact that effort can respond to the intercept of the worker’s compensation schedule \((a)\), as well as to the slope parameter, \(b\). This has now been demonstrated dozens (if not hundreds) of times in experiments that implement gift-exchange labor markets: In a gift exchange labor market, a firm first pays a worker a wage. Then the worker decides how hard to work. Effort is costly to the worker but benefits the firm. After that, the pair never meet again. Clearly the only perfect equilibrium to this game is zero effort and zero wage. But that is not what happens: in practice, workers reciprocate high wage offers by supplying high effort. Anticipating this, firms pay high wages, and both parties attain much higher payoffs than are predicted by the standard model.

One of the first papers to demonstrate the above result is Fehr, Kirchsteiger and Riedl (1993). Charness (2004) provides important evidence that workers’ behavior in these games is driven by reciprocity (as distinct from other possible causes such as inequity aversion). Overall, the results of these experiments have been seen as supportive of the branch of efficiency-wage theory associated with Akerlof and Yellen’s “fair-wage effort hypothesis” (QJE 1990).¹

The real-world phenomenon that gift-exchange labor markets are designed to represent is the fact that (unlike, say, futures contracts for pork bellies) labor contracts are incomplete. In other words, no matter how hard one might try, it is impossible to write down the quantity/quality/type of services that the worker is expected to deliver to the firm in exchange for his/her wage in a way that would be enforceable in any court. For many (if not the vast majority of) jobs, it is simply impossible to avoid relying on the worker’s goodwill to satisfy the employer’s needs, and on the employer’s honesty in assessing whether those needs have been satisfactorily met. If such unenforceable exchanges of ‘gifts’ are important, the consequences for labor markets, can be very interesting. For example, they can lead to involuntary unemployment in equilibrium (Fehr, Kirchsteiger and Riedl 1993) and to long-term attachments between firms and workers that are immune to wage underbidding by new hires (Brown, Fehr and Falk 2004).

References:


¹ Efficiency-wage models, in general, refer to models in which the overall generosity of the compensation package (i.e. the intercept, not the marginal reward to effort) affects worker effort. There are two main variants, one based on fairness and reciprocity. The other variant, sometimes called the ‘shirking’ version, assumes workers are purely egostic and rational, and that firms pursue a policy of randomly monitoring their workers’ efforts and dismissing ‘shirkers’ (Shapiro and Stiglitz 1984). Here, workers supply high effort purely to keep jobs that pay above the market-clearing wage.