Economics 2010b: Microeconomic Theory II

Michael Powell

First-Order Approach and Limited Liability

1 The First-Order Approach

Last time, we imposed a lot of structure on the Principal-Agent problem and solved for optimal affine contracts. One of the problems we identified with that approach was that there was not a particularly compelling reason for restricting attention to affine contracts. Moreover, in that particular setting, if we allowed the contracts to take more general functional forms, there in fact was no optimal contract.

Today, we will return to a slightly modified version of the more general setup of the problem and consider an alternative approach to characterizing optimal contracts without imposing any assumptions on the functional forms they might take. One change we will be making is that the Agent's preferences are now given by

$$U(w,e) = \int_{y\in\mathcal{Y}} \left[u(w(y)) - c(e)\right] dF(y|\,e) = E_y\left[u(w)|\,e\right] - c(e),$$

where u is strictly increasing and strictly concave, and the utility the Agent receives from money is additively separable from his effort costs.

Recall from last time that the Principal's problem is to choose an output-contingent contract $w \in \mathcal{W} \subset \{w : \mathcal{Y} \to \mathbb{R}\}$ and to "propose" an effort level e to solve:

$$\max_{w\in\mathcal{W}, e\in\mathcal{E}} \int_{y\in\mathcal{Y}} (py - w(y))\, dF(y|\,e)$$

subject to an incentive-compatibility constraint

$$e \in \underset{\hat{e}\in\mathcal{E}}{\operatorname{argmax}} \int_{y\in\mathcal{Y}} u(w(y))\, dF(y|\,\hat{e}) - c(\hat{e})$$

and an individual-rationality constraint

$$\int_{y\in\mathcal{Y}} u\left(w\left(y\right)\right)dF\left(y\right|e\right)-c\left(e\right)\geq\bar{u}.$$

One of the problems with solving this problem at this level of generality is that the incentive-compatibility constraint is quite a complicated set of conditions. The contract has to ensure that, of all the effort levels the Agent could potentially choose, he prefers to choose e. In other words, the contract has to deter the Agent from choosing any other effort level $\hat{e}$: for all $\hat{e}\in\mathcal{E}$, we must have

$$\int_{y\in\mathcal{Y}}\left[u\left(w\left(y\right)\right)-c\left(e\right)\right]dF\left(y\right|e\right)\geq\int_{y\in\mathcal{Y}}\left[u\left(w\left(y\right)\right)-c\left(\hat{e}\right)\right]dF\left(y\right|\hat{e}\right).$$

When effort is continuous, the incentive-compatibility constraint is actually a continuum of constraints of this form. It seems like it should be the case that if we impose more structure on the problem, we can safely ignore most of these constraints. This turns out to be true. If we impose some relatively stringent but somewhat sensible assumptions on the problem, then if it is the case that the Agent does not want to deviate *locally* to another $\hat{e}$, then he also does not want to deviate to an $\hat{e}$ that is farther away. When local constraints are sufficient, we will in fact be able to replace the Agent's incentive-compatibility constraint with the first-order condition to his problem.

Throughout, we will be focusing on models that satisfy the following assumptions.

Assumption A1 (Continuous Effort and Continuous Output). Effort is continuous and satisfies $\mathcal{E}=\mathbb{R}_+$. Output is continuous, with $\mathcal{Y}=\mathbb{R}$, and for each $e\in\mathcal{E}$, $F\left(\cdot\right|e\right)$ has support $\left[\underline{y},\bar{y}\right]$ and has density $f\left(\cdot\right|e\right)$, where $f\left(\cdot\right|e\right)$ is differentiable in e.

Assumption A2 (First-Order Stochastic Dominance—FOSD). The output distribution function satisfies $F_e\left(y\right|e\right)\leq 0$ for all $e\in\mathcal{E}$ and all y with strict inequality for some y for each e.

Assumption $(A2)$ roughly says that higher effort levels make lower output realizations less likely and higher output realizations more likely. This assumption provides sufficient conditions under which higher effort increases total expected surplus, ignoring effort costs.

We will first explore the implications of being able to replace the incentive-compatibility constraint with the Agent's first-order condition, and then we will provide some sufficient conditions under which doing so is without loss of generality. Under Assumption $(A1)$, if we replace the Agent's incentive-compatibility constraint with his first-order condition, the Principal's problem becomes:

$$\max_{w\in\mathcal{W},e\in\mathcal{E}}\int_{\underline{y}}^{\bar{y}}(py-w(y))\,f(y|\,e)\,dy$$

subject to the local incentive-compatibility constraint

$$c'(e)=\int_{\underline{y}}^{\bar{y}}u(w(y))\,f_e(y|\,e)\,dy$$

and the individual-rationality constraint

$$\int_{\underline{y}}^{\bar{y}}u(w(y))\,f(y|\,e)\,dy-c(e)\geq\bar{u}.$$

This problem is referred to as the **first-order approach** to characterizing second-best incentive contracts. It is now just a constrained-optimization problem with an equality constraint and an inequality constraint. We can therefore write the Lagrangian for this problem as

$$\begin{aligned}\mathcal{L} &= \int_{\underline{y}}^{\bar{y}}(py-w(y))\,f(y|\,e)\,dy+\lambda\left(\int_{\underline{y}}^{\bar{y}}u(w(y))\,f(y|\,e)\,dy-c(e)-\bar{u}\right)\\ &\quad+\mu\left(\int_{\underline{y}}^{\bar{y}}u(w(y))\,f_e(y|\,e)\,dy-c'(e)\right),\end{aligned}$$

where λ is the Lagrange multiplier on the individual-rationality constraint, and μ is the Lagrange multiplier on the local incentive-compatibility constraint. We can derive the conditions for the optimal contract $w^*(y)$ inducing optimal effort e^* by taking first-order conditions, point-by-point, with respect to $w(y)$. These conditions are:

$$\frac{1}{u'(w^*(y))} = \lambda + \mu \frac{f_e(y|\, e^*)}{f(y|\, e^*)}.$$

Contracts satisfying these conditions are referred to as Holmström-Mirrlees contracts (or (λ, μ) contracts as one of my colleagues calls them). There are several points to notice here. First, the left-hand side is increasing in $w(y)$, since u is concave. Second, if $\mu = 0$, then this condition would correspond to the conditions for an optimal risk-sharing rule between the Principal and the Agent. Under a Pareto-optimal risk allocation, the **Borch Rule** states that the ratio of the Principal's marginal utility to the Agent's marginal utility is equalized across states. In this case, the Principal's marginal utility is one. Any optimal-risk sharing rule will equalize the Agent's marginal utility of income across states and therefore give the Agent a constant wage.

Third, Holmström (1979) shows that under Assumption $(A2)$, $\mu > 0$, so that the right-hand side of this equation is increasing in $f_e(y|\, e^*)/f(y|\, e^*)$. You might remember from econometrics that this ratio is called the **score**—it tells us how an increase in e changes the log likelihood of e given output realization y. To prevent the Agent from choosing effort level e instead of e^*, the contract has to pay the Agent more for outputs that are more likely under e^* than under e. Since by assumption, we are looking at only local incentive constraints, the contract will pay the Agent more for outputs that are more likely under e^* than under effort levels arbitrarily close to e^*.

Together, these observations imply that the optimal contract $w^*(y)$ is increasing in the score. Just because an optimal contract is increasing in the score does not mean that it is increasing in output. The following assumption guarantees that the score is increasing in y, and therefore optimal contracts are increasing in output.

Assumption A3 (Monotone Likelihood Ratio Property—MLRP). Given any two effort levels $e, e' \in \mathcal{E}$ with $e > e'$, the ratio $f(y|e)/f(y|e')$ is increasing in y.

MLRP guarantees, roughly speaking, that higher levels of output are more indicative of higher effort levels.[1] Under Assumption $(A1)$, MLRP is equivalent to the condition that $f_e(y|e)/f(y|e)$ is increasing in y. We can therefore interpret the optimality condition as telling us that the optimal contract is increasing in output precisely when higher output levels are more indicative of higher effort levels. Put differently, the optimal contract "wants" to reward *informative* output, not necessarily *high* output.

The two statistical properties, FOSD and MLRP, that we have assumed come up a lot in different settings, and it is easy to lose track of what they each imply. To recap, the FOSD property tells us that higher effort makes higher output more likely, and it guarantees that there is always a benefit of higher effort levels, gross of effort costs. The MLRP property tells us that higher output is more indicative of higher effort, and it guarantees that optimal contracts are increasing in output. These two properties are related: MLRP implies FOSD, but not the reverse.

1.1 Informativeness Principle

Before we provide conditions under which the first-order approach is valid, we will go over what I view as the most important result to come out of this model. Suppose there is another contractible performance measure $m \in \mathcal{M}$, where y and m have joint density function $f(y, m|e)$, and the contracting space is $\mathcal{W} = \{w : \mathcal{Y} \times \mathcal{M} \to \mathbb{R}\}$. Under what conditions will an optimal contract $w(y, m)$ depend nontrivially on m? The answer is: whenever m provides additional information about e. To make this argument precise, we will introduce the following definition.

Definition 1. Given two random variables Y and M, Y is **sufficient for** (Y, M) **with**

[1] The property can also be interpreted in terms of statistical hypothesis testing. Suppose the null hypothesis is that the Agent chose effort level e', and the alternative hypothesis is that the Agent chose effort level $e > e'$. If, given output realization y, a likelihood ratio test would reject the null hypothesis of lower effort, the same test would also reject the null hypothesis for any higher output realization.

respect to $e \in \mathcal{E}$ if and only if the joint density function $f(y, m|\, e)$ is multiplicatively separable in m and e:

$$f(y, m|\, e) = g(m|\, e)\, h(y, m)\,.$$

We will say that M is **informative about** $e \in \mathcal{E}$ if Y is not sufficient for (Y, M) with respect to $e \in \mathcal{E}$.

We argued above that optimal contracts pay the Agent more for outputs that are more indicative of high effort. This same argument also extends to other performance measures, as long as they are informative about effort. This result is known as the *informativeness principle* and was first established by Holmström (1979) and Shavell (1979).

Theorem 1 (Informativeness Principle). Assume the first-order approach is valid. Let $w(y)$ be the optimal contract when m is noncontractible. If m is contractible, there exist a contract $w(y, m)$ that Pareto dominates $w(y)$ if and only if m is informative about $e \in \mathcal{E}$.

Proof. In both cases, the optimal contract gives the Agent $\bar{u}$, so we just need to show that the Principal can be made strictly better off if m is contractible.

If the first-order approach is valid, the optimality conditions for the Principal's problem when both y and m are contractible are given by

$$\frac{1}{u'(w^*(y, m))} = \lambda + \mu \frac{f_e(y, m|\, e^*)}{f(y, m|\, e^*)}.$$

The optimal contract $w^*(y, m)$ is independent of m if and only if y is sufficient for (y, m) with respect to e^*.

This result seems like it should be obvious: optimal contracts clearly should make use of all available information. But it is not ex ante obvious this would be the case. In particular, one could easily have imagined that optimal contracts should only depend on performance measures that are "sufficiently" informative about effort—after all, basing a contract on another performance measure could introduce additional noise as well. Or one could have imagined that optimal contracts should only depend on performance measures that are directly affected by the Agent's effort choice. The informativeness principle says that optimal contracts should depend on every performance measure that is even slightly informative.

This result has both positive and negative implications. On the positive and practical side, it says that optimal contracts should make use of benchmarks: a fund manager should be evaluated for her performance relative to a market index, CEOs should be rewarded for firm performance relative to other firms in their industry, and employees should be evaluated relative to their peers. On the negative side, the result shows that optimal contracts are highly sensitive to the fine details of the environment. This implication is, in a real sense, a weakness of the theory: it is the reason why the theory often predicts contracts that bear little resemblance to what we actually see in practice.

The informativeness principle was derived under the assumption that the first-order approach was valid. When the first-order approach is not valid, the informativeness principle does not necessarily hold. The reason for this is that when the first-order approach does not hold, there may be multiple binding incentive-compatibility constraints at the optimum, and just because an informative performance measure helps relax one of those constraints, if it does not help relax the other binding constraints, it need not strictly increase the firm's profits. Chaigneau, Edmans, and Gottlieb (2014) generalizes the informativeness principle to settings in which the first-order approach is not valid.

1.2 Validity of the First-Order Approach

Finally, we will briefly talk about some sufficient conditions ensuring the first-order approach is valid. Assumption $(A4)$, along with the following assumption, are sufficient.

Assumption A4 (Convexity of the Distribution Function Condition—CDFC). $F\left(\cdot|\, e\right)$ is twice differentiable, and $F_{ee}\left(\cdot|\, e\right) \geq 0$ for all e.

CDFC is a strong assumption. There is a fairly standard class of distributions that are often used in contract theory that satisfy it, but it is not satisfied by other well-known families of distributions. Let $F_H(y)$ and $F_L(y)$ be two distribution functions that have density functions $f_H(y)$ and $f_L(y)$ for which $f_H(y)/f_L(y)$ is increasing in y, and suppose

$$F(y|\,e) = eF_H(y) + (1-e)\,F_L(y).$$

Then $F(y|\,e)$ satisfies both MLRP and CDFC. In other words, MLRP and CDFC are satisfied if output is drawn from a mixture of a "high" and a "low" distribution, and higher effort increases the probability that output is drawn from the high distribution.

Theorem 2. Suppose $(A1)-(A4)$ are satisfied. If the local incentive-compatibility constraint is satisfied, the incentive-compatibility constraint is satisfied.

Proof sketch. The high-level idea of the proof is to show that MLRP and CDFC imply that the Agent's effort-choice problem is globally concave for any contract the Principal offers him. Using integration by parts, we can rewrite the Agent's expected utility as follows.

$$\begin{aligned}\int_{\underline{y}}^{\bar{y}} u(w(y))\, f(y|\,e)\, dy - c(e) &= u(w(y))\, F(y|\,e)\big|_{\underline{y}}^{\bar{y}} \\ &\quad - \int_{\underline{y}}^{\bar{y}} u'(w(y))\, \frac{dw(y)}{dy} F(y|\,e)\, dy - c(e) \\ &= u(w(\bar{y})) - \int_{\underline{y}}^{\bar{y}} u'(w(y))\, \frac{dw(y)}{dy} F(y|\,e)\, dy - c(e).\end{aligned}$$

Now, suppose $w(y)$ is increasing and differentiable. Differentiating the expression above with respect to e twice yields

$$-\int_{\underline{y}}^{\bar{y}} u'(w(y))\, \frac{dw(y)}{dy} F_{ee}(y|\,e)\, dy - c''(e) < 0$$

for every $e \in \mathcal{E}$, since $F_{ee} > 0$. Thus, the Agent's second-order condition is globally satisfied, so if the local incentive constraint is satisfied, the incentive constraint is satisfied.■

I labeled this proof as a sketch, because while it follows Mirrlees's (1976) argument, the full proof (due to Rogerson (1985)) requires showing that $w(y)$ is in fact increasing and differentiable when MLRP is satisfied. We cannot use our argument above for why MLRP implies increasing contracts, because that argument presumed the first-order approach was valid, which is exactly what we are trying to prove here. The MLRP and CDFC conditions are known as the Mirrlees-Rogerson conditions.

There are other sufficient conditions for the first-order approach to be valid that do not require such strong distributional assumptions (see, for example, Jewitt (1988)). And there are other approaches to solving the moral hazard problem that do not rely on the first-order approach. These include Grossman and Hart (1983), which decomposes the Principal's problem into two steps: the first step solves for the cost-minimizing contract that implements a given effort level, and the second step solves for the optimal effort level. We will take this approach when we think about optimal contracts under limited liability in the next section.

2 Limited Liability and the Motivation–Rent Extraction Trade-Off

We saw in the previous model that the optimal contract sometimes involved up-front payments from the Agent to the Principal. To the extent that the Agent is unable to afford such payments (or legal restrictions such as minimum wage laws prohibit such payments), the Principal will not be able to extract all the surplus that the Agent creates. Further, in order to extract surplus from the Agent, the Principal may have to put in place contracts that reduce the total surplus created. In equilibrium, the Principal may therefore offer a contract that induces effort below the first-best.

Description Again, there is a risk-neutral Principal (P). There is also a **risk-neutral** Agent (A). The Agent chooses an effort level $e \in \mathcal{E} \subset \mathbb{R}_+$ at a cost of $c(e)$, where $c : \mathbb{R}_+ \to \mathbb{R}_+$, with $c'', c' > 0$, and this effort level affects the distribution over outputs $y \in \mathcal{Y}$, with y distributed according to CDF $F(\cdot| e)$. These outputs can be sold on the product market for price p. The Principal can write a contract $w \in \mathcal{W} \subset \{w : \mathcal{Y} \to \mathbb{R}, w(y) \geq \underline{w} \text{ for all } y\}$ that determines a transfer $w(y)$ that she is compelled to pay the Agent if output y is realized. The Agent has an outside option that provides utility $\bar{u}$ to the Agent and $\bar{\pi}$ to the Principal. If the outside option is not exercised, the Principal's and Agent's preferences are, respectively,

$$\begin{aligned} \Pi(w,e) &= \int_{y \in \mathcal{Y}} (py - w(y))\, dF(y|\, e) = E_y[py - w|\, e] \\ U(w,e) &= \int_{y \in \mathcal{Y}} (w(y) - c(e))\, dF(y|\, e) = E_y[w - c(e)|\, e]. \end{aligned}$$

There are two differences between this model and the previous model. The first difference is that the Agent is risk-neutral (so that absent any other changes, the equilibrium contract would induce first-best effort). The second difference is that the wage payment from the Principal to the Agent has to exceed, for each realization of output, a value $\underline{w}$. Depending on the setting, this constraint is described as a liquidity constraint or a limited-liability constraint. In repeated settings, it is more naturally thought of as the latter—due to legal restrictions, the Agent cannot be legally compelled to make a transfer (larger than $-\underline{w}$) to the Principal. In static settings, either interpretation may be sensible depending on the particular application—if the Agent is a fruit picker, for instance, he may not have much liquid wealth that he can use to pay the Principal.

Timing The timing of the game is exactly the same as before.

1. P offers A a contract $w(y)$, which is commonly observed.

2. A accepts the contract ($d = 1$) or rejects it ($d = 0$) and receives $\bar{u}$, and the game ends. This decision is commonly observed.

3. If A accepts the contract, A chooses effort level e and incurs cost $c(e)$. e is only observed by A.

4. Output y is drawn from distribution with cdf $F(\cdot|\, e)$. y is commonly observed.

5. P pays A an amount $w(y)$. This payment is commonly observed.

Equilibrium The solution concept is the same as before. A **pure-strategy subgame-perfect equilibrium** is a contract $w^* \in \mathcal{W}$, an acceptance decision $d^* : \mathcal{W} \to \{0,1\}$, and an effort choice $e^* : \mathcal{W} \times \{0,1\} \to \mathbb{R}_+$ such that given the contract w^*, the Agent optimally chooses d^* and e^*, and given d^* and e^*, the Principal optimally offers contract w^*. We will say that the optimal contract induces effort e^*.

The Program The Principal offers a contract $w \in \mathcal{W}$ and proposes an effort level e in order to solve

$$\max_{w \in \mathcal{W}, e \in \mathcal{E}} \int_{y \in \mathcal{Y}} (py - w(y))\, dF(y|\, e)$$

subject to three constraints: the incentive-compatibility constraint

$$e \in \underset{\hat{e} \in \mathcal{E}}{\operatorname{argmax}} \int_{y \in \mathcal{Y}} (w(y) - c(\hat{e}))\, dF(y|\, \hat{e}),$$

the individual-rationality constraint

$$\int_{y \in \mathcal{Y}} (w(y) - c(e))\, dF(y|\, e) \geq \bar{u},$$

and the limited-liability constraint

$$w(y) \geq \underline{w} \text{ for all } y \in \mathcal{Y}.$$

Binary-Output Case We will impose much more structure on the problem to illustrate the main trade-off in this class of models. Innes (1990) and Jewitt, Kadan, and Swinkels (2008) explore a much more general analysis.

Assumption A1 (Binary Output). Output is $y \in \{0,1\}$, and given effort e, its distribution satisfies $\Pr[y = 1|\, e] = e$.

Assumption A2 (Well-behaved Cost). The Agent's costs have a non-negative third derivative: $c''' \geq 0$, and they satisfy conditions that ensure an interior solution: $c'(0) = 0$ and $c'(1) = +\infty$. Or for comparison across models in this module, $c(e) = \frac{c}{2}e^2$, where $p \leq c$ to ensure that $e^{FB} < 1$.

Finally, we can restrict attention to affine, nondecreasing contracts

$$\begin{aligned} \mathcal{W} &= \{w(y) = (1-y)w_0 + yw_1, w_1 \geq w_0 \geq 0\} \\ &= \{w(y) = s + by, s \geq \underline{w}, b \geq 0\}. \end{aligned}$$

When output is binary, this restriction to affine contracts is without loss of generality. Also, the restriction to nondecreasing contracts is not restrictive (i.e., any optimal contract of a relaxed problem in which we do not impose that contracts are nondecreasing will also be the solution to the full problem). This result is something that needs to be shown and is not in general true, but in this case, it is straightforward.

As Grossman and Hart (1983) highlight, in Principal–Agent models, it is often useful to break the problem down into two steps. The first step takes a target effort level, e, as given and solves for the set of cost-minimizing contracts implementing effort level e. Any cost-minimizing contract implementing effort level e results in an expected cost of $C(e)$ to the principal. The second step takes the function $C(\cdot)$ as given and solves for the optimal effort choice.

In general, the cost-minimization problem tends to be a well-behaved convex-optimization problem, since (even if the agent is risk-averse) the objective function is weakly concave, and the constraint set is a convex set (since given an effort level e, the individual-rationality constraint and the limited-liability constraint define convex sets, and each incentive constraint ruling out effort level $\hat{e} \neq e$ also defines a convex set, and the intersection of convex sets is itself a convex set). The resulting cost function $C(\cdot)$ need not have nice properties, however, so the second step of the optimization problem is only well-behaved under restrictive assumptions. In the present case, Assumptions $(A1)$ and $(A2)$ ensure that the second step of the optimization problem is well-behaved.

Cost-Minimization Problem Given an effort level e, the cost-minimization problem is given by

$$C\left(e, \bar{u}, \underline{w}\right) = \min_{s,b} s + be$$

subject to the Agent's incentive-compatibility constraint

$$e \in \underset{\hat{e}}{\operatorname{argmax}} \left\{s + b\hat{e} - c\left(\hat{e}\right)\right\},$$

his individual-rationality constraint

$$s + be - c\left(e\right) \geq \bar{u},$$

and the limited-liability constraint

$$s \geq \underline{w}.$$

I will denote a **cost-minimizing contract implementing effort level** e by (s_e^*, b_e^*).

The first step in solving this problem is to notice that the Agent's incentive-compatibility constraint implies that any cost-minimizing contract implementing effort level e must have $b_e^* = c'\left(e\right)$.

If there were no limited-liability constraint, the Principal would choose s_e^* to extract the Agent's surplus. That is, given $b = b_e^*$, s would solve

$$s + b_e^* e = \bar{u} + c\left(e\right).$$

That is, s would ensure that the Agent's expected compensation exactly equals his expected effort costs plus his opportunity cost. The resulting s, however, may not satisfy the limited-liability constraint. The question then is: given $\bar{u}$ and $\underline{w}$, for what effort levels e is the

Principal able to extract all the agent's surplus (i.e., for what effort levels does the limited-liability constraint not bind at the cost-minimizing contract?), and for what effort levels is she unable to do so? Figure 1 below shows cost-minimizing contracts for effort levels e_1 and e_2. Any contract can be represented as a line in this figure, where the line represents the expected pay the Agent will receive given an effort level e. The cost-minimizing contract for effort level e_1 is tangent to the $\bar{u} + c\left(e\right)$ curve at e_1 and its intercept is $s^*_{e_1}$. Similarly for e_2. Both $s^*_{e_1}$ and $s^*_{e_2}$ are greater than $\underline{w}$, which implies that for such effort levels, the limited-liability constraint is not binding.

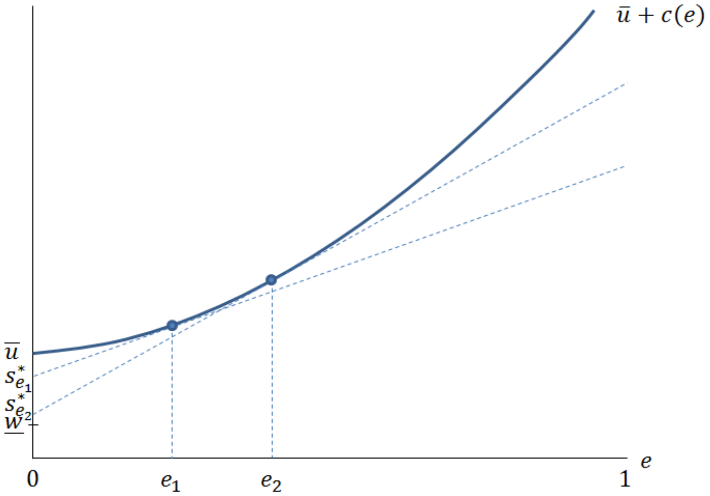

Figure 1: Cost-minimizing contracts

For effort sufficiently high, the limited-liability constraint will be binding in a cost-minimizing contract, and it will be binding for all higher effort levels. Define the threshold $\bar{e}\left(\bar{u}, \underline{w}\right)$ to be the effort level such that for all $e \geq \bar{e}\left(\bar{u}, \underline{w}\right)$, $s^*_e = \underline{w}$. Figure 2 illustrates that $\bar{e}\left(\bar{u}, \underline{w}\right)$ is the effort level at which the contract tangent to the $\bar{u} + c\left(e\right)$ curve at $\bar{e}\left(\bar{u}, \underline{w}\right)$ intersects the vertical axis at exactly $\underline{w}$. That is, $\bar{e}\left(\bar{u}, \underline{w}\right)$ solves

$$c'\left(\bar{e}\left(\bar{u}, \underline{w}\right)\right) = \frac{\bar{u} + c\left(\bar{e}\left(\bar{u}, \underline{w}\right)\right) - \underline{w}}{\bar{e}\left(\bar{u}, \underline{w}\right)}.$$

Figure 2 also illustrates that for all effort levels $e > \bar{e}(\bar{u}, \underline{w})$, the cost-minimizing contract involves giving the Agent strictly positive surplus. That is, the cost to the Principal of getting the agent to choose effort $e > \bar{e}(\bar{u}, \underline{w})$ is equal to the Agent's opportunity costs $\bar{u}$ plus his effort costs $c(e)$ plus **incentive costs** $IC(e, \bar{u}, \underline{w})$.

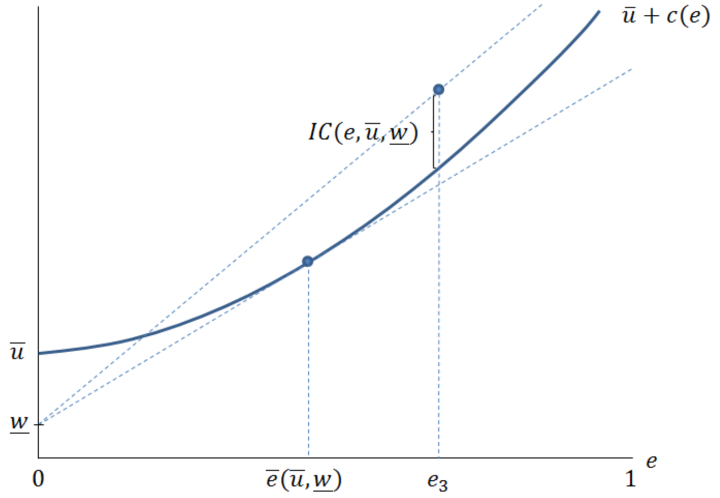

Figure 2: Incentive Costs for High Effort Levels

The incentive costs $IC(e, \bar{u}, \underline{w})$ are equal to the Agent's expected compensation given effort choice e and cost-minimizing contract (s_e^*, b_e^*) minus his costs:

$$\begin{aligned} IC(e, \bar{u}, \underline{w}) &= \begin{cases} 0 & e \le \bar{e}(\bar{u}, \underline{w}) \\ \underline{w} + c'(e)\, e - c(e) - \bar{u} & e \ge \bar{e}(\bar{u}, \underline{w}) \end{cases} \\ &= \max\{0, \underline{w} + c'(e)\, e - c(e) - \bar{u}\} \end{aligned}$$

where I used the fact that for $e \ge \bar{e}(\bar{u}, \underline{w})$, $s_e^* = \underline{w}$ and $b_e^* = c'(e)$. This incentive-cost function $IC(\cdot, \bar{u}, \underline{w})$ is the key object that captures the main contracting friction in this model. I will sometimes refer to $IC(e, \bar{u}, \underline{w})$ as the **incentive rents** required to get the

Agent to choose effort level e. Putting these results together, we see that

$$C(e, \bar{u}, \underline{w}) = \bar{u} + c(e) + IC(e, \bar{u}, \underline{w}).$$

That is, the Principal's total costs of implementing effort level e are the sum of the Agent's costs plus the incentive rents required to get the Agent to choose effort level e.

Since $IC(e, \bar{u}, \underline{w})$ is the main object of interest in this model, I will describe some of its properties. First, it is continuous in e (including, in particular, at $e = \bar{e}(\bar{u}, \underline{w})$). Next, $\bar{e}(\bar{u}, \underline{w})$ and $IC(e, \bar{u}, \underline{w})$ depend on $(\bar{u}, \underline{w})$ only inasmuch as $(\bar{u}, \underline{w})$ determines $\bar{u} - \underline{w}$, so I will abuse notation and write these expressions as $\bar{e}(\bar{u} - \underline{w})$ and $IC(e, \bar{u} - \underline{w})$. Also, given that $c'' > 0$, IC is increasing in e (since $\underline{w} + c'(e)e - c(e) - \underline{u}$ is strictly increasing in e, and IC is just the max of this expression and zero). Further, given that $c''' \geq 0$, IC is convex in e. For $e \geq \bar{e}(\bar{u} - \underline{w})$, this property follows, because

$$\frac{\partial^2}{\partial e^2} IC = c''(e) + c'''(e)e \geq 0.$$

And again, since IC is the max of two convex functions, it is also a convex function. Finally, since $IC(\cdot, \bar{u} - \underline{w})$ is flat when $e \leq \bar{e}(\bar{u} - \underline{w})$ and it is strictly increasing (with slope independent of $\bar{u} - \underline{w}$) when $e \geq \bar{e}(\bar{u} - \underline{w})$, the slope of IC with respect to e is (weakly) decreasing in $\bar{u} - \underline{w}$, since $\bar{e}(\bar{u} - \underline{w})$ is increasing in $\bar{u} - \underline{w}$. That is, $IC(e, \bar{u} - \underline{w})$ satisfies decreasing differences in $(e, \bar{u} - \underline{w})$.

Motivation-Rent Extraction Trade-off The second step of the optimization problem takes as given the function

$$C(e, \bar{u} - \underline{w}) = \bar{u} + c(e) + IC(e, \bar{u} - \underline{w})$$

and solves the Principal's problem for the optimal effort level:

$$\begin{aligned} & \max_e pe - C\left(e, \bar{u} - \underline{w}\right) \\ = \; & \max_e pe - \bar{u} - c\left(e\right) - IC\left(e, \bar{u} - \underline{w}\right). \end{aligned}$$

Note that total surplus is given by $pe - \bar{u} - c\left(e\right)$, which is therefore maximized at $e = e^{FB}$ (which, if $c\left(e\right) = ce^2/2$, then $e^{FB} = p/c$). Figure 3 below depicts the Principal's expected benefit line pe, and her expected costs of implementing effort e at minimum cost, $C\left(e, \bar{u} - \underline{w}\right)$. The first-best effort level, e^{FB} maximizes the difference between pe and $\bar{u} + c\left(e\right)$, while the equilibrium effort level e^* maximizes the difference between pe and $C\left(e, \bar{u} - \underline{w}\right)$.

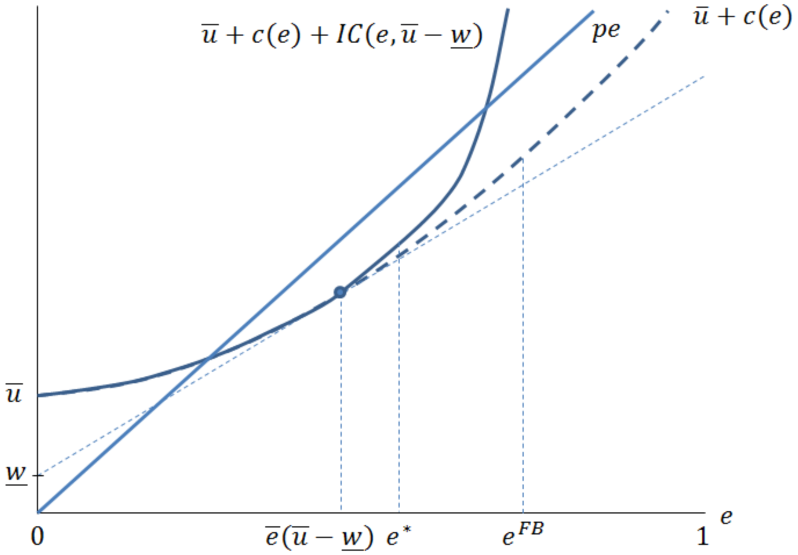

Figure 3: Optimal Effort Choice

If $c\left(e\right) = ce^2/2$, we can solve explicitly for $\bar{e}\left(\bar{u} - \underline{w}\right)$ and for $IC\left(e, \bar{u} - \underline{w}\right)$ when $e > \bar{e}\left(\bar{u} - \underline{w}\right)$. In particular,

$$\bar{e}\left(\bar{u} - \underline{w}\right) = \left(\frac{2\left(\bar{u} - \underline{w}\right)}{c}\right)^{1/2}$$

and when $e > \bar{e}(\bar{u} - \underline{w})$,

$$IC(e, \bar{u} - \underline{w}) = \underline{w} + \frac{1}{2}ce^2 - \bar{u}.$$

If $\underline{w} < 0$ and p is sufficiently small, we can have $e^* = e^{FB}$ (i.e., these are the conditions required to ensure that the limited-liability constraint is not binding for the cost-minimizing contract implementing $e = e^{FB}$). If p is sufficiently large relative to $\bar{u} - \underline{w}$, we will have $e^* = \frac{1}{2}\frac{p}{c} = \frac{1}{2}e^{FB}$. For p somewhere in between, we will have $e^* = \bar{e}(\bar{u} - \underline{w}) < e^{FB}$. In particular, $C(e, \bar{u} - \underline{w})$ is kinked at this point.

As in the risk–incentives model, we can illustrate through a partial characterization why (and when) effort is less-than first-best. Since we know that e^{FB} maximizes $pe - \bar{u} - c(e)$, we therefore have that

$$\frac{d}{de}\left[pe - \bar{u} - c(e) - IC(e, \bar{u} - \underline{w})\right]_{e=e^{FB}} = -\frac{\partial}{\partial e}IC\left(e^{FB}, \bar{u} - \underline{w}\right) \leq 0,$$

with strict inequality if the limited-liability constraint binds at the cost-minimizing contract implementing e^{FB}. This means that, even though e^{FB} maximizes total surplus, if the Principal has to provide the agent with rents at the margin, she may choose to implement a lower effort level. Reducing the effort level away from e^{FB} leads to second-order losses in terms of total surplus, but it leads to first-order gains in profits for the Principal. In this model, there is a tension between total-surplus creation and rent extraction, which yields less-than-first-best effort in equilibrium.

In my view, liquidity constraints are extremely important and are probably one of the main reasons for why many jobs do not involve first-best incentives. The logic that first-best efforts can be implemented if the firm transfers the entire profit stream to each of its members in exchange for a large up-front payment seems simultaneously compelling, trivial, and obviously impracticable. In for-profit firms, in order to make it worthwhile to transfer a large enough share of the profit stream to an individual worker to significantly affect his incentives, the firm would require a large up-front transfer that most workers cannot afford to pay. It is therefore not surprising that we do not see most workers' compensation tied directly to the firm's overall profits in a meaningful way. One implication of this logic is that firms have to find alternative instruments to use as performance measures, which we will turn to next. In principle, models in which firms do not motivate their workers by writing contracts directly on profits should include assumptions under which the firm optimally chooses not to write contracts directly on profits, but they almost never do.

Exercise 21. This exercise goes through a version of Diamond's (1998) and Barron, Georgiadis, and Swinkels's (2018) argument for why linear contracts are optimal when the Agent is able to "take on risk." Suppose the Principal and the Agent are both risk neutral, and let $\mathcal{Y} = [0, \bar{y}]$ and $\mathcal{E} = \mathbb{R}_+$. There is a limited-liability constraint, and the contracting space is $\mathcal{W} = \{w : \mathcal{Y} \to \mathbb{R}_+\}$. After the Agent chooses an effort level e, he can then choose any distribution function $F(y)$ over output that satisfies $e = \int_0^{\bar{y}} y dF(y)$. In other words, his effort level determines his *average* output, but he can then add mean-preserving noise to his output. Given a contract w, effort e, and distribution F, the Agent's expected utility is

$$\int_0^{\bar{y}} w(y)\, dF(y) - c(e),$$

where c is strictly increasing and strictly convex. The Principal's expected profits are $\int_0^{\bar{y}} (y - w(y))\, dF(y)$. The Agent's outside option gives both parties a payoff of zero.

(a) Show that a linear contract of the form $w(y) = by$ maximizes the Principal's expected profits. To do so, you will want to argue that given any contract $w(y)$ that implements effort level e, there is a linear contract that also implements effort level e but at a weakly lower cost to the Principal. [Hint: instead of thinking about all the possible distribution functions the Agent can choose among, it may be useful to just look at distributions that put weight on two levels of output, $0 \leq y_L < y_H \leq \bar{y}$ satisfying $e = (1-q) y_L + q y_H$.]

(b) Are there other contracts that maximize the Principal's expected profits? If so, how are they related to the optimal linear contract? If not, provide an intuition for why linear contracts are uniquely optimal.

3 Misaligned Performance Measures

In the previous two models, the Principal cared about output, and output, though a noisy measure of effort, was perfectly measurable. This assumption seems sensible when we think about overall firm profits (ignoring basically everything that accountants think about every day), but as we alluded to in the previous discussion, overall firm profits are too blunt of an instrument to use to motivate individual workers within the firm if they are liquidity-constrained. As a result, firms often try to motivate workers using more specific performance measures, but while these performance measures are informative about what actions workers are taking, they may be less useful as a description of how the workers' actions affect the objectives the firm cares about. And paying workers for what is measured may not get them to take actions that the firm cares about. This observation underpins the title of the famous 1975 paper by Steve Kerr called "On the Folly of Rewarding A, While Hoping for B."

As an example, think of a retail firm that hires an employee both to make sales and to provide customer service. It can be difficult to measure the quality of customer service that a particular employee provides, but it is easy to measure that employee's sales. Writing a contract that provides the employee with high-powered incentives directly on sales will get him to put a lot of effort into sales and very little effort into customer service. And in fact, he might only be able to put a lot of effort into sales by intentionally neglecting customer service. If the firm cares equally about both dimensions, it might be optimal not to offer high-powered incentives to begin with. This is what Holmström and Milgrom (1991) refers to as the "multitask problem." We will look at a model that captures some of this intuition, although not as directly as Holmström and Milgrom's model.

Description Again, there is a risk-neutral Principal (P) and a risk-neutral Agent (A). The Agent chooses an effort vector $e = (e_1, e_2) \in \mathcal{E} \subset \mathbb{R}^2_+$ at a cost of $\frac{c}{2}\left(e_1^2 + e_2^2\right)$. This effort vector affects the distribution of output $y \in \mathcal{Y} = \{0, 1\}$ and a performance measure $m \in \mathcal{M} = \{0, 1\}$ as follows:

$$\begin{aligned} \Pr\left[y = 1 \middle| e \right] &= f_1 e_1 + f_2 e_2 \\ \Pr\left[m = 1 \middle| e \right] &= g_1 e_1 + g_2 e_2, \end{aligned}$$

where it may be the case that $f = (f_1, f_2) \neq (g_1, g_2) = g$. Assume that $f_1^2 + f_2^2 = g_1^2 + g_2^2 = 1$ (i.e., the norms of the f and g vectors are unity). The output can be sold on the product market for price p. Output is noncontractible, but the performance measure is contractible. The Principal can write a contract $w \in \mathcal{W} \subset \{w : \mathcal{M} \to \mathbb{R}\}$ that determines a transfer $w(m)$ that she is compelled to pay the Agent if performance measure m is realized. Since the performance measure is binary, contracts take the form $w = s + bm$. The Agent has an outside option that provides utility $\bar{u}$ to the Agent and $\bar{\pi}$ to the Principal. If the outside option is not exercised, the Principal's and Agent's preferences are, respectively,

$$\begin{aligned} \Pi(w, e) &= f_1 e_1 + f_2 e_2 - s - b(g_1 e_1 + g_2 e_2) \\ U(w, e) &= s + b(g_1 e_1 + g_2 e_2) - \frac{c}{2}\left(e_1^2 + e_2^2\right). \end{aligned}$$

Timing The timing of the game is exactly the same as before.

1. P offers A a contract w, which is commonly observed.

2. A accepts the contract $(d = 1)$ or rejects it $(d = 0)$ and receives $\bar{u}$ and the game ends. This decision is commonly observed.

3. If A accepts the contract, A chooses effort vector e. e is only observed by A.

4. Performance measure m and output y are drawn from the distributions described above. m is commonly observed.

5. P pays A an amount $w(m)$. This payment is commonly observed.

Equilibrium The solution concept is the same as before. A **pure-strategy subgame-perfect equilibrium** is a contract $w^* \in \mathcal{W}$, an acceptance decision $d^* : \mathcal{W} \to \{0, 1\}$, and an effort choice $e^* : \mathcal{W} \times \{0, 1\} \to \mathbb{R}_+^2$ such that given the contract w^*, the Agent optimally chooses d^* and e^*, and given d^* and e^*, the Principal optimally offers contract w^*. We will say that the optimal contract induces effort e^*.

The Program The principal offers a contract w and proposes an effort level e to solve

$$\max_{s,b,e} p\left(f_1 e_1 + f_2 e_2\right) - \left(s + b\left(g_1 e_1 + g_2 e_2\right)\right)$$

subject to the incentive-compatibility constraint

$$e \in \operatorname*{argmax}_{\hat{e} \in \mathbb{R}_+^2} s + b\left(g_1 \hat{e}_1 + g_2 \hat{e}_2\right) - \frac{c}{2}\left(\hat{e}_1^2 + \hat{e}_2^2\right)$$

and the individual-rationality constraint

$$s + b\left(g_1 e_1 + g_2 e_2\right) - \frac{c}{2}\left(e_1^2 + e_2^2\right) \geq \bar{u}.$$

Equilibrium Contracts and Effort Given a contract $s + bm$, the Agent will choose

$$e_1^*\left(b\right) = \frac{b}{c} g_1; \; e_2^*\left(b\right) = \frac{b}{c} g_2.$$

The Principal will choose s so that the individual-rationality constraint holds with equality

$$s + b\left(g_1 e_1^*\left(b\right) + g_2 e_2^*\left(b\right)\right) = \bar{u} + \frac{c}{2}\left(e_1^*\left(b\right)^2 + e_2^*\left(b\right)^2\right).$$

Since contracts send the Agent off in the "wrong direction" relative to what maximizes total surplus, providing the Agent with higher-powered incentives by increasing b sends the agent farther off in the wrong direction. This is costly for the Principal because in order to get the Agent to accept the contract, she has to compensate him for his effort costs, even if they are in the wrong direction.

The Principal's unconstrained problem is therefore

$$\max_{b} p\left(f_1 e_1^*\left(b\right) + f_2 e_2^*\left(b\right)\right) - \frac{c}{2}\left(e_1^*\left(b\right)^2 + e_2^*\left(b\right)^2\right) - \bar{u}.$$

Taking first-order conditions,

$$pf_1 \underbrace{\frac{\partial e_1^*}{\partial b}}_{g_1/c} + pf_2 \underbrace{\frac{\partial e_2^*}{\partial b}}_{g_2/c} = \underbrace{ce_1^*\left(b^*\right)}_{b^* g_1/c} \underbrace{\frac{\partial e_1^*}{\partial b}}_{g_1/c} + \underbrace{ce_2^*\left(b^*\right)}_{b^* g_2/c} \underbrace{\frac{\partial e_2^*}{\partial b}}_{g_2/c},$$

or

$$b^* = p\frac{f_1 g_1 + f_2 g_2}{g_1^2 + g_2^2} = p\frac{f \cdot g}{g \cdot g} = p\frac{||f||}{||g||}\cos\theta = p\cos\theta,$$

where $\cos\theta$ is the angle between the vectors f and g. That is, the optimal incentive slope depends on the relative magnitudes of the f and g vectors (which in this model were assumed to be the same, but in a richer model this need not be the case) as well as how well-aligned they are. If m is a perfect measure of what the firm cares about, then g is a linear transformation of f and therefore the angle between f and g would be zero, so that $\cos\theta = 1$. If m is completely uninformative about what the firm cares about, then f and g are orthogonal, and therefore $\cos\theta = 0$. As a result, this model is often referred to as the **"cosine of theta model."**

It can be useful to view this problem geometrically. Since formal contracts allow for unrestricted lump-sum transfers between the Principal and the Agent, the Principal would optimally like efforts to be chosen in such a way that they maximize total surplus:

$$\max_e p\left(f_1 e_1 + f_2 e_2\right) - \frac{c}{2}\left(e_1^2 + e_2^2\right),$$

which has the same solution as

$$\max_e -\left(e_1 - \frac{p}{c}f_1\right)^2 - \left(e_2 - \frac{p}{c}f_2\right)^2.$$

That is, the Principal would like to choose an effort vector that is collinear with the vector f:

$$\left(e_1^{FB}, e_2^{FB}\right) = \frac{p}{c}\cdot\left(f_1, f_2\right).$$

This effort vector would coincide with the first-best effort vector, since it maximizes total surplus, and the players have quasilinear preferences.

Since contracts can only depend on m and not directly on y, the Principal has only limited control over the actions that the Agent chooses. That is, given a contract specifying incentive slope b, the Agent chooses $e_1^*(b) = \frac{b}{c} g_1$ and $e_2^*(b) = \frac{b}{c} g_2$. Therefore, the Principal can only indirectly "choose" an effort vector that is collinear with the vector g:

$$\left(e_1^*(b), e_2^*(b)\right) = \frac{b}{c} \cdot (g_1, g_2).$$

The question is then: which such vector maximizes total surplus, which the Principal will extract with an ex ante lump-sum transfer? That is, which point along the $k \cdot (g_1, g_2)$ ray minimizes the mean-squared error distance to $\frac{p}{c} \cdot (f_1, f_2)$?

The following figure illustrates the first-best effort vector e^{FB} and the equilibrium effort vector e^*. The concentric rings around e^{FB} are the Principal's iso-profit curves. The rings that are closer to e^{FB} represent higher profit levels. The optimal contract induces effort vector e^*, which also coincides with the orthogonal projection of e^{FB} onto the ray $k \cdot (g_1, g_2)$.

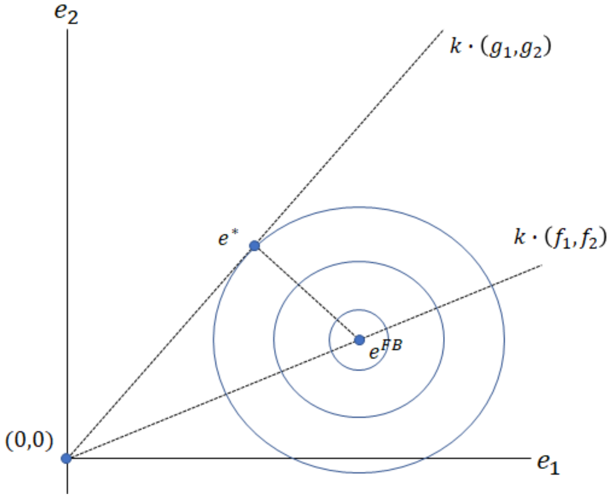

Figure 4: Optimal Effort Vector

This is a more explicit "incomplete contracts" model of motivation. That is, we are explicitly restricting the set of contracts that the Principal can offer the Agent in a way that directly determines a subset of the effort space that the Principal can induce the Agent to choose among. And it is founded not on the idea that certain measures (in particular, y) are unobservable, but rather that they simply cannot be contracted upon.

One observation that is immediate is that it may sometimes be optimal to offer incentive contracts that provide no incentives for the Agent to choose positive effort levels (i.e., $b^* = 0$). This was essentially never the case in the model in which the Agent chose only a one-dimensional effort level, yet we often see that many employees are on contracts that look like they offer no performance-based payments. As this model highlights, this may be optimal precisely when the set of available performance measures are quite bad. As an example, suppose

$$\Pr\left[y = 1 \middle| e\right] = \alpha + f_1 e_1 + f_2 e_2,$$

where $\alpha > 0$ and $f_2 < 0$, so that higher choices of e_2 reduce the probability of high output. And suppose the performance measure is again satisfies

$$\Pr\left[m = 1 \middle| e\right] = g_1 e_1 + g_2 e_2,$$

with $g_1, g_2 > 0$.

We can think of $y = 1$ as representing whether a particular customer buys something that he does not later return, which depends on how well he was treated when he went to the store. We can think of $m = 1$ as representing whether the Agent made a sale but not whether the item was later returned. In order to increase the probability of making a sale, the Agent can exert "earnest" sales effort e_1 and "shady" sales effort e_2. Both are good for sales, but the latter increases the probability the item is returned. If the vectors f and g are sufficiently poorly aligned (i.e., if it is really easy to make sales by being shady), it may be better for the firm to offer a contract with $b^* = 0$, as the following figure illustrates.

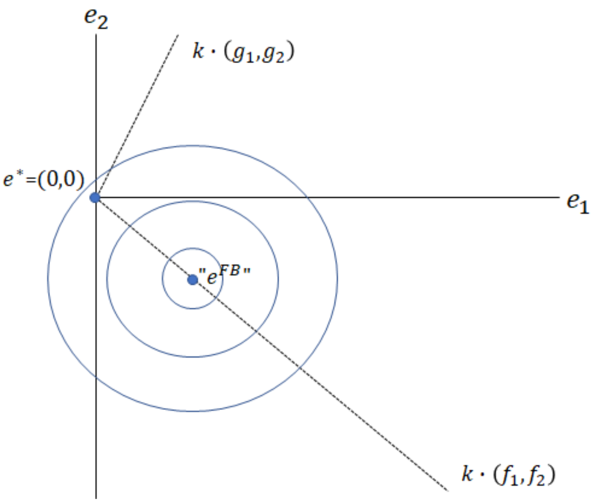

Figure 5: Sometimes Zero Effort is Optimal

This example illustrates that paying the Agent for sales can be a bad idea when what the Principal wants is *sales that are not returned.* The Kerr (1975) article is filled with many colorful examples of this problem. One such example concerns the incentives offered to the managers of orphanages. Their budgets and prestige were determined largely by the number of children they enrolled and not by whether they managed to place their children with suitable families. The claim made in the article is that the managers often denied adoption applications for inappropriate reasons: they were being rewarded for large orphanages, while the state hoped for good placements.

3.1 Limits on Activities

Firms have many instruments to help address the problems that arise in multitasking situations. We will describe two of them here in a small extension to the model. Suppose now that the Principal can put some restrictions on the types of actions the Agent is able to undertake. In particular, in addition to writing a contract on the performance measure m, she can write a contract on the dummy variables $1_{e_1>0}$ and $1_{e_2>0}$. In other words, while she cannot directly contract upon, say, e_2, she can write a contract that heavily penalizes any positive level of it. The first question we will ask here is: when does the Principal want to exclude the Agent from engaging in task 2?

We can answer this question using the graphical intuition we just developed above. The following figure illustrates this intuition. If the Principal does not exclude task 2, then she can induce the Agent to choose any effort vector of the form $k \cdot (g_1, g_2)$. If she does exclude task 2, then she can induce the Agent to choose any effort vector of the form $k \cdot (g_1, 0)$. In the former case, the equilibrium effort vector will be e^*, which corresponds to the orthogonal projection of e^{FB} onto the ray $k \cdot (g_1, g_2)$. In the latter case, the equilibrium effort will be e^{**}, which corresponds to the orthogonal projection of e^{FB} onto the ray $k \cdot (g_1, 0)$.

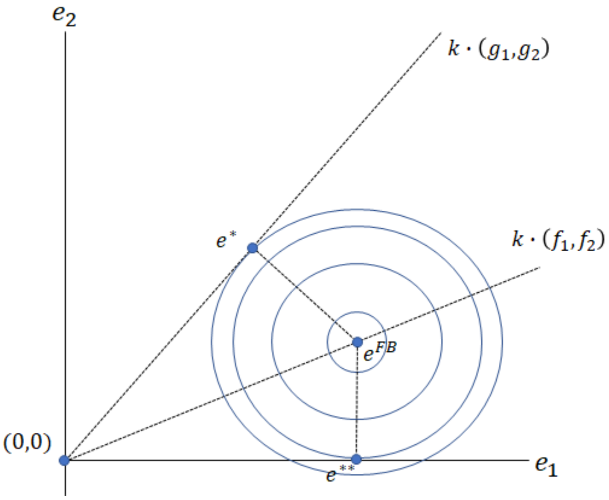

Figure 6: Excluding Task 2

This figure shows that for the particular vectors f and g it illustrates, it will be optimal for the Principal to exclude e_2: e^{**} lies on a higher iso-profit curve than e^* does. This will in fact be the case whenever the angle between vector f and g is larger than the angle between f and $(g_1, 0)$—if by excluding task 2, the performance measure m acts as if it is more closely aligned with f, then task 2 should be excluded.

3.2 Job Design

Finally, we will briefly touch upon what is referred to as job design. Suppose f and g are such that it is not optimal to exclude either task on its own. The firm may nevertheless want to hire *two* Agents who each specialize in a single task. For the first Agent, the Principal could exclude task 2, and for the second Agent, the Principal could exclude task 1. The Principal could then offer a contract that gets the first Agent to choose $\left(e_1^{FB}, 0\right)$ and the second agent to choose $\left(0, e_2^{FB}\right)$. The following figure illustrates this possibility.

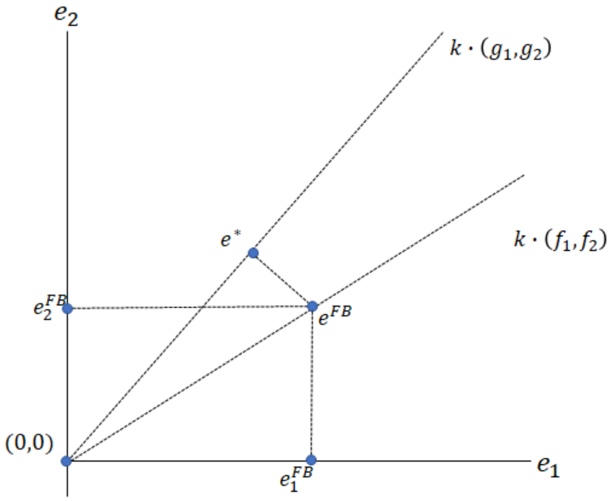

Figure 7: Job Design

When is it optimal for the firm to hire two Agents who each specialize in a single task? It depends on the Agents' opportunity cost. Total surplus under a single Agent under the optimal contract will be

$$pf \cdot e^* - \frac{c}{2} e^* \cdot e^* - \bar{u},$$

and total surplus with two specialized agents under optimal contracts will be

$$pf \cdot e^{FB} - \frac{c}{2} e^{FB} \cdot e^{FB} - 2\bar{u}.$$

Adding an additional Agent in this case is tantamount to adding an additional performance measure, which allows the Principal to choose induce any $e \in \mathbb{R}_+^2$, including the first-best effort vector. She gains from being able to do this, but to do so, she has to cover the additional Agent's opportunity cost $\bar{u}$.

4 Indistinguishable Individual Contributions

So far, we have discussed three contracting frictions that give rise to equilibrium contracts that induce effort that is not first-best. We will now discuss a final contracting friction that arises when multiple individuals contribute to a single project, and while team output is contractible, individual contributions to the team output are not. This indistinguishability gives rise to Holmström's (1982) classic "moral hazard in teams" problem.

The Model There are $I \geq 2$ risk-neutral Agents $i \in \mathcal{I} = \{1, \ldots, I\}$ who each choose efforts $e_i \in \mathcal{E}_i = \mathbb{R}_+$ at cost $c_i(e_i)$, which is increasing, convex, differentiable, and satisfies $c_i'(0) = 0$. The vector of efforts $e = (e_1, \ldots, e_I)$ determine team output $y \in \mathcal{Y} = \mathbb{R}_+$ according to a function $y(e)$ which is increasing in each e_i, concave in e, differentiable, and satisfies $\lim_{e_i \to 0} \partial y / \partial e_i = \infty$. Note that output is not stochastic, although the model can be easily extended to allow for stochastic output. Output is contractible, and each Agent i is subject to a contract $w_i \in \mathcal{W} = \{w_i : \mathcal{Y} \to \mathbb{R}\}$. We will say that the vector of contracts $w = (w_1, \ldots, w_I)$ is a **sharing rule** if

$$\sum_{i \in \mathcal{I}} w_i(y) = y$$

for each output level y. Each Agent i's preferences are given by

$$U_i(w,e) = w_i(y(e)) - c_i(e_i).$$

Each Agent i takes the contracts as given and chooses an effort level. Output is realized and each agent receives payment $w_i(y)$. The solution concept is Nash equilibrium, and we will say that w **induces** e^* if e^* is a Nash equilibrium effort profile given the vector of contracts w.

Sharing Rules and the Impossibility of First-Best Effort Since the Agents have quasilinear preferences, any Pareto-optimal outcome under a sharing rule w will involve an effort level that maximizes total surplus, so that

$$e^{FB} \in \underset{e \in \mathbb{R}_+^I}{\operatorname{argmax}}\, y(e) - \sum_{i \in \mathcal{I}} c_i(e_i).$$

Under our assumptions, there is a unique first-best effort vector, and it satisfies

$$\frac{\partial y\left(e^{FB}\right)}{\partial e_i} = c_i'\left(e_i^{FB}\right) \text{ for all } i \in \mathcal{I}.$$

First-best effort equates the social marginal benefit of each agent's effort level with its social marginal cost. We will denote the **first-best output level** $y\left(e^{FB}\right)$ by y^{FB}.

We will give an informal argument for why no sharing rule w induces e^{FB}, and then we will make that argument more precise. Suppose w is a sharing rule for which $w_i(y)$ is weakly concave and differentiable in y for all $i \in \mathcal{I}$. For any Nash equilibrium effort vector e^*, it must be the case that

$$w_i'(y) \cdot \frac{\partial y(e^*)}{\partial e_i} = c_i'(e_i^*) \text{ for all } i \in \mathcal{I}.$$

In order for e^* to be equal to e^{FB}, it has to be the case that these equilibrium conditions coincide with the Pareto-optimality conditions. This is only possible if $w_i'(y) = 1$ for all i, but because w is a sharing rule, we must have that

$$\sum_{i\in\mathcal{I}} w_i'(y) = 1 \text{ for all } y.$$

Equilibrium effort e^* therefore cannot be first-best. This argument highlights the idea that getting each Agent to choose first-best effort requires that he be given the entire social marginal benefit of his effort, but it is not possible (at least under a sharing rule) for *all* the Agents simultaneously to receive the entire social marginal benefit of their efforts.

This argument is not a full argument for the impossibility of attaining first-best effort under sharing rules because it does not rule out the possibility of non-differentiable sharing rules inducing first-best effort. It turns out that there is no sharing rule, even a non-differentiable one, that induces first-best effort.

Theorem 3 (Moral Hazard in Teams). If w is a sharing rule, w does not induce e^{FB}.

Proof. This proof is due to Stole (2001). Take an arbitrary sharing rule w, and suppose e^* is an equilibrium effort profile under w. For any $i, j \in \mathcal{I}$, define $e_j(e_i)$ by the relation $y\left(e_{-j}^*, e_j(e_i)\right) = y\left(e_{-i}^*, e_i\right)$. Since y is continuous and increasing, a unique value of $e_j(e_i)$ exists for e_i sufficiently close to e_i^*. Take such an e_i. For e^* to be a Nash equilibrium, it must be the case that

$$w_j(y(e^*)) - c_j\left(e_j^*\right) \geq w_j\left(y\left(e_{-j}^*, e_j(e_i)\right)\right) - c_j(e_j(e_j)),$$

since this inequality has to hold for all $e_j \neq e_j^*$. Rewriting this inequality, and summing up over $j \in \mathcal{I}$, we have

$$\sum_{j\in\mathcal{I}}\left(w_j(y(e^*)) - w_j\left(y\left(e_{-i}^*, e_i\right)\right)\right) \geq \sum_{j\in\mathcal{I}}\left(c_j\left(e_j^*\right) - c_j(e_j(e_i))\right).$$

Since w is a sharing rule, the left-hand side of this expression is just $y(e^*) - y\left(e_{-i}^*, e_i\right)$, so this inequality can be written

$$y\left(e^{*}\right)-y\left(e_{-i}^{*},e_{i}\right)\geq\sum_{j\in\mathcal{I}}c_{j}\left(e_{j}^{*}\right)-c_{j}\left(e_{j}\left(e_{i}\right)\right).$$

Since this must hold for all e_i close to e_i^*, we can divide by $e_i^* - e_i$ and take the limit as $e_i \to e_i^*$ to obtain

$$\frac{\partial y\left(e^{*}\right)}{\partial e_{i}}\geq\sum_{j\in\mathcal{J}}c_{j}'\left(e_{j}^{*}\right)\frac{\partial y\left(e^{*}\right)/\partial e_{i}}{\partial y\left(e^{*}\right)/\partial e_{j}}.$$

Now suppose that $e^* = e^{FB}$. Then $c_j'\left(e_j^*\right) = \partial y\left(e^*\right)/\partial e_j$, so this inequality becomes

$$\frac{\partial y\left(e^{*}\right)}{\partial e_{i}}\geq I\frac{\partial y\left(e^{*}\right)}{\partial e_{i}},$$

which is a contradiction because y is increasing in e_i.■

Joint Punishments and Budget Breakers Under a sharing rule, first-best effort cannot be implemented because in order to deter an Agent from choosing some $e_i < e_i^{FB}$, it is necessary to punish him. But because contracts can only be written on team output, the only way to deter each agent from choosing $e_i < e_i^{FB}$ is to simultaneously punish *all* the Agents when output is less than $y\left(e^{FB}\right)$. But punishing all the Agents simultaneously requires that they throw output away, which is impossible under a sharing rule. It turns out, though, that if we allow for contracts w that allow for **money burning**, in the sense that it allows for

$$\sum_{i\in\mathcal{I}}w_{i}\left(y\right)<y$$

for some output levels $y \in \mathcal{Y}$, first-best effort can in fact be implemented, and it can be implemented with a contract that does not actually burn money in equilibrium.

Proposition 1. There exist a vector of contracts w for which $\sum_{i\in\mathcal{I}} w_i\left(y^{FB}\right) = y^{FB}$.

Proof. For all i, set $w_i\left(y\right) = 0$ for all $y \neq y^{FB}$, and let $w_i\left(y^{FB}\right) > c_i\left(e_i^{FB}\right)$ for all i so that $\sum_{i\in\mathcal{I}} w_i\left(y^{FB}\right) = y^{FB}$. Such a vector of contracts is feasible, because $y^{FB} > \sum_{i\in\mathcal{I}} c_i\left(e_i^{FB}\right)$.

Finally, under w, e^{FB} is a Nash equilibrium effort profile because if all other Agents choose e_{-i}^{FB}, then if Agent i chooses $e_i \neq e_i^{FB}$, he receives $-c_i(e_i)$, if he chooses $e_i = e_i^{FB}$, he receives $w_i\left(y^{FB}\right) - c_i\left(e_i^{FB}\right) > 0.$■

Proposition 1 shows that in order to induce first-best effort, the Agents have to subject themselves to costly joint punishments in the event that one of them deviates and chooses $e_i \neq e_i^{FB}$. A concern with such contracts is that in the event that the Agents are required by the contract to burn money, they could all be made better off by renegotiating their contract and not burning money. If we insist, therefore, that w is *renegotiation-proof*, then w must be a sharing rule and therefore cannot induce e^{FB}.

This is no longer the case if we introduce an additional party, which we will call a Principal, who does not take any actions that affect output. In particular, if we denote the Principal as Agent 0, then the following sharing rule induces e^{FB}:

$$\begin{aligned} w_i(y) &= y - k \text{ for all } i = 1, \ldots, I \\ w_0(y) &= Ik - (I-1)\,y, \end{aligned}$$

where k satisfies

$$k = \frac{I-1}{I} y^{FB}.$$

This vector of contracts is a sharing rule, since for all $y \in \mathcal{Y}$,

$$\sum_{i=0}^{I} w_i(y) = Iy - (I-1)\,y = y.$$

This vector of contracts induces e^{FB} because it satisfies $\partial w_i\left(y^{FB}\right)/\partial e_i = 1$ for all $i = 1, \ldots, I$, and if we imagine the Principal having an outside option of 0, this choice of k ensures that in equilibrium, she will in fact receive 0. In this case, the Principal's role is to serve as a **budget breaker**. Her presence allows the Agents to "break the margins budget," allowing for $\sum_{i=1}^{I} w_i'(y) = I > 1$, while still allowing for renegotiation-proof contracts.

Under these contracts, the Principal essentially "sells the firm" to *each* agent for an amount k. Then, since each Agent earns the firm's entire output at the margin, each Agent's interests are aligned with society's interest. One limitation of this approach is that while each Agent earns the entire marginal benefit of his efforts, the Principal *loses* $I-1$ times the marginal benefit of each Agent's efforts. The Principal has strong incentives to collude with one of the Agents—while the players are jointly better off if Agent i chooses e_i^{FB} than any $e_i < e_i^{FB}$, Agent i and the Principal together are jointly better off if Agent i chose $e_i = 0$.