Lectures on the Theory of Contracts

Lars A. Stole[*]

First version: September 1993
Current version: September 1999

1 Preface

The contours of contract theory as a field are difficult to define. Many would argue that contract theory is a subset of Game Theory which is defined by the notion that one party to the game (typically called the principal) is given all of the bargaining power and so can make a take-it-or-leave-it offer to the other party or parties (i.e., the agent(s)). In fact, the techniques for screening contracts were largely developed by pure game theorists to study allocation mechanisms and game design. But then again, carefully defined, everything is a subset of game theory.

Others would argue that contract theory is an extension of price theory in the following sense. Price theory studies how actors interact where the actors are allowed to choose prices, wages, quantities, etc. and studies partial or general equilibrium outcomes. Contract theory extends the choice spaces of the actors to include richer strategies (i.e. contracts) rather than simple one-dimensional choice variables. Hence, a firm can offer a nonlinear price menu to its customers (i.e., a screening contract) rather than a simple uniform price and an employer can offer its employee a wage schedule for differing levels of stochastic performance (i.e., an incentives contract) rather than a simple wage.

Finally, one could group contract theory together by the substantive questions it asks. How should contracts be developed between principals and their agents to provide correct incentives for communication of information and actions. Thus, contract theory seeks to understand organizations, institutions, and relationships between productive individuals when there are differences in personal objectives (e.g., effort, information revelation, etc.). It is this later classification that probably best defines contract theory as a field, although many interesting questions such as the optimal design of auctions and resource allocation do not fit this description very well but comprise an important part of contract theory nonetheless.

The notes provided are meant to cover the rough contours of contract theory. Much of their substance is borrowed heavily from the lectures and notes of Mathias Dewatripont, Bob Gibbons, Oliver Hart, Serge Moresi, Klaus Schmidt, Jean Tirole, and Jeff Zwiebel. In addition, helpful, detailed comments and suggestions were provided by Rohan Ptichford, Adriano Rampini, David Roth, Jennifer Wu and especially by Daivd Martimort. I have relied on many outside published sources for guidance and have tried to indicate the relevant contributions in the notes where they occur. Financial support for compiling these notes into their present form was provided by a National Science Foundation Presidential Faculty Fellowship and a Sloan Foundation Fellowship. I see the purpose of these notes as (i) to standardize the notation and approaches across the many papers in the field, (ii), to present the results of later papers building upon the theorems of earlier papers, and (iii) in a few cases present my own intuition and alternative approaches when I think it adds something to the presentation of the original author(s) and differs from the standard paradigm. Please feel free to distribute these notes in their entirety if you wish to do so.

2 Moral Hazard and Incentives Contracts

2.1 Static Principal-Agent Moral Hazard Models

2.1.1 The Basic Theory

The Model We now turn to the consideration of moral hazard. The workhorse of this literature is a simple model with one principal who makes a take-it-or-leave-it offer to a single agent with outside reservation utility of $\underline{U}$ under conditions of symmetric information. If the contract is accepted, the agent then chooses an action, $a \in \mathcal{A}$, which will have an effect (usual stochastic) on an outcome, $x \in \mathcal{X}$, of which the principal cares about and is typically "informative" about the agent's action. The principal may observe some additional signal, $s \in \mathcal{S}$, which may also be informative about the agent's action. The simplest version of this model casts x as monetary profits and $s = \emptyset$; we will focus on this simple model for now ignoring information besides x.

We will assume that x is observable and *verifiable*. This latter term is used to indicate that enforceable contracts can be written on the variable, x. The nature of the principal's contract offer will be a wage schedule, $w(x)$, according to which the agent is rewarded. We will also assume for now that the principal has full commitment and will not alter the contract $w(x)$ later – even if it is Pareto improving.

The agent takes a hidden action $a \in \mathcal{A}$ which yields a random monetary return $\tilde{x} = x(\tilde{\theta}, a)$; e.g., $\tilde{x} = \tilde{\theta} + a$. This action has the effect of stochastically improving x (e.g., $\mathrm{E}_\theta[x_a(\tilde{\theta}, a)] > 0$) but at a monetary disutility of $\psi(a)$, which is continuously differentiable, increasing and strictly convex. The monetary utility of the principal is $V(x - w(x))$, where $V' > 0 \geq V''$. The agent's net utility is separable in cost of effort and money:

$$U(w(x), a) \equiv u(w(x)) - \psi(a),$$

where $u' > 0 \geq u''$.

Rather than deal with the stochastic function $\tilde{x} = x(\tilde{\theta}, a)$, which is referred to as the *state-space representation* and was used by earlier incentive papers in the literature, we will find it useful to consider instead the density and distribution induced over x for a given

action; this is referred to as the *parameterized distribution characterization.* Let $\mathcal{X} \equiv [\underline{x}, \overline{x}]$ be the support of output; let $f(x, a) > 0$ for all $x \in \mathcal{X}$ be the density; and let $F(x, a)$ be the cumulative distribution function. We will assume that f_a and f_{aa} exist and are continuous. Furthermore, our assumption that $\mathrm{E}_\theta[x_a(\tilde{\theta}, a)] > 0$ is equivalent to $\int_{\underline{x}}^{\overline{x}} F_a(x, a)dx < 0$; we will assume a stronger (but easier to use) assumption that $F_a(x, a) < 0 \ \ \forall x \in (\underline{x}, \overline{x})$; i.e., effort produces a first-order stochastic dominant shift on $\mathcal{X}$. Finally, note that since our support is fixed, $F_a(\underline{x}, a) = F_a(\overline{x}, a) = 0$ for any action, a. The assumption that the support of x is fixed is restrictive as we will observe below in remark 2.

The Full Information Benchmark Let's begin with the full information outcome where effort is observable and verifiable. The principal chooses a and $w(x)$ to satisfy

$$\max_{w(\cdot),a} \int_{\underline{x}}^{\overline{x}} V(x - w(x))f(x, a)dx,$$

subject to

$$\int_{\underline{x}}^{\overline{x}} u(w(x))f(x, a)dx - \psi(a) \geq \underline{U},$$

where the constraint is the agent's participation or individual rationality (IR) constraint. The Lagrangian is

$$\mathcal{L} = \int_{\underline{x}}^{\overline{x}} [V(x - w(x)) + \lambda u(w(x))]f(x, a)dx - \lambda \psi(a) - \lambda \underline{U},$$

where λ is the Lagrange multiplier associated with the IR constraint; it represents the shadow price of income to the agent in each state. Assuming an interior solution, we have as first-order conditions,

$$\frac{V'(x - w(x))}{u'(w(x))} = \lambda, \ \ \forall x \in \mathcal{X},$$

$$\int_{\underline{x}}^{\overline{x}} [V(x - w(x)) + \lambda u(w(x))]f_a(x, a)dx = \lambda \psi'(a),$$

and the IR constraint is binding.

The first condition is known as the Borch rule: the ratios of marginal utilities of income are equated across states under an optimal insurance contract. Note that it holds for every x and not just in expectation. The second condition is the choice of effort condition.

Remarks:

1. Note that if $V' = 1$ and $u'' < 0$, we have a risk neutral principal and a risk averse agent. In this case, the Borch rule requires $w(x)$ be constant so as to provide perfect insurance to the agent. If the reverse were true, the agent would perfectly insure the principal, and $w(x) = x + k$, where k is a constant.

2. Also note that the first-order condition for the choice of effort can be re-written as follows:

$$
\begin{aligned}
\psi'(a) &= \int_{\underline{x}}^{\overline{x}} [V(x - w(x))/\lambda + u(w(x))] f_a(x, a) dx, \\
&= [V(x - w(x))/\lambda + u(w(x))] F_a(x, a)|_{\underline{x}}^{\overline{x}} \\
&\quad - \int_{\underline{x}}^{\overline{x}} [V'(x - w(x))(1 - w'(x))/\lambda + u'(w(x))w'(x)] F_a(x, a) dx, \\
&= - \int_{\underline{x}}^{\overline{x}} u'(w(x)) F_a(x, a) dx.
\end{aligned}
$$

Thus, if the agent were risk neutral (i.e., $u' = 1$), integrating by parts one obtains

$$
\int_{\underline{x}}^{\overline{x}} x f_a(x, a) dx = \psi'(a).
$$

I.e., a maximizes $E[x|a] - \psi(a)$. So even if effort cannot be contracted on as in the full-information case, if the agent is risk neutral then the principal can "sell the enterprise to the agent" with $w(x) = x - k$, and the agent will choose the first-best level of effort.

The Hidden Action Case   We now suppose that the level of effort cannot be contracted upon and the agent is risk averse: $u'' < 0$. The principal solves the following program

$$
\max_{w(\cdot), a} \int_{\underline{x}}^{\overline{x}} V(x - w(x)) f(x, a) dx,
$$

subject to

$$
\int_{\underline{x}}^{\overline{x}} u(w(x)) f(x, a) dx - \psi(a) \geq \underline{U},
$$

$$
a \in \arg\max_{a' \in \mathcal{A}} \int_{\underline{x}}^{\overline{x}} u(w(x)) f(x, a') dx - \psi(a'),
$$

where the additional constraint is the incentive compatibility (IC) constraint for effort. The IC constraint implies (assuming an interior optimum) that

$$
\int_{\underline{x}}^{\overline{x}} u(w(x)) f_a(x, a) dx - \psi'(a) = 0,
$$

$$
\int_{\underline{x}}^{\overline{x}} u(w(x)) f_{aa}(x, a) dx - \psi''(a) \leq 0,
$$

which are the local first- and second-order conditions for a maximum.

The *first-order approach* (FOA) to incentives contracts is to maximize subject to the first-order condition rather than IC, and then check to see if the solution indeed satisfies IC ex post. Let's ignore questions of the validity of this procedure for now; we'll return to the problems associated with its use later.

Using μ as the multiplier on the effort first-order condition, the Lagrangian of the FOA program is

$$\mathcal{L} = \int_{\underline{x}}^{\overline{x}} V(x - w(x)) f(x, a) dx + \lambda \left[\int_{\underline{x}}^{\overline{x}} u(w(x)) f(x, a) dx - \psi(a) - \underline{U} \right]$$
$$+ \mu \left[\int_{\underline{x}}^{\overline{x}} u(w(x)) f_a(x, a) dx - \psi'(a) \right].$$

Maximizing $w(x)$ pointwise in x and simplifying, the first-order condition is

$$\frac{V'(x - w(x))}{u'(w(x))} = \lambda + \mu \frac{f_a(x, a)}{f(x, a)}, \quad \forall x \in \mathcal{X}. \tag{1}$$

We now have a modified Borch rule: the marginal rates of substitution may vary if $\mu > 0$ to take into account the incentives effect of $w(x)$. Thus, risk-sharing will generally be inefficient.

Consider a simple two-action case in which the principal wishes to induce the high action: $\mathcal{A} \equiv \{a_L, a_H\}$. Then the IC constraint implies the inequality

$$\int_{\underline{x}}^{\overline{x}} u(w(x))[f(x, a_H) - f(x, a_L)] dx \geq \psi(a_H) - \psi(a_L).$$

The first-order condition for the associated Lagrangian is

$$\frac{V'(x - w(x))}{u'(w(x))} = \lambda + \mu \left(\frac{f(x, a_H) - f(x, a_L)}{f(x, a_H)} \right), \quad \forall x \in \mathcal{X}.$$

In both cases, providing $\mu > 0$, the agent is rewarded for outcomes which have higher relative frequency under high effort. We now prove that this is indeed the case.

Theorem 1 (Holmström, [1979]) Assume that the FOA program is valid. Then at the optimum of the FOA program, $\mu > 0$.

Proof: The proof of the theorem relies upon first-order stochastic dominance $F_a(x, a) < 0$ and risk aversion $u'' < 0$. Consider $\frac{\partial \mathcal{L}}{\partial a} = 0$. Using the agent's first-order condition for effort choice, it simplifies to

$$\int_{\underline{x}}^{\overline{x}} V(x - w(x)) f_a(x, a) dx + \mu \left[\int_{\underline{x}}^{\overline{x}} u(w(x)) f_{aa}(x, a) dx - \psi''(a) \right] = 0.$$

Suppose that $\mu \leq 0$. By the agent's second-order condition for choosing effort we have

$$\int_{\underline{x}}^{\overline{x}} V(x - w(x)) f_a(x, a) dx \leq 0.$$

Now, define $w_\lambda(x)$ as the solution to refmh-star when $\mu = 0$; i.e.,

$$\frac{V'(x - w_\lambda(x))}{u'(w_\lambda(x))} = \lambda, \quad \forall x \in \mathcal{X}.$$

5

Note that because $u'' < 0$, w_λ is differentiable and $w'_\lambda(x) \in [0,1)$. Compare this to the solution, $w(x)$, which satisfies

$$\frac{V'(x - w(x))}{u'(w(x))} = \lambda + \mu \frac{f_a(x,a)}{f(x,a)}, \quad \forall x \in \mathcal{X}.$$

When $\mu \leq 0$, it follows that $w(x) \leq w_\lambda(x)$ if and only if $f_a(x,a) \geq As$. Thus,

$$V(x - w(x))f_a(x,a) \geq V(x - w_\lambda(x))f_a(x,a), \quad \forall x \in \mathcal{X},$$

and as a consequence,

$$\int_{\underline{x}}^{\overline{x}} V(x - w(x))f_a(x,a)dx \geq \int_{\underline{x}}^{\overline{x}} V(x - w_\lambda(x))f_a(x,a)dx.$$

The RHS is necessarily positive because integrating by parts yields

$$V(x - w_\lambda(x))F_a(x,a)\big|_{\underline{x}}^{\overline{x}} - \int_{\underline{x}}^{\overline{x}} V'(x - w_\lambda(x))(1 - w'_\lambda(x))F_a(x,a)dx > 0.$$

But then this implies a contradiction. $\square$

Remarks:

1. Note that we have assumed that the agent's chosen action is in the interior of $a \in \mathcal{A}$. If the principal wishes to implement the least costly action in $\mathcal{A}$, perhaps because the agent's actions have little effect on output, then the agent's first-order condition does not necessarily hold. In fact, the problem is trivial and the principal will supply full insurance if $V'' = 0$. In this case an inequality for the agent's corner solution must be included in the maximization program rather than a first-order condition; the associated multiplier will be zero: $\mu' = 0$.

2. The assumption that the support of x does not depend upon effort is crucial. If the support "shifts", it may be possible to obtain the first best, as some outcomes may be perfectly informative about effort.

3. Commitment is important in the implementation of the optimal contract. In equilibrium the principal knows the agent took the required action. Because $\mu > 0$, the principal is imposing unnecessary risk upon the agent, *ex post*. Between the time the action is taken and the uncertainty resolved, there generally exist Pareto-improving contracts to which the parties cloud mutually renegotiate. The principal must commit not to do this to implement the optimal contract above. We will return to this issue when we consider the case in which the principal cannot commit not to renegotiate.

4. Note that f_a/f is the derivative of the log-likelihood function, $\log f(x,a)$, an hence is the gradient for a MLE estimate of a given observed x. In equilibrium, the principal "knows" $a = a^*$ even though he commits to a mechanism which rewards based upon the informativeness of x for $a = a^*$.

5. The above FOA program can be extended to models of adverse selection and moral hazard. That is, the agent observes some private information, θ, before contracting (and choosing action) and the principal offers a wage schedule, $w(x, \hat{\theta})$. The difference now is that $\mu(\hat{\theta})$ will generally depend upon the agent's announced information and may become nonpositive for some values. See Holmström, [1979], section 6.

6. Note that although $w'_\lambda(x) \in [0, 1)$, we haven't demonstrated that the optimal $w(x)$ is monotonic. Generally, first-order stochastic dominance is not enough. We will need a stronger property.

We now turn to the question of monotonicity.

Definition 1 The **monotone likelihood ratio property (MLRP)** is satisfied for a distribution F and its density f iff

$$\frac{d}{dx}\left(\frac{f_a(x, a)}{f(x, a)}\right) \geq 0.$$

Note that when effort is restricted to only two types so f is non-differentiable, the analogous MLRP condition is that

$$\frac{d}{dx}\left(\frac{f(x, a_H) - f(x, a_L)}{f(x, a_H)}\right) \geq 0.$$

Additionally it is worth noting that MLRP implies that $F_a(x, a) < 0$ for $x \in (\underline{x}, \overline{x})$ (i.e., first-order stochastic dominance). Specifically, for $x \in (\underline{x}, \overline{x})$

$$F_a(x, a) = \int_{ux}^{x} \frac{f_a(s, a)}{f(s, a)} f(s, a) ds < 0,$$

where the latter inequality follows from MLRP (when $x = \overline{x}$, the integral is 0; when $x < \overline{x}$, the fact that the likelihood ratio is increasing in s implies that the integral must be strictly negative).

We have the following result.

Theorem 2 (Holmström, [1979], Shavell [1979]). Under the first-order approach, if F satisfies the monotone likelihood ratio property, then the wage contract is increasing in output.

The proof is immediate from the definition of MLRP and our first-order conditions above.

Remarks:

1. Sometimes you do not want monotonicity because the likelihood ratio is u-shaped. For example, suppose that only two effort levels are possible, $\{a_L, a_H\}$, and only three output levels can occur, $\{x_1, x_2, x_3\}$. Let the probabilities be given by the following $f(x, a)$:

$f(x,a)$	x_1	x_2	x_3
a_H	0.4	0.1	0.5
a_L	0.5	0.4	0.1
L-Ratio	-0.25	-3	0.8

If the principal wishes to induce high effort, the idea is to use a non-monotone wage schedule to punish moderate outputs which are most indicative of low effort, reward high outputs which are quite informative about high effort, and provide moderate income for low outputs which are not very informative.

2. If agents can freely dispose of the output, monotonicity may be a constraint which the solution must satisfy. In addition, if agent's can trade output amongst themselves (say the agents are sharecroppers), then only a linear constraint is feasible with lots of agents; any nonlinearities will be arbitraged away.

3. We still haven't demonstrated that the first-order approach is correct. We will turn to this shortly.

The Value of Information Let's return to our more general setting for a moment and assume that the principal and agent can enlarge their contract to include other information, such as an observable and verifiable signal, s. When should w depend upon s?

Definition 2 x is **sufficient** for $\{x, s\}$ with respect to $a \in \mathcal{A}$ iff f is multiplicatively separable in s and a; i.e.

$$f(x, s, a) \equiv y(x, a)z(x, s).$$

We say that s is **informative** about $a \in \mathcal{A}$ whenever x is not sufficient for $\{x, s\}$ with respect to $a \in \mathcal{A}$.

Theorem 3 (Holmström, [1979], Shavell [1979]). Assume that the FOA program is valid and yields $w(x)$ as a solution. Then there exists a new contract, $w(x, s)$, that strictly Pareto dominates $w(x)$ iff s is informative about $a \in \mathcal{A}$.

Proof: Using the FOA program, but allowing $w(\cdot)$ to depend upon s as well as x, the first-order condition determining w is given by

$$\frac{V'(x - w(x, s))}{u'(w(x, s))} = \lambda + \mu \frac{f_a(x, s, a)}{f(x, s, a)},$$

which is independent of s iff s is not informative about $a \in \mathcal{A}$. □

The result implies that without loss of generality, the principal can restrict attention to wage contracts that depend only upon a set of sufficient statistics for the agent's action. Any other dependence cannot improve the contract; it can only increase the risk the agent faces without improving incentives. Additionally, the result says that any informative signal about the agent's action should be included in the optimal contract!

8

Application: Insurance Deductibles. We want to show that under reasonable assumptions it is optimal to offer insurance policies which provide full insurance less a fixed deductible. The idea is that if conditional on an accident occurring, the value of the loss is uninformative about a, then the coverage of optimal insurance contract should also not depend upon actual losses – only whether or not there was an accident. Hence, deductibles are optimal. To this end, let x be the size of the loss and assume $f_a(0,a) = 1 - p(a)$ and $f(x,a) = p(a)g(x)$ for $x < 0$. Here, the probability of an accident depends upon effort in the obvious manner: $p'(a) < 0$. The amount of loss x is independent of a (i.e., $g(x)$ is independent of a). Thus, the optimal contract is characterized by a likelihood ratio of $\frac{f_a(x,a)}{f(x,a)} = \frac{p'(a)}{p(a)} < 0$ for $x < 0$ (which is independent of x) and $\frac{f_a(x,a)}{f(x,a)} = \frac{-p'(a)}{1-p(a)} > 0$ for $x = 0$. This implies that the final income allocation to the agent is fixed at one level for all $x < 0$ and at another for $x = 0$, which can be implemented by the insurance company by offering full coverage less a deductible.

Asymptotic First-best It may be that by making very harsh punishments with very low probability that the full-information outcome can be approximated arbitrarily closely. This insight is due to Mirrlees [1974].

Theorem 4 (Mirrlees, [1974].) Suppose $f(x,a)$ is the normal distribution with mean a and variance σ^2. Then if unlimited punishments are possible, the first-best can be approximated arbitrarily closely.

Sketch of Proof: We prove the theorem for the case of a risk-neutral principal. We have

$$f(x,a) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{(x-a)^2}{2\sigma^2}},$$

so that

$$\frac{f_a(x,a)}{f(x,a)} = \frac{d}{da}\log f(x,a) = \frac{(x-a)}{\sigma^2}.$$

That is, detection is quite efficient for x very small.

The first-best contract has a constant w^* such that $u(w^*) = \underline{U} + \psi(a^*)$, where a^* is the first-best action. The approximate first-best contract offers w^* for all $x \geq x_o$ (x_o very small), and $w = k$ (k very small) for $x < x_o$. Choose k low enough such that a^* is optimal for the agent (IC) at a given x_o:

$$\int_{-\infty}^{x_o} u(k)f_a(x,a^*)dx + \int_{x_o}^{\infty} u(w^*)f_a(x,a^*)dx = \psi'(a^*).$$

We want to show that with this contract, the agent's IR constraint can be satisfied arbitrarily closely as we lower the punishment region. Note that the loss with respect to the first-best is

$$\Delta \equiv \int_{-\infty}^{x_o} (u(w^*) - u(k))f(x,a^*)dx,$$

for the agent. Define $M(x_o) \equiv \frac{f_a(x_o,a^*)}{f(x_o,a^*)}$. Because the normal distribution satisfies MLRP, for all $x < x_o$, $f_a/f < M$ or $f > f_a/M$. This implies that the difference between the agent's

utility and $\underline{U}$ is bounded above by

$$\frac{1}{M} \int_{-\infty}^{x_o} (u(w^*) - u(k)) f_a(x, a^*) dx \geq \Delta.$$

But by the agent's IC condition, this bound is given by

$$\frac{1}{M} \left(\int_{-\infty}^{\infty} u(w^*) f_a(x, a^*) dx - \psi'(a^*) \right),$$

which is a constant divided by M. Thus, as punishments are increased, i.e, M decreased, we approach the first best. $\square$

The intuition for the result is that in the tails of a normal distribution, the outcome is very informative about the agent's action. Thus, even though the agent is risk averse and harsh random punishments are costly to the principal, the gain from informativeness dominates at any punishment level.

The Validity of the First-order Approach   We now turn to the question of the validity of the first-order approach.

The approach of first finding $w(x)$ using the relaxed FOA program and then checking that the principal's selection of a maximizes the agent's objective function is logically invalid without additional assumptions. Generally, the problem is that when the second-order condition of the agent is not be globally satisfied, it is possible that the solution to the unrelaxed program satisfies the agent's first-order condition (which is necessary) but not the principal's first-order condition. That is, the principal's optimum may involve a corner solution and so solutions to the unrelaxed direct program may not satisfy the necessary Kuhn-Tucker conditions of the relaxed FOA program. This point was first made by Mirrlees [1974].

There are a few important papers on this concern. Mirrlees [1976] has shown that MLRP with an additional convexity in the distribution function condition (CDFC) is sufficient for the validity of the first-order approach.

Definition 3 A distribution satisfies the **Convexity of Distribution Function Condition (CDFC)** iff

$$F(x, \gamma a + (1 - \gamma)a') \leq \gamma F(x, a) + (1 - \gamma) F(x, a'),$$

for all $\gamma \in [0, 1]$. (I.e., $F_{aa}(x, a) \geq 0$.)

A useful special case of this CDF condition is the *linear distribution function condition*:

$$f(x, a) \equiv a \overline{f}(x) + (1 - a) \underline{f}(x),$$

where $\overline{f}(x)$ first-order stochastically dominates $\underline{f}(x)$.

Mirrlees' theorem is correct but the proof contains a subtle omission. Independently, Rogerson [1985] determines the same sufficient conditions for the first-order approach using a correct proof. Essentially, he derives conditions which guarantee that the agent's objective function will be globally concave in action for any selected contract.

In an earlier paper, Grossman and Hart [1983] study the general unrelaxed program directly rather than a relaxed program. They find, among other things, that MLRP and CDFC are sufficient for the monotonicity of optimal wage schedules. Additionally, they show that the program can be reduced to an elegant linear programming problem. Their methodology is quite nice and a significant contribution to contract theory independent of their results.

Rogerson [1985]:

Theorem 5 (Rogerson, [1985].) The first-order approach is valid if $F(x, a)$ satisfies the MLRP and CDF conditions.

We first begin with a simple commonly used, but incorrect, "proof" to illustrate the subtle circularity of proving the validity of the FOA program.

"Proof": One can rewrite the agent's payoff as

$$\int_{\underline{x}}^{\overline{x}} u(w(x)) \, f(x,a)dx - \psi(a) \; = \; u(w(x))F(x,a)|_{\underline{x}}^{\overline{x}}$$

$$- \int_{\underline{x}}^{\overline{x}} u'(w(x)) \frac{dw(x)}{dx} F(x,a)dx - \psi(a)$$

$$= \; u(w(\overline{x})) - \int_{\underline{x}}^{\overline{x}} u'(w(x)) \frac{dw(x)}{dx} F(x,a)dx - \psi(a),$$

where we have assumed for now that $w(x)$ is differentiable. Differentiating this with respect to a twice yields

$$- \int_{\underline{x}}^{\overline{x}} u'(w(x)) \frac{dw(x)}{dx} F_{aa}(x,a)dx - \psi''(a) < 0,$$

for every $a \in \mathcal{A}$. Thus, the agent's second-order condition is globally satisfied in the FOA program if $w(x)$ is differentiable and nondecreasing. Under MLRP, $\mu > 0$, and so the first-order approach yields a monotonically increasing, differentiable $w(x)$; we are done. □

Note: The mistake is in the last line of the proof which is circular. You cannot use the FOA $\mu > 0$ result, without first proving that the first-order approach is valid. (In the proof of $\mu > 0$, we implicitly assumed that the agent's second-order condition was satisfied).

Rogerson avoids this problem by focusing on a doubly-relaxed program where the first-order condition is replaced by $\frac{d}{da} E[U(w(x), a)|a] \geq 0$. Because the constraint is an inequality, we are assured that the multiplier is nonnegative: $\delta \geq 0$. Thus, the solution to the doubly-relaxed program implies a nondecreasing, differentiable wage schedule under MLRP. The second step is to show that the solution of the doubly-relaxed program satisfies the constraints of the relaxed program (i.e., the optimal contract satisfies the agent's first-order condition with equality). This result, combined with the above "Proof", provides a complete proof of the theorem by demonstrating that the double-relaxed solution satisfies the unrelaxed constraint set. This second step is provided in the following lemma.

Lemma 1 (Rogerson, [1985].) At the doubly-relaxed program solution,

$$\frac{d}{da}E[U(w,a)|a] = 0.$$

Proof: To see that $\frac{d}{da}E[U(w,a)|a] = 0$ at the solution doubly-relaxed program, consider the inequality constraint multiplier, δ. If $\delta > 0$, the first-order condition is clearly satisfied. Suppose $\delta = 0$ and necessarily $\lambda > 0$. This implies the optimal risk-sharing choice of $w(x) = w_\lambda(x)$, where $w'_\lambda(x) \in [0,1)$. Integrating the expected utility of the principal by parts yields

$$E[V(x - w(x))|a] = V(\overline{x} - w_\lambda(\overline{x})) - \int_{\underline{x}}^{\overline{x}} V'(x - w_\lambda(x))[1 - w'_\lambda(x)]F(x,a)dx.$$

Differentiating with respect to action yields

$$\frac{\partial E[V|a]}{\partial a} = -\int_{\underline{x}}^{\overline{x}} V'(x - w_\lambda(x))[1 - w'_\lambda(x)]F_a(x,a)dx \geq 0,$$

where the inequality follows from $F_a \leq 0$. Given that $\lambda > 0$, the first-order condition of the doubly-relaxed program for a requires that

$$\frac{d}{da}E[U(w(x),a)|a] \leq 0.$$

This is only consistent with the doubly-relaxed constraint set if

$$\frac{d}{da}E[U(w(x),a)|a] = 0,$$

and so the first-order condition must be satisfied. $\square$

Remarks:

1. Due to Mirrlees' [1976] initial insight and Rogerson's [1985] correction, the MLRP-CDFC sufficiency conditions are usually called the Mirrlees-Rogerson sufficient conditions.

2. The conditions of MLRP and CDFC are very strong. It is difficult to think of many distributions which satisfy them. One possibility is a generalization of the uniform distribution (which is a type of β-distribution):

$$F(x,a) \equiv \left(\frac{x - \underline{x}}{\overline{x} - \underline{x}}\right)^{\frac{1}{1-a}},$$

where $\mathcal{A} = [0,1)$.

3. The CDF condition is particularly strong. Suppose for example that $\tilde{x} \equiv a + \tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is distributed according to some cumulative distribution function. Then the CDF condition requires that the density of the cumulative distribution is increasing in ε! Jewitt [1988] provides a collection of alternative sufficient conditions on $F(x,a)$ and $u(w)$

which avoid the assumption of CDFC. Examples include CARA utility with either (i) a Gamma distribution with mean αa (i.e., $f(x,a) = a^{-\alpha} x^{\alpha-1} e^{-x/a} \Gamma(\alpha)^{-1}$), (ii) a Poisson distribution with mean a (i.e., $f(x,a) = a^x e^{-a} / \Gamma(1+x)$), or (iii) a Chi-squared distribution with a degrees of freedom (i.e., $f(x,a) = \Gamma(2a)^{-1} 2^{-2a} x^{2a-1} e^{-x/2}$). Jewitt also extends the sufficiency theorem to situations in which there are multiple signals as in Holmström [1979]. With CARA utility for example, MLRP and CDFC are sufficient for the validity of the FOA program with multiple signals. See also Sinclair-Desgagne [1994] for further generalizations for use of the first-order approach and multiple signals.

4. Jewitt also has a nice elegant proof that the solution to the relaxed program (valid or invalid) necessarily has $\mu > 0$ when $v'' = 0$. The idea is to show that $u(w(x))$ and $1/u'(w(x))$ generally have a positive covariance and note that at the solution to the FOA program, the covariance is equal to $\mu\psi'(a)$. Specifically, note that 1 is equivalent to

$$f_a(x,a) = \left(\frac{1}{u'(w(x))} - \lambda \right) \frac{f(x,a)}{\mu},$$

which allows us to rewrite the agent's first-order condition for action as

$$\int_{\underline{x}}^{\overline{x}} u(w(x)) \left(\frac{1}{u'(w(x))} - \lambda \right) f(x,a) dx = \mu\psi'(a).$$

Since the expected value of each side of 1 is zero, the mean of $\frac{1}{u'(w(x))}$ is λ and so the righthand side of the above equation is the covariance of $u(w(x))$ and $\frac{1}{u'(w(x))}$; and since both functions are increasing in $w(x)$, the covariance is nonnegative implying that $\mu \geq 0$. This proof could have been used above instead of either Holmström's or Rogerson's results to prove a weaker theorem applicable only to risk-neutral principals.

Grossman-Hart [1983]:

We now explicitly consider the unrelaxed program following the approach of Grossman and Hart.

Following the framework of G-H, we assume that there are only a finite number of possible outputs, $x_1 < x_2 < \ldots, x_N$, which occur with probabilities: $f(x_i, a) \equiv \text{Prob}[\tilde{x} = x_i | a] > 0$. We assume that the principal is risk neutral: $V'' = 0$. The agent's utility function is slightly more general:

$$U(w, a) \equiv K(a)u(w) - \psi(a),$$

where every function is sufficiently continuous and differentiable. This is the most general utility function which preserves the requirement that the agent's preference ordering over income lotteries is independent of action. Additionally, if either $K'(a) = 0$ or $\psi'(a) = 0$, the agent's preferences over action lotteries is independent of income. We additionally assume that $\mathcal{A}$ is compact, $u' > 0 > u''$ over the interval $(\underline{I}, \infty)$, and $\lim_{w \to \underline{I}} K(a)u(w) = -\infty$. [The latter bit excludes corner solutions in the optimal contract. We were implicitly assuming

this above when we focused on the FOA program.] Finally, for existence of a solution we need that for any action there exists a $w \in (\underline{I}, \infty)$ such that $K(a)u(w) - \psi(a) \geq \underline{U}$.

When the principal cannot observe a, the second-best contract solves

$$\max_{w,a} \ \sum_i f(x_i, a)(x_i - w_i),$$

subject to

$$a \in \arg\max_{a'} \sum_i f(x_i, a')[K(a')u(w_i) - \psi(a')],$$

$$\sum_i f(x_i, a)[K(a)u(w_i) - \psi(a)] \geq \underline{U}.$$

We proceed in two steps: first, we solve for the least costly way to implement a given action. Then we determine the optimal action to implement.

What's the least costly way to implement $a^* \in \mathcal{A}$? The principal solves

$$\min_{w} \ \sum_i f(x_i, a^*)w_i,$$

subject to

$$\sum_i f(x_i, a^*)[K(a^*)u(w_i) - \psi(a^*)] \geq \sum_i f(x_i, a)[K(a)u(w_i) - \psi(a)], \quad \forall a \in \mathcal{A},$$

$$\sum_i f(x_i, a^*)[K(a^*)u(w_i) - \psi(a^*)] \geq \underline{U}.$$

This is not a convex programming problem amenable to the Kuhn-Tucker theorem. Following Grossman and Hart, we convert it using the following transformation: Let $h(u) \equiv u^{-1}(u)$, so $h(u(w)) = w$. Define $u_i \equiv u(w_i)$, and use this as the control variable. Substituting yields

$$\min_{u} \ \sum_i f(x_i, a^*)h(u_i),$$

subject to

$$\sum_i f(x_i, a^*)[K(a^*)u_i - \psi(a^*)] \geq \sum_i f(x_i, a)[K(a)u_i - \psi(a)], \quad \forall a \in \mathcal{A},$$

$$\sum_i f(x_i, a^*)[K(a^*)u_i - \psi(a^*)] \geq \underline{U}.$$

Because h is convex, this is a convex programming problem with a linear constraint set. Grossman and Hart further show that if either $K'(a) = 0$ or $\psi'(a) = 0$, (i.e., preferences over actions are independent of income), the solution to this program will have the IR constraint binding. In general, the IR constraint may not bind when wealth effects (i.e., leaving money to the agent) induce cheaper incentive compatibility. From now on, we assume that $K(a) = 1$ so that the IR constraint binds.

Further note that when $\mathcal{A}$ is finite, we will have a finite number of constraints, and so we can appeal to the Kuhn-Tucker theorem for necessary and sufficient conditions. We will do this shortly. For now, define

$$C(a^*) \equiv \begin{cases} \inf \left\{ \sum_i f(x_i, a^*) h(u_i) \right\} \text{ if } w \text{ implements } a^*, \\ \infty \text{ otherwise.} \end{cases}$$

Note that some actions cannot be feasibly implemented with any incentive scheme. For example, the principal cannot induce the agent to take a costly action that is dominated: $f(x_i, a) = f(x_i, a') \ \forall \ i$, but $\psi(a) > \psi(a')$.

Given our construction of $C(a)$, the principal's program amounts to choosing a to maximize $B(a) - C(a)$, where $B(a) \equiv \sum_i f(x_i, a) x_i$. Grossman and Hart demonstrate that a (second-best) optimum exists and that the inf in the definition of $C(a)$ can be replaced with a min.

Characteristics of the Optimal Contract

1. Suppose that $\psi(a_{FB}) > \min_{a'} \psi(a')$; (i.e., the first-best action is not the least cost action). Then the second-best contract produces less profit than the first-best. The proof is trivial: the first-best requires full insurance, but then the least cost action will be chosen.

2. Assume that $\mathcal{A}$ is finite so that we can use the Kuhn-Tucker theorem. Then we have the following program:
$$\max_u \ -\sum_i f(x_i, a^*) h(u_i),$$

   subject to

   $$\sum_i f(x_i, a^*)[u_i - \psi(a^*)] \geq \sum_i f(x_i, a_j)[u_i - \psi(a_j)], \ \ \forall a_j \neq a^*,$$

   $$\sum_i f(x_i, a^*)[u_i - \psi(a^*)] \geq \underline{U}.$$

   Let $\mu_j \geq 0$ be the multiplier on the jth IC constraint; $\lambda \geq 0$ the multiplier on the IR constraint. The first-order condition for u_i is

   $$h'(u_i) = \lambda + \sum_{a_j \in A, a_j \neq a^*} \mu_j \left(\frac{f(x_i, a^*) - f(x_i, a_j)}{f(x_i, a^*)} \right).$$

   We know from Grossman-Hart that the IR constraint is binding: $\lambda > 0$. Additionally, from the Kuhn-Tucker theorem, providing a^* is not the minimum cost action, $\mu_j > 0$ for some j where $\psi(a_j) < \psi(a^*)$.

3. Suppose that $\mathcal{A} \equiv \{a_L, a_H\}$ (i.e., there are only two actions), and the principal wishes to implement the more costly action, a_H. Then the first-order condition becomes:
   $$h'(u_i) = \lambda + \mu_L \left(\frac{f(x_i, a_H) - f(x_i, a_L)}{f(x_i, a_H)} \right).$$

15

Because $\mu_L > 0$, w_i increases with $\frac{f(x_i, a_L)}{f(x_i, a_H)}$. The condition that this likelihood ratio increase in i is the MLRP condition for discrete distributions and actions. Thus, MLRP is sufficient for monotonic wage schedules when there are only two actions.

4. Still assuming that $\mathcal{A}$ is finite, with more than two actions all we can say about the wage schedule is that it cannot be decreasing *everywhere*. This is a very weak result.

   The reason is clear from the first-order condition above. MLRP is not sufficient to prove that $\sum_{a_j \in A} \mu_j \frac{f(x_i, a_j)}{f(x_i, a^*)}$ is nonincreasing in i. Combining MLRP with a variant of CDFC (or alternatively imposing a *spanning condition*), however, Grossman and Hart show that monotonicity emerges.

   Thus, while before we demonstrated that MLRP and CDFC guarantees the FOA program is valid and yields a monotonic wage schedule, Grossman and Hart's direct approach also demonstrates that MLRP and CDFC guarantee monotonicity directly. Moreover, as Grossman and Hart discuss in section 6 of their paper, many of their results including monotonicity generalize to the case of a risk averse principal.

2.1.2 Extensions: Moral Hazard in Teams

We now turn to an analysis of multi-agent moral hazard problems, frequently referred to as moral hazard in teams (or partnerships). The classic reference is Holmström [1982].[1] Holmström makes two contributions in this paper. First, he demonstrates the importance of a budget-breaker. Second, he generalizes the notion of sufficient statistics and informativeness to the case of multi-agent situations and examines relative performance evaluation. We consider each contribution in turn.

The Importance of a Budget Breaker.

The canonical multi-agent model, has one risk neutral principal and N possibly risk-averse agents, each who privately choose $a_i \in \mathcal{A}_i$ at a cost given by a strictly increasing and convex cost function, $\psi_i(a_i)$. We assume as before that a_i cannot be contracted upon. The output of the team of agents, x, depends upon the agents' efforts $a \equiv (a_1, \ldots, a_N) \in \mathcal{A} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$, which we will assume is deterministic for now. Later, we will allow for a stochastic function as before in the single-agent model.

A contract is a collection of wage schedules $w = (w_1, \ldots, w_N)$, where each agent's wage schedule indicates the transfer the agent receives as a function of verifiable output; i.e., $w_i(x): \mathcal{X} \to \mathbb{R}^N$.

The timing of the game is as follows. At stage 1, the principal offers a wage schedule for each agent which is observable by everyone. At stage two, the agents reject the contract or accept and simultaneously and noncooperatively select their effort levels, a_i. At stage 3, the output level x is realized and participating agents are paid appropriately.

We say we have a **partnership** when there is effectively no principal and so the agents

---

[1] Mookherjee [1984] also examines many of the same issues in the context of a Grossman-hart [1983] style moral hazard model, but with many agents. His results on sufficient statistics mirrors those of Holmström's, but in a discrete environment.

split the output amongst themselves; i.e., the wage schedule satisfies budget balance:

$$\sum_{i=1}^{N} w_i(x) = x, \ \ \forall x \in \mathcal{X}.$$

An important aspect of a partnership is that the budget (i.e., transfers) are always exactly balanced on and off the equilibrium path.

Holmström [1982] points out that one frequently overlooked benefit of a principal is that she can break the budget while a partnership cannot. To illustrate this principle, suppose that the agents are risk neutral for simplicity.

Theorem 6 Assume each agent is risk neutral. Suppose that output is deterministic and given by the function $x(a)$, strictly increasing and differentiable. If aggregate wage schedules are allowed to be less than total output (i.e., $\sum_i w_i < x$), then the first-best allocation (x^*, a^*) can be implemented as a Nash equilibrium with all output going to the agents (i.e., $\sum_i w_i(x^*) = x^*$), where

$$a^* = \arg\max_a \ \ x(a) - \sum_{i=1}^{N} \psi_i(a_i)$$

and $x^* \equiv x(a^*)$.

Proof: The proof is by construction. The principal accomplishes the first best by paying each agent a fixed wage $w_i(x^*)$ when $x = x^*$ and zero otherwise. By carefully choosing the first-best wage profile, agent's will find it optimal to produce the first best. Choose $w_i(x^*)$ such that

$$w_i(x^*) - \psi_i(a_i^*) \geq 0 \ \ \text{and} \ \ \sum_{i=1}^{N} w_i(x^*) = x^*.$$

Such a wage profile can be found because x^* is optimal. Such a wage profile is also a Nash equilibrium. If all agents other than i choose their respective a_j^*, then agent i faces the following tradeoff: expend effort a_i^* so as to obtain x^* exactly and receive $w_i(x^*) - \psi_i(a_i^*)$, or shirk and receive a wage of zero. By construction of $w_i(x^*)$, agent i will choose a_i^* and so we have a Nash equilibrium. $\square$

Note that although we have budget balance on the equilibrium path, we do not have budget balance off the equilibrium path. Holmström further demonstrates that with budget balance one and off the equilibrium path (e.g., a partnership), the first-best cannot be obtained. Thus, in theory the principal can play an important role as a budget-breaker.

Theorem 7 Assume each agent is risk neutral. Suppose that $a^* \in \text{int } \mathcal{A}$ (i.e., all agent's provide some effort in the first-best allocation) and each $\mathcal{A}_i$ is a closed interval, $[\underline{a}_i, \bar{a}_i]$. Then there do not exist wage schedules which are balanced and yield a^* as a Nash equilibrium in the noncooperative game.

Holmström [1982] provides an intuitive in-text "proof" which he correctly notes relies upon a smooth wage schedule (we saw above that the optimal wage schedule may be discontinuous). The proof given in the appendix is more general but also complicated so I've

provided an alternative proof below which I believe is simpler.

Proof: Define $a_j(a_i)$ by the relation $x(a^*_{-j}, a_j) \equiv x(a^*_{-i}, a_i)$. Since x is continuous and increasing and $a^* \in$ int $\mathcal{A}$, a unique value of $a_j(a_i)$ exists for a_i sufficiently close to a^*. The existence of a Nash equilibrium requires that for such an a_i,

$$w_j(x(a^*)) - w_j(x(a^*_{-j}, a_j(a_i))) \equiv w_j(x(a^*)) - w_j(x(a^*_{-i}, a_i)) \geq \psi_j(a^*_j) - \psi_j(a_j(a_i)).$$

Summing up these relationships yields

$$\sum_{j=1}^{N} (w_j(x(a^*)) - w_j(x(a^*_{-i}, a_i))) \geq \sum_{j=1}^{N} \psi_j(a^*_j) - \psi_j(a_j(a_i)).$$

Budget balance implies that the LHS of this equation is $x(a^*) - x(a^*_{-i}, a_i)$, and so

$$x(a^*) - x(a^*_{-i}, a_i) \geq \sum_{j=1}^{N} \psi_j(a^*_j) - \psi_j(a_j(a_i)).$$

Because this must hold for all a_i close to a^*_i, we can divide by $(a^*_i - a_i)$ and take the limit as $a_i \to a^*_i$ to obtain

$$x_{a_i}(a^*) \geq \sum_{j=1}^{N} \psi'_j(a^*_j) \frac{x_{a_i}(a^*)}{x_{a_j}(a^*)}.$$

But the assumption that a^* is a first-best optimum implies that $\psi'_j(a^*_j) = x_{a_j}(a^*)$, which simplifies the previous inequality to $x_{a_i}(a^*) \geq N x_{a_i}(a^*)$ – a contradiction because x is strictly increasing. $\square$

Remarks:

1. Risk neutrality is not required to obtain the result in Theorem 6. But with sufficient risk aversion, we can obtain the first best even in a partnership if we consider random contracts. For example, if agents are infinitely risk averse, by randomly distributing the output to a single agent whenever $x \neq x(a^*)$, agents can be given incentives to choose the correct action. Here, randomization allows you to break the "utility" budget, even though the wage budget is satisfied. Such an idea appears in Rasmusen [1987] and Legros and Matthews [1993]

2. As indicated in the proof, it is important that x is continuous over an interval $\mathcal{A} \subset \mathbb{R}$. Relaxing this assumption may allow us to get the first best. For example, if x were informative as to who cheated, the first best could be implemented by imposing a large fine on the shirker and distributing it to the other agents.

3. It is also important for the proof that $a^* \in$ int $\mathcal{A}$. If, for example, $a^*_i = \min_{a' \in \mathcal{A}_i} \psi_i(a)$ for some i (i.e., i's efficient contribution is to shirk), then i can be made the "principal" and the first best can be implemented.

4. Legros and Matthews [1993] extend Holmström's budget-breaking argument and establish necessary and sufficient conditions for a partnership to implement the efficient set of actions.

   In particular, their conditions imply that partnerships with finite action spaces and a generic output function, $x(a)$, can implement the first best. For example, let $N = 3$, $\mathcal{A}_i = \{0, 1\}$, $\psi_i = \psi$, and $a^* = (1, 1, 1)$. Genericity of $x(a)$ implies that $x(a) \neq x(a')$ if $a \neq a'$. Genericity therefore implies that for any $x \neq x^*$, the identity of the shirker can be determined from the level of output. So letting

   $$w_i(x) = \begin{cases} \frac{1}{3}x & \text{if } x = x^*, \\ \frac{1}{2}(F + x) & \text{if } j \neq i \text{ deviated}, \\ -F & \text{if } i \text{ deviated}. \end{cases}$$

   For F sufficiently large, choosing a_i^* is a Nash equilibrium for agent i.

   Note that determining the identity of the shirker is not necessary; only the identity of a non-shirker can be determined for any $x \neq x^*$. In such a case, the non-shirker can collect sufficiently large fines from the other players so as to make a^* a Nash equilibrium. (Here, the non-shirker acts as a budget-breaking principal.)

5. Legros and Matthews [1993] also show that asymptotic efficiency can be obtained if (i) $\mathcal{A}_i \subset \mathbb{R}$, (ii) $\underline{a}_i \equiv \min_a \mathcal{A}_i$ and $\bar{a}_i \equiv \max_a \mathcal{A}_i$ exist and are finite, and (iii), $a_i^* \in (\underline{a}_i, \bar{a}_i)$.

   This result is best illustrated with the following example. Suppose that $N = 2$, $A_i \equiv [0, 2]$, $x(a) \equiv a_1 + a_2$, and $\psi_i(a_i) \equiv \frac{1}{2}a_i^2$. Here, $a^* = (1, 1)$. Consider the following strategies. Agent 2 always chooses $a_2 = a_2^* = 1$. Agent 1 randomizes over the set $\{\underline{a}_1, a_1^*, \bar{a}_1\} = \{0, 1, 2\}$ with probabilities $\{\delta, 1 - 2\delta, \delta\}$, respectively. We will construct wage schedules such that this is an equilibrium and show that δ may be made arbitrarily small.

   Note that on the equilibrium path, $x \in [1, 3]$. Use the following wage schedules for $x \in [1, 3]$: $w_1(x) = \frac{1}{2}(x-1)^2$ and $w_2(x) = x - w_1(x)$. When $x \notin [1, 3]$, set $w_1(x) = x + F$ and $w_2(x) = -F$. Clearly agent 2 will always choose $a_2 = a_2^* = 1$ providing agent 1 plays his equilibrium strategy and F is sufficiently large. But if agent 2 plays $a_2 = 1$, agent 1 obtains $U_1 = 0$ for any $a_1 \in [0, 2]$, and so the prescribed randomization strategy is optimal. Thus, we have a Nash equilibrium.

   Finally, it is easy to verify that as δ goes to zero, the first best allocation is obtained in the limit. One difficulty with this asymptotic mechanism is that the required size of the fine is $F \geq \frac{1 - 2\delta + 3\delta^2}{2\delta(2 - \delta)}$, so as $\delta \to 0$, the magnitude of the required fine explodes, $F \to \infty$. Another difficulty is that the strategies require very "unnatural" behavior by at least one of the agents.

6. When output is a stochastic function of actions, the first-best action profile may be sustained if the actions of the agents can be differentiated sufficiently and monetary transfers can be imposed as a function of output. Williams and Radner [1988] and Legros and Matsushima [1991] consider these issues.

Sufficient Statistics and Relative Performance Evaluation.

We now suppose more realistically that x is stochastic. We also assume that actions affect a vector of contractible variables, $y \in Y$, via a distribution parameterization, $F(y, a)$. Of course, we allow the possibility that x is a component of y.

Assuming that the principal gets to choose the Nash equilibrium which the agents play, the principal's problem is to

$$\max_{a,w} \int_Y \left(E[x|a, y] - \sum_{i=1}^N w_i(y) \right) f(y, a) dy,$$

subject to $(\forall\, i)$

$$\int_Y u_i(w_i(y)) f(y, a) - \psi_i(a_i) \geq \underline{U}_i,$$

$$a_i \in \arg\max_{a_i' \in \mathcal{A}_i} \int_Y u_i(w_i(y)) f(y, (a_i', a_{-i})) - \psi_i(a_i').$$

The first set of constraints are the IR constraints; the second set are the IC constraints. Note that the IC constraints imply that the agents are playing a Nash equilibrium amongst themselves. [Note: This is our first example of the principal designing a game for the agent's to play! We will see much more of this when we explore mechanism design.] Also note that when x is a component of y (e.g., $y = (x, s)$), we have $E[x|a, y] = x$.

Note that the actions of one agent may induce a distribution on y which is informative for the principal for the actions of a different agent. Thus, we need a new definition of statistical sufficiency to take account of this endogeneity.

Definition 4 $T_i(y)$ is **sufficient for y with respect to** a_i if there exists an $g_i \geq 0$ and $h_i \geq 0$ such that
$$f(y, a) \equiv h_i(y, a_{-i}) g_i(T_i(y), a), \; \forall\, (y, a) \in Y \times \mathcal{A}.$$

$T(y) \equiv \{T_i(y)\}_{i=1}^N$ is **sufficient for y with respect to** a if $T_i(y)$ is sufficient for y with respect to a_i for every agent i.

The following theorem is immediate.

Theorem 8 If $T(y)$ is sufficient for y with respect to a, then given any wage schedule, $w(y) \equiv \{w_i(y)\}_i$, there exists another wage schedule $\tilde{w}(T(y)) \equiv \{w_i(T_i(y))\}_i$ that weakly Pareto dominates $w(y)$.

Proof: Consider agent i and take the actions of all other agents as given (since we begin with a Nash equilibrium). Define $\tilde{w}_i(T_i)$ by

$$u_i(\tilde{w}_i(T_i)) \equiv \int_{\{y|T_i(y)=T_i\}} u_i(w_i(y)) \frac{f(y, a)}{g_i(T_i, a)} dy = \int_{\{y|T_i(y)=T_i\}} u_i(w_i(y)) h_i(y, a_{-i}) dy.$$

The agent's expected utility is unchanged under the new wage schedule, $\tilde{w}_i(T_i)$, and so IC and IR are unaffected. Additionally, the principal is weakly better off as $u'' < 0$, so by Jensen's inequality,

$$\tilde{w}_i(T_i) \leq \int_{\{y | T_i(y) = T_i\}} w_i(y) h_i(y, a_{-i}) dy.$$

Integrating over the set of T_i,

$$\int_Y \tilde{w}_i(T_i(y)) f(y, a) dy \leq \int_Y w_i(y) f(y, a) dy.$$

This argument can be repeated for N agents, because in equilibrium the actions of $N - 1$ agents can be taken as fixed parameters when examining the i agent. $\square$

The intuition is straightforward: we constructed $\tilde{w}$ so as not to affect incentives relative to w, but with improved risk-sharing, hence Pareto dominating w.

We would like to have the converse of this theorem as well. That is, if $T(y)$ is not sufficient, we can strictly improve welfare by using additional information in y. We need to be careful here about our statements. We want to define a notion of insufficiency that pertains for all a. Along these lines,

Definition 5 $T(y)$ is **globally sufficient** iff for all a, i, and T_i

$$\frac{f_{a_i}(y, a)}{f(y, a)} = \frac{f_{a_i}(y', a)}{f(y', a)}, \quad \text{for almost all } y, y' \in \{y \,|\, T_i(y) = T_i\}.$$

Additionally, $T(y)$ is **globally insufficient** iff for some i the above statement is false *for all a.*

Theorem 9 Assume $T(y)$ is globally insufficient for y. Let $\{w_i(y) \equiv \tilde{w}_I(T_i(y))\}_i$ be a collection of non-constant wage schedules such that the agent's choices are unique in equilibrium. Then there exist wage schedules $\hat{w}(y) = \{\hat{w}_i(y)\}_i$ that yield a strict Pareto improvement and induces the same equilibrium actions as the original $w(y)$.

The proof involves showing that otherwise the principal could do better by altering the optimal contract over the positive-measure subset of outcomes in which the condition for global sufficiency fails. See Holmström, [1982] page 332.

The above two theorems are useful for applications in agency theory. Theorem 8 says that randomization does not pay if the agent's utility function is separable; any uninformative noise should be integrated out. Conversely, Theorem 9 states that if $T(y)$ is not sufficient for y at the optimal a, we can do strictly better using information contained in y that is not in $T(y)$.

Application: Relative Performance Evaluation. An important application is *relative performance evaluation.* Let's switch back to the state-space parameterization where $\tilde{x} = x(a, \tilde{\theta})$ and $\tilde{\theta}$ is a random variable. In particular, let's suppose that the information system of the principal is rich enough so that

$$\tilde{x}(a, \theta) \equiv \sum_i x_i(a_i, \tilde{\theta}_i)$$

21

and *each* x_i is contractible. Ostensibly, we would think that each agent's wage should depend only upon its x_i. But the above two theorems suggest that this is not the case when the θ_i are not independently distributed. In such a case, the output of one agent may be informative about the effort of another. We have the following theorem along these lines.

Theorem 10 Assume the x_i's are monotone in θ_i. Then the optimal sharing rule of agent i depends on the individual i's output alone if and only if the θ_i's are independently distributed.

Proof: If the θ_i's are independent, then the parameterized distribution satisfies

$$f(x,a) = \prod_{i=1}^{N} f_i(x_i, a_i).$$

This implies that $T_i(x) = x_i$ is sufficient for x with respect to a_i. By theorem 8, it will be optimal to let w_i depend upon x_i alone.

Suppose instead that θ_1 and θ_2 are dependent but that w_1 does not depend upon x_2. Since in equilibrium a_2 can be inferred, assume that $x_2 = \theta_2$ without loss of generality and subsume a_2 in the distribution. The joint distribution of $x_2 = \theta_2$ conditional on a_1 is given by

$$f(x_1, \theta_2, a_1) = \tilde{f}(x_1^{-1}(a_1, x_1), \theta_2),$$

where $\tilde{f}(\theta_1, \theta_2)$ is the joint distribution of θ_1 and θ_2. It follows that

$$\frac{f_{a_1}(x_1, \theta_2, a_1)}{f(x_1, \theta_2, a_1)} = \frac{\tilde{f}_{\theta_1}(x_1^{-1}(a_1, x_1), \theta_2)}{\tilde{f}(x_1^{-1}(a_1, x_1), \theta_2)} \frac{\partial x_1^{-1}(a_1, x_1)}{\partial a_1}.$$

Since θ_1 and θ_2 are dependent, $\frac{\tilde{f}_{\theta_1}}{\tilde{f}}$ depends upon θ_2. Thus T is globally insufficient and theorem 9 applies, indicating that w_1 should depend upon information in x_2. $\square$

Remarks on Sufficient Statistics and Relative Performance:

1. The idea is that competition is not useful per se, but only as a way to get a more precise signal of and agent's action. With independent shocks, relative performance evaluation only adds noise and reduces welfare.

2. There is a literature on tournaments by Lazear and Rosen [1981] which indicates how basing wages on a tournament among workers can increase effort. Nalebuff and Stiglitz [1983] and Green and Stokey [1983] have made related points. But such tournaments are generally suboptimal as only with very restrictive technology will ordinal rankings be a sufficient statistic. The benefit of tournaments is that it is less susceptible to output tampering by the principal, since in any circumstance the wage bill is invariant.

3. All of this literature on teams presupposes that the principal gets to pick the equilibrium the agents will play. This may be unsatisfactory as better equilibria (from the agents' viewpoints) may exist. Mookherjee [1984] considers the multiple-equilibria problem in his examination of the multi-agent moral hazard problem and provides an illuminating example of its manifestation.

General Remarks on Teams and Partnerships:

1. Itoh [1991] has noted that sometimes a principal may want to reward agent i as a function of agent j's output, even if the outputs are independently distributed, when "teamwork" is desirable. The reason such dependence may be desirable is that agent i's effort may be a vector which contains a component that improves agent j's output. Itoh's result is not in any way at odds with our sufficient statistic results above; the only change is that efforts are multidimensional and so the principal's program is more complicated. Itoh characterizes the optimal contract in this setting.

2. It is possible that agents may get together and collude over their production and write (perhaps implicit) contracts amongst themselves. For example, some of the risk which the principal imposes for incentive reasons may be reduced by the agents via risk pooling. This obviously hurts the principal as it places an additional constraint on the optimal contract: the marginal utilities of agents will be equated across states.

   This cost of collusion has been noted by Itoh [1993]. Itoh also considers a *benefit* of collusion: when efforts are mutually observable by agents, the principal may be better off. The idea is that through the principal's choice of wage schedules, the agents can be made to police one another and increase effort through their induced side contracts. Or more precisely, the set of implementable contracts increases when agents can contract on each other's effort. Thus, collusion may (on net) be beneficial.

   The result that "collusion is beneficial to the principal" must be taken carefully, however. We know that when efforts are mutually observable by the agents there exist revelation mechanisms which allow the principal to obtain the first best as a Nash equilibrium (where agents are instructed to report shirking to the principal). We normally think that such first-best mechanisms are problematic *because of* collusion or coordination on other detrimental equilibria. It is not the collusion of Itoh's model that is beneficial – it is the mutual observation of efforts by the agents. Itoh's result may be more appropriately stated as "mutual effort observation may increase the principal's profits *even if* agents collude."

2.1.3   Extensions: A Rationale for Linear Contracts

We now briefly turn to an important paper by Holmström and Milgrom [1987] which provides an economic setting and conditions under which contracts will be linear in aggregates. This paper is fundamental both in explaining the simplicity of real world contracts and providing contract theorists with a rationalization for focusing on linear contracts. Although the model of the paper is dynamic in structure, its underlying stationarity (i.e., CARA utility and repeated environment) generates a static form:

> The optimal dynamic incentive scheme can be computed as if the agent were choosing the mean of a normal distribution only once and the principal were restricted to offering a linear contract.

We thus consider Holmström and Milgrom's [1987] contribution here as an examination of static contracts rather than dynamic contracts.

One-period Model:

There are $N + 1$ possible outcomes: $x_i \in \{x_0, \ldots, X_N\}$, with probability of occurrence given by the vector $p = (p_0, \ldots, p_N)$. We assume that the agent directly chooses $p \in \Delta(N)$ at a cost of $c(p)$. The principal offers a contract $w(x_i) = \{w_0, \ldots, w_N\}$ as a function of outcomes. Both principal and agent have exponential utility functions (to avoid problems of wealth effects).

$$U(w - c(p)) \equiv -e^{-r(w-c(p))},$$

$$V(x - w) \equiv \begin{cases} -e^{-R(x-w)} & \text{if } R > 0, \\ x - w & \text{if } R = 0. \end{cases}$$

Assume that $R = 0$ for now. The principal solves

$$\max_{w,p} \sum_{i=0}^{N} p_i(x_i - w_i) \quad \text{subject to}$$

$$p \in \arg\max_{p} \sum_{i=0}^{N} p_i U(w - c(p)),$$

$$\sum_{i=0}^{N} p_i U(w - c(p)) \geq U(\underline{w}),$$

where $\underline{w}$ is the certainty equivalent of the agent's outside opportunity. (I will use generally underlined variables to represent certainty equivalents.)

Given our assumption of exponential utility, we have the following result immediately.

Theorem 11 Suppose that (w^*, p^*) solves the principal's one-period program for some $\underline{w}$. Then $(w^* + \underline{w}' - \underline{w}, p^*)$ solves the program for an outside certainty equivalent of $\underline{w}'$.

Proof: Because utility is exponential,

$$\sum_{i=0}^{N} p_i U(w^*(x_i) - c(p^*)) = -U(-\underline{w} + \underline{w}') \sum_{i=0}^{N} p_i U(w^*(x_i) + \underline{w}' - \underline{w}).$$

Thus, p^* is still incentive compatible and the IR constraint is satisfied for $U(\underline{w}')$. Similarly, given the principal's utility is exponential, the optimal choice of p^* is unchanged. $\square$

The key here is that there are absolutely no wealth effects in this model. This will be an important ingredient in our proofs below.

T-period Model:

Now consider the multi-period problem where the agent chooses a probability each period after having observed the history of outputs up until that time. Let superscripts

denote histories of variables; i.e., $X^t = \{x_0, \ldots, x_t\}$. The agent gets paid at the end of the period, $w(X^t)$ and has a combined cost of effort equal to $\sum_t c(p_t)$. Thus,

$$U(w, \{p_t\}_t) = -e^{-r\left(w - \sum_t c(p_t)\right)}.$$

Because the agent observes X^{t-1} before deciding upon p_t, for a given wage schedule we can write $p_t(X^{t-1})$. We want to first characterize the wage schedule which implements an arbitrary $\{p_t(X^{t-1})\}_t$ effort function. We use dynamic programming to this end.

Let $\mathcal{U}_t$ be the agent's expected utility (ignoring past effort costs) from date t forward. Thus,

$$\mathcal{U}_t(X^t) \equiv E\left[U\left(w(X^T) - \sum_{\tau=t+1}^{T} c(p_\tau)\right) | X^t\right].$$

Note here that $\mathcal{U}_t$ differs from a standard value function by the constant $U(-\sum_t c(p_t))$. Let $\underline{w}_t(X^t)$ be the certain equivalent of income of $\mathcal{U}_t$. That is, $U(\underline{w}_t(X^t)) \equiv \mathcal{U}_t(X^t)$. Note that $\underline{w}_t(X^{t-1}, x_{it})$ is the certain equivalent for obtaining output x_i in period t following a history of X^{t-1}.

To implement $p_t(X^{t-1})$, it must be the case that

$$p_t(X^{t-1}) \in \arg\max_{p_t} \sum_{i=0}^{N} p_{it} U(\underline{w}_t(X^{t-1}, x_{it}) - c(p_t)),$$

where we have dropped the irrelevant multiplicative constant.

Our previous theorem 11 applies: $p_t(X^{t-1})$ is implementable and yields certainty equivalent $\underline{w}_{t-1}(X^{t-1})$ iff $p_t(X^{t-1})$ is also implemented by

$$\tilde{w}_t(x_{it}|p_t(X^{t-1})) \equiv \underline{w}_t(X^{t-1}, x_{it}) - \underline{w}_{t-1}(X^{t-1})$$

with a certainty equivalent of $\underline{w} = 0$.

Rearranging the above relationship,

$$\underline{w}_t(X^{t-1}, x_{it}) = \tilde{w}_t(x_{it}|p_t(X^{t-1})) + \underline{w}_{t-1}(X^{t-1}).$$

Integrating this difference equation from $t = 0$ to T yields

$$w(X^T) \equiv \underline{w}_T(X^T) = \sum_{t=0}^{T} \tilde{w}_t(x_{it}|p_t(X^{t-1})) + \underline{w}_0,$$

or in other words, the end of contract wage is the sum of the individual single-period wage schedules for implementing $p_t(X^{t-1})$.

Let $\tilde{w}_t(p_t(X^{t-1}))$ be an $N + 1$ vector over i. Then rewriting,

$$w(X^T) = \sum_{t=1}^{T} \tilde{w}_t(p_t(X^{t-1})) \cdot (A^t - A^{t-1}) + \underline{w}_0,$$

where $A^t = (A_0^t, \ldots, A_N^t)$ and A_i^t is an account that gives the number of times outcome i has occurred up to date t.

We thus have characterized a wage schedule, $w(X^T)$, for implementing $p_t(X^{t-1})$. Moreover, Holmström and Milgrom show that if c is differentiable and $p_t \in \text{int}\Delta(N)$, such a wage schedule is uniquely defined. We now wish to find the optimal contract.

Theorem 12 The optimal contract is to implement $p_t(X^{t-1}) = p^* \ \forall \ t$ and offer the wage schedule

$$w(X^T) = \sum_{t=1}^{T} w(x_t, p^*) = w(p^*) \cdot A^T.$$

Proof: By induction. The theorem is true by definition for $T = 1$. Suppose that it holds for $T = \tau$ and consider $T = \tau + 1$. Let $\mathcal{V}_T^*$ be the principal's value function for the T-period problem. The value of the contract to the principal is

$$- e^{R\underline{w}_0} E\left[V(x_{t=1} - w_{t=1}) E\left[V\left(\sum_{t=2}^{\tau+1}(x_t - w_t) \right) | X^1 \right] \right]$$

$$\leq -e^{R\underline{w}_0} E[V(x_{t=1} - w_{t=1})\mathcal{V}_\tau^*] \leq -e^{R\underline{w}_0} \mathcal{V}_1^* \mathcal{V}_\tau^*.$$

At $p_t = p^*$, $w_t = w(x_t, p^*)$, this upper bound is met. $\square$

Remarks:

1. Note very importantly that the optimal contract is *linear in accounts*. Specifically,

$$w(X^T) = \sum_{t=1}^{T} w(x_t, p^*) = \sum_{i-0}^{N} w(x_i, p^*) \cdot A_i^T,$$

   or letting $\alpha_i \equiv w(x_i, p^*) - w(x_0, p^*)$ and $\beta \equiv T - w(x_0, p^*)$,

$$w(X^T) = \sum_{i=1}^{N} \alpha_i A_i^T + \beta.$$

   This is not generally linear in profits. Nonetheless, many applied economists typically take Holmström and Milgrom's result to mean linearity in profits for the purposes of their applications.

2. If there are only two accounts, such as success or failure, then wages are linear in "profits" (i.e., successes). From above we have

$$w(X^T) = \alpha A_1^T + \beta.$$

   Not surprisingly, when we take the limit as this binomial process converges to unidimensional Brownian motion, we preserve our linearity in profits result. With more than two accounts, this is not so. Getting an output of 50 three times is not the same as getting the output of 150 once and 0 twice.

3. Note that the history of accounts is irrelevant. Only total instances of outputs are important. This is also true in the continuous case. Thus, A^T is "sufficient" with respect to X^T. This is not inconsistent with Holmström [1979] and Shavell [1979]. Sufficiency notions should be thought of as sufficient information regarding the binding constraints. Here, the binding constraint is shifting to another constant action, for which A^T is sufficient.

26

4. The key to our results are stationarity which in turn is due exclusively to time-separable CARA utility and an i.i.d. stochastic process.

Continuous Model:

We now consider the limit as the time periods become infinitesimal. We now want to ask what happens if the agent can continuous vary his effort level and observe the realizations of output in real time.

Results:

1. In the limit, we obtain a linearity in accounts result, where the accounts are movements in the stochastic process. With unidimensional Brownian motion, (i.e., the agent controls the drift rate on a one-dimensional Brownian motion process), we obtain linearity in profits.

2. Additionally, in the limit, if only a subset of accounts can be contracted upon (specifically, a linear aggregate), then the optimal contract will be linear in those accounts. Thus, if only profits are contractible, we will obtain the linearity in profits result in the limit – even when the underlying process is multinomial Brownian motion. This does not happen in the discrete case. The intuition roughly is that in the limit, information is lost in the aggregation process, while in the discrete case, this is not the case.

3. If the agent must take all of his actions simultaneously at $t = 0$, then our results do not hold. Instead, we are in the world of static nonlinear contracts. In a continuum, Mirrlees's example would apply, and we could obtain the first best.

The Simple Analytics of Linear Contracts:

To see the usefulness of Holmström and Milgrom's [1987] setting for simple comparative statics, consider the following model. The agent has exponential utility with a CARA parameter of r; the principal is risk neutral. Profits (excluding wages) are $x = \mu + \varepsilon$, where μ is the agent's action choice (the drift rate of a unidimensional Brownian process) and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. The cost of effort is $c(\mu) = \frac{k}{2}\mu^2$.

Under the full information first-best contract, $\mu^{FB} = \frac{1}{k}$, the agent is paid a constant wage to cover the cost of effort, $w^{FB} = \frac{1}{2k}$, and the principal receives net profits of $\pi = \frac{1}{2k}$.

When effort is not contractible, Holmström and Milgrom's linearity result tells us that we can restrict attention to wage schedules of the form $w(x) = \alpha x + \beta$. With this contract,

the agent's certainty equivalent[2] upon choosing an action μ is

$$\alpha\mu + \beta - \frac{k}{2}\mu^2 - \frac{r}{2}\alpha^2\sigma^2.$$

The first-order condition is $\alpha = \mu k$ which is necessary and sufficient because the agent's utility function is globally concave in μ.

It is very important to note that the utilities possibility frontier for the principal and agent is *linear* for a given (α, μ) and independent of β. The independence of β is an artifact of CARA utility (that's our result from Theorem refce above), and the linearity is due to the combination of CARA utility and normally distributed errors (the latter of which is due to the central limit theorem).

As a consequence, the principal's optimal choice of (α, μ) is independent of β; β is chosen solely to satisfy the agent's IR constraint. Thus, the principal solves

$$\max_{\alpha,\mu} \quad \mu - \frac{k}{2}\mu^2 - \frac{r}{2}\alpha^2\sigma^2,$$

subject to $\alpha = \mu k$. The solution gives us (α^*, μ^*, π^*):

$$\alpha^* = (1 + rk\sigma^2)^{-1},$$

$$\mu^* = (1 + rk\sigma^2)^{-1}k^{-1} = \alpha^*\mu^{FB} < \mu^{FB},$$

$$\pi^* = (1 + rk\sigma^2)^{-1}(2k)^{-1} = \alpha^*\pi^{FB} < \pi^{FB}.$$

The simple comparative statics are immediate. As either r, k, or σ^2 decrease, the power of the optimal incentive scheme increases (i.e., α^* increases). Because α^* increases, effort and profits also increase closer toward the first best. Thus when risk aversion, the uncertainty in measuring effort, or the curvature of the agent's effort function decrease, we move toward the first best. The intuition for why the curvature of the agent's cost function matters can be seen by totally differentiating the agent's first-order condition for effort. Doing so, we find that $\frac{d\mu}{d\alpha} = \frac{1}{C''(\mu)} = \frac{1}{k}$. Thus, lowering k makes the agent's effort choice more responsive to a change in α.

Remarks:

1. Consider the case of additional information. The principal observes an additional signal, y, which is correlated with ε. Specifically, $E[y] = 0$, $V[y] = \sigma_y^2$, and $Cov[\varepsilon, y] = \rho\sigma_y\sigma_e$. The optimal wage contract is linear in both aggregates: $w(x, y) = \alpha_1 x + \alpha_2 y + \beta$. Solving for the optimal schemes, we have

$$\alpha_1^* = (1 + rk\sigma_\varepsilon^2(1 - \rho^2))^{-1},$$

---

[2] Note that the moment generating function for a normal distribution is $M_x(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ and the defining property of the m.g.f. is that $E_x[e^{tx}] = M_x(t)$. Thus,

$$E_\varepsilon[e^{-r(\alpha\mu + \alpha\varepsilon + \beta - C(\mu)))}] = e^{-r(\alpha\mu + \beta - C(\mu)) + \frac{1}{2}\alpha^2 r^2\sigma^2}.$$

Thus, the agent's certainty equivalent is $\alpha\mu + \beta - C(\mu) - \frac{r}{2}\alpha^2\sigma^2$.

$$\alpha_2^* = -\alpha_1 \frac{\sigma_y}{\sigma_\varepsilon} \rho.$$

As before, $\mu^* = \alpha_1^* \mu^{FB}$ and $\pi^* = \alpha_1^* \pi^{FB}$. It is as if the outside signal reduces the variance on ε from σ_ε^2 to $\sigma_\varepsilon^2(1 - \rho^2)$. When either $\rho = 1$ or $\rho = -1$, the first-best is obtainable.

2. The allocation of effort across tasks may be greatly influenced by the nature of information. To see this, consider a symmetric formulation with two tasks: $x_1 = \mu_1 + \varepsilon_1$ and $x_2 = \mu_2 + \varepsilon_2$, where $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ and are independently distributed across i. Suppose also that $C(\mu) = \frac{1}{2}\mu_1^2 + \frac{1}{2}\mu_2^2$ and the principal's net profits are $\pi = x_1 + x_2 - w$. If only $x = x_1 + x_2$ were observed, then the optimal contract has $w(x) = \alpha x + \beta$, and the agent would equally devote his attention across tasks. Additionally, if $\sigma_1 = \sigma_2$ and the principal can contract on both x_1 and x_2, the optimal contract has $\alpha_1 = \alpha_2$ and so again the agent equally allocates effort across tasks.

    Now suppose that $\sigma_1 < \sigma_2$. The resulting first-order conditions imply that $\alpha_1^* > \alpha_2^*$. Thus, optimal effort allocation may be entirely determined by the information structure of the contracting environment. the intuition here is that the "price" of inducing effort on task 1 is lower for the principal because information is more informative. Thus, the principal will "buy" more effort from the agent on task 1 than task 2.

2.1.4 Extensions: Multi-task Incentive Contracts

We now consider more explicitly the implications of multiple tasks within a firm using the linear contracting model of Holmström and Milgrom [1987]. This analysis closely follows Holmström and Milgrom [1991].

The Basic Linear Model with Multiple Tasks:

The principal can contract on the following k vector of aggregates:

$$x = \mu + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \Sigma)$. The agent chooses a vector of efforts, μ, at a cost of $C(\mu)$.[3] The agent's utility is exponential with CARA parameter of r. The principal is risk neutral, offers wage schedule $w(x) = \alpha'x + \beta$, and obtains profits of $B(\mu) - w$. [Note, α and μ are vectors; $B(\mu)$, β and $w(x)$ are scalars.]

As before, CARA utility and normal errors implies that the optimal contract solves

$$\max_{\alpha,\mu} \ B(\mu) - C(\mu) - \frac{r}{2}\alpha'\Sigma\alpha,$$

such that

$$\mu \in \arg\max_{\tilde{\mu}} \ \alpha'\tilde{\mu} - C(\tilde{\mu}).$$

Given the optimal (α, μ), β is determined so as to meet the agent's IR constraint:

$$\beta = \underline{w} - \alpha'\mu + C(\mu) + \frac{r}{2}\alpha'\Sigma\alpha.$$

---

[3]Note that Holmström and Milgrom [1991] take the action vector to be t where $\mu(t)$ is determined by the action. We'll concentrate on the choice of μ as the primitive.

The agent's first-order condition (which is both necessary and sufficient) satisfies

$$\alpha_i = C_i(\mu), \ \ \forall \, i,$$

where subscripts on C denote partial derivatives with respect to the indicated element of μ. Comparative statics on this equation reveal that

$$\left[\frac{\partial \mu}{\partial \alpha} \right] = [C_{ij}(\mu)]^{-1}.$$

This implies that in the simple setting where $C_{ij} = 0 \ \forall \, i \neq j$, that $\frac{d\mu_i}{d\alpha_i} = \frac{1}{C_{ii}(\mu)}$. Thus, the marginal affect of a change in α on effort is inversely related to the curvature of the agent's cost of effort function.

We have the following theorem immediately.

Theorem 13 The optimal contract satisfies

$$\alpha^* = (I + r[C_{ij}(\mu^*)]\Sigma)^{-1} \, B'(\mu^*).$$

Proof: Form the Lagrangian:

$$\mathcal{L} \equiv B(\mu) - C(\mu) - \frac{r}{2}\alpha'\Sigma\alpha + \lambda'(\alpha - C'(\mu)),$$

where $C'(\mu) = [C_i(\mu)]$. The $2k$ first-order conditions are

$$B'(\mu^*) - C'(\mu^*) - \lambda[C_{ij}(\mu^*)] = 0,$$

$$-r\Sigma\alpha^* + \lambda = 0.$$

Substituting out λ and solving for α^* produces the desired result. $\square$

Remarks:

1. If ε_i are independent and $C_{ij} = 0$ for $i \neq j$, then

$$\alpha_i^* = B_i(\mu^*)(1 + rC_{ii}(\mu^*)\sigma_i^2)^{-1}.$$

   As r, σ_i, or C_{ii} decrease, α_i^* increases. This result was found above in our simple setting of one task.

2. Given μ^*, the cross partial derivatives of B are unimportant for the determination of α^*. Only cross partials in the agent's utility function are important (i.e., C_{ij}).

Simple Interactions of Multiple Tasks:

Consider the setting where there are two tasks, but where the effort of only the first task can be measured: $\sigma_2 = \infty$ and $\sigma_{12} = 0$. A motivating example is a teacher who teaches basic skills (task 1) which is measurable via student testing and higher-order skills such as

30

creativity, etc. (task 2) which is inherently unmeasureable. The question is how we want to reward the teacher on the basis of basic skill test scores.

Suppose that under the optimal contract $\mu^* > 0$; that is, both tasks will be provided at the optimum.[4] Then the optimal contract satisfies $\alpha_2^* = 0$ and

$$\alpha_1^* = \left(B_1(\mu^*) - B_2(\mu^*) \frac{C_{12}(\mu^*)}{C_{22}(\mu^*)} \right) \left(1 + r\sigma_1^2 \left(C_{11}(\mu^*) - \frac{C_{12}(\mu^*)^2}{C_{22}(\mu^*)} \right) \right)^{-1}.$$

Some interesting conclusions emerge.

1. If effort levels across tasks are substitutes (i.e., $C_{12} > 0$), the more positive the cross-effort effect (i.e., more substitutable the effort levels), the lower is α_1^*. (In our example, if a teacher only has 8 hours a day to teach, the optimal scheme will put less emphasis on basic skills the more likely the teacher is to substitute away from higher-order teaching.) If effort levels are complements, the reverse is true, and α_1^* is increased.

2. The above result has a flavor of the public finance results that when the government can only tax a subset of goods, it should tax them more or less depending upon whether the taxable goods are substitutes or complements with the un-taxable goods. See, for example, Atkinson and Stiglitz [1980 Ch. 12] for a discussion concerning taxation on consumption goods when leisure is not directly taxable.

3. There are several reasons why the optimal contract may have $\alpha_1^* = 0$.

   - Note that in our example, $\alpha_1^* < 0$ if $B_1 < B_2 C_{12}/C_{22}$. Thus, if the agent can freely dispose of x_1, the optimal constrained contract has $\alpha_1^* = 0$. No incentives are provided.

   - Suppose that technologies are otherwise symmetric: $C(\mu) = c(\mu_1 + \mu_2)$ and $B(\mu_1, \mu_2) \equiv B(\mu_2, \mu_1)$. Then $\alpha_1^* = \alpha_2^* = 0$. Again, no incentives are provided.

   - Note that if $C_i(0) > 0$, there is a fixed cost to effort. This implies that a corner solution may emerge where $\alpha_i^* = 0$. A final reason for no incentives.

Application: Limits on Outside Activities.

Consider the principal's problem when an additional control variable is added: the set of allowable activities. Suppose that the principal cares only about effort devoted to task 0: $\pi = \mu_0 - w$. In addition, there are N potential tasks which the agent could spend effort on and which increase the agent's personal utility. We will denote the set of these tasks by $K = \{1, \ldots, N\}$. The principal has the ability to exclude the agent from any subset of these activities, allowing only tasks or activities in the subset set $A \subset K$. Unfortunately, the principal can only contract over x_0, and so $w(x) = \alpha x_0 + \beta$.

It is not always profitable, even in the full-information setting, to exclude these tasks from the agent, because they may be efficient and therefore reduce the principal's wage

---

[4]Here we need to assume something like $C_2(\mu_1, \mu_2) < 0$ for $\mu_2 \leq 0$ so that without any incentives on task 2, the agent still allocates some effort on task 2. In the teaching example, absent incentives, a teacher will still teach some higher-order skills.

bill. As a motivating example, allowing an employee to make personal calls on the company WATTS line may be a cheap perk for the firm to provide and additionally lowers the necessary wage which the firm must pay. Unfortunately, the agent may then spend all of the day on the telephone rather than at work.

Suppose that the agent's cost of effort is

$$C(\mu) = c\left(\mu_0 + \sum_{i=1}^{N} \mu_i\right) - \sum_{i=1}^{N} u_i(\mu_i).$$

The u_i functions represent the agent's personal utility from allocating effort to task i; u_i is assumed to be strictly concave and $u_i(0) = 0$. The principal's expected returns are simply $B(\mu) = p\mu_0$.

We first determine the principal's optimal choice of $A^*(\alpha)$ for a given α, and then we solve for the optimal α^*. The first-order condition which characterizes the agent's optimal μ_0 is

$$\alpha = c'\left(\sum_{i=0}^{N} \mu_i\right),$$

and (substituting)

$$\alpha = v_i'(\mu_i), \quad \forall\, i.$$

Note that the choice of μ_i depends only upon α. Thus, if the agent is allowed an additional personal task, k, the agent will allocate time away from task 0 by an amount equal to $v_k^{-1'}(\alpha)$. The benefit of allowing the agent to spend time on task k is $v_k(\mu_k(\alpha))$ (via a reduced wage) and the (opportunity) cost is $p\mu_k(\alpha)$. Therefore, the optimal set of tasks for a given α is

$$A^*(\alpha) = \{k \in K | v_k(\mu_k(\alpha)) > p\mu_k(\alpha)\}.$$

We have the following results for a given α.

Theorem 14 Assume that α is such that $\mu(\alpha) > 0$ and $\alpha < p$. Then the optimal set of allowed tasks is given by $A^*(\alpha)$ which is monotonically expanding in α (i.e., $\alpha \leq \alpha'$, then $A^*(\alpha) \subset A^*(\alpha')$).

Proof: That the optimal set of allowed tasks is given by $A^*(\alpha)$ is true by construction. The set $A^*(\alpha)$ is monotonically expanding in α iff $v_k(\mu_k(\alpha)) - p\mu_k(\alpha)$ is increasing in α. I.e.,

$$[v_k'(\mu_k(\alpha)) - p]\frac{d\mu_k(\alpha)}{d\alpha} = [\alpha - p]\frac{1}{v_k''(\mu_k(\alpha))} > 0.$$

$\square$

Remarks:

1. The fundamental premise of exclusion is that incentives can be given by either increasing α on the relevant activity or decreasing the opportunity cost of effort (i.e, by reducing the benefits of substitutable activities).

2. The theorem indicates a basic proposition with direct empirical content: *responsibility (large α) and authority (large $A^*(\alpha)$) should go hand in hand.* An agent with high-powered incentives should be allowed the opportunity to expend effort on more personal activities than someone with low-powered incentives. In the limit when $\sigma \to 0$ or $r \to 0$, the agent is residual claimant $\alpha^* = 1$, and so $A^*(1) = K$. Exclusion will be more frequently used the more costly it is to supply incentives.

3. Note that for α small enough, $\mu(\alpha) = 0$, and the agent is not hired.

4. The set $A^*(\alpha)$ is independent of r, σ, C, etc. These variables only influence $A^*(\alpha)$ via α. Therefore, an econometrician can regress $||A^*(\alpha)||$ on α, and α on $(r, \sigma, \ldots)$ to test the multi-task theory. See Holmström and Milgrom [1994].

Now, consider the choice of α^* given the function $A^*(\alpha)$.

Theorem 15 Providing that $\mu(\alpha^*) > 0$ at the optimum,

$$\alpha^* = p \left(1 + r\sigma^2 \left(\frac{1}{c''\left(\sum_i \mu_i(\alpha^*)\right)} + \sum_{k \in A^*(\alpha^*)} \frac{1}{v_k''(\mu_k(\alpha^*))} \right) \right)^{-1}.$$

The proof of this theorem is an immediate application of our first multi-task characterization theorem. Additionally, we have the following implications.

Remarks:

1. The theorem indicates that when either r or σ decreases, α^* increases. [Note that this implication is not immediate because σ^* appears on both sides of the equation; some manipulation is required. With quadratic cost and benefit functions, this is trivial.] By our previous result on $A^*(\alpha)$, the set of allowable activities also increases as α^* increases.

2. Any personal task excluded in the first-best arrangement (i.e., $v_k'(0) < p$) will be excluded in the second-best optimal contract given our construction of $A^*(\alpha)$ and the fact that v_k is concave. This implies that there will be more constraints on agent's activities when performance rewards are weak due to a noisy environment.

3. Following the previous remark, one can motivate rigid rules which limit an agent's activities (seemingly inefficiently) as a way of dealing with substitution possibilities. Additionally, when the "personal" activity is something such as rent-seeking (e.g., inefficiently spending resources on your boss to increase your chance of promotion), a firm may wish to restrict an agent's access to such an activity or withdraw the bosses discretion to promote employees so as to reduce this inefficient activity. This idea was formalized by Milgrom [1988] and Milgrom and Roberts [1988].

4. This activity exclusion idea can also explain why firms may not want to allow their employees to "moonlight". Or more importantly, why a firm may wish to use an

internal sales force which is not allowed to sell other firms' products rather than an external sales force whose activities vis-a-vis other firms cannot be controlled.

Application: Task Allocation Between Two Agents.

Now consider two agents, $i = 1, 2$, who are needed to perform a continuum of tasks indexed by $t \in [0, 1]$. Each agent i expends effort $\mu_i(t)$ on task t; total cost of effort is $C\left(\int \mu_i(t)dt\right)$. The principal observes $x(t) = \mu(t) + \varepsilon(t)$ for each task, where $\sigma^2(t) > 0$ and $\mu(t) \equiv \mu_1(t) + \mu_2(t)$. The wages paid to the agents are given by:

$$w_i(x) = \int_0^1 \alpha_i(t)x(t)dt + \beta_i.$$

By choosing $\alpha_i(t)$, the principal allocates agents to the various tasks. For example, when $\alpha_1(.4) > 0$ but $\alpha_2(.4) = 0$, only agent 1 will work on task .4.

Two results emerge.

1. For any required effort function $\mu(t)$ defined on $[0, 1]$, it is never optimal to assign two agents to the same task: $\alpha_1^*(t)\alpha_2^*(t) \equiv 0$. This is quite natural given the teams problem which would otherwise emerge.

2. More surprisingly, suppose that the principal must obtain a uniform level of effort $\mu(t) = 1$ across all tasks. At the optimum, if $\int \mu_i(t)dt < \int \mu_j(t)dt$, then the hardest to measure tasks go to agent i (i.e., all tasks t such that $\sigma(t) \geq \bar{\sigma}$.) This results because you want to avoid the multi-task problems which occur when the various tasks have vastly different measurement errors. Thus, the principal wants information homogeneity. Additionally, the agent with the hard to measure tasks exerts lower effort and receives a lower "normalized commission" because the information structure is so noisy.

Application: Common Agency. Bernheim and Whinston [1986] were to first to undertake a detailed study of the phenomena of common agency with moral hazard. "Common agency" refers to the situation in which several principals contract with the same agent in common. The interesting economics of this setting arise when one principal's contract imposes an externality on the contracts of the others.

Here, we follow Dixit [1996] restricting attention to linear contracts in the simplest setting of n independent principals who simultaneously offer incentive contracts to a single agent who controls the m-dimensional vector t which in turn effects the output vector $x \in \mathbb{R}^m$. Let $x = t + \varepsilon$, where $x, t \in \mathbb{R}^m$ and $\varepsilon \in \mathbb{R}^m$ is distributed normally with mean vector 0 and covariance matrix Σ. Cost of effort is a quadratic form with a positive definite matrix, C.

1. The first-best contract (assuming t can be contracted upon) is simply $t = C^{-1}b$.

2. The second-best *cooperative* contract. The combined return to the principals from effort vector t is $b't$, and so the total expected surplus is

$$b't - \frac{1}{2}t'Ct - \frac{r}{2}\alpha'\Sigma\alpha.$$

This is maximized subject to $t = C^{-1}\alpha$. The first-order condition for the slope of the incentive contract, α, is

$$C^{-1}b - [C^{-1} + r\Sigma]\alpha = 0,$$

or $b = [I + rC\Sigma]\alpha$ or $b - \alpha = rC\Sigma\alpha > 0$.

3. The second-best *non-cooperative* un-restricted contract. Each principal's return is given by the vector b^j, where $b = \sum b^j$. Suppose each principal j is unrestricted in choosing its wage contract; i.e., $w^j = \alpha^{j'} + \beta^j$, where α^j is a full m-dimensional vector. Define $A^{-j} \equiv \sum_{i\neq j}\alpha^i$ and $B^{-j} \equiv \sum_{i\neq j}\beta^i$. From principal j's point of view, absent any contract from himself $t = C^{-1}A^{-j}$ and the certainty equivalent is $\frac{1}{2}A^{-j'}[C^{-1} - r\Sigma]A^{-j} + B^{-j}$. The aggregate incentive scheme facing the agent is $\alpha = A^{-j} + \alpha^j$ and $\beta = B^{-j} + \beta^j$. Thus the agent's certainty equivalent *with* principal j's contract is

$$\frac{1}{2}(A^{-j} + \alpha^j)'[C^{-1} - r\Sigma](A^{-j} + \alpha^j) + B^{-j} + \beta^j.$$

The incremental surplus to the agent from the contract is therefore

$$A^{-j'}(C^{-1} - r\Sigma)\alpha^j + \frac{1}{2}\alpha^{j'}[C^{-1} - r\Sigma]\alpha^j + \beta^j.$$

As such, principal j maximizes

$$b^{j'}C^{-1}A^{-j} - rA^{-j'}\Sigma\alpha^j + b^{j'}C^{-1}\alpha^j - \frac{1}{2}\alpha^{j'}[C^{-1} + r\Sigma]\alpha^j.$$

The first-order condition is

$$C^{-1}b^j - [C^{-1} + r\Sigma]\alpha^j - r\Sigma A^{-j} = 0.$$

Simplifying, $b^j = [I + rC\Sigma]\alpha^j + rC\Sigma A^{-j}$. Summing across all principals,

$$b = [I + rC\Sigma]\alpha + rC\Sigma(n - 1)\alpha = [I + nrC\Sigma]\alpha,$$

or $b - \alpha = nrC\Sigma\alpha > 0$. Thus, the distortion has increased by a factor of n. Intuitively, it is as if the agent's risk has increased by a factor of n, and so therefore incentives will be reduced on every margin. *Hence, unrestricted common agency leads to more effort distortions.*

Note that $b^j = \alpha^j - rC\Sigma\alpha$, so substitution provides

$$\alpha^j = b^j - rC\Sigma[I + nrC\Sigma]^{-1}b.$$

To get some intuition for the increased-distortion result, suppose that $n = m$ and that each principal cares only about output j; i.e., $b_i^j = 0$ for $i \neq j$, and $b_j^j > 0$. In such a case,

$$\alpha_i^j = -rC\Sigma[I + nrC\Sigma]^{-1}b < 0,$$

so each principal finds it optimal to pay the agent not to produce on the other dimensions!

4. The second-best *non-cooperative restricted* contract. We now consider the case in which each principal is restricted in its contract offerings so as to not pay the agent for for output on the other principals' dimensions. Specifically, let's again assume that $n = m$ and that each principal only cares about x_j: $b_i^j = 0$ for $i \neq j$, and $b_j^j > 0$. The restriction is that $\alpha_i^j = 0$ for $j \neq i$. In such a setting, Dixit demonstrates that the equilibrium incentives are higher than in the un-restricted case. Moreover, if efforts are perfect substitutes across agents, $\alpha^j = b^j$ and first-best efforts are implemented.

2.2 Dynamic Principal-Agent Moral Hazard Models

There are at least three sets of interesting questions which emerge when one turns attention to dynamic settings.

1. Can efficiency be improved in a long-run relationship with full commitment long-term contract?

2. When can short-term contracts perform as well as long-term contracts?

3. What is the effect of renegotiation (i.e., lack of commitment) between the time when the agent takes an action and the uncertainty of nature is revealed?

We consider each issue in turn.

2.2.1 Efficiency and Long-Run Relationships

We first focus on the situation in which the principal can commit to a long-run contractual relationship. Consider a simple model of repeated moral hazard where the agent takes an action, a_t, in each period, and the principal pays the agent a wage, $w_t(x^t)$, based upon the history of outputs, $x^t \equiv \{x_1, \ldots, x_t\}$.

There are two standard ways in which the agent's intertemporal preferences are modeled. First, there is time-averaging.

$$U = \frac{1}{T} \sum_{t=1}^{T} \left(u(w_t) - \psi(a_t) \right).$$

Alternatively, there is the discounted representation.

$$U = (1 - \delta) \sum_{t=1}^{T} \delta^{t-1} (u(w_t) - \psi(a_t)).$$

Under both models, it has been shown that repeated moral hazard relationships will achieve the first-best arbitrarily close as either $T \to \infty$ (in the time-averaging case) or $\delta \to 1$ (in the discounting case).

Radner [1985] shows the first-best can be approximated as T becomes large using the weak law of large numbers. Effectively, as the time horizon grows large, the principal observes the realized distribution and can punish the agent severally enough for discrepancies to prevent shirking.

Fudenberg, Holmström, and Milgrom [1990] and Fudenberg, Levine, and Maskin [1994], and others have shown that in the discounted case, as δ approaches 1, the first best is closely approximated. The intuition for their approach is that when the agent can save in the capital market, the agent is willing to become residual claimant for the firm and will smooth income across time. Although this result uses the agent's ability to use the capital market in its proof, the first best can be approximately achieved with short-term contracts (i.e., $w_t(x_t)$ and not $w_t(x^t)$). See remarks below.

Abreu, Milgrom and Pearce [1991] study the repeated partnership problem in which agents play trigger strategies as a function of past output. They produce some standard folk-theorem results with a few interesting findings. Among other things, they show that taking the limit as r goes to zero is not the same as taking the limit as the length of each period goes to zero, as the latter has a negative information effect. They also show that there is a fundamental difference between information that is good news (i.e., sales, etc.) and information that is bad news (i.e., accidents). Providing information is generated by a Poisson process, a series of unfavorable events is much more informative about shirking when news is "the number of bad outcomes" (e.g., a high number of failures) then when news is "the number of good outcomes" (e.g., a low number of successes). This latter result has to do with the likelihood function generated from a Poisson process. It is not clear that it generalizes.

2.2.2 Short-term versus Long-term Contracts

Although we have seen that when a relationship is repeated infinitely and the discount factor is close to 1 that we can achieve the first-best, what is the structure of long-term contracts when asymptotic efficiency is not attainable? Do short-run contracts perform worse than long-run contracts?

Lambert [1983] and Rogerson [1985] consider the setting in which a principal can freely utilize capital markets at the interest rate of r, *but agents have no access*. This is fundamental as we will see. In this situation, long-term contracts play a role. Just as cross-state insurance is sacrificed for incentives, intertemporal insurance is also less than first-best efficient. Wage contracts have memory (i.e., today's wage schedule depends upon yesterday's output) in order to reduce the incentive problem of the agent. The agent is generally left with a desire to use the credit markets so as to self-insure across time.

We follow Rogerson's [1985] model here. Let there be two periods, $T = 2$, and a finite set of outcomes, $\{x_1, \ldots, x_N\} \equiv \mathcal{X}$, each of which have a probability $f(x_i, a) > 0$ of occurring during a period in which action a was taken. The principal offers a long-term contract of the form $w \equiv \{w_1(x_i), w_2(x_i, x_j)\}$, where the wage subscript denotes the period in which the wage schedule is in affect and the output subscripts denote the realized output. (The first argument of w_2 is the first period output; the second argument is the second period output.) An agent either accepts or rejects this contract for the duration of the relationship. If accepted, the agent chooses an action in period 1, a_1. After observing the period 1 outcome, the agent takes the period 2 action, $a_2(x_i)$, that is optimal given the wage schedule in operation. Let $a \equiv \{a_1, a_2(\cdot)\}$ denote a temporally optimal strategy by the agent for a given wage structure, w. The principal and the agent both discount the future at rate $\delta = \frac{1}{1+r}$. (Identical discount factors is not important for the memory result).

The agent's utility therefore is

$$U = u(w_1) - \psi(a_1) + \delta[u(w_2) - \psi(a_2)].$$

The principal maximizes

$$\sum_{i=1}^{N} f(x_i, a_1) \left([x_i - w_1(x_i)] + \delta \sum_{j=1}^{N} f(x_j, a_2(x_i))[x_j - w_2(x_i, x_j)] \right).$$

We have two theorems.

Theorem 16 If (a, w) is optimal, then w must satisfy

$$\frac{1}{u'(w_1(x_j))} = \sum_{k=1}^{N} \frac{f(x_k, a_2(x_j))}{u'(w_2(x_j, x_k))},$$

for every $j \in \{1, \dots, N\}$.

Proof: We use a variation argument constructing a new contract w^*. Take the previous contract, w and change it along the x_j contingent as follows:

$$w_1^*(x_i) = w_1(x_i) \text{ for } i \neq j,,$$

$$w_2^*(x_i, x_k) = w_2(x_i, x_k) \text{ for } i \neq j, \ k \in \{1, \dots, N\},$$

but

$$u(w_1^*(x_j)) = u(w_1(x_j)) - \Delta,$$

$$u(w_2^*(x_j, x_k)) = u(w_2(x_j, x_k)) + \frac{\Delta}{\delta}, \text{ for } k \in \{1, \dots, N\}.$$

Notice that w^* only differs from w following the first-period x_j branch. By construction, the optimal strategy a is still optimal under w^*. To see this note that nothing changes following x_i, $i \neq j$ in the first period. When x_j occurs in the first period, the *relative* second-period wages are unchanged, so $a_2(x_j)$ is still optimal. Finally, the expected present value of the x_j branch also remains unchanged, so a_1 is still optimal.

Because the agent's expected present value is identical under both w and w^*, a necessary condition for the principal's optimal contract is that w minimizes the expected wage bill over the set of perturbed contracts. Thus, $\Delta = 0$ must solve the variation program

$$\min_{\Delta} \ u^{-1}(u(w_1(x_j)) + \Delta) + \delta \sum_{k=1}^{N} f(x_k, a_2(x_j)) u^{-1} \left(u(w_2(x_j, x_k)) - \frac{\Delta}{\delta} \right).$$

The necessary condition for this provides the condition in the theorem. $\square$

The above theorem provides a Borch-like condition. It says that the marginal rates of substitution between the principal and the agent should be equal across time *in expectation*. This is not the same as full insurance because of the expectation component. With the theorem above, we can easily prove that long-term contracts will optimally depend upon previous output levels. That is, contracts have memory.

Theorem 17 If $w_1(x_i) \neq w_1(x_j)$ and if the optimal second period effort conditional on period 1 output is unique, then there exists a $k \in \{1, \dots, N\}$ such that $w_2(x_i, x_k) \neq w_2(x_j, x_k)$.

Proof: Suppose not. Let w have $w_2(x_i, x_k) = w_2(x_j, x_k)$ for all k. Then the agent has as an optimal strategy $a_2(x_i) = a_2(x_j)$, which implies that $f(x_k, a_2(x_i)) = f(x_k, a_2(x_j))$ for every k. But this violates the condition in theorem 16. $\square$

Remarks:

1. Another way to understand Rogerson's result is to consider a rather loose approach using a Lagrangian and continuous outputs (this is loose because we will not concern ourselves with second-order conditions and the like):

$$\max_{w_1(x_1), w_2(x_1, x_2)} \int_{\underline{x}}^{\overline{x}} (x_1 - w_1(x_1)) f(x_1, a_1) dx_1 +$$

$$\int_{\underline{x}}^{\overline{x}} \int_{\underline{x}}^{\overline{x}} (x_2 - w_2(x_1, x_2)) f(x_1, a_1) f(x_2, a_2) dx_1 dx_2,$$

subject to

$$\int_{\underline{x}}^{\overline{x}} \int_{\underline{x}}^{\overline{x}} (u(w_1(x_1)) + \delta u(w_2(x_1, x_2))) f_a(x_1, a_1) f(x_2, a_2) dx_1 dx_2 - \psi'(a_1) = 0,$$

$$\int_{\underline{x}}^{\overline{x}} u(w_2(x_1, x_2)) f_a(x_1, a_2) dx_2 - \psi'(a_2) = 0,$$

$$\int_{\underline{x}}^{\overline{x}} u(w_1(x_1)) f(x_1, a_1) dx_1 - \psi(a_1) +$$

$$\int_{\underline{x}}^{\overline{x}} \int_{\underline{x}}^{\overline{x}} \delta u(w_2(x_1, x_2)) f(x_1, a_1) f(x_2, a_2) dx_1 dx_2 - \psi(a_2) \geq 0.$$

Let μ_1, $\mu_2(x_1)$ and λ represent the multipliers associated with each constraint (note that the second constraint – the IC constraint for period 2 – depends upon x_1, and so there is a separate constraint and multiplier for each x_1). Differentiating and simplifying one obtains

$$\frac{1}{u'(w_1(x_1))} = \lambda + \mu_1 \frac{f_a(x_1, a_1)}{f(x_1, a_1)}, \ \forall \ x_1$$

$$\frac{1}{u'(w_2(x_1, x_2))} = \lambda + \mu_1 \frac{f_a(x_1, a_1)}{f(x_1, a_1)} + \mu_2(x_1) \frac{f_a(x_2, a_2)}{f(x_2, a_2)}, \ \forall \ (x_1, x_2).$$

Combining these expressions, we have

$$\frac{1}{u'(w_2(x_1, x_2))} = \frac{1}{u'(w_1(x_1))} + \mu_2(x_1) \frac{f_a(x_2, a_2)}{f(x_2, a_2)}.$$

Because $\int_{\underline{x}}^{\overline{x}} f_a(x_2, a_2) dx_2 = 0$, the expectation of $1/u'(w_2)$ is simply $1/u'(w_1)$. Hence, $\mu_2(x) f_a(x_2, a_2)/f(x_2, a_2)$ represents the deviation of date 2 marginal utility from the first period as a function of both periods' output.

39

2. Fudenberg, Holmström and Milgrom [1990] demonstrate the importance of the agent's credit market restriction. They show that if all public information can be used in contracting and recontracting takes place with common knowledge about technology and preferences, then agent's perfect access to credit markets results in short-term contracts being as equally effective as long-term contracts. (An additional technical assumption is needed regarding the slope of the expected utility frontier of IC contracts; this is true, for example, when preferences are additively separable over time and utility is unbounded below.)

   Together with the folk-theorem results of Radner and others, this consequently implies that short term contracts in which agents have access to credit markets in a repeated setting can obtain the first best arbitrarily closely as $\delta \to 1$. F-H-M [1990] present a construction of such such-term contracts (their theorem 6). Essentially, the agent self-insures by saving in the initial periods and then smoothing income over time.

3. Malcomson and Spinnewyn [1988] show similar results to FHM [1990] where long-term contracts can be duplicated by a sequence of loan contracts.

4. Rey and Salanie [1990] consider three contracts of varying lengths: one period contracts (they call these spot contracts), two-period overlapping contracts (they call these short-term contracts), and multi-period contracts (i.e., long-term contracts). They show that for many contracting situations (including Rogerson's [1985] model), a sequence of two-period contracts that are renegotiated each period can mimic a long-term contract. Thus, even without capital market access, long-term contracts are not necessary if two-period contracts can be written and renegotiated period by period. The intuition is that a two-period contract mimics a loan/savings contract with the principal.

2.2.3   Renegotiation of Risk-sharing

Interim Renegotiation with Asymmetric Information:

Fudenberg and Tirole [1990] and Ma [1991] consider the case of moral hazard contracts where the principal has the opportunity to (i.e., the inability to commit not to) offer the agent a new Pareto-improving contract in the bf interim stage: *after the agent has supplied effort but before the outcome of the stochastic process is revealed.* Their first result is clear.

Theorem 18 *Choosing any effort level other than the lowest cannot be a pure-strategy equilibrium for the agent in any PBE in which renegotiation is allowed by the principal.*

The proof is straightforward. If the theorem were not true, at the interim stage the principal and agent will be symmetrically informed (on the equilibrium path) and so the principal will offer full insurance. But the agent will intuit that the incentive contract will be renegotiated to a full-insurance contract, and so will supply only the lowest possible effort. As a consequence, high effort chosen with certainty cannot be implemented at any cost. Given the result, the authors naturally turn to mixed-strategy equilibria to study the optimal renegotiation-constrained contract.

We will sketch Fudenberg and Tirole's [1990] analysis of the optimal mixed-strategy renegotiation proof contract. Their insight was that in a mixed-strategy equilibrium (where the agent chooses a distribution over the possible effort choices) a moral hazard setting at the ex ante stage is converted to an adverse selection setting at the interim stage in which an agent's type is chosen action. They show that it is without loss of generality for the principal to offer the agent a renegotiation-proof contract which specifies an optimal mixed strategy for the agent to follow at the ex ante stage and an a menu of contracts for the agent to choose from at the interim stage such that the principal will not wish to renegotiate the contract. Because the analysis which follows necessarily relies upon some knowledge of screening contracts (which is covered in detail in Chapter 2), the reader unfamiliar with these techniques may wish to read the first few pages of chapter 2 (up through section 2.2.1).

Consider the following setting. There are two actions, high and low. To make things interesting, we suppose that the principal wishes to implement to high effort action. The relative cost of supplying high effort to low effort is ψ and the agent is risk averse in wages. That is,

$$U(w, H) = u(w) - \psi,$$

$$U(w, L) = u(w).$$

Following Grossman and Hart [1983], let $h(\cdot)$ be the inverse of u. Thus, $h(u(w)) = w$, where h is strictly increasing and convex.

A high effort generates a distribution $\{p, 1 - p\}$ on the profit levels $\{\overline{x}, \underline{x}\}$; a low effort generates a distribution $\{q, 1 - q\}$ on $\{\overline{x}, \underline{x}\}$, where $\overline{x} > \underline{x}$ and $p > q$. The principal is risk neutral. Let μ be the probability that the agent chooses the high action at the ex ante stage. An incentive contract provides wages as a function of the outcome and the agent's reported type at the interim stage: $w \equiv \{(\overline{w}_H, \underline{w}_H), (\overline{w}_L, \underline{w}_L)\}$ Then,

$$V(w, \mu) = \mu[p\overline{w}_H + (1 - p)\underline{w}_H] + (1 - \mu)[q\overline{w}_L + (1 - q)\underline{w}_L].$$

The procedure we follow to solve for the optimal renegotiation-proof contract uses backward induction. Begin at the interim stage where the principal's beliefs are some arbitrary μ and the agent's expected utility in absence of renegotiation is $\{\overline{U}, \underline{U}\}$. We solve for the optimal renegotiation contract w as a function of μ. Then we consider the ex ante stage and maximize ex ante profits over the set of (μ, w) pairs which are generate an interim optimal contract.

Optimal Interim Contracts: In the interim period, following standard revealed-preference tricks, one can show that the incentive compatibility constraint for the low-effort agent and the individual rationality constraint for the high-effort agent will be binding while the other constraints will slack. Let $\overline{u}_H \equiv u(\overline{w}_H)$, $\underline{u}_H \equiv u(\underline{w}_H)$, $\overline{u}_L \equiv u(\overline{w}_L)$, and $\underline{u}_L \equiv u(\underline{w}_L)$. Then the principal solves the following convex programming problem:

$$\max_w \quad -\mu[ph(\overline{u}_H) + (1 - p)h(\underline{u}_H)] - (1 - \mu)[qh(\overline{u}_L) + (1 - q)h(\underline{u}_L)],$$

subject to

$$q\overline{u}_L + (1 - q)\underline{u}_L \geq q\overline{u}_H + (1 - q)\underline{u}_H,$$

$$p\overline{u}_H + (1-p)\underline{u}_H \geq \overline{U}.$$

Let γ be the IC multiplier and λ be the IR multiplier. Then by the Kuhn-Tucker theorem, we have a solution which satisfies the following four necessary first-order conditions.

$$\mu p h'(\overline{u}_H) = p\lambda - q\gamma,$$

$$\mu(1-p)h'(\underline{u}_H) = (1-p)\lambda - (1-q)\gamma,$$

$$(1-\mu)h'(\overline{u}_L) = \gamma,$$

$$(1-\mu)h'(\underline{u}_L) = \gamma.$$

Combining the last two equations implies $\overline{u}_L = \underline{u}_L = u_L$ (i.e., complete insurance for the low-effort agent), and therefore $\gamma = (1-\mu)h'(u_L)$. Using this result for γ in the first two equations, and substituting out λ, yields

$$\frac{\mu}{1-\mu} = \frac{h'(u_L)}{h'(\overline{u}_H) - h'(\underline{u}_H)}\frac{p-q}{p(1-p)}.$$

This equation is usually referred to as the renegotiation-proofness constraint. For any given wage schedule $w \equiv \{(\overline{w}_H, \underline{w}_H), (\overline{w}_L, \underline{w}_L)\}$ (or alternatively, a utility schedule $u \equiv \{(\overline{u}_H, \underline{u}_H), (\overline{u}_L, \underline{u}_L)\}$), there exists a unique $\mu^*(u)$ which satisfies the above equation, and which provides the upper bound on feasible ex ante effort choices (i.e., $\forall \ \mu \leq \mu^*$, the ex ante contract is renegotiation proof).

Optimal Ex ante Contracts: Now consider the optimal ex ante contract offer $\{\mu, u\}$. The principal solves

$$\max_{\mu,(\overline{u}_H, \underline{u}_H, u_L)} \mu[p(\overline{x} - h(\overline{u}_H)) + (1-p)(\underline{x} - h(\underline{u}_H))]$$

$$+ (1-\mu)[q(\overline{x} - h(u_L)) + (1-q)(\underline{x} - h(u_L))],$$

subject to

$$u_L = p\overline{u}_H + (1-p)\underline{u}_H - \psi,$$

$$u_L \geq 0,$$

$$\frac{\mu}{1-\mu} = \frac{h'(u_L)}{h'(\overline{u}_H) - h'(\underline{u}_H)}\frac{p-q}{p(1-p)}.$$

The first constraint is the typical binding IC constraint that the high effort agent is just willing to supply high effort. More importantly, indifference also guarantees that the agent is willing to randomize according to μ, and so it is a *mixing* constraint as well. The second constraint is the IR constraint for the low effort choice (and by the first equation, for the high-effort choice as well). The third equation is our renegotiation-proofness (RP) constraint. Note that if the principal attempts to choose μ arbitrarily close to 1, then the RP constraint implies that $\overline{u}_H \approx \underline{u}_H = u_H$. Interim IC in turn implies that $u_H \approx u_L$. But this in turn will violate the ex ante IC (mixing) constraint for $\psi > 0$. Thus, it is not feasible to choose μ arbitrarily close to 1.

Remarks:

1. In general, the RP constraint both lessens the principal's expected profit as well as reduces the principal's feasible set of implementable actions. Thus, non-commitment has two effects.

2. In general, it is not the case that the ex ante IR constraint for the agent will bind. Specifically, it is possible that by leaving the agents some rents that the RP constraint will be weakened. However, if utility displays non-increasing absolute risk aversion, the IR constraint will bind.

3. The above results are extended to a continuum of efforts and two outcomes by Fudenberg and Tirole [1990]. They also show that although it is without loss of generality to consider RP contracts, one can show that with equilibrium renegotiation, the principal can uniquely implement the optimal RP contract.

4. Our results about the cost of renegotiation were predicated upon the principal (who is uninformed at the interim stage) making a take-it-or-leave-it offer to the agent. According to Fudenberg and Tirole (who refer to Maskin and Tirole's [1992] paper on "The Principal-Agent Relationship with and Informed Principal, II: Common values"), not only is principal-led renegotiation simpler, but the same conclusions are obtained if the agent leads the renegotiation providing one requires that the original contract be "strongly renegotiation proof" (i.e., RP in *any* PBE at the interim stage). Nonetheless, in a paper by Ma [1994] it is shown that providing one is prepared to use a refinement on the principal's beliefs regarding off-the-equilibrium-path beliefs at the interim stage, agent-led renegotiation has no cost. That is, the second-best incentive contract remains renegotiation proof. Of course, according to Maskin and Tirole [1992], such a contract cannot be strongly renegotiation proof; i.e., there must be other equilibria where renegotiation does occur and is costly to the principal at the ex ante stage.

5. Matthews [1995].

Interim Renegotiation with Symmetric Information:

The previous analysis assumed that at the interim (renegotiation) stage, one party was asymmetrically informed. Hermalin and Katz [1991] demonstrate that this is the source of *costly* renegotiation. With symmetrically informed parties, full insurance can be provided and first best effort can be implemented. Hence, it is possible that "renegotiation" can improve welfare, because it can change the terms of a contract to reflect new information about the agent's performance.

Consider the following variation on standard principal-agent timing. A contract is offered by the principal to the agent at the ex ante stage, and the agent immediately takes an action. Before the interim (renegotiation), however, both parties observe a signal s regarding the agent's chosen action. This signal is observable, but not verifiable (i.e., it cannot be part of any enforceable contract). The parties can now renegotiate. Following renegotiation, the verifiable output x is observed, and the agent is rewarded according to the contract in force and the realized output (which is contractible).

For simplicity, assume that both parties observe the chosen action, a, at the renegotiation stage. Then almost trivially it follows that the principal cannot be made worse off with renegotiation when the principal makes the interim offers. The set of implementable contracts remains unchanged. Generally, however, the principal can do better by using renegotiation to her benefit.

The the following proposition for agent-led renegotiation follows immediately.

Theorem 19 Suppose that s perfectly reveals a and the agent makes a take-it-or-leave-it renegotiation offer at the interim stage. Then the first-best action is implementable at the full information cost.

Proof: By construction. The principal sells the firm to the agent at the ex ante stage for a price equal to the firm's first-best expected profit (net of effort costs). The agent is willing to purchase the firm, exert the first-best level of effort to maximize its resale value, and then sell the firm back to the principal at the interim stage, making a profit of zero. This satisfies IR and IC, and all rents go to the principal. □

A similar result is true if we give all of the bargaining power to the principal. In this case, the principal will offer the agent his certainty equivalent of the ex ante contract at the interim stage. Assume that a^{FB} is implementable with certainty equivalent equal to the agent's reservation utility. (This will be possible, for example, if the distribution vector $p(a)$ is not an element of the convex hull of $\{p(a')|a' \neq a\}$ for any a.) A principal would not normally want to do this because the risk premium the principal must give to the agent to satisfy IR is too great. But with the interim stage of renegotiation, the principal observes a and can renegotiate away all of the risk. Thus we have ...

Theorem 20 Suppose that a^{FB} is implementable and the principal makes a take-it-or-leave-it offer at the renegotiation stage. Then the first-best is implementable at the full-information cost.

Hermalin and Katz go even further to show that under some mild technical assumptions that the first-best full information allocation is obtainable with any arbitrary bargaining game at the interim stage.

Remarks:

1. Hermalin and Katz extend their results to the case where a is only imperfectly observable and find that principal-led renegotiation is still beneficial providing that the commonly observable signal s is a sufficient statistic for x with respect to (s, a). That is, having observed s, the principal can predict x as well as the agent can. Although the first-best cannot generally be achieved, renegotiation is beneficial.

2. Juxtaposing Hermalin and Katz's results with those of Fudenberg and Tirole's [1990], we find some interesting connections. In H&K, the terms of trade at the interim stage are a direct function of observed effort, a; in F&T, the dependence is obtained only through costly interim incentive compatibility conditions. Renegotiation can be bad because it undermines commitment in absence of information on a; it can be good if

renegotiation can be made conditional on a. Thus, the main differences are whether renegotiation takes place between asymmetrically or (sufficiently) symmetrically informed parties.

3 Mechanism Design and Self-selection Contracts

3.1 Mechanism Design and the Revelation Principle

We consider a setting where the principal can offer a mechanism (e.g., contract, game, etc.) which her agents can play. The agent's are assumed to have private information about their preferences. Specifically, consider I agents indexed by $i \in \{1, \ldots, I\}$.

- Each agent i observes only its own preference parameter, $\theta_i \in \Theta_i$. Let $\theta \equiv (\theta_1, \ldots, \theta_I) \in \Theta \equiv \prod_{i=1}^{I} \Theta_i$.

- Let $y \in Y$ be an allocation. For example, we might have $y \equiv (x, t)$, with $x \equiv (x_1, \ldots, x_I)$ and $t \equiv (t_1, \ldots, t_I)$, and where x_i is agent i's consumption choice and t_i is the agent's payment to the principal. The choice of y is generally controlled by the principal, although she may commit to a particular set of rules.

- Utility for i is given by $U_i(y, \theta)$; note general interdependence of utilities on θ_{-i} and y_{-i}. The principal's utility is given by the function $V(y, \theta)$. In a slight abuse of notation, if y is a distribution of outcomes, then we'll let U_i and V represent the value of expected utility after integrating with respect to the distribution.

- Let $p(\theta_{-i} | \theta_i)$ be i's probability assessment over the possible types of other agents given his type is θ_i and let $p(\theta)$ be the common prior on possible types.

Suppose that the principal has all of the bargaining power and can commit to playing a particular game or mechanism involving her agent(s). Posed as a mechanism design question, the principal will want to choose the game (from the set of all possible games) which has the best equilibrium (to be defined) for the principal. But this set of all possible games is enormous and complex. The revelation principle, due to Green and Laffont [1977], Myerson [1979], Harris and Townsend [1981], Dasgupta, Hammond, and Maskin [1979], et al., allows us to simplify the problem dramatically.

Definition: A communication *mechanism* or game,

$$\Gamma^c \equiv \{\mathcal{M}, \Theta, p, U_i(y(m), \theta)_{i=1,\ldots,I}\},$$

is characterized by a message (i.e., strategy) space for each agent, $\mathcal{M}_i$, and an allocation y for each possible message profile, $m \equiv (m_1, \ldots, m_I) \in \mathcal{M} \equiv (\mathcal{M}_1, \ldots, \mathcal{M}_I)$; i.e., $y : \mathcal{M} \mapsto Y$. For generality, we will suppose that $\mathcal{M}_i$ includes all possible mixtures over messages; thus, m_i may be a probability distribution. When no confusion would result, we sometimes indicate a mechanism Γ by the pair $\{\mathcal{M}, y\}$.

The timing of the communication mechanism game is as follows:

- Stage 1. The principal offers a communication mechanism and a Nash equilibrium to play.

- Stage 2. The agents simultaneously decide whether or not to participate in the mechanism. (This stage may be superfluous in some contexts; moreover, we can always require the principal include the message of "I do not wish to play" and the null contract, making the acceptance stage unnecessary.)

- Stage 3. Agents play the communication mechanism.

The idea here is that the principal commits to a mechanism, $y(m)$, and the agents all choose their messages in light of this. To state the revelation principle, we must choose an equilibrium concept for Γ^c. We first consider Bayesian-Nash equilibria (other possibilities include dominant-strategy equilibria and correlated equilibria).

3.1.1 The Revelation Principle for Bayesian-Nash Equilibria

Let $m^*(\theta) \equiv (m_1^*(\theta_1), \ldots, m_I^*(\theta_I))$ be a Bayesian-Nash equilibrium (BNE) of the game in stage 3, and suppose without loss of generality that all agent's participated at stage 2. Then $y(m^*(\theta))$ denotes the equilibrium allocation.

Revelation Principle (BNE): Suppose that a mechanism, Γ^c has a BNE $m^*(\theta)$ defined over all θ which yields allocation $y(m^*(\theta))$. Then there exists a direct revelation mechanism, $\Gamma^d \equiv \{\mathcal{M} \equiv \Theta, \Theta, p, U_i(\tilde{y}(\theta), \theta)_{i=1,\ldots,I}\}$, with strategy spaces $\mathcal{M}_i \equiv \Theta_i, i = 1, \ldots, I$, and an outcome function $\tilde{y}(\theta) : \Theta \mapsto Y$ such that there exists a BNE in Γ^d with $\tilde{y}(\theta) = y(m^*(\theta))$ and equilibrium strategies $m_i(\theta_i) = \theta_i \ \forall \theta$.

Proof: Because $m^*(\theta)$ is a BNE of Γ^c, for any i with type θ_i,

$$m_i^*(\theta_i) \in \arg \max_{m_i \in \mathcal{M}_i} E_{\theta_{-i}}[U_i(y(m_i, m_{-i}^*(\theta_{-i})), \theta_i, \theta_{-i})|\theta_i].$$

This implies for any θ_i

$$E_{\theta_{-i}}[U_i(y(m_i^*(\theta_i), m_{-i}^*(\theta_{-i})), \theta_i, \theta_{-i})|\theta_i]$$
$$\geq E_{\theta_{-i}}[U_i(y(m_i^*(\hat{\theta}_i), m_{-i}^*(\theta_{-i})), \theta_i, \theta_{-i})|\theta_i], \ \forall \hat{\theta}_i \in \Theta_i.$$

Let $y(\theta) \equiv y(m^*(\theta))$. Then we can rewrite the above equation as:

$$E_{\theta_{-i}}[U_i(y(\theta_i, \theta_{-i}), \theta_i, \theta_{-i})|\theta_i] \geq E_{\theta_{-i}}[U_i(y(\hat{\theta}_i, \theta_{-i}), \theta_i, \theta_{-i})|\theta_i], \ \forall \hat{\theta}_i \in \Theta_i.$$

But this implies $m_i(\theta_i) = \theta_i$ is an optimal strategy in Γ^d and $y(\theta)$ is an equilibrium allocation. Therefore, truthtelling is a BNE in the direct mechanism game. $\square$

Remarks:

1. This is an extremely useful result. If a game exists in which a particular allocation y can be implemented by the principal, there is a direct revelation mechanism with truth-telling as an equilibrium that can also accomplish this. Hence, *without loss of generality, the principal can restrict attention to direct revelation mechanisms in which truth-telling is an equilibrium.*

2. The more general formulations of this principle such as by Myerson, allow agents to take actions as well. That is, y has some components which are under the agents' control and some components which are under the principal's control. A revelation principle still holds in which the principal implements y by choosing the components it controls and making suggestions to the agents as to which actions they should take that are under their control. Truthful revelation occurs in equilibrium and suggestions are followed. Myerson refers to this as "truthtelling" and "obedience," respectively.

3. This notion of truthful implementation in a BNE is a very weak concept. There may be many other equilibria to Γ^d which are not truthful and in which the agents do better. Thus, there may be strong reasons to believe that the agents will not follow the equilibrium the principal selects. This non-uniqueness problem has spawned a large number of papers which focus on conditions for a unique equilibrium allocation which are discussed in the survey articles by Moore [1992] (regarding symmetrically informed agents) and Palfrey [1992] (regarding asymmetrically informed agents).

4. We rarely see direct revelation mechanisms being used. Economically, the indirect mechanisms are more interesting to study once we find the direct mechanism. Possible advantages from carefully choosing an indirect mechanism are uniqueness, simplicity, and robustness against collusion, etc.

5. The key to the revelation principle is commitment. With commitment, the principal can replicate the outcome of any indirect mechanism by promising to play the strategy for each player that the player would have chosen in the indirect mechanism. Without commitment, we must be careful. Thus, when renegotiation is possible, the revelation principle fails to apply.

6. In some settings, agents contract with several principal's simultaneously (common agency), so there may be one agent working for two principals where each principal has control over a component of the allocation, y. Is there a related revelation principle such as "for any BNE in the common agency game with one agent and two principals, there exists a BNE to a pair of direct-revelation mechanisms (one offered by each principal) in which the agent reports truthfully to both principals"? The answer is no. The problem is that out-of-equilibrium messages which had no use in a one-principal setting, may enlarge the set of equilibria in the original game beyond those sustainable as equilibria to the revelation game in a multi-principal setting.

7. Note we could have just as easily used correlated equilibria or dominant-strategy equilibria as our equilibrium notion. There are similar revelation principles for these concepts.

3.1.2 The Revelation Principle for Dominant-Strategy Equilibria

If the principal wants to implement an allocation, y, in dominant strategies, then she has to design a mechanism such that this mechanism has a dominant-strategy equilibrium (DSE), $m^*(\theta)$, with outcome $y(m^*(\theta))$.

Revelation Principle (DSE): Suppose that Γ^c ha a dominant-strategy equilibrium, $m^*(\theta)$ with outcome $y(m^*(\theta))$. Then there exists a direct revelation mechanism, $\Gamma^d \equiv \{\Theta, y\}$, with strategy spaces, $\mathcal{M}_i \equiv \Theta_i$, $i = 1, \ldots, I$, and an outcome function $y(\theta): \Theta \mapsto Y$ such that there exists a DSE in Γ^d in which truth-telling is a dominant strategy with DSE allocation $y(\theta) \equiv y(m^*(\theta)) \ \forall \theta \in \Theta$.

Proof: Because $m^*(\theta)$ is a DSE of Γ^c, for any i and type $\theta_i \in \Theta_i$,

$$m_i^*(\theta_i) \in \arg \max_{m_i \in \mathcal{M}_i} U_i(y(m_i, m_{-i}), \theta_i, \theta_{-i}), \ \forall \theta_{-i} \in \Theta_{-i} \text{ and } \forall m_{-i} \in \mathcal{M}_{-i}.$$

This implies that $\forall \ (\hat{\theta}_i, \theta_{-i}) \in \Theta$

$$U_i(m_i^*(\theta_i), m_{-i}^*(\theta_{-i}), \theta_i, \theta_{-i}) \geq U_i(y(m_i^*(\hat{\theta}_i), m_{-i}^*(\theta_{-i})), \theta_i, \theta_{-i}).$$

Let $y(\theta) \equiv y(m^*(\theta))$. Then we can rewrite the above equation as

$$U_i(y(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \geq U_i(y(\hat{\theta}_i, \theta_{-i}), \theta_i, \theta_{-i}), \ \forall \hat{\theta}_i \in \Theta_i, \ \forall \theta_{-i} \in \Theta_{-i}.$$

But this implies truthtelling, $m_i(\theta_i) = \theta_i$, is a DSE of Γ^d with equilibrium allocation $y(\theta)$. $\square$

Remarks:

1. Certainly this is a more robust implementation concept. Dominant strategies are more likely to be played. Additionally, if for whatever reasons you believe that the agents have different priors, then the allocation is unchanged.

2. Generically, DSE are unique, although some economically likely environments are non-generic.

3. DSE is a (weakly) stronger concept than using BNE. But when only one agent is playing, the two concepts coincide.

4. We will generally focus in the BNE revelation principle, although we will discuss the use of dominant strategy mechanisms in a few simple settings. Furthermore, Mookerjee and Reichelstein [1992] have demonstrated that under a large class of contracting environments, the outcomes implemented with BNE mechanisms can be implemented with DSE mechanisms as well.

3.2 Static Principal-Agent Screening Contracts

With the revelation principle(s) developed, we proceed to characterize the set of truthful direct-revelation mechanisms and then to find the optimal mechanism from this set in a simple single-agent setting. We proceed first by exploring a simple two-type example of non-linear pricing, and then a detailed examination of the general case.

3.2.1 A Simple 2-type Model of Nonlinear Pricing

A risk-neutral firm produces a product of quality q at a cost per unit of $c(q)$. Its profit from a sale of one unit with quality q for a price of t is is $V = t - c(q)$. There are two types of consumers, $\underline{\theta}$ and $\overline{\theta}$, with $\overline{\theta} > \underline{\theta}$ and a proportion of p of type $\overline{\theta}$. For this example, we will assume that p is sufficiently small that the firm prefers to sell to both types rather than focus only on the the $\overline{\theta}$-type customer. Each consumer has a unit demand with utility from consuming a good of quality q for a price of t equal to $U = \theta q - t$, where $\theta \in \{\underline{\theta}, \overline{\theta}\}$.

By the revelation principle, the firm can restrict attention to contracts of the form $\{(\underline{q}, \underline{t}), (\overline{q}, \overline{t})\}$ such that type $\underline{t}$ consumers find it optimal to choose the first contract pair and $\overline{t}$ consumers choose the second pair. Thus, we can write the firm's optimization program as:

$$\max_{\{(\underline{q},\underline{t}),(\overline{q},\overline{t})\}} p[\overline{t} - c(\overline{q})] + (1-p)[\underline{t} - c(\underline{q})],$$

subject to

$$
\begin{aligned}
\overline{\theta}\overline{q} - \overline{t} &\geq \overline{\theta}\underline{q} - \underline{t} & (\overline{IC}), \\
\underline{\theta}\underline{q} - \underline{t} &\geq \underline{\theta}\overline{q} - \overline{t} & (\underline{IC}), \\
\overline{\theta}\overline{q} - \overline{t} &\geq 0 & (\overline{IR}), \\
\underline{\theta}\underline{q} - \underline{t} &\geq 0 & (\underline{IR}),
\end{aligned}
$$

where (IC) refers to an incentive compatibility constraint to choose the relevant contract and (IR) refers to an individual rationality constraint to choose some contract rather than no purchase at all.

Note that the two IC constraints can be combined in the statement

$$\overline{\theta}\Delta q \geq \Delta t \geq \underline{\theta}\Delta q.$$

Among other things we find that incentive compatibility implies that $\overline{q} \geq \underline{q}$.

To simplify the maximization program facing the firm, we consider the four constraints to determine which – if any – will be binding.

1. First note that $\overline{IC}$ and $\underline{IR}$ imply that $\overline{IR}$ is slack. Hence, it will never be binding and we can ignore it.

2. A simple argument establishes that $\overline{IC}$ must always bind. Suppose otherwise and it was slack at the optimal contract offering. In such a case, $\overline{\theta}$ could be raised slightly without disturbing this constraint or $\overline{IR}$, thereby increasing profits. Moreover, this increase only eases the $\underline{IC}$ constraint. Hence, the contract cannot be optimal. $\overline{IC}$ binds.

3. If $\overline{IC}$ binds, $\underline{IC}$ must be slack if $\overline{q} - \underline{q} \geq 0$ because

$$\overline{\theta}\Delta q = \Delta t > \underline{\theta}\Delta q.$$

   Hence, we can ignore $\underline{IC}$ if we assume $\overline{q} - \underline{q} \geq 0$.

Because we have two constraints satisfied with equalities, we can use them to solve for $\bar{t}$ and $\underline{t}$ as functions of $\bar{q}$ and $\underline{q}$:

$$
\begin{aligned}
\underline{t} &= \underline{\theta}\underline{q}, \\
\bar{t} &= \underline{t} + \bar{\theta}\Delta q \\
&= \bar{\theta}\bar{q} - \Delta\theta\underline{q}.
\end{aligned}
$$

These t's are necessary and sufficient for all four constraints to be satisfied if $\bar{q} - \underline{q} \geq 0$. Substituting for the t's in the firm's objective program, the firm's maximization program becomes simply

$$
\max_{\{(\underline{q},\underline{t}),(\bar{q},\bar{t})\}} p[\bar{\theta}\bar{q} - c(\bar{q}) - \Delta\underline{q}] + (1-p)[\underline{\theta}\underline{q} - c(\underline{q})],
$$

subject to $\bar{q} \geq \underline{q}$. Ignoring the monotonicity constraint, the first-order conditions to this relaxed program imply

$$
\bar{\theta} = c'(\bar{q}),
$$

$$
\underline{\theta} = c'(\underline{q}) + \frac{p}{1-p}\Delta\theta.
$$

Hence, $\bar{q}$ is set at the first-best efficient levels of consumption but $\underline{q}$ is set at sub-optimal levels of consumption. This distortion also implies that $\bar{q} > \underline{q}$, and hence our monotonicity constraint does not bind.

Having determined the optimal $\bar{q}$ and $\underline{q}$, the firm can easily determine the appropriate prices using the conditions for $\bar{t}$ and $\underline{t}$ above and the firm's nonlinear pricing problem has been solved.

3.2.2 The Basic Paradigm with a Continuum of Types

This subsection borrows from Fudenberg-Tirole [Ch. 7, 1991], although many of the assumptions, theorems, and proofs have been modified.

There are usually two steps in mechanism design. First, characterizing the set of implementable contracts; then, selecting the optimal contract from this set.

First, some notation. For now, we consider the simpler case of a single agent, and so we have dropped subscripts. Additionally, it does not matter whether we focus on BNE or DSE allocations. The basic elements of the simple model are as follows:

1. Our allocation is a pair of non-stochastic functions, $y = (x, t)$, where $x \in \mathbb{R}_+$ is a one-dimensional activity (e.g., consumption, production, etc.) and t is a transfer (perhaps negative) from the agent to the principal. We will sometimes refer to x as the decision or activity of the agent and t as the transfer function.

2. The agent's private information is one-dimensional, $\theta \in \Theta$, where we take $\Theta = [0, 1]$ without loss of generality. The density is $p(\theta) > 0$ over $[0, 1]$, and the distribution function is $P(\theta)$.

3. The agent has quasi-linear utility: $U = u(x, \theta) - t$, where $u \in C^2$.

4. The principal has quasi-linear utility: $V = v(x, \theta) + t$, where $v \in C^2$.

5. The total surplus function is $S(x, \theta) \equiv u(x, \theta) + v(x, \theta)$; we assume that utilities are transferable.

Implementable Contracts   Unlike F&T, we begin by making the Spence-Mirrlees single-crossing property (sorting) assumption on u.

Assumption 1 $\frac{\partial u(x,\theta)}{\partial \theta} > 0$ and $\frac{\partial^2 u(x,\theta)}{\partial \theta \partial x} > 0$.

The first condition is not generally considered part of the sorting condition, but because they are so closely related economically, we make them together.

Definition 6 We say that an allocation $y = (x, t)$ is **implementable** (or alternatively, we say that x is **implementable** with transfer t) iff it satisfies the incentive-compatibility (truth-telling) constraint

$$u(x(\theta), \theta) - t(\theta) \geq u(x(\hat{\theta}), \theta) - t(\hat{\theta}), \text{ for all } (\theta, \hat{\theta}) \in [0, 1]^2. \tag{IC}$$

For notational ease, we will find it useful to consider the indirect utility function; i.e. the utility the agent of type θ receives when reporting $\hat{\theta}$: $U(\hat{\theta}|\theta) \equiv u(x(\hat{\theta}), \theta) - t(\hat{\theta})$. We will use the subscripts 1 and 2 to represent the partial derivatives of U with respect to report and type, respectively. When evaluating U in truthtelling equilibrium, we will often write $U(\theta) \equiv U(\theta|\theta)$. Note that in this case, $\frac{dU(\theta)}{d\theta} = U_1(\theta) + U_2(\theta)$. Our characterization theorem can now be presented and proved.

Theorem: Suppose $u_{x\theta} > 0$ and that the direct mechanism, $y(\theta) = (x(\theta), t(\theta))$, is is compact-valued (i.e., the set $\{(x, t) | \exists \hat{\theta} \in \Theta \text{ s.t. } (x, t) = (x(\hat{\theta}), t(\hat{\theta}))\}$ is compact). Then the direct mechanism is incentive compatible iff

$$U(\theta_1) - U(\theta_0) = \int_{\theta_0}^{\theta_1} u_\theta(x(s), s) ds, \quad \forall \theta_0, \theta_1 \in \Theta, \tag{3}$$

and $x(\theta)$ is nondecreasing.

The result in equation (3) is a restatement of the agent's first-order condition for truth-telling. Providing the mechanism is differentiable, when truth-telling is optimal we have $U_1(\theta) = 0$, and so $\frac{dU(\theta)}{d\theta} = U_2(\theta)$. Because $U_2(\theta) = u_\theta(x, \theta)$, applying the fundamental theorem of calculus yields equation (3). As the proof below makes clear, the monotonicity condition is the analog of the agent's second-order condition for truth-telling.

Proof:
Necessity: Incentive compatibility requires for any θ and $\tilde{\theta}$,

$$U(\theta) \geq U(\tilde{\theta}|\theta) \equiv U(\tilde{\theta}) + [u(x(\tilde{\theta}), \theta) - u(x(\tilde{\theta}), \tilde{\theta})].$$

Thus,

$$U(\theta) - U(\tilde{\theta}) \geq u(x(\tilde{\theta}), \theta) - u(x(\tilde{\theta}), \tilde{\theta}).$$

Reversing the θ and $\tilde{\theta}$ and combining results yields

$$u(x(\theta), \theta) - u(x(\theta), \tilde{\theta}) \geq U(\theta) - U(\tilde{\theta}) \geq u(x(\tilde{\theta}), \theta) - u(x(\tilde{\theta}), \tilde{\theta}).$$

51

Monotonicity is immediate from $u_{x\theta} > 0$.

Taking the limit as $\theta \to \hat{\theta}$ implies that

$$\frac{dU(\theta)}{d\theta} = u_\theta(x(\theta), \theta)$$

at all points at which $x(\theta)$ is continuous (which is everywhere but perhaps a countable number of points due to the monotonicity of x). Given that the set of available allocations is compact, continuity of $u(x, \theta)$ implies that $U(\theta)$ is continuous (by the Maximum theorem of Berges (1953)). The continuity of $U(\theta)$ over the compact set Θ (which implies U is uniformly continuous), combined with a bounded derivative (at all points of existence), implies that the fundamental theorem of calculus can be applied (specifically, U is Lipschitz continuous, and therefore absolutely continuous). Hence, $U(\theta)$ can be represented as in (3).

Sufficiency: Suppose not. Then there exists θ and $\hat{\theta}$ such that

$$U(\hat{\theta}|\theta) > U(\theta|\theta),$$

which implies

$$u(x(\hat{\theta}), \theta) - u(x(\hat{\theta}), \hat{\theta}) > U(\theta) - U(\hat{\theta}).$$

Integrating the lefthandside and using (1) above on the righthand side implies

$$\int_{\hat{\theta}}^{\theta} u_\theta(x(\hat{\theta}), s) ds > \int_{\hat{\theta}}^{\theta} u_\theta(x(s), s) ds.$$

Rearranging,

$$\int_{\hat{\theta}}^{\theta} [u_\theta(x(\hat{\theta}), s) - u_\theta(x(s), s)] ds > 0.$$

But the single-crossing property of A.1 that $u_{x\theta} > 0$ with the monotonicity condition implies that this is not possible. Hence, a contradiction.□

Remarks:

1. The above characterization theorem was first used by Mirrlees [1971] in his study of optimal taxation. Needless to say, it is a very powerful and useful theorem.

2. We could have used an alternative representation of the single-crossing property where type has a negative interpretation: $u_\theta < 0$ and $u_{x\theta} < 0$. This is isomorphic to our original sorting condition where θ is replaced by $-\theta$. Consequently, our characterization theorem is unchanged except that x must be nonincreasing. This alternative representation is commonly used in public regulation contexts where the type of the agent is related to marginal cost of production, and so higher types have higher costs and lower payoffs.

3. The implementation theorem above is actually easily generalized to cases of non-quasi-linear utility. In such a case, we need to make a Lipschitz assumption on the marginal rate of substitution between money and decision in order to guarantee the existence of a transfer function which can implement the decision function x. See Guesnerie and Laffont [1984] for details.

4. Related characterization theorems have been proved for the case of multi-dimensional types, multi-dimensional actions, multiple mechanisms (common agency). They all proceed in basically the same way, but rely on gradients and Hessians instead of simple one-dimensional first and second-order conditions.

5. The above characterization theorem can also be easily extended to random allocation functions, $\tilde{y} = (\tilde{x}, \tilde{t})$ by taking the appropriate expectations in the initial definition of the indirect utility function.

6. Unlike many papers in the literature, this statement uses the intergral condition (rather than the derivative condition) as the first-order condition. The reason for this is two-fold. First, the integral condition is what we ultimately would like to use for sufficiency (i.e., it is the fundamental theorem of calculus), and second, if the mechanism we are intersted in is not differentiable, the derivative condition is less useful. The difficulty over traditional proofs is then to show that the integral condition in (1) is actually a necessary condition (rather than the possibly weaker derivative condition). This is Myerson's approach in his proof in the "optimal auctions" paper. Myerson, however, leaves out the technical details in proving the necessity of (1) (i.e., that U can be integrated up to yield (3), which is more than saying that simply that U can be integrated), and in any event Myerson has the advantage of a simpler problem in that $u = \theta x$ in his framework; we do not have such a luxury. Note that to accomplish our result, I have used a requirement that the direct mechanism is compact-valued. Without such an assumption, a direct mechanism may not have an optimal report (technically, $\sup_{\hat{\theta}} U(\hat{\theta}|\theta)$ exists, but $\max_{\hat{\theta}} U(\hat{\theta}|\theta)$ may not). This seems a very unrestrictive notion to place on our contract space.

7. To reiterate, this stronger theorem for IC (without relying on continuity, etc.) is essential when looking at optimal auctions. The implementation theorem in Fudenberg and Tirole's book (ch. 7), for example, is not directly useful because they have resorted to assuming x is continuous and has first derivatives that are continuous at all but a finite number of points – far too stringent of a condition for auctions.

Optimal Contracts We now consider the optimal choice of contract by a principal. Given our characterization theorem, the principal's problem is

$$\max_{y} \; E_\theta[V(x(\theta), \theta)] \equiv E_\theta[S(x(\theta), \theta) - U(x(\theta), \theta)]$$

subject to $\frac{dU(\theta)}{d\theta} = u_\theta(x(\theta), \theta)$, x nondecreasing, and generally a participation constraint (referred to as individual rationality in the literature)

$$U(\theta) \geq \underline{U}, \text{ for all } \theta \in [0, 1]. \tag{IR}$$

[Note that we have rewritten profits as the difference between total surplus and the agent's surplus.]

Normally, one proceeds by solving the relaxed program in which monotonicity is ignored, and then checking ex post that it is in fact satisfied. If it isn't satisfied, one then either incorporates the constraint into the maximization program directly (a tedious thing to do)

or assume sufficient regularity conditions so that the resulting function is indeed monotonic. We will also follow this latter approach for now.

In solving the relaxed program, there are again two choices available. First, and when possible I think the most powerful, integrating out the agent's utility function and converting the problem to one of pointwise maximization; second, using control theoretic tools (e.g., Hamiltonians and Pontryagin's theorem) directly to solve the problem. We will begin with the former.

Note that the second term of the objective function can be rewritten using integration by parts and the fact that $\frac{dU}{d\theta} = u_\theta$.

$$
\begin{aligned}
E_\theta[U(x(\theta),\theta)] &\equiv \int_0^1 U(x(s),s)p(s)ds \\
&= -U(x(\theta),\theta)[1-P(\theta)]\big|_0^1 + \int_0^1 \frac{dU(s)}{d\theta}\frac{1-P(s)}{p(s)}p(s)ds \\
&= U(x(0),0) + E_\theta\left[u_\theta(x(\theta),\theta)\frac{1-P(\theta)}{p(\theta)}\right].
\end{aligned}
$$

Remark: It is equally true (via changing the constant of integration) that

$$
E_\theta[U(x(\theta),\theta)] = U(x(1),1) - E_\theta\left[u_\theta(x(\theta),\theta)\frac{P(\theta)}{p(\theta)}\right].
$$

We use the former representation rather than the latter because when $u_\theta > 0$, it will typically be optimal to set $U(x(0),0)$ equal to the agent's outside reservation utility, $\underline{U}$; $U(x(1),1)$ on the other hand is endogenously determined. This is true since utility is increasing in type, so the participation constraint will bind only for the lowest type. When the alternative sorting condition is in place (i.e., $u_\theta < 0$ and $u_{x\theta} < 0$), the second representation will be more useful as $U(x(1),1)$ will typically be set to the agent's outside utility.

Now we substitute the new representation of the agent's expected utility for the one in the objective function. We have our relaxed program (ignoring the monotonicity constraint) as

$$
\max_x E_\theta\left[S(x(\theta),\theta) - \frac{1-P(\theta)}{p(\theta)}u_\theta(x(\theta),\theta) - U(x(0),0)\right],
$$

subject to (IR) and (3). For notational ease, define

$$
\Phi(x,\theta) \equiv S(x,\theta) - \frac{1-P(\theta)}{p(\theta)}u_\theta(x,\theta).
$$

We make the following regularity assumption.

Assumption 2 Φ *is quasi-concave and has a unique interior maximum over* $x \in \mathbb{R}_+$ $\forall\theta$.

This assumption is uncontroversial and is met, for example, if the underlying surplus function is strictly concave, $-u_\theta$ is not too convex, and $\Phi_x(0,\theta)$ is nonnegative. We now have a theorem which characterizes the solution to our relaxed program.

Theorem 21 Suppose that A.2 holds, that x satisfies $\Phi_x(x(\theta),\theta) = 0$ $\forall\theta \in [0,1]$, and that $t(\theta) = u(x(\theta),\theta) - \left(\underline{U} + \int_0^\theta u_\theta(x(s),s)ds\right)$. Then $y = (x,t)$ solves the relaxed program.

Proof: The proof follows immediately from our simplified objective function. First, note that the objective function has been re-written independent of transfers. [When we integrated by parts, we integrated out the transfer function because U is quasi-linear in t.] Thus, we can choose x to maximize the objective function and later choose t such that the differential equation (3) is satisfied for that x; i.e., $t = u - U$, or

$$t(\theta) = u(x(\theta), \theta) - \left[U(x(0), 0) + \int_0^\theta u_\theta(x(s), s) ds \right].$$

Given that $\frac{dU}{d\theta} > 0$, the IR constraint can be restated as $U(x(0), 0) \geq \underline{U}$. From the objective function, this constraint will bind because there is never any reason to leave the lowest type rents. Hence, we have our equation for transfers given x. Finally, to obtain the optimal x, note that our choice of x solves $\max_x E[\Phi(x(\theta), \theta)]$. The optimal x will maximize $\Phi(x, \theta)$ pointwise in θ, which by A.2, is equivalent to $\Phi_x(x(\theta), \theta) = 0 \ \forall \theta$. $\square$

We still have to check that x is nondecreasing in order for $y = (x, t)$ to be an optimal contract. The necessary and sufficient condition for this to be true is given in the following regularity condition.

Assumption 3 $\Phi_{x\theta} \geq 0$ for all (x, θ).

Remark: This assumption is not standard, but is much weaker and more intuitive than that commonly made in the literature. Typically, the literature assumes that $v_{x\theta} \geq 0$, $u_{x\theta\theta} \leq 0$, and the distribution of types satisfies a monotone hazard-rate condition (MHRC). These three conditions imply A.3. The first two assumptions are straightforward, although we generally have little economic insight into third derivatives. The hazard rate is $\frac{p}{1-P}$, and the MHRC assumes that this is nondecreasing. This assumption is satisfied for several common distributions such as the uniform, normal, logistic, and exponential, for example.

Theorem 22 Suppose that A.2 and A.3 are satisfied. Then the solution to the relaxed program satisfies the original un-relaxed program.

Proof: Differentiating $\Phi_x(x(\theta), \theta) = 0$ with respect to θ implies $\frac{dx(\theta)}{d\theta} = -\frac{\Phi_{x\theta}}{\Phi_{xx}}$. By A.2, the denominator is negative. By A.3, x is nondecreasing. $\square$

Given our regularity conditions, we know that the optimal contract satisfies

$$\Phi_x(x(\theta), \theta) = 0.$$

What does this mean?

Interpretations of $\Phi_x(x(\theta), \theta) = 0$:

1. We can rewrite the optimality condition for x as

$$S_x(x(\theta), \theta) = \frac{1 - P(\theta)}{p(\theta)} u_{x\theta}(x(\theta), \theta) \geq 0.$$

Clearly, there is an under-provision of the contracted activity, x, for all but the highest type. For the highest type, $\theta = 1$, we have the full-information level of activity.

2. Alternatively, we can rewrite the optimality condition for x as

$$p(\theta)S_x(x(\theta),\theta) = [1 - P(\theta)]u_{x\theta}(x(\theta),\theta).$$

Fix a particular value of θ. The LHS represents the marginal gain in joint surplus by increasing $x(\theta)$ to $x(\theta) + dx$. It is multiplied by $p(\theta)$ which represents the probability of a type occurring between θ and $\theta + d\theta$. The RHS represents the marginal cost of increasing x at th: for all higher types, rents will be increased. Remember that the agent's rent increases by u_θ in θ. Because u_θ increases in x, a small increase in x implies that rents will be increased by $u_{x\theta}$ for all types above θ, which exist with probability $1 - P(\theta)$.

It is very similar to the distortion a monopolist introduces to maximize profits. Let the buyer's unit value of a good be distributed according to $F(v)$. The buyer buys a unit iff the value is greater than price, p; marginal cost is constant at c. Thus, the monopolist solves $\max_p[1 - F(p)](p - c)$, which yields as a first-order condition,

$$f(p)(p - c) = [1 - F(p)].$$

Lowering the price increases profits on the marginal consumer (LHS) but lowers profits on all inframarginal customers who would have purchased at the higher price (RHS).

3. Following Myerson [1981], we can redefine the agent's utility as a *virtual utility* which represents the agent's utility less information rents:

$$\tilde{u}(x,\theta) \equiv u(x,\theta) - \frac{1 - P(\theta)}{p(\theta)}u_\theta(x,\theta).$$

The principal maximizes the sum of the virtual utilities. This terminology is particular useful in the study of optimal auctions.

Remarks on Basic Paradigm:

1. If it is unreasonable to assume A.3 for the economic problem under study, one must maximize the un-relaxed program including a constraint for the monotonicity condition. The technique is straightforward, but tedious. It involves optimally "ironing out" the decision function found in the relaxed program. That is, carefully choosing intervals where x is made constant, but otherwise following the relaxed choice of x. The result is that there will be regions of pooling in the optimal contract, but in non-pooling regions, it will be the same as before. Additionally, there will typically (although not always) be no distortion at the top and insufficient x for all other types.

2. A few brave researchers have extended the optimality results to cases where there are several dimensions of private information. Roughly, the trick is to note that a multidimensional incentive problem can be converted to a single-dimensional one by defining a new type using the agent's indifference curves over the old type space. Mathematically, rather than integrating by parts, Stoke's theorem can be used. See the cites in F&T for more information, or look at Wilson's [1993] book, *Nonlinear Pricing*, or at the papers by Armstrong [1996] and Rochet [1995].

3. Risk aversion (i.e., non-quasi-linear preferences) does not affect the implementability theorem significantly, but does affect the choice of contract. Importantly, wealth effects may alter the optimal contract dramatically. Salanie [1990] and Laffont and Rochet [1994] consider these problems in detail finding that a region of pooling at the low-end of the distribution occurs for some intermediate levels of risk aversion.

4. Common agency. With two or more principals, there will be externalities in contract choice. Just like two duopolists erode some of the available monopoly profits, so will two principals. What's interesting is the conduit for the erosion. Simply stated, principals competing over substitute (complementary) activities will reduce (increase) the distortion the agent faces. See Martimort [1992,1996] and Stole [1990] for details.

5. We've assumed that the agent knows more than the principal. But suppose that its the other way around. Now it is possible that the principal will signal her private information in the contract offer. Thus, we are looking at signaling contracts rather than screening contracts. The results are useful for our purposes, however, as sometimes we may want to consider renegotiation led by a privately informed agent. We'll talk more about these issues later in the course. The relevant papers are Maskin and Tirole [1990,1992].

6. We limited our attention to deterministic mechanisms when searching for the optimal mechanism. Could stochastic mechanisms do any better? If the surplus function is concave in x, the only possible value is that a stochastic mechanism might reduce the rent term of the agent; i.e., u_θ may be concave. Most researchers assume that $u_{\theta xx} \geq 0$ which is sufficient to rule out stochastic mechanisms.

7. There is a completely different approach to optimal contract design which focuses on probability distributions over cumulative economic activity at given tariffs rather than distributions over types. This approach was first used by Goldman-Leland-Sibley [1984] and has recently been put to great use in Wilson's *Nonlinear Pricing*. Of course, it yields identical answers, but with very different mathematics. My sense is that this framework may give testable implications for demand data more directly than using the approach developed above.

8. We have derived the optimal direct revelation mechanism contract. Are there economically reasonable indirect mechanisms which yield the same allocations and that we expect to see played? Two come to mind. First, because x and t are monotonic, we can construct a nonlinear tariff, $T(x)$, which implements the same allocation; here, the agent is allowed to choose from the menu. Second, if $T(x)$ is concave (which we will see occurs frequently), an optimal indirect mechanism of the form of a menu of linear contracts exists, where the agent chooses a particular two-part tariff and can consume anywhere along it. This type of contract has a particularly nice robustness against noise, which we will see when we study Laffont-Tirole [1986], where the idea of a menu of two-part tariff was developed.

9. Many researchers use control theory to solve the problem rather than integration by parts and pointwise maximization. The cost of this approach is that the standard sufficient conditions for an optimal solution in control theory are stronger than what

we used above. Additionally, a Hamiltonian has no immediately clear economic interpretation. The benefits are that sometimes the tricks used above cannot be used. Control theory is far more powerful and general than our simple integration by parts trick. So for complicated problems, it is something which can be useful. The basic idea is to treat θ as the state variable just as engineers would treat time. The control variable is x and the co-state variable is indirect utility, U. The Hamiltonian becomes

$$\mathcal{H}(x, U, \theta) \equiv (S(x, \theta) - U)p(\theta) + \lambda(\theta)u_\theta(x, \theta).$$

Roughly speaking, providing x is piecewise-C^1 and $\mathcal{H}$ is globally strictly concave in (x, U) for any λ (this puts lots of restrictions on u and v), the following conditions are necessary and sufficient for an optimum:

$$\mathcal{H}_x(x(\theta), U(\theta), \theta) = 0,$$

$$-\mathcal{H}_U(x, U, \theta) = \lambda'(\theta),$$

$$\lambda(1) = 0.$$

Solving these equations yields the same solution as above. Weaker versions of the concavity conditions are available; see for example Seierstad and Sydsaeter's [1987] control theory book for details.

3.2.3 Finite Distribution of Types

Rather than use continuous distributions of types and deal with the functional analysis messes (Lipschitz conditions, etc.), some of the literature has used finite distributions. Most notable are Hart [1983] and Moore [1988]. The approach (first characterize implementable contracts, then optimize) is the same. The techniques of the proofs are different enough to warrant some attention.

For the purposes of this section, suppose that finite versions of A.2-A.3 still hold, but that there are only n types, $\theta_1 < \theta_2, \ldots, \theta_{n-1} < \theta_n$ with probability "density" p_i for each type and "distribution" function $P_i \equiv \sum_{j=1}^{i} p_j$. Thus, $P_1 = p_1$ and $P_n = 1$. Let $y_i = (x_i, t_i)$ represent the allocation to the agent claiming to be the ith type.

Implementable Contracts   Our principal's program is to

$$\max_{y_i} \sum_{i=1}^{n} p_i \left\{ S(x_i, \theta_i) - U(\theta_i) \right\},$$

subject to, $\forall i, j,,$

$$U(\theta_i|\theta_i) \geq U(\theta_j|\theta_i), \qquad\qquad\qquad \text{(IC(i,j))}$$

and

$$U(\theta_i|\theta_i) \geq \underline{U}. \qquad\qquad\qquad\qquad \text{(IR(i))}$$

Generally, we can eliminate many of these constraints and focus on local incentive compatibility.

Theorem 23 If $u_{x\theta} \geq 0$, then the local constraints

$$U(\theta_i|\theta_i) \geq U(\theta_{i-1}|\theta_i) \qquad\qquad\qquad \text{(DLIC(i))}$$

and

$$U(\theta_i|\theta_i) \geq U(\theta_{i+1}|\theta_i) \qquad\qquad\qquad \text{(ULIC(i))}$$

satisfied for all i are necessary and sufficient for global incentive compatibility.

Proof: Necessity is direct. Sufficiency is proven by induction.

First, note that the local constraints imply that $x_i \geq x_{i-1}$. Specifically, the local constraints imply

$$U(\theta_i|\theta_i) - U(\theta_{i-1}|\theta_i) \geq 0 \geq U(\theta_i|\theta_{i-1}) - U(\theta_{i-1}|\theta_{i-1}), \quad \forall\, i.$$

Rearranging, and using direct utilities, we have

$$u(x_i, \theta_i) - u(x_{i-1}, \theta_i) \geq u(x_i, \theta_{i-1}) - u(x_{i-1}, \theta_{i-1}).$$

Combining this inequality with the sorting condition implies monotonicity.

Consider DLIC for type i and $i - 1$. Restated in direct utility terms, these conditions are

$$u(x_i, \theta_i) - u(x_{i-1}, \theta_i) \geq t_i - t_{i-1},$$
$$u(x_{i-1}, \theta_{i-1}) - u(x_{i-2}, \theta_{i-1}) \geq t_{i-1} - t_{i-2}.$$

Adding the conditions imply

$$u(x_i, \theta_i) - u(x_{i-1}, \theta_i) + u(x_{i-1}, \theta_{i-1}) - u(x_{i-2}, \theta_{i-1}) \geq t_i - t_{i-2}.$$

By the sorting condition and monotonicity, the LHS is smaller than

$$u(x_i, \theta_i) - u(x_{i-1}, \theta_i) + u(x_{i-1}, \theta_i) - u(x_{i-2}, \theta_i) = u(x_i, \theta_i) - u(x_{i-2}, \theta_i),$$

and so IC(i,i-2) is satisfied:

$$u(x_i, \theta_i) - u(x_{i-2}, \theta_i) \geq t_i - t_{i-2}.$$

Thus, DLIC(i) and DLIC(i-1) imply IC(i,i-2). One can show that IC(i,i-1) and DLIC(i-2) imply IC(i,i-3), etc. Therefore, starting at $i = n$ and proceeding inductively, DLIC implies IC(i,j) holds for all $i \geq j$. A similar argument in the reverse direction establishes that ULIC implies IC(i,j) for $i \leq j$. $\square$

The basic idea of the theorem is that the local upward constraints imply global upward constraints, and likewise for the downward constraints. We have reduced our necessary and sufficient IC constraints from $n(n-1)$ to $2(n-1)$ constraints. We can now optimize using Kuhn-Tucker's theorem. If possible, however, it is better to check to see if we can simplify things a bit more.

We can still do better for our *particular* problem. Consider the following **relaxed program**.

$$\max_{y_i} \sum_{i=1}^{n} p_i \left\{ S(x_i, \theta_i) - U(\theta_i) \right\},$$

59

subject to DLIC(i) for every i, IR(1), and x_i nondecreasing in i.

We will demonstrate that

Theorem 24 The solution to the unrelaxed program is equivalent to the solution of the relaxed program.

Proof: The proof proceeds in 3 steps.

Step 1: The constraints of the unrelaxed program imply those of the relaxed program. It is easy to see that IC(i,j) imply DLIC(i) and IR(i) imply IR(1). Take $i > j$. By IC(i,j) and IC(j,i) we have

$$u(x_i, \theta_i) - t_i \geq u(x_j, \theta_i) - t_j,$$
$$u(x_j, \theta_j) - t_j \geq u(x_i, \theta_j) - t_i.$$

Adding and rearranging,

$$[u(x_i, \theta_i) - u(x_j, \theta_i)] - [u(x_i, \theta_j) - u(x_j, \theta_j)] \geq 0.$$

By the sorting condition, if $\theta_i > \theta_j$, then $x_i \geq x_j$.

Step 2: At the solution of the relaxed program, DLIC(i) is binding for all i. Suppose not. Take i and ε such that $[u(x_i, \theta_i) - t_i] - [u(x_{i-1}, \theta_i) - t_{i-1}] > \varepsilon > 0$. Now for all $j \geq i$, raise transfers to $t_j + \varepsilon$. No IC constraints will be violated and profit is raised by $(1 - P_{i-1})\varepsilon$, which contradicts $\{x, t\}$ being a solution to the relaxed program.

Step 3: The solution of the relaxed program satisfies the constraints of the unrelaxed program. Because DLIC(i) is binding, we have

$$u(x_i, \theta_i) - u(x_{i-1}, \theta_i) = t_i - t_{i-1}.$$

By monotonicity and the sorting condition,

$$u(x_i, \theta_{i-1}) - u(x_{i-1}, \theta_{i-1}) \leq t_i - t_{i-1}.$$

But this latter condition is ULIC(i-1). Hence, DLIC and ULIC are satisfied. By Theorem 23, this is sufficient for global incentive compatibility. Finally, it is straightforward to show that IC(i,j) and IR(1) implies IR(i). $\square$

Optimal Contracts We now solve the simpler relaxed program. Note that there are now only n constraints, all of which are binding so we can use Lagrangian analysis rather than Kuhn-Tucker conditions given appropriate assumptions of concavity and convexity.

We solve

$$\max_{x_i, U_i} \mathcal{L} = \sum_{i=1}^{n} p_i \{ S(x_i, \theta_i) - U_i \} +$$

$$\sum_{i=2}^{n} \lambda_i (U_i - U_{i-1} - u(x_{i-1}, \theta_i) + u(x_{i-1}, \theta_{i-1})) + \lambda_1 (U_1 - \underline{U}),$$

ignoring the monotonicity constraint for now. We have used $U_i \equiv U(\theta_i)$ instead of transfers as our primitive instruments, along the lines of the control theory approach. [Using indirect utilities, the DLIC constraints become $U_i - U_{i-1} - u(x_{i-1}, \theta_i) + u(x_{i-1}, \theta_{i-1}) \geq 0$.] There are $2n$ necessary first-order conditions:

$$p_i S_x(x_i, \theta_i) = \lambda_{i+1}[u_x(x_i, \theta_{i+1}) - u_x(x_i, \theta_i)], \ i = 1, \ldots, n-1,$$

$$p_n S_x(x_n, \theta_n) = 0,$$

$$-p_i + \lambda_i - \lambda_{i+1} = 0, \ i = 1, \ldots, n-1,$$

$$-p_n + \lambda_n = 0.$$

Combining the last two sets of equations, we have a first-order difference equation which we can solve uniquely for $\lambda_i = \sum_{j=i}^{n} p_j$. Thus, assuming discrete analogues of A.2 and A.3, we have the following result.

Theorem 25 In the case of a finite distribution of types, the optimal mechanism has x_i to satisfy
$$p_i S_x(x_i, \theta_i) = [1 - P_i](u_x(x_i, \theta_{i+1}) - u_x(x_i, \theta_i)), \ i = 1, ..., n,$$

t_i is chosen as the unique solution to the first-order difference equation,

$$U_i - U_{i-1} = u(x_{i-1}, \theta_i) + u(x_{i-1}, \theta_{i-1}),$$

with initial condition, $U_1 = \underline{U}$.

Remarks:

1. As before, we have no distortion at the top and a suboptimal level of activity for all lower types.

2. Step 2 of the proof of Theorem 24 is frequently skipped by researchers. Be careful because DLIC does not bind in all economic environments. Moreover, it is incorrect to assume DLIC binds (i.e., impose DLIC's with equalities in the relaxed program) and then check that the other constraints are satisfied in the solution to the relaxed program. This does not guarantee an optimum!!! Either you must show that the constraints are binding or you must use Kuhn-Tucker analysis which allows the constraints to bind or be slack. This is important.

   In many papers, it is not the case that the DLIC constraints bind. See, Hart, [1983] for example. In fact, in Stole [1996], there is an example of a price discrimination problem where the upward local IC constraints bind rather than the downward ones. This is generated by adding noise to the reservation utility of consumers. As a consequence, you may want to leave rents to some types to increase the chances that they will visit your store, but then the logic of step 2 does not work and in fact the upward constraints may bind.

3. Discrete models sometimes generate results which are quite different from those which emerge from the continuous-type settings. For example, in Stole [1996], the discrete setting with random IR constraints exhibits a downward distortion for the lowest type which is absent in the continuous setting. As another example, provided by David Martimort, in a Riley and Samuelson [1981] auction setting with a continuum of types and the seller's reservation price set equal to the buyer's lowest valuation, there is a multiplicity of symmetric equilibria (this is due to a singularity in the characterizing differential equation at $\theta = \underline{\theta}$). With discrete types, a unique symmetric equilibrium exists.

4. It is frequently convenient to use two-type models to get the feel for an economic problem before going on to tackle the n-type or continuous-type case. While this is usually helpful, some economic phenomena will only be visible in $n \geq 3$ environments. (E.g., common agency models with complements generate the first-best as an outcome with two types but not with three or more.) Fundamentally, the IC constraints in the discrete setting are much more rigid with two types than with a continuum. Nonetheless, much can be learned from two-type models, although it would certainly be better if the result could be generalized to more.

3.2.4 Application: Nonlinear Pricing

We now turn to the case of nonlinear pricing developed by Goldman, Leland and Sibley [1984], Mussa-Rosen [1978], and Maskin-Riley [1984]. Given the development for the single-agent case above, this is a simple exercise of applying our theorems.

Basic Model: Consider a monopolist who faces a population of customers with varying marginal valuations for its product. Let the consumers be indexed by type, $\theta \in [0,1]$, with utility functions $U = u(x,\theta) - t$. We will assume that a higher type customer receives both greater total and greater marginal utility from consumption, so A.1 is satisfied. The monopolist's payoffs for a given sale of x units is $V = t - C(x)$. The monopolist wishes to find the optimal nonlinear price schedule to offer its customers.

Note that $\Phi(x,\theta) \equiv u(x,\theta) - C(x) - \frac{1-P(\theta)}{p(\theta)} u_\theta(x,\theta)$ in this setting. We further assume that A.2 and A.3 are satisfied for this function.

Results:

1. Theorems 3.2.2, 21 and 22 imply that the optimal nonlinear contract satisfies

$$p(\theta)(u_x(x,\theta) - C_x(x)) = [1 - P(\theta)]u_{x\theta}(x,\theta).$$

Thus we have the result that quantity is optimally provided for the highest type and under-provided for lower types.

2. If we are prepared to assume a simpler functional form for utility, we can simplify this expression further. Let $u(x,\theta) \equiv \theta\nu(x)$. Then,

$$p(\theta)(\theta\nu_x(x) - C_x(x)) = [1 - P(\theta)]\nu_x(x).$$

Note that the marginal price a consumer of type θ pays for the marginal unit purchased is $T_x(x(\theta)) \equiv t'(\theta)/x'(\theta) = u_x(x(\theta), \theta)$. Rearranging our optimality condition, we have

$$\frac{T_x - C_x}{T_x} = \frac{1 - P(\theta)}{\theta p(\theta)}.$$

If P satisfies the monotone hazard-rate condition, then the Lerner index at the optimal allocation is decreasing in type. Note also that the RHS can be reinterpreted as the inverse virtual elasticity of demand.

3. Returning to our general model, $U = u(x, \theta) - t$, if we assume that marginal costs are constant, MHRC holds, $u_{x\theta\theta} \leq 0$, and $u_{xx\theta} \leq 0$, we can show that quantity discounts are optimal. [Note, all of these conditions are implied by the simple model above in result 2.] To show this, redefine the marginal nonlinear tariff schedule as $T_x(x) = u_x(x, x^{-1}(x))$, where $x^{-1}(x)$ gives the type θ which is willing to buy exactly x units; note that $T_x(x)$ is independent of θ. We want to show that $T(x)$ is a strictly concave function. Differentiating, $T_{xx} < 0$ is equivalent to

$$\frac{dx}{d\theta} > -\frac{u_{x\theta}(x, \theta)}{u_{xx}(x, \theta)}.$$

Because $x'(\theta) = -\Phi_{x\theta}/\Phi_{xx}$, the condition becomes

$$\frac{dx}{d\theta} = \frac{u_{x\theta} - \frac{1-P}{p} u_{x\theta\theta} - \left(\frac{d}{d\theta}\frac{1-P}{p}\right) u_{x\theta}}{-u_{xx} + \frac{1-P}{p} u_{x\theta\theta}} > -\frac{u_{x\theta}}{u_{xx}},$$

which is satisfied given our assumptions.

4. We could have instead written this model in terms of quality rather than quantity, where each consumer has unit demands but differing marginal valuations for quality. Nothing changes in the analysis. Just reinterpret x as quality.

3.2.5 Application: Regulation

The seminal papers in the theory of regulating a monopolist with unknown costs are Baron and Myerson [1982] and Laffont and Tirole [1986]. We will illustrate the main results of L&T here, and leave the results of B&M for a problem set or recitation.

Basic Model:

A regulated firm has private information about its costs, $\theta \in [\underline{\theta}, \overline{\theta}]$, distributed according to P, which we will assume satisfies the MHRC. The firm exerts an effort level, e, which has the effect of reducing the firm's marginal cost of production. The total cost of production is $C(q) = (\theta - e)q$. This effort, however, is costly; the firm's costs of effort are given by $\psi(e)$, which is increasing, strictly convex, and $\psi'''(e) \geq 0$. This last condition will imply that A.2 and A.3 are satisfied (as well as the optimality of non-random contracts). It is assumed that the regulator can observe costs, and so without loss of generality, we assume that the regulator pays the observed costs of production rather than the firm. Hence, the firm's utility is given by

$$U = t - \psi(e).$$

Laffont and Tirole assume a novel contracting structure in which the regulator can observe costs, but must determine how much of the costs are attributable to effort and how much are attributed to inherent luck (i.e., type). The regulator cannot observe e but can observe and contract upon total production costs C and output q. For any given q, the regulator can perfectly determine the firm's marginal cost $c = \theta - e$. Thus, the regulator can ask the firm to report its type and assign the firm a marginal cost target of $c(\hat{\theta})$ and an output level of $q(\hat{\theta})$ in exchange for compensation equal to $t(\hat{\theta})$. A firm with type θ that wishes to make the marginal cost target of $c(\hat{\theta})$ must expend effort equal to $e = \theta - c(\hat{\theta})$. With such a contract, the firm's indirect utility function becomes

$$U(\hat{\theta}|\theta) \equiv t(\hat{\theta}) - \psi(\theta - c(\hat{\theta})).$$

Note that this utility function is independent of $q(\hat{\theta})$ and that the sorting condition A.1 is satisfied if one normalizes θ to $-\theta$.

Theorem 3.2.2 implies that incentive compatible contracts are equivalent to requiring that

$$\frac{dU(\theta)}{d\theta} = -\psi'(\theta - c(\theta)),$$

and that $c(\theta)$ be nondecreasing.

To solve for the optimal contract, we need to state the regulator's objectives. Let's suppose that the regulator wishes to maximize a weighted average of strictly concave consumer surplus, $CS(q)$, (less costs and transfers) and producer surplus, U, with less weight afforded to the latter.

$$V = E_\theta[CS(q(\theta)) - c(\theta)q(\theta) - t(\theta) + \gamma U(\theta)],$$

or substituting out the transfer function,

$$V = E_\theta[CS(q(\theta)) - c(\theta)q(\theta) - \psi(\theta - c(\theta)) - (1 - \gamma)U(\theta)],$$

where $0 \leq \gamma < 1$.

Remarks:

1. In our above development, we could instead write $S = CS - cq - \psi$, and then the regulator maximizes $E[S - (1 - \gamma)U]$. If $\gamma = 0$, we are in the same situation as our initial framework where the principal doesn't directly value the agent's utility.

2. L&T motivate the cost of leaving rents to the firm as arising from the shadow cost of raising public funds. In that case, if $1 + \lambda$ is the cost of public funds, the regulator's objective function (after simplification) is $E[CS - (1 + \lambda)(cq + \psi) - \lambda U]$. Except for the optimal choice of q, this yields identical results as using $1 - \gamma \equiv \frac{\lambda}{1+\lambda}$.

Our regulator solves the following program:

$$\max_{q,c,t} E_\theta[CS(q(\theta)) - c(\theta)q(\theta) - \psi(\theta - c(\theta)) - (1 - \gamma)U(\theta)],$$

subject to

$$\frac{dU}{d\theta} = -\psi'(\theta - c(\theta)),$$

$c(\theta)$ nondecreasing, and the firm making nonnegative profits (i.e., $U(\theta) \geq 0$). Integrating $E_\theta[U(\theta)]$ by parts (using our previously developed techniques), we can substitute out U from the objective function (thereby eliminating transfers from the program). We then have

$$\max_{q,c} E_\theta \left[CS(q(\theta)) - c(\theta)q - \psi(\theta - c(\theta)) - (1-\gamma)\frac{P(\theta)}{p(\theta)}\psi'(\theta - c(\theta)) - (1-\gamma)U(\bar{\theta}) \right],$$

subject to transfers satisfying

$$\frac{dU}{d\theta} = -\psi'(\theta - c(\theta)),$$

$c(\theta)$ nondecreasing, and $U(\bar{\theta}) \geq 0$. We obtain the following results.

Results: Redefine for a given q

$$\Phi(c, \theta) \equiv CS(q) - cq - \psi(\theta - c) - (1 - \gamma)\frac{P(\theta)}{p(\theta)}\psi'(\theta - c).$$

A.2 and A.3 are satisfied given our conditions on P and ψ so we can apply Theorems 21 and 22.

1. The choice of effort, $e(\theta)$, satisfies

$$q(\theta) - \psi'(e(\theta)) = (1 - \gamma)\frac{P(\theta)}{p(\theta)}\psi''(e(\theta)) \geq 0.$$

   Note that the first-best level of effort, conditional on q, is $q = \psi'(e)$. As a consequence, suboptimal effort is provided for every type except the lowest. We always have no distortion on the bottom. Only if $\gamma = 1$, so the regulator places equal weight on the firm's surplus, do we have no distortion anywhere. In the L&T setting, this condition translates to $\lambda = 0$; i.e., no excess cost of public funds.

2. The optimal q is the full-information efficient production level conditional on the marginal cost $c(\theta) \equiv \theta - e(\theta)$:
   $$CS'(q) = c(\theta).$$

   This is not the result of L&T. They find that because public funds are costly, $CS'(q) = (1 + \lambda)c(\theta)$, where the RHS represents the effective marginal cost (taking into account public funds). Nonetheless, this solution still corresponds to the choice of q under full-information for a given marginal cost, because the cost of public funds is independent of informational issues. Hence, there is a dichotomy between pricing (choosing $c(\theta)$) and production.

3. Because $\psi''' \geq 0$, random schemes are never optimal.

4. L&T also show that the optimal nonlinear contract can be implemented using a realistic menu of two-part tariffs (i.e., cost-sharing contracts). Let $\overline{C}(q)$ be a total cost target. The firm chooses an output and a corresponding cost target, $\overline{C}(q)$, and then is compensated according to

$$T(q, C) = F(q) + \alpha[\overline{C}(q) - C],$$

where C is observed ex post cost. F is interpreted as a fixed payment and α is a cost-sharing parameter. If $(\bar{\theta} - \underline{\theta})$ goes to zero, α goes to one, implying a fixed-price contract in which the firm absorbs any cost overruns. Of course, this framework doesn't make much sense with no other uncertainty in the model (i.e., there would never be cost overruns which we would observe). But if some noise is introduced on the ex post cost observation, i.e. $\tilde{C} = C(q) + \varepsilon$, the mechanism is still optimal. It is robust to linear noise in observed contract variables because the firm is risk neutral and the implemented contract is linear in the observation noise.

3.2.6 Resource Allocation Devices with Multiple Agents

We now consider problems with multiple agents, and so we will reintroduce our subscripts, $i = 1, \ldots, I$ to denote the different agents. We will also assume that the agent's types are independently distributed (but not necessarily identically), and so we will denote the individual density and distribution functions for $\theta_i \in [\underline{\theta}, \bar{\theta}]$ as $p_i(\theta_i)$ and $P_i(\theta_i)$, respectively. Also, we will sometimes use $p_{-i}(\theta_{-i}) \equiv \prod_{j \neq i} p_j(\theta_j)$.

Optimal Auctions This section closely follows Myerson [1981], albeit with different notation and some simplification on the type spaces.

We restrict attention to a single-unit auction in which the good is awarded to a one individual and consider Bayesian Nash implementation. There are I potential bidders for the object. The participants' expected utilities are given by

$$U_i = \phi_i \theta_i - t_i,$$

where ϕ_i is participant i's probability of receiving the good in the auction and θ_i is the marginal valuation for the good. t_i is the payment of the ith player to the principal (auctioneer). Note that because the agents are risk neutral, it is without loss of generality to consider payments made independently of getting the good. We will require that all bidders receive at least nonnegative payoffs: $U_i \geq 0$. This setup is known as the *Independent Private Values (IPV)* model of auctions. The "private" part refers to the fact that an individual's valuation is independent of what others think. As such, think of the object as a personal consumption good that will not be re-traded in the future.

The principal's direct mechanism is given by $y = (\phi, t)$. The agent's indirect utility functions given this mechanism can be stated as

$$U_i(\hat{\theta}|\theta_i) \equiv \phi_i(\hat{\theta})\theta_i - t_i(\hat{\theta}).$$

Note that the probability of winning and the transfer can depend upon everyone's announced type (as is typical in an auction setting). To simplify notation, we will indicate that the expectation of a function has been taken by eliminating the variables of integration from the function's argument. Specifically, define

$$\phi_i(\theta_i) \equiv E_{\theta_{-i}}[\phi_i(\theta)],$$

and

$$t_i(\theta_i) \equiv E_{\theta_{-i}}[t_i(\theta)].$$

Note that there is enormous flexibility in choosing actual transfers $t_i(\theta)$ to attain a specific $t_i(\theta_i)$ for implementability. Thus, in a truth-telling Bayesian-Nash equilibrium,

$$U_i(\hat{\theta}_i|\theta_i) \equiv E_{\theta_{-1}}[U_i(\theta_{-i}, \hat{\theta}_i|\theta_i)] = \phi_i(\hat{\theta}_i)\theta_i - t_i(\hat{\theta}_i).$$

Let $U_i(\theta_i) \equiv U_i(\hat{\theta}_i|\theta_i)$. Following Theorem 3.2.2, we have

Theorem 26 An auction mechanism with $\phi_i(\theta_i)$ continuous and absolutely continuous first derivative is incentive compatible iff

$$\frac{dU_i(\theta_i)}{d\theta_i} = \phi_i(\theta_i),$$

and $\phi_i(\theta_i)$ is nondecreasing in θ_i.

The proof proceeds as in in Theorem 3.2.2. We are now in a position to examine the expected value of an incentive compatible auction.

The principal's objective function is to maximize expected payments taking into account the principal's own value of the object, which we take to be θ_0. Specifically,

$$\max_{\phi \in \Delta(I), t} E_\theta \left[\left(1 - \sum_{i=1}^{I} \phi_i(\theta) \right) \theta_0 + \sum_{i=1}^{I} \phi_i(\theta)\theta_i - \sum_{i=1}^{I} U(\theta) \right],$$

subject to IC and IR. $\Delta(I)$ is the $I - 1$ dimensional simplex.

As before, we can substitute out the indirect utility functions by integrating by parts. We are thus left the following objective function:

$$E_\theta \left[\left(1 - \sum_{i=1}^{I} \phi_i(\theta) \right) \theta_0 + \sum_{i=1}^{I} \phi_i(\theta)\theta_i - \sum_{i=1}^{I} \left(\phi_i(\theta)\frac{1 - P_i(\theta_i)}{p_i(\theta_i)} + U_i(0) \right) \right].$$

Rearranging the expression, we obtain

$$E_\theta \left[\theta_0 + \sum_{i=1}^{I} \phi_i(\theta) \left(\theta_i - \frac{1 - P_i(\theta_i)}{p_i(\theta_i)} - \theta_0 \right) - \sum_{i=1}^{I} U_i(0) \right]. \tag{9}$$

This last expression states the expected value of the auction independently of the transfer function. The expected value of the auction is completely determined by ϕ and $U(0) \equiv (U_1(0), \ldots, U_I(0))$. Any two auctions with the same functions have the same expected revenue.

Theorem 27 Revenue Equivalence. The seller's expected utility from an implementable auction is completely determined by the probability functions, ϕ, and the numbers, $U_i(0)$.

The proof follows from inspection of the expected revenue function, (9). The result is quite powerful. Consider the case of symmetric distributions of types, $p_i \equiv p$. The result implies that in the class of auctions which award the good to the highest value bidder and leave no rents to the lowest possible bidder (i.e., $U_i(0) = 0$), the expected revenue to the seller is the same. With the appropriately chosen reservation prices, the first-price auction,

the second-price auction, the Dutch auction and the English auction all belong to this class! This extends Vickrey's [1961] famous equivalence result. Note that these auctions are not always optimal, however.

Back to optimality. Because the problem requires that $\phi \in \Delta(I-1)$, it is likely that we have a corner solution which prevents us from using first-order calculus techniques. As such, we do not redefine Φ and check A.2 and A.3. Instead, we will solve the problem directly. To this end, define

$$J_i(\theta_i) \equiv \theta_i - \frac{1 - P_i(\theta_i)}{p_i(\theta_i)}.$$

This is Myerson's virtual utility or virtual type for agent i with type θ_i. The principal thus wishes to maximize

$$E_\theta \left[\sum_{i=1}^{I} \phi_i(\theta) \left(J_i(\theta_i) - \theta_0 \right) - U_i(0) \right],$$

subject to $\phi \in \Delta(I)$, monotonicity and $U_i(\theta_i) \geq 0$. We now state our result.

Theorem 28 Assume that each P_i satisfies the MHRC. Then the optimal auction has ϕ chosen such that

$$\phi_i(\theta) = \begin{cases} 1 & \text{if } J_i(\theta_i) > \max_{k \neq i} J_k(\theta_k) \text{ and } J_i(\theta_i) \geq \theta_0, \\ \in [0,1] & \text{if } J_i(\theta_i) = \max_{k \neq i} J_k(\theta_k) \text{ and } J_i(\theta_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases}$$

The lowest types receive no rents, $U_i(0) = 0$, and transfers satisfy the differential equation in Theorem 26.

Proof: Note first that the choice of ϕ in the theorem satisfies $\phi \in \Delta(I-1)$ and maximizes the value of (9). The choice of $U_i(0) = 0$ satisfies the participation constraints of the agents while maximizing profits. Lastly, the transfers are chosen so as to satisfy the differential equation in Theorem 26. Providing that $\phi_i(\theta_i)$ is nondecreasing, this implies incentive compatibility. To see that this monotonicity holds, note that $\phi_i(\theta)$ (weakly) increases as $J_i(\theta_i)$ increases, holding all other θ_{-i} fixed. The assumption of MHRC implies that $J_i(\theta_i)$ is increasing in θ_i, which implies the necessary monotonicity. $\square$

The result is that the optimal auction awards the good to the agent with the highest virtual type, providing that the type exceeds the seller's opportunity cost, θ_0.

Remarks:

1. There are two distortions which the principal introduces, underconsumption and misallocation. First, sometimes the good will not be consumed even though an agent values it more than the principal:

$$\max_i J_i(\theta_i) < \theta_0 < \max_i \theta_i.$$

Second, sometimes the wrong agent will consume the good:

$$\arg\max_i J_i(\theta_i) \neq \arg\max_i \theta_i.$$

2. The expected revenue of the optimal auction normally exceeds that of the standard English, Dutch, first-price, and second-price auctions. One reason is that if the distributions are discrete, the principal can elicit truth-telling more cheaply.

   Second, unless type distributions are symmetric, the J_i functions are asymmetric, which in turn implies that the highest value agent should not always get the item. In the four standard auctions, the highest valuation agent typically gets the good. By handicapping agents with more favorable distributions, however, you can encourage them to bid more aggressively. For example, let $\theta_0 = 0$, $I = 2$, θ_1 be uniformly distributed on $[0, 2]$ and θ_2 be uniformly distributed on $[2, 4]$. We have $J_1(\theta_1) \equiv 2(\theta_1 - 1)$ and $J_2(\theta_2) \equiv 2\theta_2 - 4 = J_1(\theta_2) - 2$. Agent 2 is handicapped by 2 relative to agent 1 in this mechanism (i.e., agent 1 must have a value in excess of agent 2 by at least 2 in order to get the good). In contrast, under a first-price auction, agent 2 always gets the good, submitting a bid of only 2.

3. With correlation, the principal can do even better. See Myerson [1981], for an example, and Crémer and McLean [1985,1988] for more details. There is also a literature beginning with Milgrom and Weber [1982] on common value (more precisely, affiliated value) auctions.

4. With risk aversion, the revenue equivalence theorem fails to apply. Here, first-price generally outperforms second-price, for example. The idea is that if you are risk averse, you will bid more aggressively in a first-price auction because bidding close to your valuation reduces the risk in your final rent. See Maskin-Riley [1984], Matthews [1983] and Milgrom and Weber [1982] for more on this subject.

5. Maskin and Riley [1990] consider multi-unit auctions where agents have multi-unit demands. The result is a combination of the above result on virtual valuations allocation and the result from price-discrimination regarding the level of consumption for those who consume in the auction.

6. Although the revenue equivalence theorem tells us that the four standard auctions are equivalent in revenue, it is in the context of Bayesian implementation. Note, however, that the second-price sealed bid auction has a unique dominant-strategy equilibrium. Thus, revenue may not be the only relevant dimension over which we should value a mechanism.

Bilateral (Multilateral) Trading   This section is based upon Myerson-Satterthwaite [1983]. Here, we reproduce their characterization of implementable trading mechanisms between a buyer and a seller with privately known valuations of trade. We then derive the nature of the optimal mechanism. We put off until later the discussion of efficiency properties of mechanisms with balanced budgets.

Basic Model of a Double Auction:

   The basic bilateral trade model of Myerson and Satterthwaite has two agents: a seller and a buyer. There is no principal. Instead think of the optimization problem as the design of a mechanism by the buyer and seller before they know their types, but under the condition that after learning their types, either party may walk away from the agreement.

Additionally, it is assumed that money can only be transferred from one party to the other. There is no outside party that can break the budget. The seller and the buyer's valuations for the single unit of good are $c \in [\underline{c}, \overline{c}]$ and $v \in [\underline{v}, \overline{v}]$ respectively; the distributions are $P_1(c)$ for the seller and $P_2(v)$ for the buyer. [We'll use v and c rather than the θ_i as they're more descriptive.]

An allocation is given by $y = (\phi, t)$ where $\phi \in [0, 1]$ is the probability of trade and t is a transfer from the buyer to the seller. Thus, the indirect utilities are

$$U_1(\hat{c}, v | c) \equiv t(\hat{c}, v) - \phi(\hat{c}, v)c,$$

and

$$U_2(c, \hat{v} | v) \equiv \phi(c, \hat{v})v - t(c, \hat{v}).$$

Using the appropriate expectations, in a truth-telling equilibrium we have

$$U_1(\hat{c} | c) \equiv t(\hat{c}) - \phi(\hat{c})c,$$

and

$$U_2(\hat{v} | v) \equiv \phi(\hat{v})v - t(\hat{v}).$$

Characterization of Implementable Contracts: M&S provide the following useful characterization theorem.

Theorem 29 For any probability function $\phi(c, v)$, there exists a transfer function t such that $y = (\phi, t)$ is IC and IR iff

$$E_{v,c}\left[\phi(c, v)\left\{\left(v - \frac{1 - P_2(v)}{p_2(v)}\right) - \left(c + \frac{P_1(c)}{p_1(c)}\right)\right\}\right] \geq 0, \tag{10}$$

$\phi(v)$ is nondecreasing, and $\phi(c)$ is nonincreasing.

Sketch of Proof: The proof of this claim is straightforward. Necessity follows from our standard arguments. Note first that substituting out the transfer function from the two indirect utility functions (which was can do sense the transfers must be equal under budget balance) and taking expectations implies

$$E_{c,v}[U_1(c) + U_2(v)] = E_{c,v}[\phi(c, v)(v - c)].$$

Using the standard arguments presented above, one can show that this is also equivalent to

$$E_{c,v}\left[U_1(\overline{c}) + \phi(c, v)\left(\frac{P_2(c)}{p_2(c)}\right) + U_2(\underline{v}) + \phi(c, v)\left(\frac{1 + P_1(v)}{p_1(v)}\right)\right].$$

Rearranging the expression and imposing individual rationality implies (10). Monotonicity is proved using the standard arguments. Sufficiency is a bit trickier. It involves finding the solution to the partial differential equations (first-order conditions) for incentive compatibility. This solution, together with monotonicity, is sufficient for truth-telling. See M&S for the details.

We now are prepared to find the optimal bilateral trading mechanism.

Optimal Bilateral Trading Mechanisms:

The "principal" wants to maximize the expected gains from trade,

$$E_{c,v}[\phi(c,v)(v-c)],$$

subject to monotonicity and (10). We will ignore monotonicity and check our solution to see that it is satisfied. Let μ be the constraint (10). Bringing the constraint into the objective function and simplifying, we have

$$\max_{c,v} E_{c,v}\left[\phi(c,v)\left\{(v-c)-\frac{\mu}{1+\mu}\left(\frac{1-P_2(v)}{p_2(v)}-\frac{P_1(c)}{p_1(c)}\right)\right\}\right].$$

Notice that trade occurs in this relaxed program iff

$$v-\frac{\mu}{1+\mu}\left(\frac{1-P_2(v)}{p_2(v)}\right)\geq c+\frac{\mu}{1+\mu}\left(\frac{P_1(c)}{p_1(c)}\right),$$

where $\mu \geq 0$. If we assume that the monotone hazard-rate condition is satisfied for both type distributions, then this ϕ is appropriately monotonic and we have a solution to the full program. Note that if $\mu > 0$, there will generally be inefficiencies in trade. This will be discussed below.

Remarks:

1. Importantly, M&S show that when $\bar{c} > \underline{v}$ and $\bar{v} > \underline{c}$ (i.e., efficient trading is state dependent), $\mu > 0$, so the full-information efficient level of trading is impossible! The proof is to show that substituting the efficient ϕ into the constraint (10) violates the inequality.

2. Chatterjee and Samuelson [1983] show that in a simple game in which each agent (buyer and seller) simultaneously submits an offer (i.e., a price at which to buy or to sell) and trade takes place at the average price iff the buyer's offer exceeds the seller's offer, if types are uniformly distributed then a Nash equilibrium in linear bidding strategies exists and achieves the upper bound for bilateral trading established by Myerson and Satterthwaite.

3. A generalization of this result to multi-lateral bargaining contexts is found in Cramton, Gibbons, and Klemperer [1987]. They look at the dissolution of a partnership, in which unlike the buyer-seller example where the seller owned the entire good, each member of the partnership may have some property claims on the partnership. The question is whether the partnership can be dissolved efficiently (i.e., the highest value partner buying out all other partners). They show that if the ownership is sufficiently distributed, that it is indeed possible.

Feasible Allocations and Efficiency   There is an old but important literature concerning the implementation of optimal public choice rules, such as when to build a bridge. Bridges cost money, so it is efficient to build one only if the sum of the individual agents' values exceed the costs. The question is how to get agents to truthfully state their valuations (e.g., no one exaggerates their value to change the probability of building the bridge to their own benefit); i.e., how do you avoid the classical "free-rider" problem.

Three important results exist. First, if one ignores budget balance and individual rationality constraints, it is possible to implement the optimal public choice rule in dominant strategies. This is the contribution of Clarke [1971] and Groves [1973]. Second, if one requires budget balance, one can still implement the optimal rule if one uses Bayesian-Nash implementability. This is the result of d'Aspremont and Gerard-Varet [1979]. Finally, if one wants budget balance and individual rationality, efficient allocation is not generally possible even under the Bayesian-Nash concept, as shown by Myerson and Satterthwaite's result [1983]. We consider the first two results now.

The basic model is that there are I agents, each with utility function $u_i(x, \theta_i) + t_i$, where x is the decision variable. Some of these agents actually build the bridge, so their utility may depend negatively on the value of x. Let $x^*(\theta)$ be the unique solution to $\max_x \sum_{i=1}^{I} u_i(x, \theta_i)$.

To be clear, the various sorts of constraints are:

$$\text{(Ex post BB)} \quad \sum_{i=1}^{I} t_i(\theta) \leq 0, \ \forall \theta$$

$$\text{(Ex ante BB)} \quad \sum_{i=1}^{I} E_\theta[t_i(\theta)] \leq 0, \ \forall \theta$$

$$\text{(BN-IC)} \quad E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i}|\theta_i)] \geq E_{\theta_{-i}}[U_i(\hat{\theta}_i, \theta_{-i}|\theta_i)], \ \forall(\theta_i, \hat{\theta}_i)$$

$$\text{(DS-IC)} \quad U_i(\theta_i, \theta_{-i}|\theta_i)] \geq U_i(\hat{\theta}_i, \theta_{-i}|\theta_i), \ \forall(\theta_i, \hat{\theta}_i, \theta_{-i}),$$

$$\text{(Ex post IR)} \quad U_i(\theta) \geq 0 \ \forall\theta,$$

$$\text{(Interim IR)} \quad E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i})] \geq 0, \ \forall\theta_i,$$

$$\text{(Ex ante IR)} \quad E_\theta[U_i(\theta)] \geq 0, \ \forall\theta_i.$$

The Groves Mechanism:

We want to implement x^* using transfers that satisfy DS-IC. The trick to implementing x^* in dominant strategies is to choose a transfer function for agent i that makes agent i's payoff equal to the social surplus.

$$t_i(\hat{\theta}) \equiv \sum_{j \neq i} u_j(x^*(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_j) + \tau_i(\hat{\theta}_{-i}),$$

where τ_i is an arbitrary function of θ_{-i}. To see that this mechanism $y = (x^*, t)$ is dominant-strategy incentive compatible, note that for any θ_{-i}, agent i's utility is

$$U_i(\hat{\theta}_i, \theta_{-i}|\theta_i) = \sum_{j=1}^{I} u_j(x^*(\hat{\theta}_i, \theta_{-i}), \theta_j) + \tau_i(\hat{\theta}_{-i}).$$

τ_i can be ignored because it is independent of agent i's report. Thus, agent i chooses $\hat{\theta}_i$ to maximize

$$U_i(\hat{\theta}_i, \theta_{-i}|\theta_i) = \sum_{j=1}^{I} u_j(x^*(\hat{\theta}_i, \theta_{-i}), \theta_j).$$

By definition of $x^*(\theta)$, the choice of $\hat{\theta}_i = \theta_i$ is optimal *for any θ_{-i}*.

Green and Laffont [1977] have shown that *any mechanism with truth-telling as a dominant strategy has the form of a Grove's mechanism.* Also, they show that in general, ex post BB is violated by a Grove's mechanism. Given the desirability of BB, we turn to a less powerful implementation concept.

The AGV Mechanism:

d'Aspremont and Gerard-Varet (AGV) [1979] show that x^* can be implemented with transfers satisfying BB if one only requires BN-IC (rather than DS-IC) to hold.

Consider the transfer

$$t_i(\hat{\theta}) \equiv E_{\theta_{-i}}\left[\sum_{j\neq i} u_j(x^*(\theta_{-i}, \hat{\theta}_i), \theta_j)\right] + \tau_i(\hat{\theta}_{-i}).$$

τ_i will be chosen to ensure BB is satisfied. Note that the agent's expected payoff given that $\hat{\theta}$ is announced is (ignoring τ_i)

$$E_{\theta_{-i}}\left[u_i(x^*(\hat{\theta}_i, \theta_{-i}), \theta_i) + \sum_{j\neq i} u_j(x^*(\theta_{-i}, \hat{\theta}_i), \theta_j)\right].$$

Providing that all the other players announce truthfully, $\hat{\theta}_{-i} = \theta_{-i}$, player i's optimal strategy is truth-telling. Hence, BN-IC is satisfied by the transfers.

Now we construct τ_i so as to achieve BB.

$$\tau_i(\hat{\theta}_{-i}) \equiv -\frac{1}{I-1}\sum_{j\neq i} E_{\theta_{-j}}\left[\sum_{k\neq j} u_k(x^*(\hat{\theta}_j, \theta_{-j}), \theta_k)\right].$$

Intuitively, the τ_i are constructed so as to have i pay off portions of the other players' subsidies (which are independent of i's report).

Remarks:

1. If the budget constraint were instead $\sum_{i=1}^{N} t_i \leq C_0$, where C_0 is the cost of provision, the same argument of budget balance for the AGV mechanism can be made.

2. Note that the AGV mechanism does not guarantee that ex post or interim individual rationality will be met. Ex ante IR can be satisfied with appropriate side payments.

3. The AGV mechanism is very useful. Suppose that two individuals can write a contract *before* learning their types. We call this the *ex ante* stage in an incomplete information game. Then the players can use an AGV mechanism to get the first best tomorrow when they learn their types, and they can transfer money among themselves today to satisfy any division of the expected surplus that they like. In particular, they can transfer money ex ante to take into account that their IR constraints may be violated ex post. [This requires that the parties can commit not to walk away ex post if they lose money.] *Therefore, ex ante contracting with risk neutrality implies no inefficiencies ex post.* Of course, most contracting situations we have considered so far involve the contract being offered at the *interim stage* where each party knows their own information, but not the information of others. At this stage we frequently have distortions emerging because of the individual rationality constraints. At the *ex post stage*, all information is known.

4. As said above, in Myerson and Satterthwaite's [1983] model of bilateral exchange, mechanisms that satisfies ex post BB, BN-IC, and interim IR are inefficient when efficient trade depends upon the state of nature.

5. Note that Myerson and Satterthwaite's [1983] inefficiency result continues to hold if we require only interim BB rather than ex post BB. Interestingly, Williams [1995] has demonstrated that if the constraints on the bilateral trading problem are all ex ante or interim in nature, if utilities are linear in money, and if the first-best outcome is implementable in a BNE mechanism, then it is also implementable in a DSE mechanism (ala' Clark-Groves). In other words, the space of efficient BNE mechanisms is spanned by Clark-Groves mechanisms. Thus, in the interim BB version of Myerson-Satterthwaite, showing that the first-best is not implementable with a Clark-Groves mechanism is sufficient for demonstrating that the first-best is also not implementable with any BNE mechanism.

6. Some authors have extended M&S by considering many buyers and sellers in a market mechanism design setting. With double auctions, as the number of agents increases, trade becomes full-information efficient.

7. With many agents in a public goods context, the limit results are negative. As Mailath and Postlewaite [1990] have shown, as the number of agent's becomes large, an agent's information is rarely pivotal in a public goods decision, and so inducing the agent to tell the truth require subsidies that violate budget balance.

3.2.7 General Remarks on the Static Mechanism Design Literature

1. The timing discussed above has generally been that the principal offers a contract to the agent at the interim stage (after the agent knows his type). We have seen that an AGV mechanism can generally obtain the first best if unrestricted contracting can occur at the ex ante stage an the agents are risk neutral. Sappington [1983] considers an interesting *hidden information* model which demonstrates that if contracting occurs at the ex ante stage among risk neutral agents *but where the agents must be guaranteed some utility level in every state* (ex post IR), the contracts are similar to the standard

interim screening contracts. Thus, if agents can always threaten to walk away from the contract after learning their type, it is as if you are in the interim contracting game.

2. The above models generally assume an IR constraint which is independent of type. This is not always plausible as high type agents may have better outside options. When one introduces type-contingent IR constraints, it is no longer clear where the IR constraint binds. The models become messy but sometimes yield considerable new economic insight. There are a series of very nice papers by Lewis and Sappington [1989a,1989b] which investigate the use of countervailing incentives which endogenously use type-dependent outside options to the principal's advantage. A nice extension and unification of their approach is given by Maggi and Rodriguez [1995], which indicates the relationship between countervailing incentives and inflexible rules and how Lewis and Sappington's results depend importantly upon whether the agent's utility is quasi-concave or quasi-convex in the information parameter. The most general paper on the techniques involved in designing screening contracts when agent's utilities depend upon their type is given by Jullien [1995]. Applications using countervailing incentives such as Laffont-Tirole [1990] and Stole [1995] consider the effects of outside options on the optimal contracts and find whole intervals of types in which the IR constraints bind.

3. Another interesting extension of the basic paradigm is the construction of general equilibrium models that endogenously determine outside options. This, combined with type-contingent outside options, has been studied by Spulber [1989] and Stole [1995] in price discrimination contexts.

3.3 Dynamic Principal-Agent Screening Contracts

We now turn to the case of dynamic relationships. We restrict our attention to two period models where the agent's relevant private information does not change over time in order to keep things simple. There are three environments to consider, each varying in terms of the commitment powers of the principal. First, there is full-commitment, where the principal can credibly commit to an incentive scheme for the duration of the relationship. Second, there is the other extreme, where no commitment exists and contracts are effectively one-period contracts: any two-period contract can be torn up by *either party* at the start of the second-period. This implies in particular that the principal will always offer a second period contract that optimally utilizes any information revealed in the first period. Third, midway between these regimes, is the case of commitment with renegotiation: Contracts cannot be torn up unless *both parties* agree to it, thus renegotiations must be Pareto improving. We consider each of these environments in turn.

This section is largely based upon Laffont and Tirole [1988,1990], which are nicely presented in Chapters 9 and 10 of their 1993 book. The main difference is in the economic setting; here we study nonlinear pricing contracts (ala' Mussa-Rosen [1978]) rather than the procurement environment. Also of interest for the third regime of renegotiation with commitment discussed in subsection 3.3.4 is Dewatripont [1989]. Lastly, Hart and Tirole [1988] offer another study of dynamic screening contracts across these three regimes in an intertemporal price discrimination framework, discussed below in section 3.3.5.

3.3.1   The Basic Model

We will utilize a simple model of price discrimination (ala' Mussa-Rosen [1978]) throughout the analysis but our results do not depend on this model directly. [Laffont-Tirole [1993; Chapters 1, 9 and 10] perform a similar analysis in the context of their regulatory framework.] Suppose that the firm's unit cost of producing a product with quality q is $C(q) \equiv \frac{1}{2}q^2$. A customer of type θ values a good of quality q by $u(q,\theta) \equiv \theta q$. Utilities of both actors are linear and transferable in money: $V \equiv t - C(q)$, $U \equiv \theta q - t$, and $S(q,\theta) \equiv \theta q - C(q)$. We will consider two different informational settings. First, the continuous case, where θ is distributed according to $P(\theta)$ on $[\underline{\theta}, \overline{\theta}]$; second, the two-type case, where $\overline{\theta}$ occurs with probability p and $\underline{\theta}$ occurs with probability $1 - p$. Under either setting, the first-best full-information solution is to set $q(\theta) = \theta$. Under a one-period private information setting, our results for the choice of quality under each setting are, for the continuous case,

$$q(\theta) = \theta - \frac{1 - P(\theta)}{p(\theta)}, \ \forall \ \theta,$$

and for the two-type case,

$$\overline{q} \equiv q(\overline{\theta}) = \overline{\theta},$$

and

$$\underline{q} \equiv q(\underline{\theta}) = \underline{\theta} - \frac{p}{1 - p}\Delta\theta,$$

where $\Delta\theta \equiv \overline{\theta} - \underline{\theta}$. We assume throughout that $\underline{\theta} > p\overline{\theta}$, so that the firm wishes to serve the agent (i.e., $q(\underline{\theta}) \geq 0$).

3.3.2   The Full-Commitment Benchmark

Suppose that the relationship lasts for two contracting periods, where the common discount factor for the second period is $\delta > 0$. We allow either $\delta < 1$ or $\delta > 1$ to generally reflect the relative importance of the payoffs in the two periods. We have the following immediate result.

Theorem 30 Under full-commitment, the optimal long-term contract commits the principal to offering the optimal static contract in each period.

Proof: The proof is simple. Because $u_{\theta qq} = 0$, there is no value to introducing randomization in the static contract. In particular, there is no value to offering one allocation with probability $\frac{1}{1+\delta}$ and the other with probability $\frac{\delta}{1+\delta}$. But this also implies that the principal's contract should not vary over time. $\square$

3.3.3   The No-Commitment Case

Now consider the other extreme of no commitment. We show two results. First, in the continuum case for any implementable first-period contract pooling must occur almost everywhere. Second, in the two-type case, although separation may be feasible, separation is generally not optimal and some pooling will be induced in the first period. Of particular

interest is that in the second case, both upward and downward incentive compatibility constraints may bind because the low-type agent can always "take-the-money-and-run."

Continuous Case:

With a continuum of types, it is impossible to get separation almost everywhere. Let $U_2(\hat{\theta}|\theta)$ be the continuation equilibrium payoff that type θ obtains in period 2 given that the principal believes the agent's type is actually $\hat{\theta}$. Note that if there is full separation for some type θ, then $U_2(\theta|\theta) = 0$. Additionally, if there is separation in equilibrium between θ and $\hat{\theta}$, then out of equilibrium, $U_2(\hat{\theta}|\theta) = \max\{(\theta - \hat{\theta})q(\hat{\theta}), 0\}$.

The basic idea goes like this. Suppose that type θ is separated from type $\hat{\theta} = \theta - d\theta$ in the period 1 contract. By lying downward in the first period, the high type suffers only a second-order loss; but being thought to be a lower type in the second period raises the high type's payoff by a first-order amount: $\delta U_2(\theta - d\theta|\theta)$. Specifically, we have the following theorem.

Theorem 31 For *any* first-period contract, there exists no non-degenerate subinterval of $[\underline{\theta}, \overline{\theta}]$ in which full separation occurs.

Proof: Suppose not and there is full sorting over $(\theta_0, \theta_1) \subset [\underline{\theta}, \overline{\theta}]$.

Step 1. (q, t) increases in θ implying that the functions are almost everywhere differentiable. Take $\theta > \hat{\theta}$, both in the subinterval. By incentive compatibility,

$$\theta q(\theta) - t(\theta) \geq \theta q(\hat{\theta}) - t(\hat{\theta}) + \delta U_2(\hat{\theta}|\theta),$$

$$\hat{\theta} q(\hat{\theta}) - t(\hat{\theta}) \geq \hat{\theta} q(\theta) - t(\theta).$$

The first equation represents that the high type agent will receive a positive rent from deceiving the principal in the first period. This rent term is not in the second equation because along the equilibrium path (truthtelling), no agent makes rents in the second period, and so the lower type will prefer to quit the relationship rather than consume the second-period bundle for the high-type which would produce negative utility for the low type. [This is the "take-the-money-and-run" strategy.] Given the positive rent from lying in the second period, adding these equations together imply $(\theta - \hat{\theta})(q(\theta) - q(\hat{\theta})) > 0$, which implies that $q(\theta)$ is strictly increasing over the subinterval (θ_0, θ_1). This result, combined with either inequality above, implies that $t(\theta)$ is also strictly increasing.

Step 2. Consider a point of differentiability, θ, in the subinterval. Incentive compatibility between θ and $\theta + d\theta$ implies

$$\theta q(\theta) - t(\theta) \geq \theta q(\theta + d\theta) - t(\theta + d\theta),$$

or

$$t(\theta + d\theta) - t(\theta) \geq \theta[q(\theta + d\theta) - q(\theta)].$$

Dividing by $d\theta$ and taking the limit as $d\theta \to 0$ yields

$$\frac{dt(\theta)}{d\theta} \geq \theta \frac{dq(\theta)}{d\theta}.$$

Now consider incentive compatibility between θ and $\theta - d\theta$ for the type θ agent. An agent of type θ who is mistaken for an agent of type $\theta - d\theta$ will receive second period rents of $U_2(\theta - d\theta | \theta) = q(\theta - d\theta)d\theta > 0$. As a consequence,

$$\theta q(\theta) - t(\theta) \geq \theta q(\theta - d\theta) - t(\theta - d\theta) + \delta U_2(\theta - d\theta | \theta),$$

or

$$t(\theta) - t(\theta - d\theta) \leq \theta[q(\theta) - q(\theta - d\theta)] - \delta q(\theta - d\theta)d\theta,$$

or taking the limit as $d\theta \to 0$,

$$\frac{dt(\theta)}{d\theta} \leq \theta \frac{dq(\theta)}{d\theta} - \delta q(\theta).$$

Combining the two inequalities yields

$$\theta \frac{dq(\theta)}{d\theta} \leq \frac{dt(\theta)}{d\theta} \leq \theta \frac{dq(\theta)}{d\theta} - \delta q(\theta),$$

which is a contradiction. $\square$

In their paper, Laffont-Tirole [1988] also characterize the nature of equilibria (in particular, they provide necessary and sufficient conditions for partition equilibria for quadratic utility functions in their regulatory context). Instead of studying this, we turn to the simpler two-type case.

Two-type Case:

Before we begin, some remarks are in order.

Remarks:

1. Note a very important distinction between the continuous case and the two-type case. For the *continuous type case*, full separation over any subinterval *is not implementable*; in the *two-type case* we will see that separation may be implementable but typically it *is not optimal*.

2. Because we shall look at the principal's optimal choice of contracts, we need to make clear the notion of continuation equilibria. At the end of the first period, several equilibria may exist in the continuation of the game. We will generally consider the best outcome from the principal's point of view, but if one is worried about uniqueness, one needs to be more careful here.

3. We will restrict our attention to menus with only two contracts. We do not know if this is without loss of generality.

4. We will also restrict our attention to parameters values such that the principal always prefers to serve both types in each period.

5. Let ν be the principal's probability assessment that the agent is of type $\bar{\theta}$. Then the conditionally optimal contract has $\bar{q} = \bar{\theta}$ and $\underline{q} = \underline{\theta} - \frac{\nu}{1-\nu}\Delta\theta$. Thus, the low-type's allocation decreases with the principal's assessment.

We will proceed by restricting our attention to two-part menus. There will be three cases of interest. We will analyze the optimal contracts for these cases.

Consider the following contract offer: $(\underline{q}_1, \underline{t}_1)$ and $(\bar{q}_1, \bar{t}_1)$. We denote periods by the subscript on the contracting variables. Without loss of generality, assume that $\underline{\theta}$ chooses $(\underline{q}_1, \underline{t}_1)$ with positive probability and $\bar{\theta}$ chooses $(\bar{q}_1, \bar{t}_1)$ with positive probability. Let $U_2(\nu|\bar{\theta})$ be the rent the high type receives in the continuation game where the principal's belief that the agent is of type $\bar{\theta}$ is ν. We will let $\bar{\nu}$ represent the belief of the principal after observing the contract choice $(\bar{q}_1, \bar{t}_1)$ and $\underline{\nu}$ represent the belief of the principal after observing the contract choice $(\underline{q}_1, \underline{t}_1)$. Then the principal will design an allocation which maximizes profit subject to four constraints.

$$
\begin{array}{rll}
(\overline{IC}) & \bar{\theta}\bar{q}_1 - \bar{t}_1 + \delta U_2(\bar{\nu}|\bar{\theta}) \geq & \bar{\theta}\underline{q}_1 - \underline{t}_1 + \delta U_2(\underline{\nu}|\bar{\theta}), \\
(\underline{IC}) & \underline{\theta}\underline{q}_1 - \underline{t}_1 + \delta U_2(\underline{\nu}|\underline{\theta}) \geq & \underline{\theta}\bar{q}_1 - \bar{t}_1 + \delta U_2(\bar{\nu}|\underline{\theta}), \\
(\overline{IR}) & \bar{\theta}\bar{q}_1 - \bar{t}_1 + \delta U_2(\bar{\nu}|\bar{\theta}) \geq & 0, \\
(\underline{IR}) & \underline{\theta}\underline{q}_1 - \underline{t}_1 \geq & 0.
\end{array}
$$

As is usual, $\overline{IR}$ is implied by the $\overline{IC}$ and $\underline{IR}$. Additionally, $\underline{IR}$ must be binding (providing the principal gets to choose the continuation equilibrium). To see this, note that if it were not binding, both $\underline{t}$ and $\bar{t}$ could be raised by equal amounts without violating the IC constraints and profits could be increased. Thus, the principal can substitute out $\underline{t}$ from the objective function using $\underline{IR}$ (thereby imposing this constraint with an equality). Now the principal's problem is to maximize profits subject to the two IC constraints.

There are three cases to consider. Contracts in which only the high-type's IC constraint is binding (Type I); contracts in which only the low-type's IC constraint is binding (Type II); and contracts in which both IC constraints bind (Type III). In turns out that Type II contracts are never optimal for the principal, so we will ignore them. [See Laffont and Tirole, 1988, for the argument.] Type I contracts are the simplest to study; Type III contracts are more complicated because the take-the-money-and-run strategy of the low type causes the low-type's IC constraint to bind upward.

Type I Contracts: Let the $\bar{\theta}$ type customer choose the high-type contract, $(\bar{q}_1, \bar{t}_1)$ with probability $1 - \alpha$ and $(\underline{q}_1, \underline{t}_1)$ with probability α. In a type I equilibrium, whenever the $(\bar{q}_1, \bar{t}_1)$ contract is chosen, the principal's beliefs are degenerate: $\bar{\nu} = 1$. Following Bayes' rule, when $(\underline{q}_1, \underline{t}_1)$ is chosen, $\underline{\nu} = \frac{\alpha p}{\alpha p + (1-p)} < p$. Note that $\frac{d\nu}{d\alpha} > 0$. Define the single-period expected asymmetric information profit level of the principal with belief ν who offers the optimal static contract as

$$
\Pi(\nu) \equiv \max_{\bar{q}, \underline{q}} \ \nu(\bar{\theta}\bar{q} - C(\bar{q}) - \Delta\theta\underline{q}) + (1 - \nu)(\underline{\theta}\underline{q} - C(\underline{q})).
$$

Consider the second period. Given any belief ν, the principal will choose the conditionally optimal contract. This implies a choice of $\bar{q}_2 = \bar{\theta}$, $\underline{q}_2(\alpha) = \underline{\theta} - \alpha\frac{p}{1-p}\Delta\theta$, and profit

$\Pi(\underline{\nu}(\alpha))$, where we have directly acknowledged the dependence of $\underline{q}_2$ and second period profits on α.[5] Thus, we are left with calculating the first-period choices of $(\bar{q}_1, \underline{q}_1, \alpha)$ by the principal. Specifically, the principal solves

$$\max_{(\bar{q}_1, \bar{t}_1),(\underline{q}_1, \underline{t}_1),\alpha} p\left\{(1-\alpha)[\bar{t}_1 - C(\bar{q}_1) + \delta\Pi(1)] + \alpha[\underline{t}_1 - C(\underline{q}_1) + \delta\Pi(\underline{\nu}(\alpha))]\right\}$$
$$+ (1-p)[\underline{t}_1 - C(\underline{q}_1) + \delta\Pi(\underline{\nu}(\alpha))],$$

subject to

$$\bar{\theta}\bar{q}_1 - \bar{t}_1 = \bar{\theta}\underline{q}_1 - \underline{t}_1 + \delta\Delta\theta\underline{q}_2(\alpha),$$

and

$$\underline{\theta}\underline{q}_1 - \underline{t}_1 = 0.$$

The first constraint is the binding IC constraint for the high type where $U_2(\underline{\nu}(\alpha)|\bar{\theta}) = \Delta\theta\underline{q}_2(\alpha)$ given our previous discussion of price discrimination. The second constraint is the binding IR constraint for the low type. Note that by increasing α, the principal directly decreases profit by learning less information, but simultaneously credibly lowers $\underline{q}_2$ in the second period which weakens the high-type's IC constraint in the first period. Thus, there is a tradeoff between separation (which is good per se because this improves efficiency) and the rents which must be given to the high type to obtain the separation. Substituting the constraints into the principal's objective function and simplifying yields

$$\max_{\bar{q}_1, \underline{q}_1, \alpha} p(1-\alpha)[\bar{\theta}\bar{q}_1 - C(\bar{q}_1) - \Delta\theta\underline{q}_1 - \delta\Delta\theta\underline{q}_2(\alpha)]$$
$$+ (1-p+\alpha p)[\underline{\theta}\underline{q}_1 - C(\underline{q}_1)] + p(1-\alpha)\delta\Pi(1) + (1-p+\alpha p)\delta\Pi(\underline{\nu}(\alpha)).$$

The first-order conditions for output are $\bar{q}_1 = \bar{\theta}$ and

$$\underline{q}_1 = \underline{\theta} - \frac{p-\alpha p}{(1-p)+\alpha p}\Delta\theta.$$

The latter condition indicates that $\underline{q}_1$ will be chosen at a level *above* that of the static contract if $\alpha > 0$. To see the choice of $\underline{q}_1$ in a different way, note that by rearranging the terms of the above objective function we have

$$\underline{q}_1 = \arg\max_q \ \alpha p(\bar{\theta}q - C(q)) + (1-p)(\underline{\theta}q - C(q)) - p\Delta\theta q.$$

Here, $\underline{q}_1$ is chosen taking into account the efficiency costs of pooling and the rent which must be left to the high type.

Finally, one must maximize subject to α, which will take into account the tradeoff between surplus-increasing separation and reducing the rents which the high-type receives in order to separate in period one with a higher probability. Generally, we will find that

---

[5]In general, it is not enough to assume that $\Delta\theta$ is small enough that the principal always prefers to serve both types in the static equilibrium, because we might suspect that in a dynamic model low types may not be served in a later period. For Type I contracts, this is not in issue as $\underline{\nu} < p$, so low types are even more attractive in the second period than in the first when $(\underline{q}_1, \underline{t}_1)$ is chosen. When the other contract is chosen, no low types exist, and so this is irrelevant. Unfortunately, this is not the case with Type III contracts.

the pooling probability, α, is increasing in δ. Thus, as the second-period becomes more important, less separation occurs in the first period.

One can demonstrate that if δ is sufficiently large, the IC constraint for the low type will be binding, and so one must check the solution to the Type I program to verify that it is indeed a type I equilibrium (i.e., the IC constraint for the low type is slack at the optimum). For high δ, this will not be the case, and so we have a type III contract.

Type III Contracts:

Let the high type choose $(\underline{q}_1, \underline{t}_1)$ with probability α as before, but let the low type choose $(\bar{q}_1, \bar{t}_1)$ with probability β. Now, by Bayes' rule, we have

$$\bar{\nu}(\alpha, \beta) \equiv \frac{p(1-\alpha)}{p(1-\alpha) + (1-p)(1-\beta)},$$

and

$$\underline{\nu}(\alpha, \beta) \equiv \frac{p\alpha}{p\alpha + (1-p)\beta}.$$

As before, the second-period contract will be conditionally optimal. This implies that $\bar{q}_2 = \bar{\theta}$, $\underline{q}_2(\nu) = \underline{\theta} - \frac{\nu}{1-\nu}\Delta\theta$, and $U_2(\nu|\bar{\theta}) = \Delta\theta \underline{q}_2(\nu)$.

The principal's type III program is

$$\max \ [p(1-\alpha) + (1-p)\beta][\bar{t}_1 - C(\bar{q}_1) + \delta\Pi(\bar{\nu}(\alpha, \beta)]$$
$$+ [p\alpha + (1-p)(1-\beta)][\underline{t}_1 - C(\underline{q}_1) + \delta\Pi(\underline{\nu}(\alpha, \beta)),]$$

subject to

$$\bar{\theta}\bar{q}_1 - \bar{t}_1 + \delta U_2(\bar{\nu}(\alpha, \beta)) = \bar{\theta}\underline{q}_1 - \underline{t}_1 + \delta U_2(\underline{\nu}(\alpha, \beta)),$$
$$\underline{\theta}\underline{q}_1 - \underline{t}_1 = \underline{\theta}\bar{q}_1 - \bar{t}_1,$$
$$\underline{\theta}\underline{q}_1 - \underline{t}_1 = 0.$$

The first and second constraints are the binding IC constraints; the third constraint is the binding IR constraint for the low type. Manipulating the three constraints implies that $\Delta\theta(\bar{q}_1 - \underline{q}_1) = \delta[U_2(\underline{\nu}(\alpha, \beta)) - U_2(\bar{\nu}(\alpha, \beta))]$. Substituting the three constraints into the objective function to eliminate t and simplifying, we have

$$\max_{\bar{q}_1, \underline{q}_1, \alpha, \beta} \ [p(1-\alpha) + (1-p)\beta][\underline{\theta}\bar{q}_1 - C(\bar{q}_1) + \delta\Pi(\bar{\nu}(\alpha, \beta)]$$
$$+ [p\alpha + (1-p)(1-\beta)][\underline{\theta}\underline{q}_1 - C(\underline{q}_1) + \delta\Pi(\underline{\nu}(\alpha, \beta)),]$$

subject to

$$\bar{q}_1 - \underline{q}_1 = \delta[\underline{q}_2(\underline{\nu}(\alpha, \beta)) - \underline{q}_2(\bar{\nu}(\alpha, \beta))].$$

First-order conditions imply that $\bar{q}_1$ is not generally efficient because of the IC constraint for the low type. In fact, given Laffont and Tirole's simulations, it is generally possible that first-period allocation may be decreasing in type: $\bar{q}_1 < \underline{\theta} < \underline{q}_1$! [Note that a nondecreasing allocation is no longer a necessary condition for incentive compatibility because of the second period rents.]

Remarks for the Two-type Case (Type I and III):

1. For δ small, the low-type's IC constraint will not bind and so we will have a type I contract. For sufficiently high δ, the reverse is true.

2. As δ goes to ∞, $\underline{q}_1$ goes to $\overline{q}_1$ and there is complete pooling in the first period. The idea is that by pooling in the first period, the principal can commit not to learn anything and therefore impose the statically optimal separation contract in the second period.

3. Note that for any finite δ, complete pooling in period 1 is never optimal. That is, the above result is a limit result only. To see why, suppose that full pooling were undertaken. Then the second period output levels are statically optimal. By reducing pooling in the first period by a small degree, surplus is increased by a first-order amount while there is only a second-order effect on profits in the second period (since it was previously statically optimal).

4. Clearly, the firm always prefers commitment to non-commitment. In addition, for δ small, the buyer prefers non-commitment to commitment. The intuition is that the low type always gets zero, but the high type gets more rents when there is full separation. For δ close to zero, the high type gets $U(p|\overline{\theta}) + \delta U(0|\overline{\theta})$ instead of $(1+\delta)U(p|\overline{\theta})$.

5. The main difference between the two-type case and the continuum is that separation is possible but not usually optimal in the two-type case, while it is impossible in the continuous-type case. Intuitively, as $\Delta\theta$ becomes small, Type III contracts occur, requiring that both IC constraints bind. With more than two types, these IC constraints cannot be satisfied unless there is pooling almost everywhere.

3.3.4 Commitment with Renegotiation

We now consider commitment, but with the possibility that Pareto-improving renegotiation takes place between periods 1 and 2. The seminal work is Dewatripont's dissertation published in 1989. We will follow Laffont and Tirole's [1990] article, but in a nonlinear pricing framework.

The fundamental difference between non-commitment and commitment with renegotiation is that the "take-the-money-and-run" strategy of the low type is not possible in the latter. That is, a low-type agent that takes the high-type's contract can be forced to continue with the high-type's contract resulting in negative payoffs (even if it is renegotiated). Because of this, the low type's IC constraint is no longer problematic. Generally, full separation is possible even in a continuum, although it may not be optimal.

The Two-Type Case: We first examine the two-type case.

We assume that at the renegotiation stage, the principal makes all contract renegotiation offers in a take-it-or-leave-it fashion.[6] Given that parties have rational expectations, the principal can restrict attention to renegotiation-proof contracts.

---

[6]If one is prepared to restrict attention to strongly-renegotiation-proof contracts (contracts which do not have renegotiation as *any* equilibrium), this is without loss of generality as shown by Maskin and Tirole [1992].

It is straightforward to show that in any contract offer by the principal, it is also without loss of generality to restrict attention to a two-part menu of contracts, where each element of the menu specifies a first-period allocation, (q_1, t_1), and a second-period menu continuation menu, $\{(\underline{q}_2, \underline{t}_2), (\overline{q}_2, \overline{t}_2)\}$, conditional on the first-period choice, (q_1, t_1).

Renegotiation-proofness requires that for a given probability assessment of the high type following the first period choice, the solution to the program below is the continuation menu, where the expected continuation utilities of the high and low type under the continuation menu are $\{\overline{U}^o, 0\}$.[7]

$$\max \ \nu(\overline{\theta}\overline{q}_2 - \overline{U} - \frac{1}{2}\overline{q}_2^2) + (1 - \nu)(\underline{\theta}\underline{q}_2 - \frac{1}{2}\underline{q}_2^2),$$

subject to

$$\overline{U} \geq \Delta\theta\underline{q}_2,$$

$$\overline{U} \geq \overline{U}^o,$$

where $\overline{U}$ is the utility of the high type in the solution to the above program. The two constraints are IC and interim-IR for the high-type, respectively.

The following partial characterization of the optimal contract, proven in Laffont-Tirole [1990], simplifies our problem considerably.

Theorem 32 The firm offers a menu of two allocations in the first period in which the low-type choose one for sure and the high-type randomizes between them with probability α on the low-type's contract. The second period continuation contracts are conditionally optimal given beliefs derived from Bayes' rule, $\underline{\nu} = \frac{\alpha p}{\alpha p + (1-p)} < p$.

We are thus in the case of Type I contracts discussed above with non-commitment. As a consequence, we know that $\overline{q}_1 = \overline{q}_2 = \overline{\theta}$ and $\underline{q}_2 = \underline{\theta} - \frac{\nu}{1-\nu}\Delta\theta$. The principal chooses $\underline{q}_1$ and α jointly to maximize profit. As before our results are ...

Results for Two-type Case:

1. $\underline{q}_1$ is chosen between the full-information and static-optimal levels. That is,

$$\underline{\theta} - \frac{p}{1-p}\Delta\theta \leq \underline{q}_1 \leq \underline{\theta}.$$

2. The probability of pooling is nondecreasing in the discount factor, δ. For δ sufficiently low, the full separation occurs ($\alpha = 0$).

3. As $\delta \to \infty$, $\alpha \to 1$, but for any finite δ the principal will choose $\alpha < 1$.

4. By using a long-term contract for the high-type and a short-term contract for he low-type, it is possible to generate the optimal contract above as a unique renegotiation equilibrium. See Laffont-Tirole [1990].

---

[7]We have normalized the continuation payoff for the low type to be $\underline{U}^o = 0$. This is without loss of generality, as first-period transfers can be adjusted accordingly.

Continuous Type Case: We look at the continuous type case briefly to note that full separation is now possible. The following contract is renegotiation-proof. Offer the optimal static contract for the first-period component and a sales contract for the second-period component (i.e., $q(\theta) = \theta$, and $t_2(\theta) \equiv C(\theta)$.) Because the second-period allocation is Pareto efficient in a first-best sense, it is necessarily renegotiation-proof. Additionally, no information discovered in the first period can be used against the agent in the second period (because the original contract guarantees them the maximal level of information rents) , so the first period allocation is incentive compatible. Without commitment, the principal cannot guarantee not to use the information against the agent.

Conclusion:

The main result to note is that commitment with renegotiation typically lies between the full-information contract and the non-commitment contract in terms of the principal's pay-off. In the two-type case, this is clear as the lower IC constraint, which binds in the Type III non-commitment case, disappears in the commitment with renegotiation environment. In addition, the set of feasible contracts is enlarged in both the two-type and continuous-type cases.

3.3.5 General Remarks on the Renegotiation Literature

Intertemporal Price Discrimination: Following Hart and Tirole [1988] (and also Laffont and Tirole [1993, pp. 460-464], many of the above results can be applied to the case of intertemporal price discrimination.

Restrictions to Linear Contracts:
Freixas, Guesnerie, and Tirole [1985], consider the "ratchet effect" in non-commitment environments, but they restrict attention to two-part tariffs rather than nonlinear contracts. The idea is that after observing the output choice of the first period, the principal will offer a lower-rent tariff in the second-period. Their analysis yields similar insights in a far simpler manner. The nature of two-part tariffs effectively eliminates problems of "take-the-money-and-run" strategies and simplifies the mathematics of contract choice (a contract is just an intercept and a slope). The result is that the principal can only obtain more separation in the first period by offering more efficient contracts (higher-powered contracts). The optimal contract will induce pooling or semi-separating for some parameter values, and in these cases contracts are less distortionary in the first period.

Common Agency as a "Commitment" Device:
Restricting attention to linear contracts (as in Freixas, et al. [1985]), Olsen and Torsvik [1993], show how common agency can be a blessing in disguise. When two principals contract with the same agent and the agent's actions are complements, common agency has the effect of introducing greater distortions and larger rent extraction in the static setting. Within a dynamic setting, the agent's expected reduction of second-period rents from common agency reduces the high type's benefit of consuming the low type's bundle. It is therefore cheaper to get separation, and so the optimal contract has more information revealed in the first-period. Common agency effectively commits the principal to a second-

period contract offer that lowers the high-types gain from lying. Martimort [1996] has also found a similar effect, but in a common agency setting with nonlinear contracts. Again, the existence of common agency lowers the bite of renegotiation.

Renegotiation as a Commitment Device vis-à-vis Third Parties:

Dewatripont [1988] studies a model in which in order to deter entry, a firm and its workers sign a contract providing for high severance pay (and therefore reducing the opportunity cost of the firm's production). Would-be entrants realize that the severance pay will induce the incumbent to maintain employment and output at high levels after entry has occurred, and therefore may deter entry. Nonetheless, there is an incentive for workers and the firm to renegotiate away the severance payments once entry has occurred, so normally this threat is not credible. But if asymmetric information exists, information may be revealed only slowly because of pooling, and so there is still some commitment value against the entrant (i.e. a third party). A related analysis is performed by Caillaud, Jullien and Picard [1995] in their study of agency contracts in a competitive environment (ala' Fershtman-Judd, [1987]) where two competing firms each contract with their own agents for output, but where secret renegotiation is possible. As in Dewatripont, they find that with asymmetric information between agents and principals, there is some pre-commitment effect.

Organizational Design as a Commitment Device against Renegotiation.

Dewatripont and Maskin [1995] consider the beneficial effects of designing institutions to prevent renegotiation. Decentralization of creditors may serve as a commitment device to cancel ex ante unprofitable projects at the renegotiation stage, but at the cost of some long-run profitable projects not being undertaken. In related work, Dewatripont and Maskin [1992] suggest that sometimes institutions should be developed in which the principal commits to less information so as to relax the renegotiation-proofness constraint.