Lectures in Contract Theory[1]

Steve Tadelis and Ilya Segal[2]
UC Berkeley and Stanford University

Preliminary and Incomplete

December 2005

[1] These notes were prepared for a second year graduate course in Contract theory at Stanford University (Econ 282/291) and UC Berkeley (Econ 206). They are preliminary and incomplete. Thanks also to Debbie Johnston for typing the first version and to Scott Hemphill for excellent research assistance.

[2]

Contents

Part I

Introduction

Economics is about gains from trade. Contracts are essential to realizing these gains.

Competitive Equilibrium Paradigm: Contracts are anonymous trades with the market (auctioneer) at given prices:

- Spot contracts: Short-term
- Arrow-Debreu contracts: Long-term, contingent on resolution of uncertainty).

The First Welfare Theorem establishes that a Competitive Equilibrium with complete markets is Pareto optimal.

This paradigm ignores hazards of real-world contracting, and as a result misses a lot of institutional detail. The ignored issues:

0.1 Trade with small numbers of agents

Even when we start with a large number of agents (competitive), after two (or more) agents engage in a relationship and make relationship-specific investments they enter some sort of small-number bargaining situation. The simplest example is given by bilateral monopoly bargaining. Example: regulation of cable franchise. Williamson's "Fundamental Transformation": Ex ante perfect competition but ex post imperfect competition.

Does non-competitive contracting among small numbers of agents necessarily give rise to inefficiency? As argued by Coase, the answer is "no". The Coase "Theorem" says that in the absence of "transaction costs," the outcome of private bargaining is Pareto efficient. The idea is that the parties will always implement mutually beneficient exchanges. (We can think of the Coase theorem as an analog of the First Welfare Theorem for small-numbers situation.) This "theorem" can be viewed as a definition of "transaction costs". Numerous sources of "transaction costs" have been suggested, but they can all be classified into two broad categories: incentives and bounded rationality. Contract theory studies contracting under such "transaction costs."

0.2 Incentives

Consider a state-contingent Arrow-Debreu delivery contract that obliges a seller to deliver a high-quality good in those states in which his cost is low. The contract may not be implementable when either the state of the world or the quality of the delivered good is, while privately observed by the seller, not contractible, e.g. because it is not observed by anybody else. This situation is called *asymmetric information*. First, if the cost is only observed by the seller, he may have the incentive to misrepresent the state - say that the cost is high and not deliver even when the actual cost is low. When the quality of the delivered good is only observed by the seller, he may have the incentive to deliver a low quality good instead of a high quality good. Thus, asymmetric information creates incentive problems of two kinds arise:

1. **Hidden Information** (Adverse Selection): Agents may not reveal the state truthfully. A contract in these circumstances tries to elicit agents' information. This will be Part I of the course.

2. **Hidden Action** (Moral Hazard): Agents may not deliver on their promises due to imperfect monitoring. This will be the second part of the course.

One remarkable achievement of the economics of asymmetric information is that many different economic phenomena can be understood using the same theoretical tools, as will be demonstrated throughout this text.

Examples of Asymmetric Information Problems:

- *Taxation and Welfare Schemes*: individuals may claim their ability to produce is low, and may work less hard than socially optimal, so as to reduce tax payments and/or increase welfare assistance

- *Monopolistic Price Discrimination and Auctions*: buyers may claim to have a lower willingness to pay so as to pay a lower price

- *Regulation of a natural monopoly*: the monopoly may claim to have high costs, or under-invest in cost reduction, so it can charge a high price

- *Employee compensation*: employees may misrepresent their ability or may engage in low effort

- *Financial Structure of Firms*: managers may misrepresent their abilities or their firm's potentials, or fail to act in the interest of shareholders.

0.3 Bounded Rationality

One property of the Arrow-Debreu economy with complete markets is that all trades can take place at date zero, and from then on the existing contracts are executed but the markets need not reopen. The same is true in contracting situations, even in the presence of asymmetric information. This conclusion is unrealistic since the parties may be unable to foresee all possible states (contingencies) far ahead and contract upon them. Thus, similarly to "incomplete markets" models , we could have "incomplete contracts" models, in which not all contingent contracts can be written. Such incomplete contracts should ideally be explained as optimal when the parties are boundedly rational, e.g. cannot foresee future states of the world, or cannot write complex contracts. Unfortunately, there is no consensus on modeling such bounded rationality, but some issues are discussed in the third part of the course. These issues are believed to be important for understanding economic institutions, more specifically firms.

0.4 Contracts, Mechanisms, Institutions

The three are to a large extent synonymous, meaning "rules of the game". Specifically, they describe what actions the parties can undertake, and what outcomes these actions would entail. If we think about this, we see that indeed almost all games observed in everyday life are not physical, but defined by "institutions" - legal and social norms. Without such institutions, we would be in the "natural state" - anarchy, and play physical games, e.g. someone robs you in the street. Fortunately, in most cases the rules for games are written by someone: the rules of chess, football, and even wars (e.g. the Geneva convention) are written by someone to achieve "better" outcomes.

Thus, Mechanism design = normative economics (relative to game theory). Knowledge of game theory is important because it predicts how a given game will be played by agents. Mechanism design goes one step back: given the physical environment and the constraints faced by the designer (incentive constraints and bounded rationality constraints), what mechanisms are feasible? What mechanisms are optimal?

The idea of Mechanism design can be traced back to market socialism. It realized that the unfettered market is just one possible mechanism for the economy. It is fairly efficient in many cases but not in all cases. Can we improve upon it? Ask people: e.g. how many shoes do you want? Then plan how many shoe factories are needed. This could be better than blind market forces that may cause uncertain outcomes, resulting in a crises of over, or under-production.

However, now we know that in designing mechanisms we must take into account important incentive constraints (e.g. people may not report truthfully how many shoes they need if the price is not right) and bounded rationality constraints (think e.g. how many goods there are in the economy and how difficult it is to collect all information and calculate the optimal plan).

Contract theory vs. **Mechanism design**. Mechanism design theory has been mathematically elegant but has been unable to address the "big" questions, such as "socialism vs. capitalism". Instead it proved useful for more manageable smaller questions, specifically business practices - contracts among agents. In this reincarnation, as "Contract theory", it has these additional features:

- A contract is designed by (one of) the parties themselves

- Parties may refuse to participate ($\Rightarrow$ participation constraints).

- Parties may be able to renegotiate the contract later on.

- Theoretical shortcuts are taken, e.g. restricting attention to particular simple "incomplete" contracts (general optimal mechanisms often predict unrealistically complicated contracts, e.g. infinitely lived contracts)

Part II

Hidden Information

This part considers the design of contracts in which one or many agents have private information. The agent who designs the contract will be called the *Principal*, while the other agents will be simply called *agents*. For the most part we will focus on the situation where the Principal has no private information and the agents do. This framework is called *screening*, because the principal will in general try to screen different types of agents by inducing them to choose different bundles. The opposite situation, in which the Principal has private information and the agents do not, is called *signaling*, since the Principal could signal his type with the design of his contract. This situation will be briefly considered in Section ??. (Of course, a general contracting situation could involve elements of both signaling and screening).

Chapter 1 analyzes a static model with one agent, which is often called the principal-agent model. Chapter 2 analyzes a static model with many agents, and Chapter 2 analyzes a dynamic principal-agent model.

Chapter 1

The Principal-Agent Model

This chapter deals with a situation in which one *Principal* contracts with one *Agent* who possesses private information. The best known application of this model is where the Principal is a monopolistic seller and the Agent is a buyer. The seller tries to *screen* buyers of different types, i.e., induce them to select different consumption bundles. This situation is known as *monopolistic screening,* or *second-degree price discrimination.* For concreteness, we will focus on this setting, and later we show how the same model can be applied to many other settings.

1.1 Setup

1.1.1 Preferences

The Principal is a seller who may choose to sell some quantity $x \in X \subset \Re_+$, in exchange for a payment $t \in \Re$. The Principal's profit is given by

$$t - c(x),$$

where $c(\cdot)$ is her cost function. The Agent is a buyer; if he consumes $x \in X$ and pays a transfer $t \in \Re$, his utility is given by

$$v(x, \theta) - t$$

where $\theta \in \Theta \subset \Re$ is the Agent's "type". We think of $\theta \in \Theta$ as a random variable whose realization is the Agent's private (hidden) information. That is, θ is not observed by the Principal, and it is not contractible, i.e., a contract cannot directly depend upon it. The principal has a prior probability distribution over θ.

1.1.2 Contracting

We assume that the Principal has all the bargaining power in contracting; i.e., he makes a take-it-or-leave-it offer to the agent. The Agent can accept or reject the contract. If the agent rejects, the outcome is $(x, t) = (0, 0)$. This is the Agent's *reservation bundle.* The Agent's *reservation utility* is therefore $v(0, \theta)$.

What kind of contracts can the Principal offer the Agent? One possibility is to offer one bundle (x, t), and ensure that all types of agent accept it. This outcome is called *pooling (bunching).* In general, however, the Principal will do better by offering a contract in which different types of agent *separate* themselves by choosing different bundles. For simplicity, we will consider contracts that take the form of a *tariff*:

Definition 1 *A **tariff** is a function $T : X \to \Re$, which specifies the payments $T(x)$ that the Agent has to make in order to receive different amounts $x \in X$ of the good.*

Notice that disallowing a certain trade x is equivalent to setting $T(x) = +\infty$, and since the agent always has the option to reject the tariff, without loss of generality we constrain the Principal to offer $T(0) = 0$, and assume that the Agent always accepts. Thus, the contractual form of a tariff is quite general, and as we will later see we lose nothing by restricting attention to this form of a contract.

1.2 Single-Crossing and the Monotonicity of Choice

Faced with a tariff $T(\cdot)$, the agent of type will θ will choose the quantity that maximizes his utility, that is, he will select

$$x \in \arg\max_{x \in X} \left[v(x, \theta) - T(x) \right].$$

It turns out that the analysis of contracting is dramatically simplified when the Agent's types can be *ordered* so that higher types choose a higher consumption when faced with any tariff. In this subsection we describe assumptions under which this is always the case. More generally, we identify when solutions to the parametrized maximization program

$$\max_{x \in X} \varphi(x, \theta)$$

are nondecreasing (or strictly increasing) in the parameter θ. In the case of an agent choosing from a tariff, the objective function is $\varphi(x, \theta) = v(x, \theta) - T(x)$; however, the general results of this subsection will be applied to other settings as well. This question is in the realm of a field called "monotone comparative statics," which investigates conditions under which equilibria (in our case, maximizer) of a system respond to changes in parameter in a monotonic way (i.e., the solution is always either nonincreasing or nondecreasing in the parameter).

A key property to ensure monotone comparative statics is the following:

Definition 2 *A function $\varphi : X \times \Theta \to \Re$, where $X, \Theta \subset \Re$, has the **Single-Crossing Property** (SCP) if $\varphi_x(x, \theta)$ exists and is strictly increasing in $\theta \in \Theta$ for all $x \in X$.*

The Single-Crossing Property was first suggested by Spence (1972) and Mirrlees (1971), in application to the agent's value function $v(x, \theta)$.[1] Intuitively, $v(x, \theta)$ satisfies SCP when the marginal utility of consumption, v_x, is increasing in type θ, i.e., higher types always have "steeper" indifference curves in the $x - t$ space. SCP also implies that *large* increases in x are also more valuable for higher parameters θ:

Definition 3 *A function $\varphi : X \times \Theta \to \Re$, where $X, \Theta \subset \Re$, has **Increasing Differences**[2] (ID) if $\varphi(x'', \theta) - \varphi(x', \theta)$ is strictly increasing in $\theta \in \Theta$ for all $x'', x' \in X$ such that $x'' > x'$.*

[1] Spence and Mirrlees formulated a more general version of SCP that works for agent's utilities $u(x, t, \theta)$. This property is sometimes called the "sorting condition" or the "Spence-Mirrlees condition". Our definition is a simplified version for preferences that are quasilinear in transfers t. Our SCP was introduced by Edlin and Shannon [1998] under the name "increasing marginal returns".

[2] This property is more precisely called *strictly* increasing differences, see e.g. Topkis [1998].

Lemma 1 *If $\varphi(x,\theta)$ is continuously differentiable and satisfies SCP, and X is an interval, then φ satisfies ID.*

Proof. For $\theta'' > \theta'$,

$$\begin{aligned}\varphi(x'',\theta'') - \varphi(x',\theta'') &= \int_{x'}^{x''} \varphi_x(x,\theta'')dx \\ &> \int_{x'}^{x''} \varphi_x(x,\theta')dx \\ &= \varphi(x'',\theta') - \varphi(x',\theta'). \qquad \blacksquare\end{aligned}$$

Note that if the Agent's value function $v(x,\theta)$ satisfies ID, then the indifference curves for two different types of the Agent, θ' and $\theta'' > \theta'$, cannot intersect more than once. Indeed, if they intersected at two points $(x',t')\,,(x'',t'')$ with $x'' > x'$, this would mean that the benefit of increasing x from x' to x'' exactly equals $t'-t''$ for both types θ' and θ'', which contradicts ID. This observation justifies the name of "single-crossing property" (which as we have just seen is stronger than ID).

A key result in monotone comparative statics says that when the objective function satisfies ID, maximizers are nondecreasing in the parameter value θ. Moreover, if SCP holds and maximizers are interior, they are strictly increasing in the parameter. Formally,

Theorem 1 (Topkis, Edlin-Shannon) *Let $\theta'' > \theta'$, $x' \in \arg\max_{x\in X} \varphi(x,\theta')$ and $x'' \in \arg\max_{x\in X} \varphi(x,\theta'')$. Then*

(a) If φ has ID, then $x'' \geq x'$.

(b) If, moreover, φ has SCP, and either x' or x'' is in the interior of X, then $x'' > x'$.

Proof. (a) is proven by revealed preference. By construction

$$\begin{aligned}\varphi(x',\theta') &\geq \varphi(x'',\theta') \\ \varphi(x'',\theta'') &\geq \varphi(x',\theta'')\end{aligned}$$

Adding up and rearranging terms, we have

$$\varphi(x'',\theta'') - \varphi(x',\theta'') \geq \varphi(x'',\theta') - \varphi(x',\theta').$$

By ID, this inequality is only possible when $x'' \geq x'$.

For (b), suppose for definiteness that x' is in the interior of X. Then the following first-order condition must hold:

$$\varphi_x(x', \theta') = 0.$$

But then by SCP

$$\varphi_x(x', \theta'') > \varphi_x(x', \theta') = 0,$$

and therefore x' cannot be optimal for parameter value θ'' - a small increase in x would increase φ. Since by (a) $x'' \geq x'$, we must have $x'' > x'$. ■.

We can apply Theorem 1 to the agent's problem of choosing from a tariff $T(\cdot)$, assuming that the agent's value $v(x, \theta)$ satisfies SCP. In this case, the agent's objective function

$$\varphi(x, \theta) = v(x, \theta) - T(x)$$

satisfies ID, since $v(x, \theta)$ satisfies ID $T(x)$ trivially satisfies ID, and ID is an additive property. Therefore, by Theorem 1(a) the agent's consumption choice from any tariff is nondecreasing in his type. This explains why SCP is also known under the name "sorting condition". However, this does not rule out the possibility that the agent's consumption is constant over some interval of types, i.e., we have some *pooling.* Indeed, when the tariff $T(\cdot)$ has a kink at some $\widehat{x}$, we should expect consumption $\widehat{x}$ to be chosen by an interval of types. In order to ensure full separation of types, we need to assume that the tariff $T(\cdot)$ is differentiable. In this case, $\varphi(x, \theta)$ will satisfy SCP, and Theorem 1(b) implies that consumption is strictly increasing in type, i.e., we have *full separation.*

Figure Here

1.3 The Full Information Benchmark

As a benchmark, we consider the case in which the Principal observes the Agent's type θ. Given θ, she offers the bundle (x, t) to solve:

$$\text{FB} \left\{ \begin{array}{lll} \max\limits_{(x,t) \in X \times \Re} & t - c(x) & \\ \text{s.t.} & v(x, \theta) - t \geq v(0, \theta) & \text{(IR)} \end{array} \right.$$

(IR) is the Agent's *Individual Rationality* or *Participation* Constraint, which ensures that the agent prefers to accept the contract.[3] Is is easy to see that (IR) must bind at a solution — if it did not, the principal can raise profits by raising t while still satisfying (IR). Expressing t from (IR), substituting into the objective function, and discarding the constant $v(0, \theta)$, we see that the Principal solves

$$\max_{x \in X} v(x, \theta) - c(x).$$

This is exactly the *total surplus-maximization* problem, hence the resulting consumption level is socially optimal (also called *first-best*). Intuitively, since the participation constraint binds regardless of the Agent's type, the Principal extracts all the surplus above the agent's reservation utility, and therefore has the incentive to maximize it. This situation is known as *first-degree price discrimination.* It can be illustrated in a diagram that resembles an Edgeworth-box in the (x, t) space (without actually limiting the size of the box):

Figure Here

This setting offers the simplest possible demonstration of the *Coase Theorem*, which says that private bargaining among parties in the absence of "transaction costs" results in socially efficient outcomes. One view of the theorem is as a definition of "transaction costs" as everything that could prevent the parties from achieving efficiency. We will see that private information will indeed constitute a "transaction cost," i.e., give rise to inefficiency.

In what follows, we will make some assumptions that will guarantee a well behaved mathematical model of monopolistic screening. First, assume that efficient consumption levels exist and are unique for each type:

A1: For each θ, $\arg\max_{x \in X} [v(x, \theta) - c(x)] = \{x^*(\theta)\}$.[4]

Next, using Theorem 1(b), we also formulate assumptions that ensure that the first-best consumption is strictly increasing in type:

[3] The term "Individual Rationality" is obsolete, but, unfortunately, well-established.

[4] Uniqueness of efficient trades simplifies the exposition, but all results can be restated for multiple efficient trades. Indeed, note that Theorem 1 holds with multiple maximizers, for any selection of maximizers.

One way to ensure single-valuedness is by assuming that $v(x, \theta) - c(x)$ is strictly concave in x, which is often done. We will not need concavity, but our graphical illustrations will assume it.

A2: $v(x,\theta)$ satisfies SCP.

A3: $c(x)$ is differentiable in x.

A4: For each θ, $x^*(\theta)$ is in the interior of X.[5]

Assumptions A2 and A3 imply that the total surplus $v(x,\theta)-c(x)$ satisfies SCP, which together with A4 allows us to use Theorem 1(b) to conclude that the efficient consumption $x^*(\theta)$ is strictly increasing in θ.

1.4 The Revelation Principle

We now consider the contracting problem with private information. Suppose the principal offers a tariff $T : X \to \Re$, with $T(0) = 0$. Let $x(\theta) \in \arg\max_{x\in X}\left[v(x,\theta) - T(x)\right]$ be the Agent's choice from the tariff when his type is θ, and let $t(\theta) = T(x(\theta))$. Then the following inequalities must hold:

$$\begin{array}{ll} v\left(x(\theta),\theta\right) - t(\theta) \geq v\left(0,\theta\right) & \forall\theta\in\Theta \ (\text{IR}_{\theta}) \\ v\left(x(\theta),\theta\right) - t(\theta) \geq v\left(x(\widehat{\theta}),\theta\right) - t(\widehat{\theta}) & \forall\theta,\widehat{\theta}\in\Theta \ (\text{IC}_{\theta\widehat{\theta}}) \end{array}$$

(IR) stands for the familiar Individual Rationality (or Participation) constraints. They inequalities (IR) reflect the fact that the agent of type θ has the option of choosing $x = 0$, i.e., rejecting the tariff, but prefers to choose $x(\theta)$. (IC) stands for incentive-compatibility. The inequalities (IC) reflect the fact that the agent of type θ has the option of choosing $x(\widehat{\theta})$, the equilibrium consumption of type $\widehat{\theta}$, but prefers to choose $x(\theta)$.

Now consider a different mechanism in which the principal asks the agent to make an announcement $\widehat{\theta}$ and then supplies the agent with the quantity $x(\widehat{\theta})$ in exchange for the payment $t(\widehat{\theta})$. Since the inequalities (IC) are satisfied, each agent will prefer to announce his true type $\widehat{\theta} = \theta$, rather than lying. Since the inequalities (IR) are satisfied, each agent will accept this mechanism. Therefore, the new mechanism will have an equilibrium in which the consumption bundles of all types coincide with those in the original tariff.

[5] The role of the last two assumptions is to ensure strict comparative statics. Without them, and replacing A2 with the weaker assumption that $v(\cdot,\cdot)$ satisfies ID, Theorem 1(a) can still be used to show that $x^*(\theta)$ is nondecreasing in θ, but it may now be constant over some region of types.

A mechanism in which the Agent is asked to announce his type and receives bundle $\left(x(\widehat{\theta}), t(\widehat{\theta})\right)$, and which satisfies inequalities (IR) and (IC), is called a *Direct Revelation Contract.* The above discussion establishes the following result:

Proposition 1 (The Revelation Principle for Tariffs) *Any tariff can be replaced with a Direct Revelation Contract that has an equilibrium giving rise to the same equilibrium consumption bundles for all types.*

This is one the most important principles of mechanism design. It can be established not only for tariffs, but for any other contract we can think of. For example, the agent could announce his type $\widehat{\theta}$, receive a type-contingent tariff $T(x|\widehat{\theta})$, and choose consumption. All such mechanisms can be replaced with Direct Revelation mechanisms.??[6] The Revelation Principle will later be extended to multi-agent situations.

It turns out that every Direct Revelation Mechanism $(x(\theta), t(\theta))_{\theta \in \Theta}$ can be replaced with a tariff

$$T(x) = \begin{cases} t(\theta) & \text{if } x = x(\theta) \text{ for some } \theta \in \Theta, \\ +\infty & \text{otherwise.} \end{cases}$$

Thus, if every contract can be replaced with a direct revelation mechanism, it can also be replaced with a tariff. This observation is known as the "**Taxation Principle**," by analogy with the Revelation Principle. In real life, we observe tariffs more often than direct revelation mechanisms, probably because the set of possible type spaces may be hard to describe in reality. For most theoretical purposes, however, it is more convenient to represent a contract as a direct revelation contract than as a tariff. The taxation principle also implies that restricting attention to tariffs is without loss.

The Principal's problem for choosing the profit-maximizing Direct Revelation Contract under private information can be written as:

$$\left\{ \begin{array}{lll} \displaystyle\max_{x:\Theta\to X,\ t:\Theta\to\Re} & E_\theta\left[t(\theta) - c(x(\theta))\right] & \\ \text{s.t.} & v(x(\theta),\theta) - t(\theta) \geq v(x(\hat{\theta}),\theta) - t(\hat{\theta}) \ \ \forall\,\theta,\hat{\theta} & (\text{IC}_{\theta\widehat{\theta}})\ \forall\theta,\widehat{\theta} \in \Theta \\ & v(x(\theta),\theta) - t(\theta) \geq v(0,\theta)\ \forall\,\theta & (\text{IR}_\theta)\ \forall\theta \in \Theta \end{array} \right.$$

[6] We do not consider the possibility of randomization. Such randomization may sometimes be useful for the principal - see Section ?? below.

This is called the *Second-Best* program, since, in contrast to the first-best (full-information) program, the principal now has to honor the agent's incentive constraints.

1.5 Solution with Two Types

This subsection solves the second-best program in the case in which the agent can have one of two possible types: $\Theta = \{\theta_L, \theta_H\}$, with $\theta_H > \theta_L$. Type θ_H will be called the "high type" and type θ_L the "low type". The Principal's prior is given by $\Pr\{\theta = \theta_L\} = \pi \in (0,1)$ and $\Pr\{\theta = \theta_H\} = 1 - \pi$. The Principal's problem of choosing the optimal Direct Revelation Contract $\langle (x(\theta_L), t(\theta_L)), (x(\theta_H), t(\theta_H)) \rangle = \langle (x_L, t_L), (x_H, t_H) \rangle$ can be written as:

$$\text{(SB)} \left\{ \begin{array}{rll} \max\limits_{\langle (x_L,t_L),(x_H,t_H)\rangle} & \pi[t_L - c(x_L)] + (1-\pi)[t_H - c(x_H)] & \\ \text{s.t.} & v(x_L,\theta_L) - t_L \geq v(0,\theta_L) & (\mathrm{IR_L}) \\ & v(x_H,\theta_H) - t_H \geq v(0,\theta_H) & (\mathrm{IR_H}) \\ & v(x_L,\theta_L) - t_L \geq v(x_H,\theta_L) - t_H & (\mathrm{IC_{LH}}) \\ & v(x_{H,}\theta_H) - t_H \geq v(x_L,\theta_H) - t_L & (\mathrm{IC_{HL}}) \end{array} \right.$$

Note that if we ignored the ICs, the problem would break down into the first-best problems for each type, and we would obtain the First-Best allocation $\langle (x^*(\theta_L), t^*(\theta_L)), (x^*(\theta_H), t^*(\theta_H)) \rangle = \langle (x_L^*, t_L^*), (x_H^*, t_H^*) \rangle$. However, this allocation is not incentive-compatible, i.e., it violates the IC constraints. To see this, recall that in the first-best allocation both types receive their reservation utilities. However, if the θ_H type deviates and claims to be an θ_L type, he achieves a positive surplus:

$$v(x_L^*, \theta_H) - v(0,\theta_H) - t_L^* > v(x_L^*, \theta_L) - v(0,\theta_L) - t_L^* = 0$$

by SCP whenever $x_L^* > 0$. Intuitively, the high type is wiling to pay more than the low type for an increase of trade from 0 to x_L^*. Thus, the first-best contract is not incentive-compatible when types are privately observed. This suggests intuitively that it is the incentive constraint ($\mathrm{IC_{HL}}$) that will be binding in the second-best problem. This intuition is confirmed formally in the following Lemma, which characterizes which constraints bind and which ones do not:

Lemma 2 *At the solution to (SB), constraints (IR_L) and (IC_{HL}) bind, whereas constraints (IR_H) and (IC_{LH}) are redundant.*

Proof.

IR$_\text{H}$ is redundant. We show that [(IR$_\text{L}$) and (IC$_\text{HL}$)] ⇒ (IR$_\text{H}$):

$$v(x_H,\theta_H)-t_H-v(0,\theta_H) \overset{(\text{IC}_\text{HL})}{\geq} v(x_L,\theta_H)-t_L-v(0,\theta_H)$$
$$\overset{\text{SCP}}{\geq} v(x_L,\theta_L)-t_L-v(0,\theta_L) \overset{(\text{IR}_\text{L})}{\geq} 0$$

Thus, (IR$_\text{H}$) can be discarded.

(IR$_\text{L}$) binds. Otherwise increasing both t_L and t_H by a small $\varepsilon > 0$ would preserve (IR$_\text{L}$), not affect (IC$_\text{HL}$) and (IC$_\text{LH}$), and raise profits.

(IC$_\text{HL}$) binds. Otherwise, increasing t_H by a small $\varepsilon > 0$ would preserve (IC$_\text{HL}$), not affect (IR$_\text{L}$), relax (IC$_\text{LH}$), and raise profits.

(IC$_\text{LH}$) is redundant: Assume not, and let $\{(x'_H,t'_H),(x'_L,t'_L)\}$ be a solution to the *reduced problem* subject to *only* (IR$_\text{L}$) and (IC$_\text{HL}$). If (IC$_\text{LH}$) is not redundant then this solution violates (IC$_\text{LH}$): type θ_L strictly prefers (x'_H,t'_H) to (x'_L,t'_L). If $t'_H - c(x'_H) \geq t'_L - c(x'_L)$, the principal can raise profits by giving both types $(x_H, t_H+\varepsilon)$ (recall that (IR$_\text{H}$) can be discarded). If $t'_H - c(x'_H) < t'_L - c(x'_L)$, she can do better by giving both types (x_L, t_L). Note that (IR$_\text{L}$) is satisfied in both cases because (x_L, t_L) satisfies it and (x_H, t_H) is strictly preferred by type θ_L. (IC$_\text{HL}$) is trivially satisfied in both cases. Thus, we have a contradiction, and (IC$_\text{LH}$) can be discarded. ■

Using Lemma 2, we can qualitatively characterize a solution to the program (SB). First, we consider the two types' rents, i.e., their utilities in excess of the reservation utilities. Lemma 2 implies that:

S1: No rent for the low type: (IR$_\text{L}$) binds.

From this we can calculate the low type's payment as

$$t_L = v(x_L,\theta_L) - v(0,\theta_L).$$

As for the high type, according to Lemma 2, we can calculate his rent from the binding ($\mathrm{IC_{HL}}$):

$$\begin{aligned} v(x_H,\theta_H)-t_H-v(0,\theta_H) &= v(x_L,\theta_H)-t_L-v(0,\theta_H) \\ &= [v(x_L,\theta_H)-v(x_L,\theta_L)]-[v(0,\theta_H)-v(0,\theta_L)]\geq 0 \end{aligned}$$

by SCP, and the inequality is strict when $x_L > 0$. Thus, we have

S2: High type has a positive rent when $x_L > 0$: ($\mathrm{IR_H}$) does not bind.

Intuitively, the high type can always get a positive rent just by choosing the low type's bundle, and this is exactly his rent in the optimal contract. We call this his *information rent* because it is due to the agent's hidden information, more precisely, to the principal's not knowing that she faces the high type.

Expressing transfers from the binding constraints ($\mathrm{IR_L}$) and ($\mathrm{IC_{HL}}$), substituting them into the Principal's objective function, and ignoring the terms that do not depend on x_L,x_H, the Principal's program can be written as

$$\max_{x_H,x_L\in X} \overbrace{\pi[v(x_L,\theta_L)-c(x_L)]+(1-\pi)[v(x_H,\theta_H)-c(x_H)]}^{\text{total expected surplus}} \\ -\underbrace{(1-\pi)[v(x_L,\theta_H)-v(x_L,\theta_L)]}_{\text{information rent of high type}}$$

We see that the objective function is additively separable in x_L and x_H, hence the program can be broken into two:

$$\begin{aligned} &\max_{x_H\in X}(1-\pi)[v(x_H,\theta_H)-c(x_H)] \\ &\max_{x_L\in X}\pi[v(x_L,\theta_L)-c(x_L)]-(1-\pi)[v(x_L,\theta_H)-v(x_L,\theta_L)] \end{aligned}$$

From the first line we see that x_H maximizes the total surplus for the high type:

S3.: Efficiency at the top: $x_H = x_H^*$.

Graphical intuition: Since ($\mathrm{IC_{LH}}$) does not bind, we can move type H's bundle along his indifference curve to the tangency point.

As for x_L, let $z \in [-1, 0]$ and consider the parametrized objective function

$$\varphi(x_L, z) = \pi \left[v(x_L, \theta_L) - c(x_L)\right] + z\,(1 - \pi)\left[v(x_L, \theta_H) - v(x_L, \theta_L)\right].$$

Here $z = 0$ corresponds to surplus-maximization and $z = -1$ corresponds to the Principal's second-best problem. Note that

$$\frac{\partial^2 \varphi(x_L, z)}{\partial x_L \partial z} = (1 - \pi)\left[v_x(x_L, \theta_H) - v_x(x_L, \theta_L)\right] > 0$$

since by SCP of the agent's value function $v(x, \theta_H)$ the square bracket is positive. Therefore, the function $\varphi(x_L, z)$ has the SCP in (x_L, z). Since by assumption x_L^* is in the interior of X, Theorem 1(b) implies

S4: Downward distortion at the bottom: $x_L < x_L^*$.

Figure Here

Graphical intuition: Start at the optimal First-Best choices of x. A small reduction x_L results in a *second-order reduction* in total surplus for the low type, but a *first-order reduction* on the high type's rent through relaxing the $(\mathrm{IC_{HL}})$ constraint and allowing the Principal to raise t_H.

Conclusions S3-S4 imply that a monopolistic seller, in order to screen the buyers of different types with lower information rents, *enlarges the quantity spectrum* relative to the efficient quantity spectrum. This result is quite general.

Remark 1 *Here we have separation of the two types, since $x_L < x_L^* < x_H^* = x_H$. However, as will be shown below, this finding does not always generalize to more than two types - some pooling is possible.*

Remark 2 *With a different reservation bundle $\left(\widehat{x}, \widehat{t}\right) \neq (0, 0)$, we could have different binding constraints. For example, suppose $\max X = \widehat{x}$, i.e., the reservation trade is the maximum rather than minimum possible trade. We can apply all the theory for $y = \overline{x} - x$. All the arguments above are inverted, and we have (IR_H) and (IC_{LH}) bind, $x_L = x_L^*$ and $x_H > x_H^*$. For a reservation bundle $\left(\widehat{x}, \widehat{t}\right)$ that satisfies $\widehat{x} \in [x_L^*, x_H^*]$ we even obtain the First-Best solution (Check!).*

Figure Here

1.5.1 Many Discrete Types

This setup was explored by Maskin and Riley (1984). There are many types, $\theta \in \{\theta_1, \theta_2, ..., \theta_n\}$ which are ordered such that $\theta_i > \theta_{i-1}$ for all $i > 2$. Let $\pi_i = \Pr\{\theta = \theta_i\}$, and assume that all types have the same reservation utility normalized to 0. Then, the principal's problem is:

$$\begin{array}{lll} \max\limits_{\{(t_i,x_i)\}_{i=1}^n} & \sum\limits_{i=1}^{n} \pi_i(t_i - c(q_i)) & \\ \text{s.t.} & v(x_i, \theta_i) - t_i \geq 0 \ \forall\, i & \text{(IR)} \\ & v(x_i, \theta_i) - t_i \geq v(x_j, \theta_i) - t_j \ \ \forall\, i \neq j & \text{(IC)} \end{array}$$

which is the straightforward extension of the two type case. This is, however, a complicated problem, especially as n grows large: There are a total of n (IR) constraints and another $n(n-1)$ (IC) constraints. It turns out that we can reduce the problem here too, as before, in a very appealing way:

Proposition 3.1: (Maskin-Riley) *The principal's problem reduces to*:

$$\begin{array}{lll} \max\limits_{\{(t_i,x_i)\}_{i=1}^n} & \sum\limits_{i=1}^{n} \pi_i(t_i - c(x_i)) & \\ \text{s.t.} & v(q_1, \theta_1) - t_1 \geq 0 & (\text{IR}_1) \\ & v(x_i, \theta_i) - t_i \geq v(x_{i-1}, \theta_i) - t_{i-1} \ \ \forall\, i = 2, ..., n & \text{(DIC)} \\ & x_i \geq x_{i-1} \ \forall\, i = 2, ..., n & \text{(MON)} \end{array}$$

That is, there is one (IR) constraint, $(n-1)$ "Downward" (IC) constraints, and another $(n-1)$ "Monotonicity" constraints. These features are features of the solution, that must hold for any solution under the assumptions that we usually make. We do not prove this proposition, but it is interesting to go over the features of the solution*:*

1. **Monotonicity:** We have this as a constraint, but it turns out to be a feature of the solution.

2. **No rents for the low type**: All other types will have some information rents.

3. **Efficiency at the top**: $x_n^{SB} = x_n^{FB}$ and $x_i^{SB} < x_i^{FB}$ for all $i = 1, ..., n-1$.

4. **“Bunching”:**We may have two or more *adjacent types* that will get the same (x, t). (This depends on the distribution of types. Separating two adjacent types may give too much rents to a the higher type of the two, and to all types that are higher. In such a case it may be better to “pool” these two adjacent types and offer them the same contract.)

5. **“Reversals”:** Different reservations can cause switching of the results, as we saw for the two type case.

1.6 Solution with a Continuum of Types

Now let the type space be continuous, $\Theta = [\underline{\theta}, \overline{\theta}]$, with the cumulative distribution function $F(\cdot)$, and with a strictly positive density $f(\theta) = F'(\theta)$.

The principal's problem is:

$$\begin{cases} \max\limits_{\langle x(\cdot),t(\cdot)\rangle} & \int\limits_{\underline{\theta}}^{\overline{\theta}} [t(\theta) - c(x(\theta))] f(\theta) d\theta \\ \text{s.t.} & v(x(\theta), \theta) - t(\theta) \geq v(x(\hat{\theta}), \theta) - t(\hat{\theta}) \ \forall \, \theta, \hat{\theta} \quad (\text{IC}_{\theta,\widehat{\theta}}) \ \forall \theta, \widehat{\theta} \in \Theta \\ & v(x(\theta), \theta) - t(\theta) \geq v(0, \theta) \ \forall \, \theta \quad (\text{IR}_{\theta}) \ \forall \theta \in \Theta \end{cases}$$

For simplicity we only consider contracts $\langle x(\cdot), t(\cdot)\rangle$ that are piecewise continuously differentiable.[7]

Just as in the two-type case, out of all the participation constraints, only the lowest type's IR binds:

Lemma 3 *At a solution $(x(\cdot), t(\cdot))$, all IR$_\theta$ with $\theta > \underline{\theta}$ are not binding, IR$_{\underline{\theta}}$ is binding.*

Proof. As in the two-type case, by SCP, $\text{IR}_{\underline{\theta}}$ and $\text{IC}_{\underline{\theta},\theta}$ imply IR_θ.

Now, if $\text{IR}_{\underline{\theta}}$ were not binding, we could increase $t(\theta)$ by $\varepsilon > 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, which would preserve all incentive constraints and increase the Principal's profit. ■

As for the analysis of ICs, it appears very difficult, because with a continuum of types there is a double continuum of incentive constraints. Mirrlees

[7] That is, there is a discrete number of points where the function is not differentiable, and it is continuously differentiable everywhere else. This restriction is not necessary. In fact, any incentive-compatible contract can be shown to be integrable, which suffices for the subsequent derivation.

in his Nobel prize - winning work suggested a way to reduce these constraints to a much smaller number, by replacing them with the corresponding First-Order Conditions. The argument is as follows: If we think of the agent's problem as choosing an announcement $\widehat{\theta} \in \Theta$, where the parameter is his true type $\theta \in \Theta$, then his maximization problem can be written as,

$$\max_{\hat{\theta}\in\Theta} \Phi(\hat{\theta}, \theta) = v(x(\hat{\theta}), \theta) - t(\hat{\theta}) \,,$$

and the (IC) constraints can be written as,

$$\theta \in \arg\max_{\hat{\theta}\in\Theta} \Phi(\hat{\theta}, \theta).$$

For all $\theta \in \left(\underline{\theta}, \overline{\theta}\right)$ at which the objective function is differentiable (which is by assumption almost everywhere), the following First-Order Condition must therefore hold:

$$0 = \frac{\partial}{\partial\hat{\theta}}\Phi(\theta, \theta) \,. \qquad (\text{ICFOC}_\theta)$$

That is, truth telling, or incentive compatibility, implies that the FOC of $\Phi(\hat{\theta}, \theta)$ is satisfied when $\hat{\theta} = \theta$.

Define the Agent's *equilibrium utility* as $U(\theta) \equiv \Phi(\theta, \theta)$, which depends on θ in two ways — through the agent's true type and through his announcement. Differentiating with respect to θ, we have

$$U'(\theta) = \Phi_\theta(\theta, \theta) + \Phi_{\hat{\theta}}(\theta, \theta).$$

Since the second term equals zero by (ICFOC$_\theta$), we have

$$U'(\theta) = \Phi_\theta(\theta, \theta) = v_\theta(x(\theta), \theta).$$

This is an equivalent representation of (ICFOC$_\theta$). The result is the *Envelope Theorem* — the full derivative of the value of a maximization program with respect to the parameter θ equals to the partial derivative (holding fixed the maximizer - the agent's announcement).

Notice that (ICFOC$_\theta$) incorporates *local* incentive constraints, i.e., the Agent does not gain by misrepresenting θ around the neighborhood of θ. By itself, it does not ensure that the Agent does not want to misrepresent θ by a *large* amount. For example, (ICFOC$_\theta$) could be consistent with the truthful announcement announcement $\widehat{\theta} = \theta$ being a local maximum, but not a global one. It is even conceivable that the truthful announcement is a local minimum!

Fortunately, these situations can be easily ruled out. For this purpose, recall that by SCP, 1(b) establishes that the Agent's consumption choices from any tariff (and therefore in any incentive-compatible contract) are nondecreasing in type (implying that it is differentiable almost everywhere). Thus, any piecewise differentiable IC contract must satisfy

$$x'(\theta) \geq 0 \text{ a.e. } \theta \ . \tag{M}$$

It turns out that under SCP, ICFOC in conjunction with (M) do ensure that truth-telling is a global maximum, i.e., all ICs are satisfied:

Proposition 2 $(x(\cdot), t(\cdot))$ *is Incentive Compatible if and only if both of the following hold:*

(M) $x'(\theta) \geq 0$ for a.e. θ.

(ICFOC) $v_x(x(\theta), \theta)x'(\theta) - t'(\theta) = 0$ for a.e. θ.

Proof. The "only if" part was established above. It remains to show that monotonicity and local IC imply global IC. Note that

$$\frac{\partial}{\partial \hat{\theta}} \Phi(\hat{\theta}, \theta) = v_x(x(\widehat{\theta}), \theta) \cdot x'(\widehat{\theta}) - t'(\widehat{\theta})$$

By (M) $x'(\widehat{\theta}) \geq 0$. For $\widehat{\theta} > \theta$, by SCP $v_x(x(\widehat{\theta}), \theta) \leq v_x(x(\widehat{\theta}), \widehat{\theta})$, thus the above equality implies

$$\frac{\partial}{\partial \hat{\theta}} \Phi(\hat{\theta}, \theta) \leq v_x(x(\widehat{\theta}), \widehat{\theta}) \cdot x'(\widehat{\theta}) - t'(\widehat{\theta}) = \frac{\partial}{\partial \hat{\theta}} \Phi(\hat{\theta}, \hat{\theta}) = 0$$

by ICFOC. Thus, the function $\Phi(\hat{\theta}, \theta)$ is nonincreasing in $\hat{\theta}$ for $\widehat{\theta} > \theta$. Similarly, we show that $\Phi(\hat{\theta}, \theta)$ is nondecreasing in $\hat{\theta}$ for $\widehat{\theta} < \theta$.[8] This implies that $\theta \in \arg\max_{\hat{\theta} \in \Theta} \Phi(\hat{\theta}, \theta)$. ■

[8] Such a function is called pseudoconcave in $\hat{\theta}$.

1.6.1 The Relaxed Problem

From the analysis above we can rewrite the principal's problem as

$$\begin{cases} \max\limits_{x(\cdot),t(\cdot)} & \int\limits_{\underline{\theta}}^{\overline{\theta}} [t(\theta) - c(x(\theta))] f(\theta) d\theta & \\ \text{s.t.} & x'(\cdot) \geq 0 & \text{(M)} \\ & v_x(x(\theta), \theta) x'(\theta) - t'(\theta) = 0 \ \ \forall \theta & \text{(ICFOC)} \\ & v(x(\underline{\theta}), \underline{\theta}) - t(\underline{\theta}) = v(0, \underline{\theta}) & (\underline{\text{IR}}) \end{cases}$$

To solve this program in general requires optimal control theory, but this can sometimes be avoided by the following *Shortcut:* We solve the *relaxed program* obtained by ignoring the monotonicity constraint (M). If it turns out that the resulting solution satisfies (M), then we are done.

To solve the relaxed problem, ICFOC$_\theta$ can equivalently be written as

$$U(\theta) = U(\underline{\theta}) + \int\limits_{\underline{\theta}}^{\theta} v_\theta(x(s), s) ds, \tag{1.1}$$

and the binding $(\underline{\text{IR}})$ means $U(\underline{\theta}) = v(0, \underline{\theta})$, thus (ICFOC$_\theta$) and $(\underline{\text{IR}})$ together are equivalent to

$$U(\theta) = v(0, \theta) + \int\limits_{\underline{\theta}}^{\theta} v_\theta(x(s), s) ds. \tag{1.2}$$

Thus, in equilibrium the information rent of a type θ Agent equals to $\int\limits_{\underline{\theta}}^{\theta} v_\theta(x(s), s) ds$.

This implies that we can substitute transfers $t(\theta) = v(x, \theta) - U(\theta)$ into the Principal's objective function. Eliminating the constant term $v(0, \theta)$, the objective function takes the familiar form as the expected difference between total surplus and the Agent's information rent:

$$\max_{x(\cdot)} \int\limits_{\underline{\theta}}^{\overline{\theta}} \left[\underbrace{v(x(\theta), \theta) - c(x(\theta))}_{\text{Total Surplus}} - \underbrace{\int\limits_{\underline{\theta}}^{\theta} v_\theta(x(s), s) ds}_{\text{Information rent of } \theta} \right] f(\theta) d\theta \tag{1.3}$$

We can rewrite the *expected* information rents using integration by parts:

$$\begin{aligned}
\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} [v_\theta(x(s),s)ds] f(\theta)d\theta &= \left[\int_{\underline{\theta}}^{\theta} v_\theta(x(s),s)ds \cdot F(\theta)\right|_{\underline{\theta}}^{\overline{\theta}} - \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(x(\theta),\theta)F(\theta)d\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(x(\theta),\theta)d\theta - \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(x(\theta),\theta)F(\theta)d\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(x(\theta),\theta)\frac{1-F(\theta)}{f(\theta)} f(\theta)d\theta \qquad (1.4)
\end{aligned}$$

where the second equality follows from $F(\overline{\theta}) = 1$, and $F(\underline{\theta}) = 0$, and the third equality is obtained by multiplying each one of the integrands by $\frac{f(\theta)}{f(\theta)}$.

With the expected information rents given as in (1.4) above, we can rewrite the principal's problem as,

$$\max_{x(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} \underbrace{\left[v(x(\theta),\theta) - c(x(\theta)) - v_\theta(x(\theta),\theta)\frac{1-F(\theta)}{f(\theta)}\right]}_{\text{virtual surplus}} f(\theta)d\theta. \qquad (1.5)$$

Thus, the Principal will maximize the expected value of the expression within square brackets, which is called the *virtual surplus*. This expected value is maximized by simultaneously maximizing the virtual surplus at (almost) every state θ, i.e., by pointwise maximization which implies that for (almost) all θ,

$$x(\theta) \in \arg\max v(x,\theta) - c(x) - \left[\frac{1-F(\theta)}{f(\theta)}\right] v_\theta(x,\theta).$$

This implicitly gives us the consumption rule $x(\cdot)$ in the optimal contract for the relaxed program. However, for this solution of the relaxed program to be a solution to the original problem we need to check that this solution satisfies the monotonicity constraint (M), i.e., that the resulting consumption rule $x(\theta)$ is nondecreasing. We can check this using Theorem 1. Letting

$$\varphi(x,\theta) \equiv v(x,\theta) - c(x) - \left[\frac{1-F(\theta)}{f(\theta)}\right] v_\theta(x,\theta)$$

represent the objective function, and assuming that $v(x, \theta)$ is sufficiently smooth, we have

$$\frac{\partial^2 \varphi(x, \theta)}{\partial x \partial \theta} = v_{x\theta} - \frac{v_{x\theta\theta}}{h(\theta)} + v_{x\theta} \frac{h'(\theta)}{[h(\theta)]^2}.$$

where $h(\theta) \equiv f(\theta) / [1 - F(\theta)] > 0$ is called the *hazard rate* of type θ.[9]

By Theorem 1, a sufficient condition for $x(\theta)$ to be nondecreasing is for the cross-derivative to be positive, which under SCP ($v_{x\theta} > 0$) can be ensured with the additional assumptions $v_{x\theta\theta} \leq 0$ and $h'(\theta) \geqslant 0$ (increasing hazard rate). Thus, under the two added assumptions, any solution to the relaxed problem (1.3) satisfies (M), and solves the full problem. Without the two added assumptions, this may not be the case, i.e., constraint (M) may bind.

Assuming that the relaxed solution satisfies monotonicity, we can consider the properties of the solution. In particular, consider the FOC of $\varphi(x, \theta)$ with respect to x,

$$v_x(x(\theta), \theta) - c'(x(\theta)) - \frac{1}{h(\theta)} \cdot v_{x\theta}(x(\theta), \theta) = 0\,. \tag{1.6}$$

Since at $\overline{\theta}$ we have $\frac{1}{h(\overline{\theta})} = 0$, we get the familiar *efficiency at the top* result: $v_x(x(\overline{\theta}), \overline{\theta}) = c'(x(\overline{\theta}))$. Furthermore, since $\frac{1}{h(\theta)} > 0$ for all $\theta < \overline{\theta}$, we also get *distortion everywhere else:* $v_x(x(\theta), \theta) > c'(x(\theta))\ \forall\, \theta < \overline{\theta}$. The intuition is the same as for the two-type case we analyzed before. If we take $x(\theta)$ for some $\theta < \overline{\theta}$ and increase it, we gain an increase in total surplus through type θ, but we have to give higher information rents to all the types $\theta' > \theta$. Using the definition of $h(\cdot)$, we can rewrite (1.6) as,

$$f(\theta)\left[v_x(x(\theta), \theta) - c'(x(\theta))\right] = (1 - F(\theta)) v_{x\theta}(x(\theta), \theta)\,,$$

which can be interpreted as follows: The left-hand side of the equation is the total surplus generated when a type θ gets an infinitesimal increase in $x(\theta)$, whereas the right-hand side of the equation is the sum of the rents that an increase in $x(\theta)$ gives to all $\theta' > \theta$ (which have a total measure of $(1 - F(\theta))$ agents). The increase in surplus for type θ is what the monopolist can extract from this type, but this is at the cost of paying informational rents to all the "higher" type agents.

[9] The term comes from Actuary studies. Intuitively, if $f(t)$ is the probability of dying at time t, then $h(t)$ is the probability of dying at time t conditional on not dying before time t. In other words, it is just Bayes updating.

1.6.2 What to do if monotonicity binds (technical)

Incorporating ICFOC$_\theta$ and ($\underline{\text{IR}}$) into the objective function of the principal, the original program can be written as follows,

$$\begin{array}{ll} \max\limits_{\{x(\cdot),\mu(\cdot)\}} & \int\limits_{\underline{\theta}}^{\overline{\theta}} \left[v(x(\theta),\theta) - c(x(\theta)) - v_\theta(x(\theta),\theta)\frac{1}{h(\theta)}\right] f(\theta)d\theta \\ \text{s.t.} & x'(\theta) = \mu(\theta) \ \forall\theta \ \ (MU) \\ & \mu(\theta) \geq 0 \ \forall\theta \qquad (M) \end{array}$$

which now appears in the format of an optimal control problem in dynamic optimization, with state variable $x(\cdot)$ and control variable $\mu(\cdot)$. (See Kamien-Schwartz part II, sections 17-18.) We solve this by applying known methods from control theory.

The Hamiltonian is then

$$H(\theta, x, \mu, \lambda) = [v(x(\theta),\theta) - c(x(\theta)) - v_\theta(x(\theta),\theta))\frac{1}{h(\theta)}]f(\theta) + \lambda(\theta)\mu(\theta)\,,$$

where $\lambda(\cdot)$ is the multiplier of (MU). If at the solution $\overline{x}(\cdot)$ we get $\overline{x}'(\cdot) > 0$, then $\mu(\theta) > 0$ and $\lambda(\theta) = 0$ so that the FOC is identical to the unconstrained program. Finding the intervals over which $\overline{x}(\cdot)$ is constant is trickier.

Let $x^*(\cdot)$ be the solution to the unconstrained program (1.5) above, and let $\overline{x}(\cdot)$ be the solution to the real constrained problem. The solution will look like the following figure:

Figure Here

At the solution, there will be an interval, $[\theta_1, \theta_2]$, (or more than one such interval, depending on how many "humps" we get from $x^*(\cdot)$) such that:

$$\overline{x}(\theta) = \begin{cases} x^*(\theta) & \forall\theta \in [\underline{\theta}, \theta_1] \\ x^*(\theta_1) & \forall\theta \in [\theta_1, \theta_2] \\ x^*(\theta) & \forall\theta \in [\theta_2, \overline{\theta}] \end{cases}$$

Similarly, $t(\cdot)$ will coincide with $t^*(\cdot)$ for $\theta \notin (\theta_1, \theta_2)$ and will be constant for $\theta \in (\theta_1, \theta_2)$. Salanie gives good "intuition" (math intuition) for why $x^*(\cdot)$ and $\overline{x}(\cdot)$ coincide over the areas where $\overline{x}(\cdot)$ is strictly increases:

?????????????????????????

1.7 Applications of the Model

1.7.1 Second Degree Price Discrimination

This is the example we investigated where x was quantity, and this was developed by Maskin-Riley (1984). They also generalize the Mussa-Rosen (1978) model to address quality and quantity discrimination, and they develop the discrete-type solution we discussed with n-types. They also demonstrate that all the previous "sorting" models fit under the general framework of optimal sorting mechanisms.

1.7.2 Vertical Differentiation: Quality

A monopoly manufactures goods in 1-unit quantities each, but they can differ in quality. Just take x to be quality of a unit of good, and $c(x)$ to be the cost of producing one unit at quality x, and we are back in the model we analyzed. This model was analyzed by Mussa-Rosen (1978) and they note the connection to Mirrlees' work but just apply it to this problem. (Examples: train/plane classes, "olives" and "figs" restaurants in Charlestown, 486SX and 486DX computer chips.)

1.7.3 Optimal Income Tax

This is the seminal hidden-information paper by Mirrlees (1971) that earned him the Nobel Prize for 1996. A Government wishes to do some benevolent act (e.g., raise taxes for public project, or redistribute consumption given some social welfare function.) Each agent can produce output according to the production function

$$y(\ell,\theta) = \theta\ell\,,$$

where ℓ is labor and θ is the marginal product of the agent. We assume that both ℓ and θ are private information (not observed by the government) while y is observable by the government. The utility of an individual from consuming c units of the output good and working ℓ units of labor is given by $v(c)-\ell$. Let τ denote a tax that the government can impose on an agent's production. Then, given a production level y, the agent's utility (given his type) is:

$$v(x-\tau(x)) - \frac{x}{\theta}\,,$$

or, using a monotonic transformation (multiply by θ),

$$u(x,\tau,\theta) = \theta v(x-\tau) - x\,.$$

We can now redefine the variables so as to put this problem in the notation of our original model. That is, let

$$\begin{aligned} x(\theta) &\equiv x(\theta) - \tau(\theta), \\ t(\theta) &\equiv x(\theta)\,. \end{aligned}$$

Here, instead of the IR constraints (taxes are not voluntary!) we will have an integral constraint for a balanced budget of the government. (See Salanie or Mirrlees for a detailed account of this problem.)

Dual problem!!!

1.7.4 Regulating a (Natural) Monopolist

A natural monopolist has costs $\psi(x,\theta)$ where x is output produced (e.g., electricity) and θ is a private cost parameter measuring efficiency: $\psi_x > 0, \psi_\theta < 0, \psi_{\theta x} < 0$ (higher θ implies more efficiency and lower marginal costs). Given a subsidy s from the government, the firm maximizes profits:

$$\pi(x,\theta,s) = p(x)x - \psi(x,\theta) + s.$$

The government (regulator) maximizes social welfare:

$$B(x) - (1-\lambda)s + s - \psi(x,\theta)\,,$$

where $B(x) = \int\limits_0^x p(x)dx$ is the social surplus from producing x, and $\lambda > 0$ is the "shadow cost" of distortionary taxes (taxes are needed to collect the subsidy s). (Everything is common knowledge except θ.) The government can offer the firm a menu: $(x(\theta), s(\theta))$ (or, using the taxation principle, the government can offer a subsidy $s(x)$ and the firm just chooses x) and the firm's profits are,

$$u(x,s,\theta) = -\psi(x(\theta),\theta) + s(\theta)$$

(That is, we can redefine the subsidy s to include the revenues that the government can collect and transfer to the firm.) The government must assure that $u(x,s,\theta) \geq 0$, which is the IR constraint, and must also respect the IC constraints of truthful revelation. We can now redefine the variables so as to put this problem in the notation of our original model. That is, let

$$\begin{aligned} x(\theta) &\equiv x(\theta) \\ t(\theta) &\equiv -s(\theta) \\ v(x,\theta) &\equiv -[\psi(x,\theta)+F] \end{aligned}$$

and assume that $v_{\theta x} = -\psi_{\theta x} > 0$ (SC is satisfied). Letting $c(x(\theta)) \equiv \psi(x(\theta),\theta) - B(x(\theta))$, the government maximizes:

$$\max_{x(\cdot),t(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} [\lambda t(\theta) - c(x(\theta))] f(\theta) d\theta$$

subject to the standard IR and IC.

This model was initially introduced by Baron-Myerson (1982). It is interesting that they don't mention either Mirrlees or Mussa-Rosen. They take the Mechanism-Design approach and basically show that this is an application of it. (That is, a *Bayesian Incentive Compatible* mechanism with IR.) We get the same features:

1. Highest type $\overline{\theta}$ (lowest cost) produces efficiently (marginal costs equal marginal benefits).

2. Lowest type $\underline{\theta}$ (highest cost) gets no rents.

3. All $\theta > \underline{\theta}$ get informational rents.

4. All $\theta < \overline{\theta}$ produce less than the first-best production level.

Figure Here

Notes:

1. This problem is a *common values* problem: The government's utility depends on θ.

2. See Laffont-Tirole (1993) for an exhaustive analysis of regulation using mechanism design.

1.7.5 Optimal Labor Contracts

Consider the case where the manager-owner of a firm is risk neutral and the employee is risk averse to the amount of labor input. That is, assume that the worker's utility is given by,

$$u(\ell, w, \theta) = w - \psi(\ell, \theta)\,,$$

and the owner's utility is given by,

$$\pi(\ell, w, \theta) = \theta\ell - w$$

where θ is that marginal product of the worker, w is the wage the worker receives, and ℓ is the worker's labor input. As in the Mirrlees taxation model, ℓ, θ are assumed to be private information of the worker, and the employer only observes the output $\ell\theta$. We can now redefine the variables so as to put this problem in the notation of our original model. That is, let

$$\begin{aligned} x(\theta) &\equiv -\ell\theta, \\ t(\theta) &\equiv -w(\theta), \\ v(x,\theta) &\equiv -\psi(\ell,\theta) = -\psi\left(\frac{-x}{\theta}, \theta\right), \end{aligned}$$

which yields the exact same problem.

Notes:

1. Here again, like in the regulation example, we have common values.
2. This is different from Hart (1983) where firm is risk averse, workers are risk neutral, and the firm observes a demand parameter. In Hart's model we get first-best employment at the "top" (best state of demand) and under-employment for all other states of demand.

1.7.6 Monopolistic Insurance

Consider a simple monopolistic insurance market where there are two types of individuals, *low risk* (L) and *high risk* (H), each with an endowment of $\omega = (\omega_g, \omega_b)$ where ω_s is the endowment of the private good in state $s \in \{g, b\}$. The probability of a bad (b) state is $p_\theta \in \{p_H, p_L\}$ where $1 > p_H > p_L > 0$.

Figure Here

Here in general non-quasilinear utilities, the single-crossing property takes a more general form.

1.8 Extensions

1.8.1 Randomizations

1.8.2 Type-Dependent (IR) and Countervailing Incentives

This was first investigate by Lewis and Sappington (1989).

- If IR is type dependent then we don't know which IR binds (this was discussed earlier).
- If the principal can *endogenously* set the different types' outside option (????????), maybe she will be better off: e.g., fixed costs negatively correlated with MC will potentially have two effects:
  1. A type with low MC will want to overestimate MC, but then he gets low FC rebate,
  2. A type with high MC will want to underestimate MC to get high FC rebate
- These two effects may cancel out, so an incentive scheme will have higher efficiency (less distortions). Also, with a continuum of types we get pooling "in the middle."
- This relates directly to what we saw earlier in the two-type model with different reservation bundles: with low ones, we got our original solution. With high ones, the inefficiencies were reversed. With intermediate ones, we can get efficiency with asymmetric information.

1.8.3 Ex Ante Participation Constraint

Suppose Agent's ex ante vNM utility is $u()$. (IR) constraints are replaced with one ex ante (IR):

$E_\theta\left[u(v(x(\theta),\theta)-t(\theta))\right] \geq E_\theta\left[v(0,\theta)\right]$.

When agent is risk-neutral, get 1st-best. In the limiting case where agent is infinitely risk-averse: $E_\theta\left[u(w(\theta))\right] = \min_{\theta\in\Theta} w(\theta)$, and reservation utility $v(0,\theta) = \overline{v}$ does not depend on type, then ex ante are reduced ex post participation constraints.

Developed in "Implicit labor contracts" literature - firm is risk-averse.

1.8.4 Common Values

Example here.

The analysis of constraints that does not use the properties of P's objective function is same as before. Among participation constraints, only the lowest type's binds. ICFOC holds. But qualitative difference is that the likelihood of pooling may increase. For example pooling is now possible with two types - the last argument in Lemma ?? does not go through.

Exercise with ex ante participation constraint.

1.8.5 Multidimensional Characteristics

- Analysis is much more complicated, not as "nice"
- Hard to get monotonicity (Matthews-Moore 1987)
- Bunching is a big problem (See references in Salanié)

1.8.6 Informed Principal (Maskin and Tirole 1990, 1992)

- Principal has private information too, e.g., a type λ
- Contract can be contingent on Principal's announcements as well: $\{x(\theta, \lambda), t(\theta, \lambda)\}$
- We get a signalling model: The principal may signal his type with the contract she offers
- In the private values case the principal can do as good as in the full information case (where the principal's type is public information) just by offering the menu of second-best contracts for the full information case. Generically, the principal can do better because IR need only hold in expectations (using Bayesian Nash Equilibrium). (Note: in the quasilinear case there are no gains from the principal's private information.)
- In the common values case it is not necessarily true that the principal can do as well. (costly signalling).

Chapter 2

The Multi-Agent Model

2.1 General Setup

- Set of *Agents*, $I = \{1, 2, ...I\}$ (like in Mas-Colell, Whinston and Green use abuse of notation: I is used for both number of agents and for denoting the set of agents)

- Set of *Outcomes* or *Alternatives* X

- Each agent i observes a *private signal* which determines his *preferences* over alternatives $x \in X$, the signal for each $i : \theta_i \in \Theta_i$

- Each agent *maximizes expected utility* with a vNM utility over outcomes $u_i(x, \theta_i)$. (also referred to as a *Bernoulli* utility function)

- **Note**: This is the *private values case* for which θ_i can represent some signal of the agents "willingness to pay" for an object. There is also the *common values* case in which utilities are given by $u_i(x, \theta)$, and θ consists of signals that reflect the true, or absolute value of an object. (e.g., oil well site)

- The vector of types, $\theta = (\theta_1, \theta_2, ..., \theta_I) \in \Theta_1 \times \Theta_2 \times ... \times \Theta_I \equiv \Theta$ is drawn from a *prior distribution* with density $\phi(\cdot)$ [can be probabilities for *finite* Θ]. θ is also called the *state of the world.*

- **Information**:

$$\begin{array}{l} (1)\ \theta_i \text{ is } \textit{privately observed} \text{ by agent } i \\ (2)\ \{u_i(\cdot,\cdot)\}_{i=1}^{I} \text{ is } \textit{common knowledge} \\ (3)\ \phi(\cdot) \text{ is } \textit{common knowledge} \end{array}$$

- **Social Choice Function**: $f : \Theta \to X$
- **Goals**: Designer wants to implement $f(\cdot)$
- **Problem***:* Designer doesn't know θ, and if asked directly agents may misrepresent themselves given $f(\cdot)$ and their beliefs about what others will announce.

Example 1.1: A public project can be built ($k = 1$) or not ($k = 0$) where the cost of the project is 1 if built and 0 if not. Assume two agents with type spaces $\Theta_1 = [0, 1]$ and $\Theta_2 = \{\frac{3}{4}\}$ (singleton), and utilities are given by $u_i = \theta_i \cdot k - t_i$, where $t_i \in \Re$ is the payment to center. The set of outcomes is

$$X = \{(k, t_1, t_2) : k \in \{0, 1\}, t_i \in \Re, t_1 + t_2 \geq k \cdot 1\}$$

where the weak inequality implies that there is no outside funding. Assume, for example, that the center wishes to maximize a socially efficient, egalitarian (equal sharing) SCF

$$f(\theta) = \langle k(\theta), t_1(\theta), t_2(\theta)\rangle$$

which is given by

$$k(\theta) = \begin{cases} 1 \text{ if } \theta_1 > \frac{1}{4} \\ 0 \text{ otherwise} \end{cases}$$

$$t_i(\theta) = \begin{cases} \frac{1}{2} \text{ if } \theta_1 > \frac{1}{4} \\ 0 \text{ otherwise} \end{cases}$$

(Since agents have quasilinear utility functions then social efficiency is achieved by the decision rule "build if and only if $\theta_1 > \frac{1}{4}$.) Will agent 1 reveal his true θ_1? For $\theta_1 \in [0, \frac{1}{4}] \cup [\frac{1}{2}, 1]$ the answer is yes, whereas for $\theta_1 \in (\frac{1}{4}, \frac{1}{2})$ the answer is no! ■

- **Question:** Can the designer find some clever way to get the truth out of the agents?

Definition 1.1: A *Mechanism* (or Game Form) $\Gamma = (S_1, S_2, ..., S_I, g(\cdot))$ is a collection of I strategy sets (or *message spaces*) $S_1, S_2, ..., S_I$ and an outcome function $g : S_1 \times S_2 \times ... \times S_I \to X$.

- Each agent sends a message $s_i \in S_i$ to the center, and the outcome is then determined according to $g(\cdot)$. A *pure strategy* for agent i will be a function $s_i : \Theta_i \to S_i$ (We restrict attention to pure strategies. We will later consider mixed strategies.*)*

- Since each agent only knows θ_i and the outcome of mechanism depends on θ_{-i} , then the elements $\Gamma, \Theta, \phi(i)$ and $\{u_i\ (\cdot, \cdot)\}_{i=1}^{I}$ define a *Bayesian Game of Incomplete Information*, which is given in strategic form by

$$\left\langle I, \{S_i\}_{i=1}^{I}, \{\tilde{u}_i(\cdot)\}_{i=1}^{I}, \Theta, \phi(\cdot) \right\rangle$$

  where

$$\tilde{u}_i(s_1, ..., s_I, \theta_i) \equiv u_i(g(s_1, ..., s_I), \theta_i) \text{ for all } i \in I.$$

- Let $E(\Gamma, \theta) \subset S_1 \times S_2 \times ... \times S_I$ denote the set of *equilibria plays* of the game in state $\theta \in \Theta$.

  Note: $E(\Gamma, \cdot)$ depends on the equilibrium concept which we will later define. Basically, given an equilibrium concept, then these are all the equilibria for the state of the world θ.

- Two main concepts:

(1) $E^D(\Gamma, \cdot)-$ set of *Dominant Strategy* Equilibria plays
(2) $E^B(\Gamma, \cdot)$ - set of *Bayesian Nash* Equilibria plays

- Let $g(E(\Gamma, \theta)) \equiv$ set of *equilibrium supported outcomes* in state θ given the mechanism Γ.

Definition 1.2: A mechanism Γ *fully implements* the SCF $f(\cdot)$ if $g(E(\Gamma, \theta)) = f(\theta)$ for all $\theta \in \Theta$.

figure 1.1 here

- We can have *Partial Implementation* which is the case where there exists θ such that:

(1) $f(\theta) \in g(E(\Gamma, \theta))$, and

(2) $\exists\ x \in g(E(\Gamma, \theta))$ such that $x \neq f(\theta)$. That is, we encounter the problem of *Multiple Equilibria*.

Example 1.2: Majority voting: Consider $I = 3, X = \{a, b\}$, and the type θ_i determines whether $a \succsim_i b$ or $b \succsim_i a$. Suppose that $\theta_1 = \theta_2 = \theta_3 = \theta$, and that $u(a, \theta) > u(b, \theta)$ (i.e., they all have the same preferences). Yet all voting for b is a NE.

Notes:

(1) Using stronger equilibrium concept (undominated Nash) will help.

(2) Using social choice *correspondence* rather than function will help (the designer is willing to accept one of several outcomes).

We won't be concerned with multiple equilibria problems.

2.2 Dominant Strategy Implementation

2.2.1 Definition and the Revelation Principle

This is the strongest equilibrium requirement:

Definition 1.3: The strategy profile $s^*(\cdot) = (s_1^*(\cdot), ..., s_I^*(\cdot))$ is a *Dominant Strategy Equilibrium* of the mechanism Γ if for all i and for all $\theta_i \in \Theta_i$,

$$\begin{aligned} u_i(g(s_i^*(\theta), s_{-i}), \theta) &\geq u_i(g(s_i', s_{-i}), \theta) \\ \text{for all } s_i' &\in S_i \text{ and for all } s_{-i} \in S_{-i}. \end{aligned}$$

That is, no matter what i thinks that the other players are doing, $s_i^*(\theta_i)$ is the best for him when he is θ_i, and this is true whatever his type is. The beauty of this equilibrium concept is in the weak informational structure it imposes on the agents: Each player **need not forecast** what the others are doing, or what their types are.

Note: Multiple equilibria is practically not a problem (it must come from indifference between different strategies which is not a generic property).

BIG QUESTION: How do we know if $f(\cdot)$ is implementable in Dominant Strategies?

That is, **can we find** a mechanism Γ that implements $f(\cdot)$ in dominant strategies? (By *implement* we mean *partial implementation* in the sense that we do not care about the multiple equilibria problem.) To answer this question we first have to ask: **Do we have to consider all** (∞) **mechanisms**?

Let's assume that we were "lucky" and for some $f(\cdot)$ we found a mechanism Γ that implements $f(\cdot)$. Recalling Figure 1.1 we can consider the implementing equilibrium as one where each player figured out $s_i^*(\cdot)$. Now let each player i use a "computer" to compute his equilibrium message. That is, the agent feeds the computer with $\theta_i \in \Theta_i$ and the computer prints out $s_i^*(\theta_i)$.

figure 1.2 here

Question: Will the agent want to lie to the program?
Answer: No! The program *is* $s_i^*(\cdot)$.
But Note: The designer *knows* that $s_i^*(\cdot)$ is the equilibrium strategy since *she found* Γ to implement $f(\cdot)$!

Now we can consider the following alteration of the mechanism Γ made by the designer: Let Γ^* be a mechanism which calls for the agents to reveal there types, that is, $S_i = \Theta_i$ for all i. Let $g^*(\cdot)$ be the outcome function of Γ^* so that it incorporates the computer programs $s_i^*(\cdot)$ for all i *into* the mechanism Γ. That is, given announcements $\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_I$, $g^*(\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_I) \equiv g(s_1^*(\widehat{\theta}_1), s_2^*(\widehat{\theta}_2), ..., s_I^*(\widehat{\theta}_I))$. That is, one can view the new mechanism as one that works in the following way: The agents report there types θ_i (they are not restricted to tell the truth!) to the designer, the designer's mechanism Γ^* first calculates the optimal strategies for Γ, and finally uses $g(\cdot)$ for the choice of outcome.

Claim: If Γ^* is the mechanism, every agent will reveal θ_i truthfully to the designer.

The proof is trivial by the construction of Γ^*, and by the assumption that Γ implemented $f(\cdot)$ in dominant strategies. This is the idea behind the *Revelation Principle*, a strong and beautiful result. First define:

Definition 1.4: $\Gamma = (S_1, ..., S_I, g(\cdot))$ is a *Direct Revelation Mechanism* if $S_i = \Theta_i$ for all $i \in I$ and if $g(\theta) = f(\theta)$ for all $\theta \in \Theta$.

Definition 1.5: The SCF $f(\cdot)$ is *Truthfully Implementable in Dominant Strategies* if for all θ, the direct revelation mechanism $\Gamma = (\Theta_1, ..., \Theta_I, f(\cdot))$ has $s_i^*(\theta_i) = \theta_i$ for all i, and $s^* \in E^D(\Gamma, \theta)$. Equivalently, for all i and for all $\theta_i \in \Theta_i$,

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i) \text{ for all } \hat{\theta}_i \in \Theta_i \text{ and } \theta_{-i} \in \Theta_{-i} \ .$$

That is, $f(\cdot)$ is truthfully implementable in dominant strategies if truth telling is a dominant strategy in the direct revelation mechanism.

Note: *Truthfully Implementable in Dominant Strategies* is also called:
(i) dominant strategy incentive compatible
(ii) strategy proof
(iii) straightforward.

Proposition 1.1: (The Revelation Principle for Dominant Strategy Equilibria) *$f(\cdot)$ is implementable in dominant strategies if and only if it is truthfully implementable in dominant strategies.*

proof: $\underline{\Leftarrow}$: by definition.

$\underline{\Rightarrow}$**:** Suppose $\Gamma = (S_1, ..., S_I, g(\cdot)$ implements $f(\cdot)$ in dominant strategies. This implies that there exists $s^*(\cdot) = (s_1^*(\cdot), ..., S_I^*(\cdot))$ such that $g(s^*(\theta)) = f(\theta)$ for all $\theta \in \Theta$, and for all i and all $\theta_i \in \Theta_i$:

$$u_i(g(s_i^*(\theta_i), s_{-i}), \theta_i) \geq u_i(g(s_i', s_{-i}), \theta_i) \qquad (2.1)$$
$$\text{for all } s_i' \in S_i \text{ and } s_{-i} \in S_{-i}$$

In particular, for all $\theta_i \in \Theta_i$

$$u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i) \geq u_i(g(s_i^*(\hat{\theta}), s_{-i}^*(\theta_{-i})), \theta_i) \qquad (2.2)$$
$$\text{for all } \hat{\theta}_i \in \Theta_i, \text{ and } \theta_{-i} \in \Theta_{-i}$$

(That is, (2.2) is a special case of (2.1).) But $g(s^*(\theta)) = f(\theta)$ for all $\theta \in \Theta$ implies that:

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i) \text{ for all } \hat{\theta}_i \in \Theta_i, \text{ and } \theta_{-i} \in \Theta_{-i}$$

which is the condition for $f(\cdot)$ to be truthfully implementable. *Q.E.D.*

Notes:

1. The Revelation Principle does **not apply** to full implementation: Direct mechanisms may have multiple equilibria which can be eliminated by more elaborate mechanisms.

2. The Revelation Principle does **not apply** without *commitment.* (We will explore this in a dynamic setting.)

Conclusion: To check if $f(\cdot)$ is implementable in dominant strategies we only need to check that $f(\cdot)$ is truthfully implementable in dominant strategies. That is, checking that

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i) \text{ for all } \hat{\theta}_i \in \Theta_i, \text{ and } \theta_{-i} \in \Theta_{-i}$$

2.2.2 Universal Domain: the Gibbard-Satterthwaite Theorem

Question: What SCF's can we implement in DS?

Answer: Not much if we allow for any preferences over the alternatives, or, if we allow for *Universal Domain.*

We first move "backward" from the utility function $u_i(\cdot, \theta_i)$ to the underlying preferences, denoted by the *preference relation* $\succsim_i (\theta_i)$. That is, a type θ_i induces a preference relation $\succsim_i (\theta_i)$, which in turn can be represented by the utility function $u_i(\cdot, \theta_i)$. Let

$$\mathcal{R}_i = \{\succsim_i : \ \succsim_i = \ \succsim_i (\theta_i) \text{ for some } \theta_i \in \Theta_i\},$$

i.e. the set of all possible preference relations that are associated with the possible types in Θ_i. Each preference relation $R_i \in \mathcal{R}_i$ is associated with a particular type in Θ_i, and the following notation is commonly used:

$$\begin{aligned} \text{for type } \theta_i \quad &: \quad xR_iy \Leftrightarrow x \succsim_i (\theta_i)\ y \Leftrightarrow u_i(x, \theta_i) \geq u_i(y, \theta_i), \\ \text{or, for type } \theta_i' \quad &: \quad yR_i'x \Leftrightarrow y \succsim_i (\theta_i')\ x \Leftrightarrow u_i(y, \theta_i') \geq u_i(x, \theta_i'). \end{aligned}$$

The *Universal Domain* of preferences, denoted by $\mathcal{P}$, is the set of *all possible* strict rational preferences over X (this is the "Domain" on which $f(\cdot)$ operates). Denote by $f(\Theta)$ the image of $f(\cdot)$. We say that the SCF exhibits *full range* if $f(\Theta) = X$. (That is, the range X is fully covered by f when the domain is Θ.)

Definition 1.6: The SCF $f(\cdot)$ is *DICTATORIAL* if there is an agent i such that for all $\theta \in \Theta$

$$f(\theta) \in \{x \in X : u_i(x, \theta_i) \geq u_i(y, \theta_i) \;\; \forall\, y \in X\}$$

We call agent i a **dictator** if $f(\cdot)$ always chooses (one of) his most preferred alternative(s).

Proposition 1.2: (The Gibbard-Satterthwaite Theorem) *Suppose that X is finite, $|X| > 2, \mathcal{R}_i = \mathcal{P}\ \forall i \in I$ and $f(\Theta) = X$. Then, $f(\cdot)$ is truthfully implementable in Dominant Strategies if and only if it is Dictatorial.*

"proof": $\Leftarrow$ Trivial

$\Rightarrow$ Not trivial (Truthfully implementable $\Rightarrow$ Monotonic; Monotonic + [$f(\Theta) = X$] $\Rightarrow$ Paretian; Monotonic + Paretian $\Rightarrow$ Arrow's Impossibility Theorem)

Bottom Line*:* In the most general case the result is very pessimistic.

Consolation: It turns out that in some "reasonable" economic environments we can obtain more optimistic results.

2.2.3 Quasilinear Environment

The Quasilinear general environment is given by the following ingredients:

- $k \in K$ denotes the actual level of public choice parameter. (For now K is as abstract as we want it to be, and we will assume for now that K is *finite.*)
- $t_i \in \Re$ transfer to agent i.
- $x = (k, t_1, t_2, ..., t_I)$ is an *alternative*

- $u_i(x, \theta_i) = v_i(k, \theta_i) + t_i$ (Quasilinear *utility)*

Suppose we have *no outside sources for transfers.* Then this setup defines the set of alternatives as:

$$X = \{(k, t_1, ..., t_I) : k \in K,\ t_i \in \Re\ \forall\, i \in I,\ \sum_{i=1}^{I} t_i \leq 0\}.$$

We will also assume (for now) that there are *no wealth constraints. K* can have a "public" flavor or a "private" flavor. We have seen an example of a *Public Project* described in Example 1.1 above, where the project was either built or not, i.e., $K = \{0, 1\}$.

Example 1.3: Allocation of a Private good*:* Let y_i denote the probability that agent i gets an indivisible good.

$$K = \{(y_1, y_2, ..., y_I) : y_i \in \{0, 1\},\ \sum_{i=1}^{I} y_i = 1\}$$

and $v_i(k, \theta_i) = v_i(y_i, \theta_i) = \theta_i y_i$ where θ_i is i's valuation for the good. (Note that y_i is restricted to be either 1 or 0 to fit with our assumption that K is finite. The more general case would have $y_i \in [0, 1]$.) ■

A *SCF* in the quasilinear environment takes the form of:

$$f(\theta) = (k(\theta), t_1(\theta), ..., t_I(\theta))$$

where $k(\theta) \in K$ and $\sum_{i=1}^{I} t_i(\theta) \leq 0$. (From the definition of X.)

Let $k^*(\theta)$ be the *ex-post* efficient choice of k given state-of-the-world θ (or the "First-Best" choice of k). This implies that for all $\theta \in \Theta$ we must have,

$$k^*(\theta) \in \arg\max_{k \in K} \sum_{i=1}^{I} v_i(k, \theta_i)\,,$$

or,

$$\sum_{i=1}^{I} v_i(k^*(\theta), \theta_i) \geq \sum_{i=1}^{I} v_i(k, \theta_i) \text{ for all } k \in K\,.$$

- **Fact:** It is trivial to see that any Pareto optimal SCF must have $k(\theta) = k^*(\theta)$. (Quasilinearity implies that we can do interpersonal transfers through t_i without affecting total social surplus.)

- **Question:** Can we find some transfer functions $t_i(\cdot)$ such that the SCF $(k^*(\cdot), t_1(\cdot), ..., t_I(\cdot))$ can be implemented in Dominant Strategies?

- **Answer:** Use the Revelation Principle: each agent announces $\hat{\theta}_i \in \Theta_i$, which generates an announcement profile, $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_I)$, and each agent's payoff is given by $u_i(x, \theta) = v_i(k^*(\hat{\theta}_i, \hat{\theta}_{-i}), \theta_i) + t_i(\hat{\theta}_i, \hat{\theta}_{-i})$, where $\hat{\theta}_i$ is agent i's announcement, $\hat{\theta}_{-i}$ are the announcements of all the other agents, and θ_i is agent i's true type. The question that needs to be answered is, *When will agents announce truthfully?* The answer is, of course, **only** if we can get each agent to choose $\hat{\theta}_i$ so that his own utility is maximized when he tells the truth. In other words, we need to align *individual utilities* with *social welfare*!

Groves-Clarke Mechanisms

Groves (1973) suggested the above by setting the functions $t_i(\cdot)$ so that each agent *internalizes the externality* that he imposes on the others:

Proposition 1.3: $f(\theta) = (k^*(\theta), t_1(\theta), ..., t_I(\theta))$ is truthfully implementable in dominant strategies if for all $i = 1, ..., I$

$$t_i(\hat{\theta}) = \sum_{j \neq i} v_j(k^*(\hat{\theta}), \hat{\theta}_j) + h_i(\hat{\theta}_{-i}),$$

where $h_i(\hat{\theta}_{-i})$ is an arbitrary function of $\hat{\theta}_{-i}$.

proof: Every i solves:

$$\max_{\hat{\theta}_i \in \Theta_i} v_i(k^*(\hat{\theta}_i, \hat{\theta}_{-i}), \theta_i) + \sum_{j \neq i} v_j(k^*(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_j) + h_i(\hat{\theta}_{-i})$$

Note that $h_i(\hat{\theta}_{-i})$does not affect i's choice so i maximizes the sum above by setting $\hat{\theta}_i = \theta_i$ (this follows from the definition of $k^*(\cdot)$). *Q.E.D.*

Notes:

1. $v_j(k^*(\hat{\theta}_i, \theta_{-i}), \hat{\theta}_j)$, $j \neq i$, is the value that agent j gets from the project *were he* (agent j) *telling the truth.* That is, the first part of agent i's transfer is calculated using the sum of the other agents' valuations were they to be truth-tellers.

2. From the above it follows that agent i's payoff (through the transfer) depends on $\hat{\theta}_i$ *only through* its effect on the choice $k^*(\cdot)$, and this is *exactly equal to his externality* imposed on all the other agents (if they were telling the truth).

Notice that there are *many* $h_i(\hat{\theta}_i)$ that work (as long as it is not a function of $\hat{\theta}_i$). Clarke (1971) suggested a *particular* Groves mechanism known as a **Clarke mechanism** (or "Pivotal" mechanism). Define:

$$k^*_{-i}(\theta_{-i}) \in \arg\max_{k \in K} \sum_{j \neq i} v_j(k, \theta_j),$$

which is just the optimal choice of k for the $j \neq i$ agents when their types are θ_{-j}. Now let,

$$h_i(\hat{\theta}_{-i}) = -\sum_{j \neq i} v_j(k^*_{-i}(\hat{\theta}_{-i}), \hat{\theta}_j),$$

which implies that,

$$t_i(\hat{\theta}_i) = \sum_{j \neq i} v_j(k^*(\hat{\theta}_i, \hat{\theta}_{-i}), \hat{\theta}_j) - \sum_{j \neq i} v_j(k^*_{-i}(\hat{\theta}_{-i}), \hat{\theta}_j)$$

Interpretation:

- *Case 1:* when $k^*(\hat{\theta}_i, \hat{\theta}_{-i}) = k^*_{-i}(\hat{\theta}_{-i})$. In this case agent i's announcement does *not change* what *would have happened* if he did not exist and everyone was announcing the truth.

- *Case 2:* when $k^*(\hat{\theta}_i, \hat{\theta}_{-i}) \neq k^*_{-i}(\hat{\theta}_{-i})$. In this case agent i is "pivotal"; he *changes what would have happened without him*, and he is taxed for that *exactly* at the level of the externality that this change imposes on the other agents.

Example 1.4: Public Project*:* Simple application of the above: each agent pays a transfer only if he is pivotal. (**Note**: Formally, we must have $c(k)$ divided between the agents so that $v(\cdot,\cdot)$ reflects *net* value; if this is not the case, then we can have two agents who together value the project at more that its cost, none of which is pivotal, and there is no financing of the project. Another alternative is to have t_i include a share of the cost.)■

Example 1.5: Allocation of one unit of a Private good*:* As before, let

$$k = (y_1, ..., y_I),\ y_i \in \{0,1\}\ \ \forall\, i \in I,\ \sum_{i=1}^{I} y_i = 1.$$

The First best allocation is to give the good to the agent with the highest valuation, or simply choose k to maximize,

$$\max_k \sum_{i=1} \theta_i y_i$$

Let i^* be the agent with the highest valuation. Then $k^*(\cdot)$ sets $y_{i^*} = 1$ and $y_j = 0$ for all $j \neq i^*$. If we apply the **Clarke mechanism** to this setup:

$$t_i(\hat\theta_i, \hat\theta_{-i}) = \underbrace{\sum_{j\neq i} v_j(k^*(\hat\theta_i, \hat\theta_{-i}), \hat\theta_{-i})}_{A_i} - \underbrace{\sum_{j\neq i} v_j(k^*_{-i}(\hat\theta_{-i}), \hat\theta_j)\,,}_{B_i}$$

and since $\hat\theta_i = \theta_i$ is a DS for all i in a Groves-Clarke mechanism, then $A_i = B_i = \theta_{i^*}$ for all $i \neq i^*$, and for i^*, $A_{i^*} = 0$, and $B_{i^*} = \theta_{j^*}$, where j^* denotes the agent with the second-highest valuation. Thus, applying the Clarke Mechanism to the problem of allocating a private good yields the same outcome as a second price sealed bid auction, also known as a **Vickrey Auction**.■

Remarks:

1. In much of the literature such mechanisms are known as **Groves-Clarke** mechanisms, and sometimes are referred to as **Groves-Clarke-Vickrey** mechanisms.

2. When the set of $v(\cdot,\cdot)$ functions is "rich" (something like a "universal domain" of the quasilinear set-up) only the Groves mechanisms can be implemented in DS. (This result was proven by J. Green and J.J. Laffont, and is restated as Proposition 23.C.5 in MWG).

Groves-Clarke & Balanced Budgets

The function $k^*(\cdot)$, as we defined it, is a *necessary condition* for ex-post optimality, but it is not *sufficient:* we must also guarantee that there is a *Balanced Budget* (no waste of resources, or, no money is being "burnt")*:*

$$\sum_{i=1}^{I} t_i(\theta) = 0 \text{ for all } \theta \in \Theta \qquad \text{(BB)}$$

We can now state a *second negative result:* When the set of $v(\cdot,\cdot)$ functions is "rich" then there is no SCF $f(\cdot) = (k^*(\cdot), t_1(\cdot), ..., t_I(\cdot))$ (where $k^*(\cdot)$ is *ex post* optimal) that is truthfully implementable in DS, *and* that satisfies (BB). A trivial (and not very interesting) case for which we can achieve full *ex post* optimality is given in the following example:

Example 1.6: If there exists some agent i such that $\Theta_i = \{\theta_i\}$ a singleton, then we have no incentive problem for agent i, and we can set

$$t_i(\theta) = -\sum_{j \neq i} t_j(\theta) \text{ for all } \theta \in \Theta\,.$$

This trivially guarantees a balanced budget. Note, however, that we are **forcing** the agents to participate in this decision. If, for example, the agent who "balances the budget" (agent i in this example) gets negative utility, we may consider his ability to "walk out" on the project. This will be thoroughly discussed in the next chapter of Adverse Selection.■

Do example with nonbalanced budget?

2.3 Bayesian Implementation

Recall that $\Gamma, \Theta, \phi(\cdot)$ and $\{u_i(\cdot)\}_{i=1}^{I}$ define a Bayesian Game of Incomplete information.

Definition 1.7: The strategy profile $s^*(\cdot) = (s_1^*(\cdot), ..., s_I^*(\cdot))$ is a *Bayesian Nash Equilibrium* (BNE) if $\forall\, i \in I$ and $\forall\, \theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i]$$
$$\geq E_{\theta-i}[u_i(g, (s_i', s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \;\; \forall\, s_i' \in S_i$$

That is, if player i believes that other players are playing according to $s_{-i}^*(\theta_{-i})$ then he is best off by following the behavior prescribed by $s_i^*(\theta_i)$. Applying this as an equilibrium to mechanism design we have,

Definition 1.8: The mechanism $\Gamma = (S_1, ..., S_I, g(\cdot))$ implements the SCF $f(\cdot)$ in BNE if there is a BNE of Γ, $s^*(\cdot) = (s_1^*(\cdot), ..., s_I^*(\cdot))$ s.t. $g(s^*(\theta)) = f(\theta)$ for all $\theta \in \Theta$.

Definition 1.9: The SCF $f(\cdot)$ is *Truthfully Implementable in BNE (*or, is *Bayesian Incentive Compatible)* if for all $\theta_i \in \Theta_i$ and $i \in I$, $s_i^*(\theta_i) = \theta_i$ is a BNE of the direct revelation mechanism $\Gamma = (\Theta_1, ..., \Theta_i, f(\cdot))$, i.e., for all i :

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\theta_i', \theta_{-i}), \theta_i)|\theta_i] \;\; \forall\, \theta_i' \in \Theta_i\,.$$

Note: Bayesian Nash is a weaker form of equilibria than Dominant Strategy equilibria. Thus we should expect to be able to implement a larger set of SCF's.

Proposition 1.4: (The Revelation Principle for BNE*) Suppose that there exists a mechanism $\Gamma = (S_1, ..., S_I, g(\cdot))$ that implements the SCF $f(\cdot)$ in BNE. Then $f(\cdot)$ is truthfully implementable in BNE.*

proof: Same idea as before: $\exists\, s^*(\cdot) = (s_1^*(\cdot), ..., s_I^*(\cdot))$ such that $\forall\, i$ and $\forall \theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(s_i', s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \;\; \forall\, s_i' \in S_i\,,$$

in particular, $\forall\, i$ and $\forall\, \theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i'), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \;\forall\, \theta_i' \in \Theta_i\,.$$

But since Γ implements $f(\cdot)$ then $g(s^*(\theta)) = f(\theta) \;\; \forall\, \theta \in \Theta \Rightarrow \forall\, i$ and $\forall\, \theta_i \in \Theta_i$

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\theta_i', \theta_{-i}), \theta_i)|\theta_i] \;\; \forall\, \theta_i' \in \Theta_i\,.$$

Q.E.D.

Intuition: Same story with the computers...

Notes:

1. Must have commitment on behalf of the planner as before.
2. Each agent maximizes expected utility taking the strategies of the others as given. That is, each agent optimizes *on average* and not *for any* s_{-i}. This makes it easier to implement social choice functions as we will see.
3. But, this requires more sophistication from the agents. (Beliefs must be correct and $\phi(\cdot)$ must be known).

2.3.1 Expected Externality Mechanisms

Consider now quasilinear environments. Suppose we want to implement the *ex post efficient* SCF $f(\cdot) = (k^*(\cdot), t_1(\cdot), ..., t_I(\cdot))$, where $\forall\, \theta \in \Theta$,

$$\sum_{i=1}^{I} v_i(k^*(\theta), \theta_i) \geq \sum_{i=1}^{I} v_i(k, \theta_i) \;\; \forall\, k \in K$$

and there is a balanced budget,

$$\sum_{i=1}^{I} t_i(\theta) = 0 \ .$$

We saw that generally we cannot implement ex-post efficient SCF's in Dominant Strategies in quasilinear environments - could not balance the budget. This will be possible however when we relax the equilibrium requirement by choosing BNE as the equilibrium concept.

The mechanism was developed independently by D'Aspermont and Gerard-Varet (1979) and Arrow (1979) and are also called AGV mechanisms in the literature. It turns out that we can get ex-post efficient implementation if we assume the following:

Assumption 1.1: Types are distributed independently: $\phi(\theta) = \phi_1(\theta_1) \times \phi_2(\theta_2) \times \cdots \times \phi_I(\theta_I)$, $\forall\, \theta \in \Theta$.

Now, extend Groves transfers as follows:

$$t_i(\widehat{\theta}) = E_{\theta_{-i}}[\sum_{j\neq i} v_j(k^*(\widehat{\theta}_i, \theta_{-i}), \theta_j)] + h_i(\widehat{\theta}_{-i})$$

Note that unlike in Groves-Clarke mechanisms:

1. The first term only depends on $\widehat{\theta}_i$, i.e., on the announcement of agent i.
2. The first term sums the *expected benefits* of agents $j \neq i$ assuming that they tell the truth and given that i announced $\widehat{\theta}_i$: it is not a function of the actual announcements of agents $j \neq i$. This implies that $t_i(\cdot)$ is less "variable" (j's announcements do not affect it), but on average it causes i's incentives to be well aligned with the social benefits.

To see that incentive compatibility is satisfied given that agents $j \neq i$ announce truthfully, we observe that agent i maximizes:

$$\begin{aligned}\underset{\widehat{\theta}_i\in\Theta_i}{\text{Max}} \quad & E_{\theta_{-i}}[v_i(k(\hat{\theta}_i, \theta_{-i}), \theta_i) + t_i(\hat{\theta}_i, \theta_{-i})] \\ &= E_{\theta_{-i}}[\sum_{j=1}^{I} v_j(k(\hat{\theta}_i, \theta_{-i}), \theta_j)] + E_{\theta_{-i}}[h_i(\theta_{-i})]\end{aligned}$$

We know that for each $\theta \in \Theta$ this is maximized when $\hat{\theta}_i = \theta_i$ by the definition of $k(\cdot)$, which implies that $\hat{\theta}_i = \theta_i$ maximizes this expectations (since it is an expectations expression). Thus, incentive compatibility is satisfied.

Note: It is *not a dominant strategy* to announce the truth: If the other agents lie then the realized distribution of $\hat{\theta}_{-i}$ (the announcements) differs from the prior distribution of types dictated by $\phi(\cdot)$. This implies that we cannot write the problem as written above if we consider $\widehat{\theta}_{-i} \neq \theta_{-i}$!

We see that we can implement $k^*(\cdot)$ in BNE, but to "improve" upon DS implementation we now need to show that we can achieve a balanced budget when BNE implementation is considered. It will be convenient to write the transfer of each agent as the sum of two expressions. Denote:

$$\xi_i(\theta_i) = E_{\theta_{-i}}[\sum_{j\neq i} v_j(k^*(\theta_i, \theta_{-i}), \theta_j)]\,,$$

so that each agent i that announces truthfully (which we showed is an equilibrium) will receive a transfer of $t_i(\theta) = \xi_i(\theta_i) + h_i(\theta_{-i})$. Using this decomposition we will show that we can use the $h(\cdot)$ function to "finance" the $\xi(\cdot)$ functions in the following way: each agent $j \neq i$ will pay (through his $h(\cdot)$ function),

$$\frac{1}{I-1}\xi_i(\theta_i)\,,$$

so each agent i will pay:

$$h_i(\theta_{-i}) = -\sum_{j\neq i}\frac{1}{I-1}\xi_j(\theta_j) = -\frac{1}{I-1}\sum_{j\neq i}\xi_j(\theta_j)\,,$$

and summing this up over all agents yields,

$$\sum_{i=1}^{I} h_i(\theta_{-i}) = -\frac{1}{I-1}\sum_{i=1}^{I}\sum_{j\neq i}\xi_j(\theta_j) = -\frac{I-1}{I-1}\sum_{i=1}^{I}\xi_i(\theta_i) = -\sum_{i=1}^{I}\xi_i(\theta_i)\,,$$

which clearly implies that,

$$\sum_{i=1}^{I} t_i(\theta) = \sum_{i=1}^{I}\xi_i(\theta_i) + \sum_{i=1}^{I} h_i(\theta_{-i}) = 0.$$

Notes:

1. We are still ignoring participation constraints (i.e., participation is not voluntary in this model).

2. Correlated types can help to implement social welfare functions. (A paper by Cremer and McLean (Econometrica, 1985) shows that when the realization of types is correlated across agents then the principal/designer can take advantage of this correlation and implement a riches set of social choice functions.)

3. We cannot do the "trick" of financing as above in Dominant Strategy implementation because $\xi_j(\widehat{\theta}_j)$ in the Groves mechanism is $\xi_j(\widehat{\theta}_j, \widehat{\theta}_{-j})$, that is, it is a function of what the other agents announce because we are not using expectations over the truthful announcements of the other agents. This in turn implies that $h_i(\widehat{\theta})$ would be a function of $\widehat{\theta}_i$ through the ξ_j's of the other agents.

2.4 Bayesian and Dominant Strategy Implementation: An Equivalence Result

Moving from dominant strategy implementation to Bayesian implementation was crucial to get implementation with a budget balanced mechanism. Before continuing with the imposition of other desirable properties on mechanisms, it is useful to recognize an important relationship between dominant strategy implementation and Bayesian implementation.

To do this, consider the quasilinear environment we have discussed above, with agents $i \in I$, type sets $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i]$ with distribution $F_i(\cdot)$, and public choices $k \in K$. Given an allocation of a public choice and transfers, $x = (k, t_1, ..., t_I)$, agents' utilities are given by $u_i(x, \theta_i) = v_i(k, \theta_i) + t_i$. It will be useful to distinguish between the *mechanism*, $\Gamma = (k(\cdot), t_1(\cdot), ..., t_I(\cdot))$, and the *fundamentals of the economy*, which include the types and their distributions, the preferences (utility functions) and the set of public choices, K.

2.4.1 The Equivalence Theorem

Recall that for the principal-agent model we introduced in chapter 1 there was a very appealing structure for incentive compatible contracts that was based on the envelope theorem. Namely, equation (1.1) demonstrated that in equilibrium, the utility of every type is the utility of the lowest type, plus an integral term that was a direct consequence of the envelope theorem. It turns out that such a relationship will reappear for implementation with Bayesian incentive compatibility, which will lead to a powerful result.

First, it is useful to define some notation. Let $\Gamma = (k(\cdot), t_1(\cdot), ..., t_I(\cdot))$ be a direct revelation mechanism, and consider an agent i who believes that all the other agents are reporting truthfully. If i's type is θ_i and he chooses to report that his type is θ_i', then define his *Expected Interim Valuation* as,

$$\overline{v}_i(\theta_i', \theta_i) = E_{\theta_{-i}}[v_i(k(\theta_i', \theta_{-i}), \theta_i)] ,$$

his *Expected Interim Transfer* as,

$$\overline{t}_i(\theta_i') = E_{\theta_{-i}}[t_i(\theta_i', \theta_{-i})] ,$$

his *Expected Interim Utility* as,

$$\Phi_i(\theta_i', \theta_i) = \overline{v}_i(\theta_i', \theta_i) + \overline{t}_i(\theta_i') .$$

Finally, define his *Equilibrium Expected Interim Utility* as his interim expected utility from truth telling, and like in the principal agent model denote this as,

$$U_i(\theta_i) \equiv \Phi_i(\theta_i, \theta_i) = \overline{v}_i(\theta_i, \theta_i) + \overline{t}_i(\theta_i)\,.$$

Now notice that the mechanism is Bayesian incentive compatible if and only if,

$$\Phi_i(\theta_i, \theta_i) \geq \Phi_i(\theta_i', \theta_i) \ \text{ for all } \theta_i', \theta_i \in \Theta_i\,, \tag{2.3}$$

and if we assume appropriate differentiability assumptions on the mechanism's transfer functions and on the utility functions, then (2.3) is equivalent to the following first order condition for the agent when he is truth telling,

$$\frac{\partial \overline{v}_i(\theta_i, \theta_i)}{\partial \theta_i'} + \overline{t}_i'(\theta_i) = 0 \ a.e. \tag{2.4}$$

As it turns out, Bayesian incentive compatibility imposes very strong restrictions on the mechanism that will result in differentiability of the transfer functions, and in turn the representation of (2.4), with weaker assumptions as the following theorem claims:

Theorem 2 *If the direct revelation mechanism* $\Gamma = (k(\cdot), t_1(\cdot), ..., t_I(\cdot))$ *is Bayesian Incentive Compatible (BIC), and if* $\overline{v}_i(\theta_i', \theta_i)$ *is continuously differentiable in* (θ_i', θ_i) *at all points* $\theta_i' = \theta_i$*, then for any* $\theta_i' > \theta_i$*,*

$$U_i(\theta_i) = \int_{\theta_i}^{\theta_i'} \frac{\partial \overline{v}_i(r, r)}{\partial \theta_i} dr\ .$$

Proof. Let $\theta_i' = \theta_i + \varepsilon$, where $\varepsilon > 0$. BIC implies that,

$$\begin{aligned} U_i(\theta_i') &\geq \overline{v}_i(\theta_i, \theta_i') + \overline{t}_i(\theta_i), \\ &= U_i(\theta_i) + \overline{v}_i(\theta_i, \theta_i') - \overline{v}_i(\theta_i, \theta_i)\,, \end{aligned}$$

where the equality follows from adding and subtracting $\overline{v}_i(\theta_i, \theta_i)$. This can be rewritten as,

$$U_i(\theta_i') - U_i(\theta_i) \geq \overline{v}_i(\theta_i, \theta_i') - \overline{v}_i(\theta_i, \theta_i)\,. \tag{2.5}$$

Similarly, BIC implies that,

$$\begin{aligned} U_i(\theta_i) &\geq \overline{v}_i(\theta_i', \theta_i) + \overline{t}_i(\theta_i'), \\ &= U_i(\theta_i') + \overline{v}_i(\theta_i', \theta_i) - \overline{v}_i(\theta_i', \theta_i')\,, \end{aligned}$$

where the equality follows from adding and subtracting $\overline{v}_i(\theta_i', \theta_i')$. This can be rewritten as,

$$U_i(\theta_i') - U_i(\theta_i) \leq \overline{v}_i(\theta_i', \theta_i') - \overline{v}_i(\theta_i', \theta_i) \,. \tag{2.6}$$

After adding and subtracting $\overline{v}_i(\theta_i, \theta_i)$ from the right side of (2.6), then() and () together imply,

$$\frac{\overline{v}_i(\theta_i + \varepsilon) - \overline{v}_i(\theta_i)}{\varepsilon} - \frac{\overline{v}_i(\theta_i + \varepsilon, \theta_i) - \overline{v}_i(\theta_i, \theta_i)}{\varepsilon} \geq \frac{U_i(\theta_i') - U_i(\theta_i)}{\varepsilon} \geq \frac{\overline{v}_i(\theta_i, \theta_i + \varepsilon) - \overline{v}_i(\theta_i, \theta_i)}{\varepsilon} \,,$$

where $\overline{v}_i(\theta_i) \equiv \overline{v}_i(\theta_i, \theta_i)$. Taking limits on $\varepsilon \to 0$, we get by the definition of the total and partial derivatives of $\overline{v}_i(\theta_i, \theta_i)$,

$$\frac{d\overline{v}_i(\theta_i)}{d\theta_i} - \frac{\partial \overline{v}_i(\theta_i, \theta_i)}{\partial \theta_i'} \geq U_i'(\theta_i) \geq \frac{\partial \overline{v}_i(\theta_i, \theta_i)}{\partial \theta_i} \,.$$

but since $\frac{d\overline{v}_i(\theta_i)}{d\theta_i} = \frac{\partial \overline{v}_i(\theta_i,\theta_i)}{\partial \theta_i'} + \frac{\partial \overline{v}_i(\theta_i,\theta_i)}{\partial \theta_i}$, we have $U_i'(\theta_i) = \frac{\partial \overline{v}_i(\theta_i,\theta_i)}{\partial \theta_i}$. ■

As with the single agent case, this theorem is a consequence of the envelope theorem. That is, since an agent i of type θ_i chooses θ_i' to maximize,

$$\max_{\theta_i' \in \Theta_i} \Phi_i(\theta_i', \theta_i) = \overline{v}_i(\theta_i', \theta_i) + \overline{t}_i(\theta_i') \,,$$

then if $\overline{v}_i(\theta_i', \theta_i)$ and $\overline{t}_i(\theta_i')$ were differentiable, we would get (2.4), and thus the envelope theorem implies that

$$U_i'(\theta_i) = \frac{\partial \overline{v}_i(\theta_i, \theta_i)}{\partial \theta_i}.$$

However, the proof above shows that this result holds without assuming differentiability of the transfer function. This means that BIC itself implies that the transfer functions must be differentiable, and they must satisfy,

$$\begin{aligned} \overline{t}_i(\theta_i) &= U_i(\theta_i) - \overline{v}_i(\theta_i) \\ &= U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \frac{\partial \overline{v}_i(r, r)}{\partial \theta_i} dr - \overline{v}_i(\theta_i) \,. \end{aligned}$$

Before stating the equivalence result, we need to define what it means for two mechanisms to be equivalent:

Definition 4 *Let $\Gamma = (k(\cdot), t_1(\cdot), ..., t_I(\cdot))$ and $\Gamma' = (k'(\cdot), t'_1(\cdot), ..., t'_I(\cdot))$ be two mechanisms. We say that Γ and Γ' are* ***equivalent*** *if they implement the same public choice, $k(\theta) = k'(\theta)$ for all $\theta \in \Theta$, and give the same interim expected utility to every agent, $U_i(\theta_i)$ is the same in both mechanisms for all $i \in I$, and all $\theta_i \in \Theta_i$.*

The following then, is a corollary of the envelope theorem proved above:

Corollary 1 *Let $\Gamma = (k(\cdot), t_1(\cdot), ..., t_I(\cdot))$ and $\Gamma' = (k'(\cdot), t'_1(\cdot), ..., t'_I(\cdot))$ be two BIC mechanisms with $k(\theta) = k'(\theta)$. If both mechanisms give the same interim expected utility $U_i(\underline{\theta}_i)$ for the lowest type of each agent i, then they are equivalent.*

It is easy to see from this corollary that interim expected utilities are uniquely determined up to a constant by the implications of BIC.

2.4.2 BIC and Groves Equivalence

Recall that any groves mechanism has two features. First, for any profile of types θ, it implements the efficient public choice, and second, for announcements $(\theta'_i, \theta'_{-i})$ it includes transfers only of the form,

$$t_i(\theta'_i, \theta'_{-i}) = \sum_{j \neq i} v_j(k(\theta'_i, \theta'_{-i}), \theta'_j) + h_i(\theta'_{-i}) \, .$$

Define the **basic groves mechanism** as the groves mechanism with $h_i(\theta'_{-i}) = 0$ for all i. It is easy to see that at the interim stage, where types are privately known, the interim expected utility of any groves mechanism can be replaced with a basic groves mechanism plus some constant, yielding transfers,

$$t_i(\theta'_i, \theta'_{-i}) = \sum_{j \neq i} v_j(k(\theta'_i, \theta'_{-i}), \theta'_j) + m_i \, ,$$

where $m_i = E_{\theta_{-i}}[h_i(\theta_{-i})]$ is the expectation of what i gets from $h_i(\cdot)$ given the truthful announcements of the other agents.

Now notice that any Groves mechanism is trivially a BIC mechanism since dominant strategy incentive compatibility immediately implies BIC. Since any Groves mechanism is efficient, we immediately have,

Proposition 3 *If a mechanism Γ is BIC and efficient, then it is equivalent to some Groves mechanism.*

This is a very powerful result, and as the discussion above indicates, it is a direct consequence of the envelope theorem we have proved. It says that if an efficient public choice can be implemented by some BIC mechanism, then there is an efficient Groves mechanism that implements the same interim expected utilities for all agent and all types.

this implies, for example, that if we take an expected externality (AGV) mechanism that is efficient and ex post balances the budget, then the agents would be equally happy with some Groves mechanism since there is one that implements the same public choice outcomes, and the same interim (and therefore ex ante) expected utilities. The twist is that the groves mechanisms may not, and generally will not be balanced budget.

- Describe the history: Mookherjee and Reichelstein (1992), Williams (1999), Krishna and Perry (2000) - and how the result extends to multidimensional types.

- ????derive Laffont-Maskin result on Groves-Clarke mechanisms (MWG pp. 881-882)

2.4.3 An Application: Auctions

(to be added)

1. Revenue Equivalence

Revenue-maximizing auctions

2. Efficient and revenue-maximizing auctions can be equivalently done in dominant strategies - as 2nd-price (Vickrey) auctions

2.5 Participation Constraints

Up until now we *forced* the agents to play, and got truthful implementation. In other words, the "center's" lack of knowledge had *no social cost* (in BNE with quasilinear utilities).

Question: What happens if agents can "walk away" to some individual status quo?

This will cause further restrictions, so we should anticipate further "implementation" problems. To consider this problem of adding a participation constraint, we first must define *when* agents can withdraw from the mechanism, and *what* they can get. Let $\overline{u}_i(\theta_i)$ be the utility of agent i when he withdraws from the mechanism.

Definition 5 *The direct revelation mechanism* $\Gamma = (\Theta_i, ..., \Theta_I, f(\cdot))$ *is:*

1. ***ex-post Individually Rational*** *if it satisfies the ex-post participation constraint: for all* i,

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq \overline{u}_i(\theta_i) \text{ for all } (\theta_i, \theta_{-i}) \in \Theta.$$

2. ***Interim Individually Rational*** *if it satisfies the interim participation constraint: for all i,*

$$U_i(\theta_i|f) \equiv E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq \overline{u}_i(\theta_i) \text{ for all } \theta_i \in \Theta_i\,.$$

3. ***ex-ante Individually Rational*** *if it satisfies the ex-ante participation constraint: for all* i,

$$U_i(f) \equiv E_{\theta_i}[U_i(\theta_i|f)] \geq E_{\theta_i}[\overline{u}_i(\theta_i)]\,.$$

First, with the *ex-post* participation constraint the agent can withdraw *after* the announcement of the outcome by the central planner. This is the hardest case to satisfy. Second, with the *Interim* participation constraint the agent can withdraw after learning θ_i, but before participating in the mechanism. That is, once the agent decides to participate then the outcome can be imposed on him. This case is easier to satisfy. Finally, with the *ex-ante* participation constraint the agent can commit to "playing" even before his type is realized, and he can be bound to his commitment. This is the easiest case to satisfy.

Will focus on INTERIM - explain.

As it turns out, when agents can voluntarily choose not to participate in the mechanism, then the consequence of asymmetric information will be loss in efficiency. Thus, voluntary participation can be described as putting additional constraints on the mechanism designer, and these constraints will have an influence on the mechanism's desirable properties.

2.5.1 Inefficient Trade: The Myerson-Satterthwaite Theorem

One of the most important results that follow from the imposition of participation constraints is known as the Myerson-Satterthwaite theorem.

Consider a simple bilateral trade setting in which a seller owns a good that he values at $\theta_s \in [\underline{\theta}_s, \overline{\theta}_s]$ and only he knows his value. There is a potential buyer who values the good at $\theta_b \in [\underline{\theta}_b, \overline{\theta}_b]$ and similarly, only she knows her value.

Efficient trade would entail that the buyer gets the good whenever $\theta_b > \theta_s$, but private information will make this non-trivial, and thus we assume that first, $\underline{\theta}_s < \overline{\theta}_b$, implying that trade is sometimes desirable, and that $[\underline{\theta}_b, \overline{\theta}_b] \cap [\underline{\theta}_s, \overline{\theta}_s] \neq \emptyset$, implying that trade is not always desirable. Thus, for a mechanism to be efficient, it will have to elicit truthful revelation of valuations. We will simplify here and assume that $[\underline{\theta}_b, \overline{\theta}_b] = [\underline{\theta}_s, \overline{\theta}_s]$, though the results hold in much more general settings (See Williams (1999) for the generalized version of the proof below, which is based on Williams's paper.)

An outcome of a direct revelation mechanism will then be $x = (k(\theta), t_s(\theta), t_b(\theta))$ where $k = 1$ is "trade", $k = 0$ is "no trade", and t_i is the transfer received by i. This mechanism will result in utilities for the seller and buyer as follows,

$$\begin{aligned} u_s(x, \theta_s) &= -k\theta_s + t_s\,, \\ u_b(x, \theta_b) &= k\theta_b + t_b\,. \end{aligned}$$

A mechanism will be ex post efficient if it satisfies the following two conditions,

(FB) The trade decision is first-best efficient:

$$k(\theta) = 1 \Longleftrightarrow \theta_b \geq \theta_s$$

(BB) The transfers are ex post budget balanced:

$$t_s(\theta) + t_b(\theta) = 0 \text{ for all } \theta \in [\underline{\theta}_b, \overline{\theta}_b] \times [\underline{\theta}_s, \overline{\theta}_s]$$

Before we proceed to describe the main result, it is useful to revisit some of the definitions we have provided above.

First, we introduce a weaker form of a budget balanced mechanism. We say that a mechanism is *ex ante budget balanced* (EABB) if,

(EABB) $E_\theta[t_s(\theta)+t_b(\theta)]\leq 0$,

which implies that the transfers can be financed by the two traders, and surpluses that leave the relationship are allowed. Note that (BB) implies (EABB), a relationship that we will use below.

Second, we revisit some notation from the previous section. Letting $U_i(\theta_i)$ denote the interim expected utility of type θ_i, we know that,

$$U_i(\theta_i)=\overline{v}_i(\theta_i)+\bar{t}_i(\theta_i)\,.$$

In this set-up, $v_s(k,\theta_s)=-k\theta_s$, and $v_b(k,\theta_b)=k\theta_b$. We can now rewrite (EEAB) as,

(EABB) $E_\theta\left[[U_b(\theta_b)+U_s(\theta_s)]-[v_b(k(\theta_b,\theta_s),\theta_b)+v_s(k(\theta_b,\theta_s),\theta_s)]\right]\leq 0$,

Finally, since we have normalize the "no trade - no transfers" decision to yield utilities of zero for each agent, and since we will focus on interim participation, we know that a mechanism will be interim individually rational if and only if it satisfies,

(IIR) $U_i(\theta_i)\geq 0$ for all $\theta_i\in\Theta_i$, $i\in\{b,s\}$.

Theorem 3 *(Myerson-Satterthwaite) In the bilateral trade setting there is no BIC mechanism that satisfies (FB), (BB), and (IIR). That is, there is no ex post efficient mechanism that satisfies interim participation.*

Proof. Assume in negation that such a mechanism exists, which implies that some Groves mechanism that satisfies (FB) exists that is equivalent to it, and we call it the equivalent Groves mechanism. Consider first the "basic Groves mechanism" we have defined in the previous section, which in this set-up is $t_i(\theta)=v_j(k(\theta),\theta_j)$. The basic groves mechanism is then:

$$(k(\theta),t_s^{BG}(\theta),t_b^{BG}(\theta))=\begin{array}{cl}(1,\theta_b,-\theta_s) & \text{if }\theta_b\geq\theta_s\\ (0,0,0) & \text{if }\theta_b<\theta_s\end{array}.$$

The equivalent Groves mechanism can be given by

$$(k(\theta),t_s(\theta),t_b(\theta))=(k(\theta),t_s^{BG}(\theta)-m_s,t_b^{BG}(\theta)-m_b),$$

where m_s and m_b are the utility shifting constants. The *ex ante* expected utilities of the buyer and the seller from participating in this mechanism are,

$$\begin{aligned} E_{\theta_b}\left[U_b(\theta_b)\right] &= E_\theta\left[v_b(k(\theta_b,\theta_s),\theta_b) + v_s(k(\theta_b,\theta_s),\theta_s) - m_b\right], \\ &= GT - m_b\,, \end{aligned}$$

and,

$$\begin{aligned} E_{\theta_s}\left[U_s(\theta_s)\right] &= E_\theta\left[v_s(k(\theta_b,\theta_s),\theta_s) + v_b(k(\theta_b,\theta_s),\theta_b) - m_s\right], \\ &= GT - m_s\,, \end{aligned}$$

where GT denotes the expected gains from trade since these are realized only upon trade. (BB) implies (EABB), which in turn implies that

$$E_\theta\left[U_b(\theta_b) + U_s(\theta_s) - (v_b(k(\theta_b,\theta_s),\theta_b) + v_s(k(\theta_b,\theta_s),\theta_s))\right] \le 0,$$

or,

$$\begin{gathered} GT - m_b + GT - m_s - GT \le 0 \\ \Longleftrightarrow GT \le m_b + m_s\,, \end{gathered}$$

and since by assumption $GT > 0$ it must be that $m_b + m_s > 0$. For (IIR) to hold it must be that both types $\overline{\theta}_s$ and $\underline{\theta}_b$ are willing to participate. But for these two types we have

$$v_b(k(\underline{\theta}_b,\theta_s),\underline{\theta}_b) + v_s(k(\underline{\theta}_b,\theta_s),\theta_s) = 0 \ \forall\theta_s$$

and

$$v_s(k(\theta_b,\overline{\theta}_s),\overline{\theta}_s) + v_b(k(\theta_b,\overline{\theta}_s),\theta_b) = 0 \ \forall\theta_b\,,$$

which implies that for $U_b(\underline{\theta}_b) \ge 0$ we must have $m_b \le 0$, and for $U_s(\overline{\theta}_s) \ge 0$ we must have $m_s \le 0$, contradicting that $m_b + m_s > 0$. ■

The interpretation of this theorem can be viewed through the celebrated Coase theorem. Here, we have a violation of perfect information, and if we interpret a mechanism as the most general form of unrestricted bargaining, we see how the Coase theorem fails in the face of asymmetric information.

2.6 Mechanism Design: Summary and Comments

I. **Revelation Principle:** makes life easy.

II. **Dominant Strategies:**

- Universal Domain ⇒ impossibility result (Gibbard-Satterthwaite Theorem)
- Quasilinear preferences ⇒ get ex-post optimal choice implemented in dominant strategies. BUT may need to burn money! (VCG mechanisms)

III. **BNE implementation**:

- AGV-Arrow ⇒ get ex-post optimality including Balanced Budget in BNE.
- Participation Constraints ⇒ adding interim IR causes impossibility of ex-post efficient BIC mechanisms (Myerson-Satterthwaite Theorem).

Notes:

1. We only considered *deterministic mechanisms.* See discussion at the end of Fudenberg-Tirole (ch. 7) for some points on randomized mechanisms (doesn't matter for the linear model because of risk neutrality).
2. *Risk aversion:* Makes things much more complicated, same flavor remains, and for some cases the same type of analysis goes through with "harder" math. (Maskin-Riley, Matthews - see Fudenberg-Tirole)

2.7 Appendix: the Linear Model

Here consider quasilinear model with many agents and social choice functions of the form $f(\theta) = (k(\theta), t_1(\theta), ..., t_I(\theta))$, but now we take $v_i(\cdot)$ to be of a particular form:

$$v_i(k, \theta_i) = \theta_i v_i(k)\,,$$

which in turn implies that,

$$u_i(x, \theta_i) = \theta_i v_i(k) + t_i\,.$$

We assume that the type space is compact, $\theta_i \in \Theta_i = [\underline{\theta}_i, \overline{\theta}_i] \subset \Re$, $\underline{\theta}_i < \overline{\theta}_i\,.$ We also assume that types are distributed independently with full support, that is, $\theta_i \sim \phi_i(\cdot)$ and $\phi_i(\theta_i) > 0$ for all $\theta_i \in \Theta_i\,.$ We will look for necessary and sufficient conditions for $f(\cdot)$ to be Bayesian Incentive Compatible (BIC). We first introduce some helpful notation:

- $\bar{t}_i(\hat{\theta}_i) = E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]$; This is i's expected transfer (including the $h(\cdot)$ component).
- $\overline{v}_i(\hat{\theta}_i) = E_{\theta_{-i}}[v_i(k(\hat{\theta}_i, \theta_{-i}))]$; This is i's expected "benefit." So, when i is type θ_i and he announces $\hat{\theta}_i$, his expected utility is:

$$\theta_i \overline{v}_i(\hat{\theta}_i) + \bar{t}_i(\hat{\theta}_i).$$

- Define $U_i(\theta_i)$ as i's expected utility when he is type θ_i and he tells the truth,

$$U_i(\theta_i) = \theta_i \overline{v}_i(\theta_i) + \bar{t}_i(\theta_i)\,.$$

Proposition 4 *The SCF $f(\cdot)$ is Bayesian Incentive Compatible if and only if for all $i \in I$,*

(I) $\overline{v}_i(\hat{\theta}_i)$ *is non-decreasing in* $\hat{\theta}_i$ *, and,*

(II) $U_i(\theta_i) = U_i(\underline{\theta}_i) + \int\limits_{\underline{\theta}_i}^{\theta_i} \overline{v}_i(s)ds$ *for all* $\theta_i \in \Theta_i\,.$

proof: $\underline{\Rightarrow :}$ Let $\theta_i' > \theta_i$. If $f(\cdot)$ is BIC then a type θ_i' should not want to lie, in particular should not choose θ_i :

$$\begin{aligned} U_i(\theta_i') &\geq \theta_i'\overline{v}_i(\theta_i) + \bar{t}_i(\theta_i) \\ &= U_i(\theta_i) + \overline{v}_i(\theta_i)(\theta_i' - \theta_i) \end{aligned}$$

$$\Rightarrow \overline{v}_i(\theta_i) \leq \frac{U_i(\theta_i') - U_i(\theta_i)}{\theta_i' - \theta_i} \tag{2.7}$$

Also, type θ_i should not want to announce $\theta_i^{'}$:

$$\begin{aligned} U_i(\theta_i) &\geq \theta_i\overline{v}_i(\theta_i') + t_i(\theta_i') \\ &= U_i(\theta_i') - \overline{v}_i(\theta_i')(\theta_i' - \theta_i) \end{aligned}$$

$$\Rightarrow \overline{v}_i(\theta_i') \geq \frac{U_i(\theta_i') - U_i(\theta_i)}{\theta_i' - \theta_i} \tag{2.8}$$

Both (2.7) and (**??**) together imply that $\overline{v}_i(\theta_i') \geq \overline{v}_i(\theta_i)$, which is condition (I) of the proposition. This formulation also allows us to write:

$$U_i'(\theta_i) = \lim_{\theta_i' \to \theta_i} \frac{U_i(\theta_i') - U_i(\theta_i)}{\theta_i' - \theta_i} = \overline{v}_i(\theta_i) \tag{(ET)}$$

$$\Rightarrow U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \overline{v}_i(s)ds \text{ for all } \theta_i \in \Theta_i$$

$\underline{\Leftarrow:}$ Consider *any* pair $\theta_i' \neq \theta_i$. We will show that if (I) and (II) are satisfied then neither type will choose to announce the other. Without loss of generality let $\theta_i' > \theta_i$. We have,

$$U_i(\theta_i') - U_i(\theta_i) = \int_{\theta_i}^{\theta_i'} \overline{v}_i(s)ds \geq \int_{\theta_i}^{\theta_i'} \overline{v}_i(\theta_i)ds = (\theta_i' - \theta_i)\overline{v}_i(\theta_i)\,,$$

where the first equality follows from (II) and the inequality follows from (I) since $\theta_i < s$ for all $s \in (\theta_i, \theta_i']$. But this can be rewritten as,

$$U_i(\theta_i') \geq U_i(\theta_i) + \overline{v}_i(\theta_i)(\theta_i' - \theta_i)\,.$$

But this is just inequality (2.7) in the first part of the proof, which implies that θ_i' won't imitate θ_i. Similarly we get inequality (**??**). *Q.E.D.*

Note that the result in equation (ET) that we obtained in the proof is just the *envelope theorem*: Assume that all functions are differentiable. We therefore have,

$$U(\theta_i) = \theta_i\overline{v}_i(\theta_i) + \overline{t}_i(\theta_i)\,,$$

$$\Rightarrow U'(\theta_i) = \overline{v}_i(\theta_i) + \theta_i\frac{d\overline{v}_i(\theta_i)}{d\theta_i} + \frac{d\overline{t}_i(\theta_i)}{d\theta_i}\,. \tag{2.9}$$

Then, since the agent maximizes $\theta_i\overline{v}_i(\hat{\theta}_i) + t_i(\hat{\theta}_i)$ over $\hat{\theta}_i \in \Theta_i$, and we know that truth telling implies that this is maximized at $\hat{\theta}_i = \theta_i$, then the following is no other than the first-order necessary condition,

$$\theta_i\frac{d\overline{v}_i(\theta_i)}{d\theta_i} + \frac{d\overline{t}_i(\theta_i)}{d\theta_i} = 0\,,$$

which in turn implies that (2.9) reduces to $U'(\theta_i) = \overline{v}_i(\theta_i)$. (This characterization of the problem is just like Mirrlees (1971) for the single agent case.)

2.7.1 The Myerson-Satterthwaite Theorem

Consider a Seller (S) and Buyer (B) of an indivisible good. Each agent's (linear) valuation is $\theta_i \in \Theta_i = [\underline{\theta}_i, \overline{\theta}_i] \subset \Re$, where $\theta_i \sim \phi_i(\cdot)$ are independent, and $\phi_i(\cdot) > 0$ for all $\theta_i \in \Theta_i$. Finally, let $\underline{\theta}_i < \overline{\theta}_i$ so that we are not in a degenerate case. Let $f(\theta) = (y_1(\theta), y_2(\theta), t_1(\theta), t_2(\theta))$ be a "trading rule" (this is our SCF for this setting) where $y_i(\theta)$ denotes the probability that agent i gets the good, and $t_i(\theta)$ is his transfer. Therefore, given θ, i's expected utility is $\theta_i y_i(\theta) + t_i(\theta)$. Now define:

$$y(\theta_B, \theta_S) = \begin{cases} 1 \text{ if } \theta_B > \theta_S \\ \frac{1}{2} \text{ if } \theta_B = \theta_S \\ 1 \text{ if } \theta_B < \theta_S \end{cases}$$

It is easy to see that for any ex-post efficient trading rule we must have: $y_B(\theta) = y(\theta)$, $y_S(\theta) = 1 - y(\theta)$ and $t_S(\theta) = -t_B(\theta)$.

We know from AGV that we can implement an ex-post efficient trading rule in BNE. However, if we add *Interim IR* we get the following negative result:

Proposition 5 *(The Myerson-Satterthwaite Theorem) In the bilateral trade setting above with $\Theta_B \cap \Theta_S \neq \phi$ (there are expected gains from trade) there is no BIC trading rule that is ex-post efficient and satisfies interim IR.*

Proof: For the general proof where $[\underline{\theta}_B, \overline{\theta}_B] \cap [\underline{\theta}_S, \overline{\theta}_S] \neq \phi$, see MWG p. 895. Here take $[\underline{\theta}_i.\overline{\theta}_i] = [0, 1]$, for both $i \in \{B, S\}$. From the Linear-Model analysis we have:

$$U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_0^{\theta_i} \overline{v}_i(\tau) d\tau$$

for this case, $v_i(k(\theta)) = y_i(\theta)$, so we use $\overline{y}_i(\theta)$ instead of $\overline{v}_i(\theta)$

$$\begin{aligned}
\overline{y}_B(\theta_B) &= \int_0^1 y(\theta_B, \theta_S)\underbrace{\phi_S(\theta_S)}_{=1} d\theta_S = \int_0^{\theta_B} 1 \cdot d\theta_S + \int_{\theta_B}^1 0 \cdot d\theta_S = \theta_B \\
\overline{y}_S(\theta_S) &= \int_0^1 [1 - y(\theta_B, \theta_S)]\underbrace{\phi_B(\phi_B)}_{=1} d\theta_B \\
&= \int_0^1 d\theta_B - \int_0^{\theta_S} 0 \cdot d\theta_B - \int_{\theta_S}^1 1 \cdot d\theta_S = \cdot\theta_S
\end{aligned}$$

We can now compute the *expected ex-ante surplus* for each agent. First, given the buyer's type, his expected utility is,

$$U_B(\theta_B) = U_B(0) + \int_0^{\theta_B} \tau d\tau = U_B(0) + \frac{\theta_B^2}{2} ,$$

which implies that his expected ex-ante utility is,

$$\begin{aligned}
E_{\theta_B}[U_B(\theta_B)] &= U_B(0) + \int_0^1 \frac{\theta_B^2}{2} \phi(\theta_B) d\theta_B \\
&= U_B(0) + \left[\frac{\theta_B^3}{6} \right|_0^1 = U_B(0) + \frac{1}{6} .
\end{aligned}$$

Similarly (same $\overline{y}(\cdot)$ function): $E_{\theta_S}[U_S(\theta_S)] = U_S(0) + \frac{1}{6}$. So, total ex-ante expected surplus from the mechanism that satisfies ex-post efficiency, BIC and Interim IR must be,

$$E[TS] = \underbrace{U_B(0)}_{\geqslant 0} + \underbrace{U_S(0)}_{\geqslant 0} + \frac{2}{6} \geq \frac{2}{6} .$$

Now we can compute the expected gains from trade. Gains from trade are realized (through trade) when $\theta_B > \theta_S$, and are equal to $\theta_B - \theta_S$. If $\theta_B < \theta_S$ no trade should happen and there are no gains from trade.

Thus, expected gains from trade are,

$$\begin{aligned}
E[GT] &= \int_0^1 \int_0^1 (\theta_B - \theta_S) \cdot y(\theta_B, \theta_S) d\theta_S d\theta_B = \\
&= \int_0^1 \int_0^{\theta_B} (\theta_B - \theta_S) d\theta_S d\theta_B = \int_0^1 \left[\int_0^{\theta_B} \theta_B d\theta_S - \int_0^{\theta_B} \theta_S d\theta_S \right] d\theta_B \\
&= \int_0^1 [\theta_B^2 - \frac{\theta_B^2}{2}] d\theta_S = \left[\frac{\theta_B^3}{6} \right|_0^1 = \frac{1}{6}
\end{aligned}$$

Thus, for the BIC mechanism to be ex-post efficient and satisfy interim IR, we need to supply more expected surplus than we possibly have - a contradiction. *Q.E.D.*

Intuition: To satisfy IC we may need to make some types worse off at the interim stage. If we cannot do that, then to satisfy both IC and IR we need more surplus than the efficient trading rule could produce.

- Cramton-Gibbons-Klemperer : use countervailing incentives

2.7.2 Provision of Public Goods

Add this???????

Chapter 3

Dynamic Contracts

3.1 The Problem of Commitment

Use Fudenberg-Tirole 7.6.4 and Salanie - dynamic mechanism can be replaced with a static lottery - same in each period.

Consider again the two type linear utility model with:

$$u(x,t,\theta)=\theta x-t,$$

where $\theta\in\{\theta_H,\theta_L\}$, $\theta_H>\theta_L$, and $\Pr\{\theta=\theta_H\}=\mu$. Let $c(x)$ be the cost of producing x, with $c'>0$ and $c''>0$. We know that IC_H and IR_L will bind (and the other two constraints are redundant) so that the principal maximizes,

$$\begin{cases}\max\limits_{(t_H,x_H)(t_L,x_L)} & (1-\mu)[t_L-c(x_L)]+\mu[t_H-c(x_H)] & \\ \text{s.t.} & t_L=\theta_L x_L & (\text{IR}_\text{L}) \\ & t_H=\theta_L x_L+\theta_H(x_H-x_L) & (IC_\text{HL})\end{cases}$$

and the solution is solved using the FOC's as before:

$$\begin{aligned}c'(x_H) &= \theta_H\,,\\ c'(x_L^{SB}) &= \theta_L-\frac{\mu}{1-\mu}(\theta_H-\theta_L)<\theta_L\,.\end{aligned}$$

These equations solve for the x's, and we have $x_H^{SB}=x_H^{FB}$ and $x_L^{SB}<x_L^{FB}$, and the transfers are given by,

$$\begin{aligned}t_L &= \theta_L x_L^{SB}\\ t_H &= \theta_H x_H^{SB}-\underbrace{x_L^{SB}(\theta_H-\theta_L)}_{\text{information rents}}\end{aligned}$$

We now ask ourselves what happens if this buyer-seller relationship is repeated twice? (or more...)

- **Setup***:*

  1. Relationship is repeated twice.
  2. Buyer's type is determined *before* the first period and *does not change* over time.
  3. *What kind of contracts can the seller offer, and commit to?* This is a very important point that will later change the results of the model. We begin by assuming that the seller can commit to a contract ex-ante. That is, he can offer a menu $\{(t_{H\tau}, x_{H\tau}), (t_{L\tau}, x_{L\tau})\}_{\tau=1}^{2}$ where τ denotes the period. This menu is offered in advance, before transactions begin, and is "written in stone."
  4. Let $\delta > 0$ be the discount factor between periods. In subsequent parts of this section we will assume a rather uncommon assumption on the values of δ, and distinguish between two cases: (1) $\delta < 1$ (which is the standard case in economics,) and (2) $\delta > 1$ which is *very unusual,* but tries to capture the following idea: The second period is more valuable, for example, it is "longer" in terms of enjoying the good. (but θ is not changed!). This, in a rough way, represents a model with more than one period, where the contract distinguishes between the first period and all other periods.

Under the above setup, we can use the **Revelation Principle.** This follows from the fact that the principal can commit to the menu of contracts, and this menu (which is a mechanism) is defined before players reveal their types. Thus, the principal's problem in the 2-period model is:

$$\max_{\substack{\{(t_{L\tau},x_{L\tau}),\\ (t_{H\tau},x_{H\tau})\}_{\tau=1}^{2}}} \quad \begin{aligned}&(1-\mu)[t_{L1}+\delta t_{L2}-c(x_{L1})-\delta c(x_{L2})]\\ &+\mu[t_{H1}+\delta t_{H2}-c(x_{H1})-\delta c(x_{H2})]\end{aligned}$$

$$\begin{array}{lll}\text{s.t.} & \theta_H(x_{H1}+\delta x_{H2})-t_{H1}-\delta t_{H2}\geq 0 & (IR_{\text{H}})\\ & \theta_L(x_{L1}+\delta x_{L2})-t_{L1}-\delta t_{L2}\geq 0 & (IR_{\text{L}})\\ & \theta_H(x_{H1}+\delta x_{H2})-t_{H1}-\delta t_{H2}\geq \theta_H(x_{L1}+\delta x_{L2})-t_{L1}-\delta t_{L2} & (IC_{\text{HL}})\\ & \theta_L(x_{L1}+\delta x_{L2})-t_{L1}-\delta t_{L2}\geq \theta_L(x_{H1}+\delta x_{H2})-t_{H1}-\delta t_{H2} & (IC_{\text{LH}})\end{array}$$

If we define,

$$\begin{aligned}\tilde{x}_i &= x_{i1} + \delta x_{i2}\,, \\ \tilde{t}_i &= t_{i1} + \delta t_{i2}\,,\end{aligned}$$

then the constraints of the two period problem reduce to those of a standard static one-period problem. This implies that we get the same form of IR and IC as before, in which case we know that only ($\mathrm{IR_L}$) and ($\mathrm{IC_{HL}}$) are relevant, and will bind at a solution.

Constraint ($\mathrm{IR_L}$) binding yields,

$$t_{L1} + \delta t_{L2} = \theta_L(x_{L1} + \delta x_{L2}), \tag{3.1}$$

and substituting this into ($\mathrm{IC_{HL}}$) yields,

$$t_{H1} + \delta t_{H2} = \theta_H(x_{H1} + \delta x_{H2}) - (x_{L1} + \delta x_{L2})(\theta_H - \theta_L) \tag{3.2}$$

We now can substitute (3.1) and (3.2) into the objective function to solve for $(x_{H1}^{SB}, x_{H2}^{SB}, x_{L1}^{SB}, x_{L2}^{SB})$. The FOCs with respect to x_{H1} and x_{H2}:

$$\frac{\partial}{\partial x_{H1}} = \mu[\theta_H - c'(x_{H1}^{SB})] = 0$$

$$\frac{\partial}{\partial x_{H2}} = \mu[\delta\theta_H - \delta c'(x_{H2}^{SB})\} = 0$$

which together imply that

$$x_{H1}^{SB} = x_{H2}^{SB} = x_H^{SB} = x_H^{FB}\,.$$

Similarly, the FOCs with respect to x_{L1} and x_{L2} are,

$$\begin{aligned}\tfrac{\partial}{\partial x_{L1}} &= (1-\mu)[\theta_L - c'(x_{L1}^{SB}) - \mu(\theta_H - \theta_L) = 0 \\ &\Leftrightarrow c'(x_{L1}^{SB}) = \theta_L - \tfrac{\mu}{1-\mu}(\theta_H - \theta_L)\,, \\ \tfrac{\partial}{\partial x_{L2}} &= (1-\mu)[\delta\theta_L - \delta c'(x_{L2}^{SB})] - \mu\delta(\theta_H - \theta_L) = 0 \\ &\Leftrightarrow c'(x_{L1}^{SB}) = \theta_L - \tfrac{\mu}{1-\mu}(\theta_H - \theta_L)\,,\end{aligned}$$

which together imply that,

$$x_{L1}^{SB} = x_{L2}^{SB} = x_L^{SB} < x_L^{FB}\,.$$

Note: This analysis implies that $t_{i\tau} = t_i$, $i \in \{\theta_H, \theta_L\}$. That is, transfers are the same as in the one-period model too.

We see that if the principal can commit ex ante to long-term contracts then we get the second-best solution for the static model replicated in each of the two periods. It is easy to see that this analysis extends to more than two periods and we can conclude that in the case of **commitment** we could use the Revelation Principle for *any number* of periods. In a T period model with commitment we will get,

$$x_{i\tau}^{SB} = x_i^{SB} \; \forall i \in \{\theta_L, \theta_H\}, \text{ for all } \tau = 1, ..., T.$$

If, however, the principal cannot commit to stick to the contract, then we can expect the outcomes to vary. The focus of this section is to see what are the consequences of *lack of commitment* on behalf of the principle, and how this affects payoffs and welfare. In particular, two problems arise from the lack of long-term commitment as follows:

Problem 1: What happens when the principal *cannot commit* to the transfers above? (That is, transfers as functions of quantities, using the taxation principle.) In other words, what happens if at $\tau = 2$ the principal can ignore the original contract?

Consider a high type who chooses (t_{H1}^{SB}, x_H^{FB}) at $\tau = 1$, which is illustrated by point A in figure XX below. (recall that the high type receives the FB quantity in both periods.)

Figure Here

In this case, the principal knows that the buyer is θ_H, and in $\tau = 2$ he can extract *all the surplus* by reneging on (x_H^{FB}, t_{H2}^{SB}) (also point A) and offering $(x_H^{FB}, \hat{t}_H)$ instead (point B). Similarly, after a low type reveals himself by choosing $(x_{L1}^{SB}, t_{L1}^{SB})$ (point C), the principal can extract *more surplus* by reneging on $(x_{L2}^{SB}, t_{L2}^{SB})$ (also point C) and offering $(x_L^{FB}, \hat{t}_L)$ instead (point D). Since the nature of the game is common knowledge, the θ_H buyer will anticipate this, choose $(x_{L1}^{SB}, t_{L1}^{SB})$ which does not change his rents in the first period (recall that $(\mathrm{IC_{HL}})$ binds), and then he will be offered a θ_L contract (point D) which gives him further additional rents compared to the first period outcome (since $x_L^{FB} > x_{L1}^{SB}$).

This problem is called the *Rachet Effect. The* informal idea was around for quite some time: If a person reveals himself to value something very much, then in subsequent transactions he will be charged more for the same good, assuming some monopoly is supplying the good. (An even better example is in centrally planned economies, in which "target" incentives were "ratcheted up" with dynamics. This can also apply to demands from civil servants, and other such relationships.) The formal analysis was offered by Frexais et al (1985), and Laffont and Tirole (1988).

Problem 2: What happens if the principal can *commit not to change* contracts in $\tau = 2$ unilaterally, but *cannot commit not to renegotiate* the contract?

To see what happens we can look back at the Figure XX above. The principal cannot change the high-type's point from A to B, and the high-type is served efficiently after revealing himself (as he was before). As for the low-type, after revealing himself he is supposed to still be served inefficiently in the second period (since $x_L^{FB} > x_{L1}^{SB}$). The principal can then offer him the point D instead of point C and the low-type will accept. (In fact, the low-type buyer will be indifferent between the new and old contract since both lie on his zero-utility indifference curve. If we want the low-type buyer to strictly prefer the change, the principal can offer ε less in the transfer.) This point D gives the high type more rents than his (efficient) point A. Thus, the high-type will anticipate this, claim to be a low-type, get the same rents in $\tau = 1$ (since (IC_{HL}) binds) and higher rents in $\tau = 2$. This is the *problem of renegotiation*. (First introduced by Dewatripont (1988, 1989)).

Following the two problems described above, we will focus attention on three cases of commitment power:

1. **Full Commitment:** The Revelation Principle works, since the principal can commit to any contract ex-ante. In this case we get a simple replication of the static model.

2. **Long Term Renegotiable Contracts:** The Revelation Principle breaks down. The principal can commit *not to make the agents worse off* (the long term contract binds the principal against unilateral deviations). But after information is revealed, the principal can offer a new contract which may ex-post benefit both parties. That is, a contract that maximizes the extraction of surplus using ex-post inefficiencies is not credible if it reveals these inefficiencies.

3. **No commitment:** The Revelation Principle breaks down. The principal can change contracts unilaterally, which implies that after information is revealed, the principal can offer a new contract which extracts all the rents from the buyer. That is, if information is revealed, then a contract that gives buyers some ex-post rents is not credible.

Note: Salanie mentions a fourth type of **Short-Term commitment** that we won't discuss.

How can we think about the inability to commit? Basically, this limits the sets of contracts that the principal can offer *ex-ante*. With commitment, the principal can choose *any* contract that is incentive compatible and individually rational, even if it implies future inefficiencies or rents to the different types of agents. In particular, if there is a contract that the principal can implement without some commitment power, then he can implement the same contract with full commitment power. This implies that more contracts can be implemented with commitment, which in turn implies that lack of commitment can be viewed as a constraint on the set of contracts that the principal can implement.

Thus, we can expect the inability to commit to impose ex-ante costs on the principal, and reduce his ex-ante (expected) profits. It seems intuitive that a complete lack of commitment is a stronger restriction on the set of credible contracts relative to the inability to commit not to renegotiate. This is in fact true, and hurts the principal more.

Question: How "fast" will high types reveal themselves truthfully?

That is, since the revelation of hidden information is at the heart of commitment problems, will the principal choose to maximize the information revelation at the beginning of the relationship? The formal analysis is fairly complicated since we need to resort to Perfect-Bayesian Equilibrium (PBE) with possibly different types of agents using mixed strategies.

3.2 A Simple Model of Information Revelation

Consider a situation in which a monopolist can produce one unit of a good in each period at zero cost. As in the previous subsection, assume that

$u(x,t,\theta) = \theta x - t$, where x is defined as the probability of having the good, $\theta \in \{\theta_L, \theta_H\}$, and $\mu = \Pr\{\theta = \theta_H\}$.

When considering the static model, this is a particular version of what we had earlier. The constraints (IC_{HL}) and (IR_{L}) bind, which together yield (we use (IR_{L}) binding to rewrite (IC_{HL})),

$$\begin{array}{ll} t_L = \theta_L x_L & (IR_L) \\ t_H = \theta_H x_H - x_L(\theta_H - \theta_L) & (IC_{\text{HL}}) \end{array}$$

The monopolist therefore solves,

$$\max_{\{(t_L,x_L),(t_H,x_H)\}} \quad \mu t_H + (1-\mu)t_L$$

$$\begin{array}{ll} t_L = \theta_L x_L & (IR_{\text{L}}) \\ t_H = \theta_H x_H - x_L(\theta_H - \theta_L) & (IC_{\text{HL}}) \end{array}$$

and substituting the binding constraints into the objective yields,

$$\max_{x_L,x_H} \quad \mu\theta_H x_H + (\theta_L - \mu\theta_H)x_L$$

Solving this trivially yields,

$$\begin{aligned} x_H &= \begin{cases} 1 \text{ if } \mu\theta_H > 0 \\ 0 \text{ if } \mu\theta_H < 0 \end{cases} \\ x_L &= \begin{cases} 1 \text{ if } \mu < \frac{\theta_L}{\theta_H} \\ 0 \text{ if } \mu > \frac{\theta_L}{\theta_H} \end{cases} \end{aligned}$$

This is the typical "Bang-Bang" solution of a linear model with a fixed quantity. Since $\mu\theta_H > 0$ we must have that $x_H = 1$.[1] As for determining x_L, we have two cases:

1. $\mu > \frac{\theta_L}{\theta_H}$. In this case there are "enough" high types (high probability of high type), so it is optimal to charge only a high price and only sell to the high types (make a sale with probability μ).

2. $\mu < \frac{\theta_L}{\theta_H}$. In this case there are "too few" high types (low probability of high type), so it is better selling to both types (sell with probability 1) at a low price.

[1] Otherwise, either $\mu = 0$, which means there is no uncertainty, or $\theta_H = 0$, which again is not "interesting" since $\theta_H > \theta_L$ cannot be satisfied with positive valuations.

Using the *Taxation Principle* and assuming that in a static model type θ_i rationally buys if and only if $t \leqslant \theta_i$, the monopolist sets,

$$t^*(\mu) = \begin{cases} \theta_H \text{ if } \mu > \frac{\theta_L}{\theta_H} \\ \theta_L \text{ if } \mu < \frac{\theta_L}{\theta_H} \end{cases} \tag{3.3}$$

3.2.1 Two Periods: Commitment With $\delta > 0$

The monopolist's objective function is,

$$\max\ \mu(t_{H1} + \delta t_{H2}) + (1-\mu)(t_{L1} + \delta t_{L2})$$

and the constraints are,

$$\begin{array}{l} \theta_i(x_{i1} + \delta x_{i2}) - (t_{i1} + \delta t_{i2}) \geq 0, \;\; i \in \{\theta_H, \theta_L\} \\ \theta_i(x_{i1} + \delta x_{i2}) - (t_{i1} + \delta t_{i2}) \geq \theta_i(x_{ji} + \delta x_{j2}) - (t_{ji} + \delta t_{j2}), \;\; i \neq j. \end{array}$$

where the first represents the two IR constraints and the second the two IC constraints. As before, redefine

$$\begin{aligned} \tilde{x}_j &\equiv x_{j1} + \delta x_{j2} \in [0, 1+\delta]\,, \\ \tilde{t}_j &\equiv t_{j1} + \delta t_{j2}\,, \end{aligned}$$

and we are back to the same problem with the solution,

$$\begin{aligned} \tilde{x}_H &= 1+\delta\,; \\ \tilde{x}_L &= \begin{cases} 1+\delta \text{ if } \mu < \frac{\theta_L}{\theta_H} \\ 0 \text{ if } \mu > \frac{\theta_L}{\theta_H} \end{cases} \end{aligned}$$

where $\tilde{x}_j = 1+\delta$ is equivalent to $x_{j1} = x_{j2} = 1$. This implies that charging a price of $t^*(\mu)$ (as defined in (3.3) above) in each period will be optimal.

From now we only consider $\mu > \frac{\theta_L}{\theta_H}$, since this is the "interesting" case, (When $\mu < \frac{\theta_L}{\theta_H}$ both types are served and no information is revealed in the optimal commitment contract.) From the analysis above, if the principal can commit then the optimal contract calls for the high type to reveal himself immediately, and all the surplus is extracted from him in both periods.

Question: What happens if the principal cannot commit to the original contract?

There are clearly two cases:

1. If at $\tau = 1$ the agent buys at $t = \theta_H$, then he must be a high type so that $t_2 = \theta_H$ is optimal in the second period, and the high-type gets no rents in either period.

2. If at $\tau = 1$ the agent doesn't buy, then he must be a low type, so the principal is tempted to charge $t_2 = \theta_L > 0$, thus changing the contract. Anticipating this, the high type prefers not to buy in period $\tau = 1$, and he will get positive rents equal to $\delta(\theta_H - \theta_L)$ in the second period (from buying at a price θ_L.)

Question: Is this ratcheting or renegotiation failure?

It turns out that in this simple 2-period model we cannot distinguish between them if $\delta < 1$ (though, it looks more like a renegotiation problem).

To distinguish between the two cases we will use a trick of having $\delta < 1$ and $\delta > 1$. When $\delta > 1$ there is a difference between Short-Term (ST) contracts (no-commitment), and Long-Term (LT) renegotiable contracts.

3.2.2 The Case of No Commitment (ST contracts):

We solve the problem by looking for the principal's preferred PBE in this dynamic game of incomplete information. The game has two stages:

- At $\tau = 1$ the principal offers t_1, then the buyer chooses $x_1 \in [0, 1]$.

- Following the buyer's choice, let $\mu_2(x_1, t_1) \equiv \Pr\{\theta = \theta_H | x_1, t_1\}$ denote the posterior equilibrium belief of the principal that the agent is a high type after x_1 was chosen following an offer of t_1. From the previous analysis of the static model we can use backward induction and conclude that,
$$t_2(x_1, t_1) = \begin{cases} \theta_H \text{ if } \mu_2(x_1, t_1) > \frac{\theta_L}{\theta_H} \\ \\ \theta_L \text{ if } \mu_2(x_1, t_1) < \frac{\theta_L}{\theta_H} \end{cases}$$
To save on notation we will slightly abuse notation and use $\mu_2(x_1)$ and $t_2(x_1)$ since the principle always knows t_1.

Option 1: No Revelation of Information in Period 1.

In this case we have both types of buyers choosing the same action in the first period, implying that in any such equilibrium, $\mu_2(x_1, t_1) = \mu$. Observe that once we restrict attention to equilibria with no revelation of information, it can easily be established that we cannot have a PBE in which both types do not buy at $\tau = 1$. This follows because if the equilibrium path requires both types not to buy at $\tau = 1$, then $\mu_2(x_1 = 0, t_1) = \mu$, in which case $t_2 = \theta_H$, profits are achieved only in the second period, and ex ante expected profits are $\mu\delta\theta_H$. Also, in this situation sequential rationality implies that the high type would be willing to buy in the first period at any price below θ_H. (This follows since not buying will cause $t_2 = \theta_H$, whereas buying will cause $t_2 \leqslant \theta_H$, depending on beliefs.) But then, this strategy of the high type implies that the principal is better off charging $t_1 = \theta_H - \varepsilon$, that will cause the high type to buy in the first period (anticipating $t_2 = \theta_H$), and then the principal's profits are $\mu[(\theta_H - \varepsilon) + \delta\theta_H]$. Thus, we cannot have an equilibrium in which both types do not buy in the first period.

Clearly, in any equilibrium that has both types buying in the first period, we must have $t_1 \leq \theta_L$. Thus, if this can be achieved as an equilibrium, the following sequence of events must be true:

- Principal charges $t_1 = \theta_L$,and,

$$t_2(x_1) = \begin{cases} \theta_H \text{ if } \mu_2(x_1) > \frac{\theta_L}{\theta_H} \\ \\ \theta_L \text{ if } \mu_2(x_1) < \frac{\theta_L}{\theta_H} \end{cases}$$

- Both types buys in period 1
- High type buys in period 2 if and only if $t_2 \leq \theta_H$

Such an outcome would cause the principal's expected profits to be,

$$\pi_0 = \theta_L + \delta\mu\theta_H \ .$$

Thus, if this could be supported as a PBE, the low type buyer buys only in the first period, and a high type buyer buys in both periods. There is no revelation of information since both types pool their actions in the first period.[2]

[2] For the example at hand this cannot be a PBE, and the next section will show why.

Option 2: Full Revelation of Information in Period 1.

We saw that in the commitment case we get full revelation of information in period 1. If we have full revelation then by Bayes rule we must have that $\mu_2(x_1 = 0) = 0$ and $\mu_2(x_1 = 1) = 1$ since $x_{H1} = 1$ and $x_{L1} = 0$. This implies that $t_2(x_1 = 1) = \theta_H, \quad t_2(x_1 = 0) = \theta_L$. Then, we must have $t_1 < \theta_H$ so that the high type will have some rents and be willing to reveal himself. In particular, the first period (IC_{H}) will be,

$$\underbrace{\theta_H - t_1}_{\text{buy at } \tau=1} + \underbrace{\delta(\theta_H - \theta_H)}_{\text{no rents at } \tau=2} \geq \underbrace{0}_{\text{don't buy at } \tau=1} + \underbrace{\delta(\theta_H - \theta_L)}_{\text{rents at } \tau=2},$$

or,

$$t_1 \leq (1-\delta)\theta_H + \delta\theta_L \tag{3.4}$$

If $\delta \in (0,1)$ then(3.4) implies that $t_1 \in (\theta_L, \theta_H)$, while if $\delta > 1$ then $t_1 < \theta_L$. But then, low types will choose to buy in the first period and get $(\theta_L - t_1) > 0$ rather than not buy and get 0 in both periods. (Often called the "Take the money and run" phenomenon.) Therefore, we cannot get full revelation of information in the first period if $\delta > 1$. The intuition in the model is simple. When the future is more important, we need to give huge rents in period 1 for the high type to reveal himself, and these benefits actually make the low type want to purchase, violating the separating incentives.

If $\delta < 1$, we need to satisfy IC_{L} in the first period $\tau = 1$:

$$\underbrace{0}_{\text{don't buy in } \tau=1} + \underbrace{\delta(\theta_L - \theta_L)}_{\text{no rents}} \geq \underbrace{\theta_L - t_1}_{\text{buy at } \tau=1} + \underbrace{0}_{\text{don't buy at } \tau=2} \tag{3.5}$$

or,

$$t_1 \geq \theta_L \,.$$

Thus, for $\delta < 1$ there is a continuum of t_1 that satisfy both IC constraints (3.4) and (3.5), but clearly the principal will set the highest t_1 which is calculated from (3.4) holding with equality:

$$t_1 = (1-\delta)\theta_H + \delta\theta_L \,,$$

we will get full revelation, and profits are:

$$\begin{aligned} \pi_F &= \mu[(1-\delta)\theta_H + \delta\theta_L + \delta\theta_H] + (1-\mu)\delta\theta_L \\ &= \mu\theta_H + \delta\theta_L \end{aligned}$$

Comparing this with the profits from no revelation we have,

$$\begin{aligned}\pi_F - \pi_0 &= \mu\theta_H + \delta\theta_L - \theta_L - \delta\mu\theta_H \\ &= (\mu\theta_H - \theta_L)(1-\delta) \\ &> 0\end{aligned}$$

Therefore, if $\delta < 1$ then full revelation can be supported as a PBE and it is better than no revelation of information.[3] Thus, even if no revelation could be supported as a PBE, it would not be the principal's preferred PBE.

Option 3: Partial Revelation of Information in Period 1.

In this scenario some type of agent will play a mixed strategy in period 1, and this will prevent the principal from "learning" the true type of the agent after the first period choice of purchase. There are three cases here:

Case 3.1: One type always buys at $\tau = 1$, and the other type buys at $t = 1$ with some positive probability ρ. In this case we must have $t_1 \leqslant \theta_L$ (for both types to be willing to buy), and $t_2(x_1) \leqslant \theta_H$ regardless of x_1. But then, no revelation will strictly dominate this case (from the principal's point of view) for all $\rho < 1$, and full revelation will also dominate this case. Thus, this type of partial revelation can never be the preferred equilibrium for the principal.

Case 3.2: High types never buy at $\tau = 1$, Low types buy at $\tau = 1$ with some positive probability. This implies that $\mu_2(x_1 = 1) = 0$, and $\mu_2(x_1 = 0) > \frac{\theta_L}{\theta_H}$, which in turn implies that $t_2(x_1 = 1) = \theta_L$, and $t_2(x_1 = 0) = \theta_H$. However, this situation will violate IC for the High types because in order to get the Low types to buy at $\tau = 1$ we need $t_1 \leq \theta_L$, so that High-types will get positive rents in each period from imitating Low-types. So, case 3.1 is ruled out as an equilibrium.

Case 3.3: Low types never buy at $\tau = 1$, High types buy with positive probability. This implies that $\mu_2(x_1 = 1) = 1$, $\mu_2(x_1 = 0) \geq \frac{\theta_L}{\theta_H}$, $t_1 = \theta_H$, $t_2(x_1 = 0) = t_2(x_1 = 1) = \theta_H$. In this case High types use a mixed

[3] The PBE is described as follows: The strategies in the first period for the low and high types are given by the IC constraints. Namely, for the high type $x_1(t_1) = 1$ if and only if $t_1 \leq (1-\delta)\theta_H + \delta\theta_L$, and $x_1 = 0$ otherwise, and for the low type $x_1(t_1) = 1$ if and only if $t_1 \leq \theta_L$, and $x_1 = 0$ otherwise. The beliefs are well defined for any history. This makes the principal act optimally in the second period, and choose $t_1 = (1-\delta)\theta_H + \delta\theta_L$ in the first.

strategy in the first period: buy with probability $\rho > 0$, and don't buy with probability 1− ρ. Using Bayes rule,

$$\begin{aligned}\mu_2(x_1 &= 1) = 1, \\ \mu_2(x_1 &= 0) = \frac{(1-\rho)\mu}{(1-\rho)\mu + (1-\mu)}\end{aligned}$$

Notice that in equilibrium, we first we need $\mu_2(x_1 = 0) \geq \frac{\theta_L}{\theta_H}$ (so that the principal will be willing to stick to $t_2 = \theta_H,$) and second, the principal would like to separate "as many" H-types as possible (i.e., have highest probability of separation) in first period since this maximizes profits. Higher separation in the first period implies lower $\mu_2(x_1 = 0)$, which means that in equilibrium we must have,

$$\begin{aligned}\mu_2(x_1 &= 0) = \frac{\theta_L}{\theta_H} \\ &\Leftrightarrow \frac{(1-\rho)\mu}{(1-\rho)\mu + (1-\mu)} = \frac{\theta_L}{\theta_H} \\ &\Leftrightarrow (1-\rho) = \frac{(1-\mu)\theta_L}{\mu(\theta_H - \theta_L)} \\ &\Leftrightarrow \rho = \frac{\mu\theta_H - \theta_L}{\mu(\theta_H - \theta_L)} < \frac{\mu\theta_H - \mu\theta_L}{\mu(\theta_H - \theta_L)} = 1\end{aligned}$$

and indeed $\rho < 1$ is consistent with the case of partial revelation. Now the principal's profits are:

$$\pi_P = \mu[\rho\theta_H + \delta\theta_H]$$

When is this better than full revelation?

$$\pi_P \geq \pi_F$$

$$\begin{aligned}
&\Leftrightarrow\quad \mu\rho\theta_H + \mu\delta\theta_H \geq \mu\theta_H + \delta\theta_L \\
&\Leftrightarrow\quad \mu(\delta+\rho-1) \geq \delta\frac{\theta_L}{\theta_H} \\
&\Leftrightarrow\quad \mu\delta + \frac{(1-\mu)\theta_L}{\theta_H-\theta_L} \geq \delta\frac{\theta_L}{\theta_H} \\
&\Leftrightarrow\quad \mu(\delta\theta_H - \delta\theta_L + \theta_L) \geq \theta_L\left(1+\delta-\delta\frac{\theta_L}{\theta_H}\right) \\
&\Leftrightarrow\quad \mu \geq \frac{\theta_L}{\theta_H}\cdot\underbrace{\left[\frac{\theta_H+\delta(\theta_H-\theta_L)}{\theta_L+\delta(\theta_H-\theta_L)}\right]}_{>1} > \frac{\theta_L}{\theta_H}
\end{aligned}$$

Conclusion for ST contracts (no commitment):

1. If $\mu < \frac{\theta_L}{\theta_H}$ then $t_1 = t_2 = \theta_L$ is best for principal, and we have no inefficiencies and no revelation of information.

2. If $\frac{\theta_L}{\theta_H} \leq \mu < \frac{\theta_L}{\theta_H}\cdot\frac{\theta_H+\delta(\theta_H-\theta_L)}{\theta_L+\delta(\theta_H-\theta_L)}$ then

$$\begin{aligned}
t_1 &= (1-\delta)\theta_H + \delta\theta_L, \\
t_2 &= \begin{cases} \theta_H & \text{If } x_1 = 1 \\ \theta_L & \text{If } x_1 = 0 \end{cases}.
\end{aligned}$$

   Furthermore, we get the following features of the equilibrium:

   - If the buyer is a high type then he buys in both periods.
   - If the buyer is a low type then he buys only in the second period. Therefore, low types consume inefficiently in $\tau = 1$.
   - Full revelation of information in the first period.
   - Not Possible for $\delta > 1$!

3. If $\mu \geq \frac{\theta_L}{\theta_H}\cdot\frac{\theta_H+\delta(\theta_H-\theta_L)}{\theta_L+\delta(\theta_H-\theta_L)}$ then $t_1 = t_2 = \theta_H$. Furthermore, we get the following features of the equilibrium:

   - The high type buys in the first period with probability $\rho = \frac{\mu\theta_H-\theta_L}{\mu(\theta_H-\theta_L)} < 1$, and buys for sure in the second period. Therefore, high types consume inefficiently in first period ($\rho < 1$).
   - A low type buyer never buys. Therefore, low types consume inefficiently in both periods.
   - We have "slow" revelation of information.

3.2.3 Long Term Renegotiable Contracts

Now consider the case where the principal can commit to prices $t_2(x_1 = 0)$ and $t_2(x_1 = 1)$, but cannot commit not to renegotiate:

Figure Here

Figure XX depicts the time line of a very simple "renegotiation game." We will look for the PBE of this game. A question one can ask is, will we have renegotiation happening in equilibrium?

Definition 6 *A Renegotiation-Proof (RNP) contract is one where there is no renegotiation in equilibrium.*

Turning back to the question of whether we have renegotiation in equilibrium, the following result (in the spirit of the revelation principle) is useful:

Proposition 6 *(RNP Principle) If a contract is a PBE in which renegotiation occurs in equilibrium, then there exists a RNP contract that implements the same outcome.*

The intuition is simple: We can think of the process of renegotiation as a game on its own (or a mechanism) where renegotiation is part of the equilibrium. Thus, if we think of a contract as a "Grand Mechanism" that includes a renegotiation mechanism as part of the equilibrium, then we can incorporate the outcome of the renegotiation mechanism into the grand mechanism. That is, in equilibrium we anticipate the renegotiation, and replicate it by a contract that "skips" the renegotiation and implements the anticipated outcome. The implication is that we can think of the inability to commit not to renegotiate as *a constraint* on the set of contracts: That is, contracts must satisfy IC, IR and RNP.

Example 1 *With full commitment above, $t_1 = t_2(x_1 = 0) = t_2(x_1 = 1) = \theta_H$, but since $\mu_2(0) = 0$ then the principal will renegotiate to $t_2^{'} = \theta_L$, that is, this contract is not RNP. However, the contract $t_1 = \theta_H, \quad t_2(x_1 = 0) = t_2(x_1 = 1) = \theta_L$ is RNP (and also IC and IR).*

First, we wish to investigate what can be implemented using PBE that are IC, IR and RNP.

Claim: If a succession of Short-Term contracts is a PBE in the no-commitment case, then there is a Long Term RNP contract that implements the same outcome when the principal can commit to Long Term contracts but cannot commit not to renegotiate.

proof: Trivial: Any succession of Short-Term contracts is a series of (contingent) prices and probabilities of separation. Take such a series, and specify all contingent prices in advance. Since this will be robust against the principal's unilateral deviations, and the principal has all the bargaining power, it must be RNP. *Q.E.D.*

This claim implies that all of our analysis for the no-commitment case follows through. That is, for $\delta < 1$, full separation, no separation and partial separation as given in the analysis of Short-Term contracts, are all Long-Term RNP contracts where the prices $t_1, t_2(x_1)$ are specified *ex ante.* (Note that the probability of separation, ρ, will be part of the equilibrium specification. e.g., $t_1 = \theta_H, \quad t_2(1) = t_2(0) = \theta_H$ and ρ as before are a Long-Term RNP contract with partial revelation of information.) Thus, for $\delta < 1$ the conclusions of the no-commitment case coincide with those of Long-Term RNP contracts where the Long-Term contract is properly defined.

Note: For $\delta < 1$ there are no Long-Term RNP contracts that make the principal better off compared to the Short-Term contracts. This follows from the fact that when $\mu > \frac{\theta_L}{\theta_H}$, then for all levels of separation, the unilateral deviation of the principal that we tried to prevent in Short-Term contracts is *not to lower the price*, which is exactly the same type of deviation we worry about in the case when the principal is unable to commit not to renegotiate.

We now turn to the case $\delta > 1$. This "trick" is meant to capture a relationship where the second period is more than one period.

Example 3.2: Assume that when the agent gets the good in the second period then it lasts for two periods so at $\tau = 2$ he gets $\theta + \delta\theta$, and looking back at $\tau = 1$ he gets (from consumption at time $\tau = 2$): $\delta\theta + \delta^2\theta$. For $\delta \tilde{>} 0.62, \quad (\delta + \delta^2) > 1$ so redefine $\hat{\delta} = (\delta + \delta^2)$.

1. Of course, if this were the case then the principal can charge more than θ at $\tau = 2$. This is just an interpretation and not the true way to capture a 3-period model.

2. If we solved a 3-period model we would exactly see the difference between no-commitment Short-Term contracts and Long-Term RNP contracts, but that would be at the cost of a long and time consuming analysis.

Recall that for $\delta > 1$ we could not have full revelation without commitment since the principal could not commit to $t_2(x_1 = 1) < \theta_H$. Note, however, that now she can commit to a "low" price in the second period, so we can reconsider the incentive constraints for full separation where $t_1 = \theta_H$. We will have to specify $t_2(x_1 = 1)$ and $t_2(x_1 = 0)$ ex-ante, and beliefs $\mu_2(x_1 = 1) = 1\,,\ \mu_2(x_1 = 0) = 0$ will ensure that the contract is RNP. For simplicity of notation, redefine $t_2(1) \equiv t_2(x_1 = 1)$ and $t_2(0) \equiv t_2(x_1 = 0)$.

We begin with the high type's incentive constraint. If the high type chooses to purchase in both periods, his utility is $\theta_H - t_1 + \delta(\theta_H - t_2(1))$. In contrast, if he does not buy then he gets zero in the first period, and after "fooling" the principal he gets $\delta(\theta_H - t_2(0))$. Thus, the high types IC is given by,

$$\theta_H - t_1 + \delta(\theta_H - t_2(1)) \geq 0 + \delta(\theta_H - t_2(0))\,.$$

It is easy to see that any RNP contract that has $\mu_2(x_1 = 0) = 0$ in equilibrium, must also have $t_2(0) = \theta_L$, so we can rewrite ($\mathrm{IC_{HL}}$) as,

$$t_1 + \delta t_2(1) \leq \theta_H + \delta\theta_L$$

As for the low type, ($\mathrm{IC_{LH}}$) is the same as before:

$$t_1 \geq \theta_L$$

Thus, to implement full revelation in the first period with the highest possible price, $t_1 = \theta_H$, the unique Long-Term RNP, IC and IR contract is:

$$\begin{aligned} t_1 &= \theta_H, \\ t_2(1) &= t_2(0) = \theta_L. \end{aligned}$$

Clearly, this is a Long-Term RNP contract even for $\delta > 1$. So, if $\delta > 1$ and $\frac{\theta_L}{\theta_H} \leq \mu < \frac{\theta_L}{\theta_H} \cdot \frac{\theta_H + \delta(\theta_H - \theta_L)}{\theta_L + \delta(\theta_H - \theta_L)}$ the principal can induce full separation which is the *optimal* Long-Term RNP contract. (get "faster" revelation than no commitment). To see that this is indeed the optimal RNP contract.

note that if the principal wants a Long-Term RNP contract with partial separation, then it must be the same one as the no-commitment case. This follows since if she wants to separate a larger ρ than that specified for the Short-Term contract case, then the resulting beliefs after no purchase must be $\mu_2(0) < \frac{\theta_L}{\theta_H}$, and the only RNP price is $t_2(0) = \theta_L$. But then the principal should move to $\rho = 1$ which increases her profits, and this is the case of full separation.

3.3 The Durable Goods Model: Sales vs. Rent Contracts

We could interpret the previous model as one with a durable good, and the contracts are for the right to use the good for each period at a time, that is, these were *rental contracts.* Assume that the principal can offer *sales contracts* as follows:

Sales Contract: "pay s_τ in period τ for the use of the good from period τ until the end."

Then, with Short-Term sales contracts the principal can *get the same outcome* as with Long-Term RNP rent contracts as follows: At $\tau = 1$, let $s_1 = \theta_H + \delta\theta_L$ for use in both periods, and at $\tau = 2$, let $s_2 = \theta_L$ for those who did not buy in $\tau = 1$. It is easy to see that (IC_{HL}) binds and (IC_{LH}) does not bind. This is true for all $\delta > 0$ in the sales-game, and ex ante expected profits are

$$\begin{aligned}\pi &= \mu[\theta_H + \delta\theta_L] + (1 - \mu)\delta\theta_L \\ &= \mu\theta_H + \delta\theta_L.\end{aligned}$$

This and other issues are the focus of Hart-Tirole (1988).

Note: Hart-Tirole (1988) show that in durable good case, Long-Term RNP rental contracts give the same outcome as Short-Term sales contracts with no commitment. They also show:

1. This is a Coasian-Dynamics price path where price decreases over time to the lowest valuation. (Gul-Sonneshein-Wilson 1988, Coase 1972).

2. In the case of sales contracts, LT-RNP adds nothing to Short-Term contracts with no commitment.
3. In the case of rental contracts, Short-Term contracts with no commitment leads to a non-Coasian outcome price path (prices increase) in which there is no initial separation, and then partial separation occurs.

In summary, the paper is interesting but *very complicated.*

3.4 Concluding Remarks on Dynamics

I. *No Commitment: the continuous type case:*

In the static model with $\theta \in [\underline{\theta}, \overline{\theta}]$ we had full separation (for the types being served, assuming that the solution to the reduced program resulted in a monotonic choice function) and this follows through for the dynamic case with commitment. However, when the principal cannot commit we have a strong version of no separation:

Proposition 7 *(Laffont-Tirole 1988) For any first period contract there exists no subinterval of* $\left[\underline{\theta}, \overline{\theta}\right]$ *with positive measure in which full separation occurs.*

We don't provide a proof of this proposition, but the intuition is quite straightforward. If a type $\theta \in [\theta', \theta''] \subseteq [\underline{\theta}, \overline{\theta}]$ is fully revealed then he has some first period rents and *no second period rents.* (This follows because after his type is revealed, all future rents will be extracted by the principal.) By not revealing truthfully, and by announcing $\hat{\theta} = \theta - \varepsilon$ for ε close to zero, there is only a second order loss on rents of the first period, and a first order gain on the rents of the second period, so θ will not announce truthfully.

Laffont and Tirole continue to analyze the procurement set-up with quadratic costs of effort, and they characterize "partition equilibria" in which the type space is partitioned into a (countable) number of pooling subintervals. (Note that their model has an effort parameter following their 1986 paper in the JPE. This is, however, not really a different type of hidden information, because the choice is effort, and in equilibrium the principal knows what effort was chosen.)

II. *No commitment and upward binding IC's:*

In a more smooth setting than the one we analyzed, (e.g., continuous quantities or types,) we might have both low types and high types IC constraints binding: In our model this happened only for $\delta = 1$ and full separation. For $\delta > 1$ we could not fully separate precisely because $\mathrm{IC_L}$ was violated, what we termed the "Take-the-money-and-run" situation where the L-type imitates the H-types in the first period, thus getting rents, and he then declines to participate in second period. (This is one reason the general analysis is very complicated.)

III. *Long-Term RNP contracts: The continuous case*

Proposition 8 *If the principal is constrained to offer long term RNP contracts then,*

1. *There exists an IC, IR and RNP contract (incentive scheme) that separates all types in the first period. The optimal contract in this class yields the optimal static contract in the first period, and the first best allocation in the second period.*

2. *A fully separating contract is never optimal for the principal.*

This result is due to Laffont-Tirole in a succession of articles. For an analysis of this problem the reader is referred to chapters 1, 9, & 10 in Laffont and Tirole, 1993. Instead of proving this, we again resort to describing the intuition of this result.

Consider a two-type setting as before, but let the quantity x be continuous. For part (1) above, consider the case where in period 1 the optimal static contract is offered, and in period 2 the optimal (first-best) choice of quantity is offered with transfers that give all the rents to the agents. (The principal gets 0 profits in the second period.) Laffont and Tirole refer to this incentive scheme as the "sell-out contract."

Figure Here

This is clearly IC because in period 2 it is IC (which follows from $c'' > 0$ and $\theta_H > \theta_L$) and anticipating this, the static contract is IC in period 1. (We may be able to find a Long-Term fully separating contract which takes some rents away from the agents and transfers these to the principal, keeping IC, IR and RNP satisfied. For example, if in the sell-out contract $\mathrm{IR_L}$ is slack, then we can move to a second-period contract with the same first-best quantities, and with larger transfers to the principal.)

More generally, we can separate a continuum of types in a RNP contract (consider the types $\theta_H > \theta_L$ to be two types from the continuum $[\underline{\theta}, \overline{\theta}]$. It is easy to see that the above intuition of the sell-out contract will work with a continuum of contracts, where each type θ chooses the point along the monopolists zero-profit iso-profit curve where the tangent is equal to θ.)

Turning to part (2) of the proposition, first note that in any fully separating equilibrium there is FB optimality in the second period. Given this, the principal gets second best optimality ex-ante using the RNP constraint and full separation. However, by adding some pooling in the first period, the principal can create a new contract that generates second order losses in the first period, and first order gains from extracting rents through inefficiencies in the second period. That is, partial revelation will be optimal.

Chapter 4

Contracting with Externalities

Segal (1999)

Part III

Hidden Action

Chapter 5

The Principal-Agent Model

5.1 Setup

- 2 players: The principal, owner of the firm; and the agent, manager/worker in the firm
- The principal hires the agent to perform a task
- The agent chooses an action, $a \in A$
- Each $a \in A$ yields a distribution over payoffs (revenues) for the firm, q, according to the function $q = f(a, \varepsilon)$ where ε is some random variable with known distribution (Note that the principal "owns" the rights to the revenue q.)
- $a \in A$ is *not observable* to the principal; q is *observable and verifiable* (so q can be a basis for an enforceable contract)

Figure Here (time line)

Examples:

	Principal	Agent	Action
1	Firm owner	Manager	choice of (risky) project
2	Employer/Manager	employee/worker	effort in job
3	Regulator	Regulated firm	cost reduction research
4	Insurer	Insuree	care effort

Question: What makes the problem interesting?

A conflict of interest over $a \in A$ between the principal and the agent.

5.1.1 Utilities

We assume that the principal is risk neutral, and the agent is risk averse. This is the standard formulation of the P-A problem.

- **Agent's utility:** The agent has a vNM utility function defined on his action, a, and the income he receives, I,

  $$U(a, I) = v(I) - g(a),$$

  where $v' > 0$, and $v'' < 0$ guarantee risk aversion. (As for $g(\cdot)$, we assume that effort is costly, $g' > 0$, and we usually assume that the marginal cost of effort increases with effort, $g'' > 0$, which seems "reasonable.") Note that this utility function is additively separable. This simplification is very helpful for the analysis and its implication is that the agent's preference over income lotteries is independent of his choice of action Finally, assume that the agent's reservation utility is given by $\overline{u}$ (which is determined by some alternative option.)

- **Principal's utility:** The principal's utility is just revenue less costs, and if we assume that the only cost is compensating the agent then this is trivially given by $q - I$.

Question: Why is it reasonable to assume that the principal is risk neutral while the agent is risk averse?

1. If the agent is risk neutral, and the agent has no limits on wealth, the problem will be trivial (as we will later see why).

2. If both are risk averse the analysis is more complicated, but we get the same general issues and results, so it is unnecessarily more complex.

3. An appealing rationale is that the principal is wealthy and has many investments, so this firm is only a small fraction of his portfolio and risk is idiosyncratic. (Caveat: usually shareholders don't design incentive schemes!)

4. **Another modelling possibility:** Agent is risk neutral but has limited wealth - say, in the form of limited liability - so that he cannot suffer large losses. This gives a "kink" at an income level of zero, which gives the necessary concavity of the agent's utility function which yields the same kind of results.

Further Assumptions

We begin our analysis by following the work of Grossman and Hart (1983) (G-H hereafter), and later go back to the earlier formulations of the principal-agent problem.

- $\tilde{q}$ (the random variable) is discrete: $q \in \{q_1, q_2, ..., q_n\}$, and w.l.o.g., $q_1 < q_2 < \cdots < q_n$.
- $A \subset \Re^k$ is compact and non-empty.
- given $a \in A$, the mapping $\pi : A \longrightarrow S$ maps actions into distributions over outcomes, where $S = \{x \in \Re^n : x_i \geq 0 \; \forall i \text{ and } \sum\limits_{i=1}^{n} x_i = 1\}$ is the n-dimensional simplex. $\pi_i(a)$ denotes the probability that q_i will be realized given that $a \in A$ was chosen by the agent. As usual, we assume that the distribution functions are common knowledge.
- $v(\cdot)$ is continuous, $v' > 0$, $v'' < 0$, and $v(\cdot)$ is defined over $(\underline{I}, \infty) \in \Re$ where $\lim\limits_{I \to \underline{I}} v(I) = -\infty$ (This guarantees that we need not worry about corner solutions.) For example: $v(\cdot) = \ln(\cdot)$, in which case $\underline{I} = 0$. (This is Assumption A1 in G-H.)
- $g(\cdot)$ is only assumed to be continuous.

5.2 First Best Benchmark: Verifiable Actions

Assume that $a \in A$ is *observable and verifiable* so that the principal can basically "choose" the best $a \in A$, and contract on it while promising the agent some compensation. Assume that the principal has all the bargaining power. (This is a standard simplification, which can be relaxed without affecting the qualitative results.) She then solves:

$$F.B. \begin{cases} \max\limits_{\substack{a \in A \\ I \in \Re}} & \sum\limits_{i=1}^{n} \pi_i(a) q_i - I \\ \text{s.t.} & v(I) - g(a) \geq \overline{u} \quad \text{(IR)} \end{cases}$$

where $I \in \Re$ is the payoff to the agent. Note that we considered I to be a fixed payoff, which brings us to the following question: Could a random payment $\tilde{I}$ be optimal? The answer is clearly no, which follows from the agent's risk aversion. We can replace any $\tilde{I}$ with its *Certainty Equivalent*, which will be less costly to principal since she is risk neutral.

Claim: (IR) binds at the solution

This claim is clearly trivial, for if (IR) would not bind we can reduce I, and lower the principal's costs.

We can define

$$C_{FB}(a) \equiv v^{-1}(\overline{u} + g(a))$$

to be the cost to the principal of compensating the agent for choosing $a \in A$. This is the *First-Best* (FB) cost of implementing an action a, since risk is optimally shared between the risk neutral principal and the risk averse agent. Using this formalization, the principal solves:

$$\max_{a \in A} \sum_{i=1}^{n} \pi_i(q) q_i - C_{FB}(a)$$

The *First-Best Solution* yields a FB action, $a^*_{FB} \in A$ (which may not be unique), and this action is implemented by the FB contract: (a^*_{FB}, I^*_{FB}) where $I^*_{FB} = C_{FB}(a^*_{FB})$.

5.3 Second Best: non-observable action

Once the action is not observable, we can no longer offer contracts of the form (a^*_{FB}, I^*_{FB}). One can then ask, if it is not enough just to offer I^*_{FB} and anticipate the agent to perform a^*_{FB} ? The answer is no, since if I^*_{FB} is offered, then the agent will choose $a \in A$ to minimize $g(a)$, and it may not be that "likely" that a^*_{FB} will achieve the agent's goal.

Question: What can the principal do to implement an action $a \in A$?

The principal can resort to offering an *Incentive Scheme* which rewards the agent according to the level of revenues, since these are assumed to be observable and verifiable. That is, the agent's compensation will be a function, $I(q)$.

By the finiteness of $\tilde{q}$, restrict attention to $I \in \{I_1, ..., I_n\}$. The principal will choose some a^*_{SB}, together with an incentive scheme that implements this action. (That is, an incentive scheme that causes the agent to choose a^*_{SB} as his "best response.") Clearly, at the optimum it must be true that a^*_{SB} is implemented at the *lowest possible cost* to the principal. This implies that we can decompose the principal's problem to a two stage problem as follows:

1. First, look at the lowest cost to implement any $a \in A$ (i.e., for each $a \in A$, find $(I^*_1(a), ..., I^*_n(a))$ which is the lowest cost incentive scheme needed to implement $a \in A$)
2. Given $\{(I^*_1(a), ..., I^*_n(a))\}_{a \in A}$ choose a^*_{SB} to maximize profits.

The Second Stage

If the principal has decided to implement a^*, then it must be implemented at the lowest cost. That is, $(I_1, ..., I_n)$ must solve

$$\begin{cases} \min\limits_{I_1,...,I_n} & \sum\limits_{i=1}^{I} \pi_i(a^*)I_i \\ \text{s.t.} & \sum\limits_{i=1}^{n} \pi_i(a^*)v(I_i) - g(a^*) \geq \overline{u} & \text{(IR)} \\ & \sum\limits_{i=1}^{n} \pi_i(a^*)v(I_i) - g(a^*) \geq \sum\limits_{i} \pi_i(a)v(I_i) - g(a) \;\; \forall a \in A & \text{(IC)} \end{cases}$$

Note that we have a program with a linear objective function, and concave constraints. For mathematical convenience, we can transform the program into one with a convex objective function and with linear constraints. To do this, we work with "utils" instead of income: Let $h(\cdot) \equiv v^{-1}(\cdot)$ be the inverse utility function with respect to the agent's income, and consider the principal's choice of $(v_1, ..., v_n)$, where $I_i = h(v_i)$. (The existence of such an inverse function $h(\cdot)$ is guaranteed by G-H assumption A2.) We know that

$h(\cdot)$ is convex since, $v' > 0$ and $v'' < 0$ imply that $h' > 0$ and $h'' > 0$, and the assumption that $\lim_{I \to \underline{I}} v(I) = -\infty$ implies that $v_i \in (-\infty, \overline{v})$ where $\overline{v}$ can be ∞. The program can therefore be written as,

$$\begin{cases} \min\limits_{v_1,\ldots,v_n} & \sum\limits_{i=1}^{I} \pi_i(a^*)h(v_i) & \\ \text{s.t.} & \sum\limits_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) \geq \overline{u} & \text{(IR)} \\ & \sum\limits_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) \geq \sum\limits_{i} \pi_i(a)v_i - g(a) \;\; \forall a \in A & \text{(IC)} \end{cases} \qquad (5.1)$$

which is a "well behaved" program. (If A is finite, we can use Kuhn-Tucker.)

Question: When will we have a solution?

If the constrained set is empty, we clearly won't have one, so this is not an interesting case. Assuming that the constrained set is non-empty, then if we can show that it is closed and bounded, then a solution exists (Weirstrass). The question is, therefore, when will it be true?

Assumption 4.1: (A3 in G-H) $\pi_i(a) > 0$ for all $i \in \{1, ..., n\}$ and for all $a \in A$.

Proposition 4.1: Under the assumptions stated above, a solution is guaranteed.

Sketch of Proof: The idea is that we can bound the set of v's, thus creating a compact constrained set. Assume in negation that we cannot: $\exists$ an unbounded sequence $\{v_1^k, ..., v_n^k\}_{k=1}^{\infty}$ such that some components go to $-\infty$. This implies that I_i^k will also be unbounded for some i where $I_i^k = h(v_i^k)$. Since $\pi_i(a) > 0$ for all i, then if some $v_i^k \to -\infty$ without some $v_j^k \to +\infty$ then the agent's utility will be going to $-\infty$. So, if $v_i \in (-\infty, \overline{v})$, where $\overline{v} < \infty$ we are done since we cannot have some $v_j^k \to +\infty$. Assume, therefore, that $v_i \in (-\infty, \infty)$ so that we could have some components going to $-\infty$ and others going to $+\infty$. Now risk aversion will come into play: the variance of the incentive scheme I^k goes to infinity, and to compensate the agent for this risk, the mean must go to infinity as well. Thus, the principal's expected payment to agent goes to $+\infty$, which is worse than not implementing the action at all. Therefore, we can put an upper bound on $(-\infty, \infty)$, which is calculated so that if payoffs reach the bound then the principal prefers no action. But then we're back in $(-\infty, \overline{v})$. □

Since we have guaranteed a solution, we can now state some facts about the solution itself. Let $C_{SB}(a^*)$ be the value function of the program, i.e., the second best cost of implementing a^*. This function is well defined, and has the following features:

1. $C_{SB}(a^*)$ can be $+\infty$ for some $a^* \in A$ (if the constrained set is empty for this a^*)
2. $C_{SB}(a^*)$ is *lower semi-continuous,* which implies that it attains a minimum in the constrained set. (A function $f(\cdot)$ is Lower Semi-Continuous at x if $\liminf_{k\to\infty} f(x_k) \geq f(x)$.)
3. If $v'' < 0$ then $C_{SB}(a^*)$ has a unique minimizer.

We also have the following straightforward result:

Lemma 4.1: At the optimum IR binds.

Proof: Assume not. Then, we can reduce all v_i's by $\varepsilon > 0$ such that (IR) is still satisfied. Notice that this does not affect the incentive constraint, and thus still implements a^* at a lower cost to the principal - a contradiction. *Q.E.D.*

Notes:

1. If $U(a, I) = v(I)g(a)$ (multiplicative separability) then (IR) still binds at a solution.. (If not, scale v_i's down by some proportion $\alpha < 1$ and the same logic goes through.)
2. If $U(a, I) = g(a) + v(I)k(a)$ then (IR) may not bind and we may have the agents expected utility exceeding $\overline{u}$. (We can't use any of the above arguments.) In this case we get some "efficiency wage."

The First Stage

After finding $C_{SB}(a)$ the principal solves:

$$\max_{a\in A} B(a) - C_{SB}(a) \tag{5.2}$$

where $B(a) \equiv \sum_{i=1}^{n} \pi_i(a)q_i$ is continuous and $-C_{SB}(a)$ is upper-semi-continuous (because $C_{SB}(\cdot)$ is lower semi-continuous). Thus, since A is compact, we have a "well behaved" program. Let a^*_{SB} solve (5.2) above, we call a^*_{SB} the *second best optimal solution* where $\{I^*_i\}_{i=1}^n$ is given by the solution to the first program, (5.1) above.

Question: When is the SB solution also the FB solution?

Each one of the following is a sufficient condition (from parts of Proposition 3 in G-H):

1. $v'' = 0$ and the agent has unlimited wealth. In this case the principal and agent share the same risk-attitude, and the principal can "sell" the project (or firm) to the agent. Since the agent would then maximize

   $$\max_{a\in A} \sum_{i=1}^{n} \pi_i(a)q_i - v^{-1}(g(a) + \overline{u}) \ ,$$

   which is the expected profits less the cost of effort (and less the outside option), then the principal can ask for a price equal to the value of the agent's maximization program. This results in the principal getting the same profits as in a FB situation, and a^*_{FB} solves the agent's problem after he purchases the firm.

2. If a^*_{FB} is also the solution to $\min_{a\in A} g(a)$. In this case there is no "conflict of interest" between the principal's objectives and the agent's cost-minimizing action.

3. If A is finite, and there exists some a^*_{FB} such that for some i, $\pi_i(a^*_{FB}) = 0$ and $\pi_i(a) > 0$ for all $a \neq a^*_{FB}$. In this case it happens to be that if a^*_{FB} is chosen, then there is some outcome q_i that cannot occur, and if the agent chooses any other $a \in A$, then q_i can happen with some positive

probability. Thus, by letting $I_i = -\infty$ and $I_j = v^{-i}(g(a^*_{FB}) + \overline{u})$ for all $j \neq i$, we implement a^*_{FB} at the FB cost of $v^{-i}(g(a^*_{FB}) + \overline{u})$. This is called the case of "shifting support", since the support of the probability distribution changes with the chosen action.

In cases (1)-(3) above there is no trade-off between optimal risk-sharing and giving the agent incentives to choose a^*_{FB}. (There are 3 more cases in proposition 3 of G-H). But in general, it turns out that we will have such a trade-off, as the following result easily demonstrates, (part of proposition 3 in G-H)

Proposition 4.2: If $\pi_i(a) \geq 0$ for all i and for all $a \in A$, if $v''(\cdot) < 0$, and if $g(a^*_{FB}) > \min_{a \in A} g(a)$ for all a^*_{FB} , then the principal's profits in the SB solution are lower than in FB solution.

Proof: $C_{SB}(a^*_{FB}) > C_{FB}(a^*_{FB})$ because we need incentives ($I_i \neq I_j$ for some i, j) for the agent to choose $a^*_{FB} \neq \arg\min_{a \in A} g(a)$. Therefore, we cannot have optimal risk sharing, which implies that the solution to (5.2) is worse than the FB solution. *Q.E.D.*

5.3.1 The Form of the Incentive Scheme

Assume for simplicity that A is finite and contains m elements, $|A| = m$, so that the Lagrangian of program (5.1) (the cost-minimization program) is:

$$\begin{aligned}
\mathcal{L} &= \sum_{i=1}^{n} \pi_i(a^*)h(v_i) \\
&\quad - \sum_{a_j \neq a^*} \mu_j \left[\sum_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) - \sum_i \pi_i(a_j)v_i + g(a_j) \right] \\
&\quad -\lambda \left[\sum_{i=1}^{n} \pi_i(a^*)v_i - g(a^*) - \overline{u} \right]
\end{aligned}$$

(Note that there are $m-1$ (IC) constraints and one (IR) constraint.) Assume that the program is convex so that the FOCs are necessary & sufficient for a solution, then we have,

$$\pi_i(a^*)h'(v_i) - \sum_{a_j \neq a^*} \mu_j[\pi_i(a^*) - \pi_i(a_j)] - \lambda\pi_i(a^*) = 0 \;\; \forall\, i = 1, ..., n$$

where $\mu_j \geq 0\ \forall j \neq a^*$ and $\lambda \geq 0$ (where $\mu_j > 0$ implies that IC_j binds.) Since Assumption 4.1 guarantees that $\pi_i(a^*) > 0$, we can divide the FOC by $\pi_i(a^*)$ to obtain,

$$h'(v_i) = \lambda + \sum_{a_j \neq a^*} \mu_j - \sum_{a_j \neq a^*} \mu_j \frac{\pi_i(a_j)}{\pi_i(a^*)} \qquad \forall i = 1, ..., n$$

The following result is proposition 6 in G-H:

Proposition 4.3: If $a^*_{SB} \notin \arg\min_{a \in A} g(a)$ and if $v'' < 0$, then $\mu_j > 0$ for some j with $g(a_j) \leq g(a^*_{SB})$.

Proof: Assume not, i.e., the agent strictly prefers a^*_{SB} to all a_j which satisfy $g(a_j) \leq g(a^*_{SB})$. Define the set

$$A' = A \backslash \{a_j : a_j \neq a^*_{SB} \text{ and } g(a_j) \leq g(a^*_{SB})\},$$

and solve the program for $a \in A'$. Since none of the elements in $A \backslash A'$ were chosen when we solved for $a \in A$, we must still get a^*_{SB} as the solution when $a \in A'$. But observe that now $a^*_{SB} \in \arg\min_{a \in A'} g(a)$, which implies that $I_1 = I_2 = \cdots = I_n$ at the solution (full insurance). But this is the case for the FB solution, which contradicts proposition 4.2 above. *Q.E.D.*

This proposition implies that the agent will be indifferent between choosing a^* and choosing some a_j such that $g(a_j) \leq g(a^*_{SB})$. That is, will be indifferent between working "optimally" and working "less", if higher costs of effort are associated with higher levels of effort. Thus, the main point of Proposition is that we must have some "downward" binding incentive constraints.

1. The proof depends on the finiteness of A. In the infinite case this result holds only locally.
2. This result does not rule out some "upward" binding IC's which will be somewhat "problematic" as we will see shortly.

Recall that $h' > 0$, and $h'' > 0$. Thus, we have,

Corollary 4.1: If $\mu_j > 0$ for only one $a_j \neq a^*_{SB}$ then $I_i > I_k$ if and only if $\frac{\pi_i(a^*_{SB})}{\pi_i(a_j)} > \frac{\pi_k(a^*_{SB})}{\pi_k(a_j)}$. That is, I is monotonic in the likelihood ratio.

This follows from the fact that $h'(\cdot)$ is increasing ($h'' > 0$) and from the FOC:

$$h'(v_i) = \lambda + \mu_j - \mu_j \frac{\pi_i(a_j)}{\pi_i(a^*_{SB})}\,. \tag{5.3}$$

If $\frac{\pi_i(a^*_{SB})}{\pi_i(a_j)} > \frac{\pi_k(a^*_{SB})}{\pi_k(a_j)}$, then clearly $\frac{\pi_i(a_j)}{\pi_i(a^*_{SB})} < \frac{\pi_k(a_j)}{\pi_k(a^*_{SB})}$, which implies that for i, the last term in (5.3) becomes less negative compared to k. This implies that $h'(v_i) > h'(v_k)$, which implies that $v_i > v_k$ (since $h'' > 0$.)

We can relate this result to an appealing property of the probability distributions induced by the different actions as follows:

Definition 4.1: Assume that $\pi_i(a) > 0$ for all $i \in \{1, ..., n\}$ and for all $a \in A$. (This was assumption 4.1 above.) The *monotone likelihood ratio condition (MLRC)* is satisfied if $\forall\, a, a' \in A$ such that $g(a') \leq g(a)$, we have $\frac{\pi_i(a)}{\pi_i(a')}$ is nondecreasing in i.

Corollary 4.2: Assume MLRC. Then, $I_{i+1} \geq I_i \;\forall\, i = 1, ..., n-1$ if either,

1. $\mu_j > 0$ for only one $a_j \neq a^*_{SB}$ ($\Rightarrow g(a_j) < g(a^*_{SB})$ from Proposition 4.3)
2. $A = \{a_L, a_H\}$, $g(a_L) < g(a_H)$ and $a^*_{SB} = a_H$

Case (1) follows immediately from Corollary 4.1. Case (2) does as well but it is worth mentioning since this is the "simple" 2-action case. We focus on this kind of "monotonicity" since it seems realistic in the sense that higher output leads to higher payments. We are therefore interested in exploring under what assumptions this kind of result prevails.

1. The solution to the principal-agent problem seems to have a flavor of a statistical-inference problem (the MLRC result). Note, however, that this is *not* a statistical inference problem, but rather an *equilibrium model* for which we found a subgame-perfect equilibrium. In equilibrium the principal has *correct beliefs* as to what the agent chooses and does not need to infer it from the outcome.

2. MLRC $\Rightarrow$FOSD but FOSD $\nRightarrow$ MLRC: Recall that the distribution $\pi(a^*)$ *First Order Stochastically Dominates* the distribution $\pi(a)$ if

$$\sum_{i=1}^{k} \pi_i(a) \geq \sum_{i=1}^{k} \pi_i(a^*) \quad \forall\, k = 1, ..., n$$

i.e., lower output is more likely under a than it is under a^*.(Don't go through the following in class.)

Claim: MLRC$\Rightarrow$FOSD. That is, $\frac{\pi_i(a^*)}{\pi_i(a)}$ increases (weakly) in i implies that

$$\sum_{i=1}^{k} \pi_i(a) \geq \sum_{i=1}^{k} \pi_i(a^*) \;\; \forall\, k = 1, ..., n$$

Proof: (I) we can't have $\frac{\pi_i(a^*)}{\pi_i(a)} > 1 \;\forall\, i$. To see this, assume in negation that $\frac{\pi_i(a^*)}{\pi_i(a)} > 1 \;\forall\, i$. $\Rightarrow$

$$1 = \sum \pi_i(a^*) = \sum \frac{\pi_i(a^*)}{\pi_i(a)} \cdot \pi_i(a) > \sum \pi_i(a) = 1 \text{ a contradiction.}$$

(II) Let $k^* = \max\{i = 1, ..., n \,|\, \frac{\pi_i(a^*)}{\pi_i(a)} \leq 1\}$. Define

$$\varphi_k = \begin{cases} 0 \text{ for } k = 0 \\ \sum\limits_{i=1}^{k} \pi_i(a) - \sum\limits_{i=1}^{k} \pi_i(a^*) \text{ for } k = 1, ..., n \end{cases}$$

note that $\varphi_0 = \varphi_n = 0$, for all $k \leq k^*$ φ_k is increasing in k, and for all $k \geq k^*$ φ_k is decreasing in k. $\Rightarrow \sum\limits_{i=1}^{k} \pi_i(a) - \sum\limits_{i=1}^{k} \pi_i(a^*)$ $\geq 0 \forall k = 1, ..., n$. *Q.E.D.*

Claim: FOSD $\nRightarrow$MLRC.

Example: 3 outcomes: q_1, q_2, q_3, two actions: a, a^*, with probabilities

$$\begin{array}{ll} \pi_1(a) = 0.4 & \pi_1(a^*) = 0.2 \\ \pi_2(a) = 0.4 & \pi_2(a^*) = 0.6 \\ \pi_3(a) = 0.2 & \pi_3(a^*) = 0.2 \end{array}$$

Figure Here

$\pi(a^*)$ FOSD's $\pi(a)$ by taking some probability from q_1 to q_2 without changing the probability of q_3. However, MLRC is violated:

$$\frac{\pi_1(a^*)}{\pi_1(a)} = \frac{1}{2} < \frac{\pi_2(a^*)}{\pi_2(a)} = \frac{3}{2} > \frac{\pi_3(a^*)}{\pi_3(a)} = 1$$

(Again, this follows the intuition of an inference problem - trying to verify that a^* was chosen is by identifying which outcome has a higher likelihood ratio.)

3. If $\mu_j > 0$ for more than one j then MLRC is *not enough* for monotonicity. In this case the IC's can bind in different directions and this causes "trouble" in the analysis. G-H use the *Spanning Condition* as a sufficient condition for monotonicity (see G-H proposition 7.)

4. *Robustness of Monotonicity:* Without MLRC or the Spanning condition, G-H are still are able to show that some monotonicity exists:

   Proposition 5 (G-H): In the SB solution without MLRC and without the Spanning condition, when the SB solution is worse than the FB, then:

   (i) $\exists\ i,\ 1 \leq i < n$ such that $I_{i+1} > I_i$,

   (ii) $\exists\ j,\ 1 \leq j < n$ s.t. $q_j - I_j < q_{j+1} - I_{j+1}$

   Idea: (i) I is monotonic somewhere (that is, we can't have "perverse" incentive schemes.) (ii) "$I'(\cdot) < 1$" somewhere: There is some increase in output that induces an increase in the principal's share (again, can't have "perverse" profit sharing).

5. *Enriching the agent's action space restricts the set of incentive schemes:*

   (I) First, allowing for *free disposal*: implies that the slope of the incentive scheme must be non-negative. That is, if the agent can "destroy" output q, then we must have $I'(\cdot) \geq 1$. This implies monotonicity.

   (II) Second, allowing the agent to borrow q with no restrictions (from a third party, say a bank) implies that $I'(\cdot) \leq 1$ (or else, the agent will borrow, present a higher output to the principal, get more than his loan and repay the loan at a profit, assuming negligible interest rates for very short loans.) Therefore, the principal's share must increase in i.

So, in remark 3 we saw that $I'(\cdot) \geq 0$, and $I'(\cdot) \leq 1$ must hold somewhere, and with a realistic enrichment of the agent's action space we get these conditions holding everywhere.

6. *Random incentive schemes don't help.* This is straightforward: If $\tilde{I}_i$ is the random income that the agent faces after q_i is realized, define $\tilde{v}_i = v(\tilde{I}_i)$ and $\overline{v}_i = E\tilde{v}_i$ so that $\overline{I}_i = h(\overline{v}_i)$ is the *certainty equivalent* of $\tilde{I}_i$. Now have the principal offer $\{\overline{I}_1, ..., \overline{I}_n\}$ instead of $\{\tilde{I}_1, ..., \tilde{I}_n\}$. This contract has no effect on (IC) or (IR), and since $\overline{I}_i = h(\overline{v}_i) < Eh(\tilde{v}_i) = E\tilde{I}_i$, then the principal implements the same action at a lower cost.

5.4 The Continuous Model

Let the utilities be specified as before, but now assume that $a \in [\underline{a}, \overline{a}]$ is a continuous, one-dimensional effort choice, and assume that q is continuous with density $f(q|a)$. (That is, the density function is conditional on the choice of a.) The principal's problem is now:

$$\begin{cases} \max\limits_{a, I(\cdot)} & \int\limits_q [q - I(q)] f(q|a) dq \\ \text{s.t.} & \int\limits_q v(I(q)) f(q, a) dq - g(a) \geq \overline{u} \qquad \text{(IR)} \\ & a \in \arg\max \left\{ \int\limits_q v(I(q)) f(q, a') dq - g(a'),\ a' \in A \right\} \quad \text{(IC)} \end{cases}$$

This is the general (and *correct*) way of writing the problem. However, this is not a form that we can do much with. As we will see, there is a simple way of reducing this problem to a "manageable" program, but this will require some extra assumptions for the solution to be correct. We begin by analyzing the first-best benchmark.

5.4.1 First Best Benchmark: Verifiable Actions

As before, in this case we only have (IR), which will bind at a solution, so the Lagrangian is,

$$\mathcal{L} = \int_q [q - I(q) + \lambda v(I(q))] f(q|a) dq - \lambda g(a) - \lambda \overline{u}$$

Assuming interior solution, the (point-wise) FOC with respect to $I(\cdot)$ yields,

$$\frac{1}{v'(I(q))} = \lambda \;\forall q. \tag{5.4}$$

Denote by $f_a \equiv \frac{\partial f(q|a)}{\partial a}$, then the FOC with respect to a yields,

$$\int_q [q - I(q) + \lambda v(I(q))] f_a(q|a) dq = \lambda g'(a) \tag{5.5}$$

and the (IR) constraint will bind (the usual argument.)

1. The first FOC with respect to. $I(\cdot)$ is known as the Borch rule where optimal risk sharing occurs. If the principal were not risk neutral but would rather have some risk averse utility function $u(\cdot)$, then the numerator of the LHS of the FOC (5.4) above would be $u'(q - I(q)]$.

2. Notice that the Borch rule is satisfied for all q and not on average since this is what risk sharing is all about.

5.4.2 Second Best: non-observable action

The First Order Approach:

In the hidden information framework we saw that under some conditions we can replace global incentive compatibility with local incentive compatibility. The question is, can we restrict attention to local incentive compatibility in the moral hazard framework?

This is known as the *first order approach,* an approach that was popular in the 70's until Mirrlees (1975) showed that it is flawed, unless we impose additional restrictions on $f(\cdot|\cdot)$. The first order approach simplifies the problem by replacing (IC) above with the agent's FOC of his optimization problem, that is, of choosing his optimal action $a \in A$. The agents FOC condition is,

$$\int_q v(I(q)) f_a(q|a) dq - g'(a) = 0 \qquad ((\text{IC}^F))$$

We proceed to solve the principal's problem subject to (IC^F) and (IR), and later we will check to see if the agents SOC is satisfied, that is, if

$$\int_q v(I(q)) f_{aa}(q|a) dq - g''(a) \leq 0 \,.$$

Also, we will have to check for global IC, a condition for which the FOC and SOC of the agent are neither necessary nor sufficient.

The Lagrangian of the principal's problem is,

$$\begin{aligned} \mathcal{L} &= \int_q [q - I(q)] f(q|a) dq + \lambda \left[\int v(I(q)) f(q|a) dq - g(a) - \overline{u} \right] \\ &\quad + \mu \left[\int v(I(q)) f_a(q|a) dq - g'(a) \right] \end{aligned}$$

and maximizing with respect to $I(\cdot)$ point-wise yields the FOC,

$$\frac{1}{v'(I(q))} = \lambda + \mu \frac{f_a(q|a)}{f(q|a)} \quad \text{a.e.} \qquad (5.6)$$

Notice that this looks very similar to the FOC for the case of A being finite, with only one (downward) binding IC. (This is also like the 2-action case with $a^*_{SB} = a_H$.) In the previous formulation we had $h'(\cdot) = \frac{1}{v'(\cdot)}$, and $\frac{f_a}{f}$ is "similar" to the continuous version of the likelihood ratio for small changes in a, which is $\frac{f(q|a+\delta)}{f(q|a)}$. This can be seen as follows: Notice that $\frac{f(q|a+\delta)}{f(q|a)} - 1$ is a monotonic transformation of $\frac{f(q|a+\delta)}{f(q|a)}$, and dividing this by δ is yet another monotonic transformation, which then yields,

$$\lim_{\delta \to 0} \frac{f(q|a+\delta) - f(q|a)}{\delta f(q|a)} = \frac{f_a(q|a)}{f(q|a)} \,.$$

Definition 4.2: $f(q|a)$ satisfies MLRC if $\frac{f_a(q|a)}{f(q|a)}$ increases in q.

Proposition 4.4: Assume that the first order approach is valid, that MLRC is satisfied, and that $\mu > 0$. Then $I'(q) \geq 0$.

Proof: Follows directly from the FOC (**??**) that we derived above and from $v'' < 0$: As q increases, $\mu \frac{fa}{f}$ increases, $\Rightarrow \frac{1}{v'}$ increases, $\Rightarrow v'$ decreases, $\Rightarrow I(q)$ increases. *Q.E.D.*

Therefore, in addition to assuming MLRC, and that the first order approach is valid, we need to ask ourselves when is $\mu > 0$ guaranteed? This is answer is that if the first order approach is valid, then we must have $\mu > 0$. This is demonstrated in the following proposition (Holmstrom (1979) proposition 1):

Proposition 4.5: If the first order approach is valid, then at the optimum, $\mu > 0$.

Proof: (i) Assume in negation that $\mu < 0$. If the first order approach is valid, then from the FOC (**??**) above, we get the solution $I^*(q)$ which is decreasing in $\frac{fa}{f}$. Define $\widehat{I} = I(q)$ for those q which satisfy $\frac{f_a(q|a)}{f(q|a)} = 0$. That is, since I is a function of q, and given a each q determines $\frac{f_a(q|a)}{f(q|a)}$, then we can think of I as a function of $\frac{f_a(q|a)}{f(q|a)}$, s shown in the following figure:

Figure Here

Now consider the first term of the agent's FOC (IC^F) above:

$$\begin{aligned} &\int_q v(I(q))f_a(q|a)dq \\ = &\int_{\{q:\, f_a \geq 0\}} v(I(q))f_a(q|a)dq + \int_{\{q:\, f_a < 0\}} v(I(q))f_a(q|a)dq \\ < &\int_{\{q|fa \geq 0\}} v(\hat{I})f_a(q|a)dq + \int_{\{q:\, f_a < 0\}} v(\hat{I})f_a(q|a)dq \\ &= v(\hat{I}) \int_q f_a(q|a)dq \\ &= 0 \end{aligned}$$

(The last equality follows from $\int_q f(q|a)dq = 1 \ \forall a$.) But this contradicts

$$\int_q v(I(q))f_a(q|a)dq = g'(a) > 0.$$

(ii) Assume in negation that $\mu = 0$. From the FOC (**??**) above we get,

$$\frac{1}{v'(I(q))} = \lambda \ \forall q \,,$$

which implies that the agent is not exposed to risk. This in turn implies that the agent chooses his action to minimize his cost $g(a)$, which is generally not the solution to the principal's program. *Q.E.D.*

Caveat: It turns out that we could have $\mu = 0$ and the FB is almost achieved. The following example is due to Mirrlees (1974):

Example 4.1: Assume that output is distributed according to $q = a + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$. That is, for two distinct actions $a_1 < a_2$, we get the distributions of q to "shift" as shown in the following figure:

Figure Here

Assume also that $a \in [\underline{a}, \overline{a}]$, and that the agents utility is given by

$$U(I, a) = \ell n I - a \,.$$

This looks like a "well behaved" problem but it turns out that the principal can achieve profits that are arbitrarily close to the FB profits: Let $a^*_{FB} \in (\underline{a}, \overline{a})$, and calculated $\overline{I}$ such that

$$\ell n \overline{I} - a^*_{FB} = \overline{u}.$$

Set $\underline{q}$ to be "very" negative, and let the incentive scheme be

$$I(q) = \begin{cases} \overline{I} > \underline{q} \\ \delta \leq \underline{q} \end{cases}$$

where δ is "very" small. Mirrlees showed that

$$\lim_{\underline{q} \to -\infty} F(\underline{q}) = \lim_{\underline{q} \to -\infty} \Pr\{q < \underline{q}\} = 0,$$

but that

$$\lim_{\underline{q}\to-\infty} F_a(\underline{q}) = \lim_{\underline{q}\to-\infty} \frac{d}{da}\left(\Pr\{q < \overline{q}\}\right) \neq 0,$$

which implies that the agent will not "slack," his expected utility is close to $\overline{u}$, and he is almost not exposed to risk (To get (IR) satisfied, we need to add ε to $\overline{I}$, which measures the departure from FB costs to the principal.) This example is a continuous approximation of the "shifting support" discrete case.

1. Again, this looks like a statistical inference problem but it is not. (Notice that $\frac{f_a}{a}$ is the derivative of the log-likelihood function, $\ell n f(q,a)$, with respect to a. This turns out to be the gradient for the Maximum Likelihood Estimator of a given q, which is the best way to infer a from the "data" q, if this *were* a statistical inference problem.)
2. Holmstrom (1979, section 6) shows that this approach is valid when the agent has private information (i.e., a type) as in the hidden-information models. The wage schedule will then be $I(q,\theta)$, where $\theta \in \Theta$ is the agent's type. Now the multiplier will be $\mu(\hat{\theta})$ and it may be negative at a solution.
3. Intuitively we may think that $a^*_{SB} < a^*_{FB}$. This is not a generic property.

Validity of the First Order Approach

Mirrlees (1974) observed that the solution to the "Relaxed program," i.e., using the first order approach, *may not be* the true SB solution to the correct unrelaxed program. The problem is more serious than the standard one, in which case the FOC's (of the principal's program, not of the agent's program) being necessary but not sufficient. It turns out that the first order approach suffers from a much deeper problem; it may be the case that when the first order approach is used, then the principal's FOC is not only insufficient, but it *may not be necessary.*

This can be seen using the following graphical interpretation of incentive schemes.

2 figures here

Given $I(\cdot)$, the agent maximizes

$$EU(a|I) = \int_q v(I(q))f(q|a)dq - g(a)\,.$$

Then, using the first order approach, we maximize the principal's profits subject to the constraint,

$$\frac{dEU(a|I(\cdot))}{da} = 0\,,$$

that is, given an incentive scheme $I(\cdot)$, the agent's FOC determines his choice of a. But notice that $EU(a|I(\cdot))$ need *not be* concave in a, which implies that there may be more than one local maximizer to the agent's problem (we ignore the problem of no solutions.) To see the problem, imagine that we can "order" the space of incentive schemes $I(\cdot)$ using the ordering "<"as follows: Consider a scheme $\tilde{I}$ such that for all schemes $I < \tilde{I}$, there is only one local maximizer for $Eu(a|I)$. Also consider a scheme $\hat{I}$ such that for all schemes $I > \hat{I}$ there is only one local maximizer. Assume that $\tilde{I}$ and $\hat{I}$ have one inflection point, (a point which is not a local maximizer or minimizer, but has the FOC of the agent satisfied,) so that all schemes I such that $\tilde{I} < I < \hat{I}$ have two local maximizers, and one local minimizer. In particular, consider some scheme $\overline{I}(\cdot)$, $\hat{I}(\cdot) < \overline{I}(\cdot) < I$, such that the agent is indifferent between the two local maximizers $\overline{a}_0$ and $\overline{a}_1$, but the principal prefers $\overline{a}_1$. Then, it may be optimal for principal to move to $\langle I^*(\cdot), a_1^*\rangle$ at which both the principal's *and* agent's FOC's are satisfied. But under $I^*(\cdot)$, the action a_1^* is a local maximizer, while a_0^* is the *global maximizer*, in which case $\langle I^*(\cdot), a_1^*\rangle$ is not implementable. When we solve the *true program* we maximize the principal's profits subject to the agents *true* choices, so given the graphical description, $\left\langle \overline{I}(\cdot), \overline{a}_1\right\rangle$ is the solution to the true program, and at this point the principal's FOC in the reduced first order approach program is not even satisfied.

There are two ways of dealing with this problem. The first, is to ignore it and check if the solution to the first order approach is a true solution to the SB problem, and if it is not, then the true problem needs to be solved. The second, is to find cases for which the first order approach is valid. We will explore two such cases:

Case 1: LDFC: Let $a \in A = [0,1]$. We say that the *Linear Distribution Function Condition* (LDFC) is satisfied if there exist two density functions, $f_H(q)$, and $f_L(q)$, such that

$$f(q|a) = af_H(q) + (1-a)f_L(q)$$

(this is also called the *Spanning Condition* - see G-H 1983.) In this case

$$EU(a|I(\cdot)) = \int v(I(q))f(q|a)dq - g(a)$$

$$= a\int v(I(q))f_H(q)dq + (1-a)\int v(I(q))f_L(q)dq - g(a)$$

$$= ak_1 + (1-a)k_2 - g(a)$$

where k_1 and k_2 are constants (given the incentive scheme $I(\cdot)$,)and since $g' > 0$ and $g'' > 0$, this is a well behaved concave function which guarantees that there is no problem, and LDFC is a sufficient condition for the first order approach to work.

Case 1: MLRC + CDFC: We say that the cumulative distribution function $F(q|a)$ satisfies the Convexity of the Distribution Function Condition (CDFC) if it is convex in a; that is, for all $\lambda \in [0,1]$,

$$F(q|\lambda a + (1-\lambda)a') \leq \lambda F(q|a) + (1-\lambda)F(q|a')\,.$$

Recall that MLRC implies that $\frac{f_a(q|a)}{f(q|a)}$ increases in q. Take the agent's expected utility and integrate it by parts:

$$EU(a|I(\cdot)) = \int_{\underline{q}}^{\overline{q}} v(I(q))f(q|a)dq$$

$$= [v(I(q))\cdot F(q|a)|_{\underline{q}}^{\overline{q}} - \int_{\underline{q}}^{\overline{q}} v'(I(q))I'(q)F(q|a)dq$$

$$= v(I(\overline{q})) - \int_{\underline{q}}^{\overline{q}} v'(I(q))I'(q)F(q|a)dq - g(a)$$

where the last equality follows from $F(\overline{q}|a) = 1$ and $F(\underline{q}|a) = 0$ (note that $\underline{q}$ or $\overline{q}$ need not be bounded.) Now consider the second derivative of $EU(a|I(\cdot))$ with respect to a:

$$[EU(a|I(\cdot))]'' = -\int_{\underline{q}}^{\overline{q}} v'(I(q))F_{aa}(q|a)dq - g''(a) < 0$$

which follows from $v' > 0$, $I'(q) > 0$ (which is implied by MLRC,) $F_{aa} > 0$ (which is implied by CDFC,) and finally $g'' > 0$. Thus, MLRC and CDFC together are sufficient for the first order approach to be valid.

1. Recall that in our proof that MLRC $\Rightarrow I'(\cdot) > 0$ we used the fact that $\mu > 0$. But this fact is true only if the first order approach is valid. Thus, the proof we have given (for MLRC and CDFC together to be sufficient for the first order approach to be valid) contains a "circular" mistake. See Rogerson (1985) for a complete and correct proof.
2. Jewitt (1988) provides alternative sufficient conditions for validity of the first order approach which avoids CDFC (CDFC turns out to be very restrictive), and puts conditions on $v(\cdot)$, namely that $-\frac{v''(\cdot)}{[v'(\cdot)]^3}$ is non-decreasing, that is, risk aversion does not decrease too quickly. CARA is an example that works with various "commonly used" distributions.

5.5 The Value of Information: The Sufficient-Statistic Result

Assume now that there are more signals above and beyond q. For example, let y denote some other parameter of the project (say, some intermediate measure of success) and assume that q and y are jointly distributed given the agent's action a.

Question: When should y be part of the contract in addition to q?

This question was answered independently by Holmstrom (1979) and Shavel (1979). Rewrite the Lagrangian (assuming that the first order approach is valid) by taking (q, y) to be the two-dimensional signal, and $I(q, y)$

is the incentive scheme:

$$\begin{aligned}\mathcal{L} &= \int\limits_{y}\int\limits_{q}[q - I(q,y)]f(q,y|a)dqdy \\ &\quad +\lambda\left[\int\limits_{y}\int\limits_{q} v(I(q,y))f(q,y|a)dqdy - g(a) - \overline{u}\right] \\ &\quad +\mu\left[\int\limits_{y}\int\limits_{q} v(I(q,y)f_a(q,y|a)dqdy - g'(a)\right]\end{aligned}$$

and the FOC with respect to $(I\cdot,\cdot)$ is:

$$\frac{1}{v'(I(q,y))} = \lambda + \mu\frac{f_a(q,y|a)}{f(q,y|a)}$$

Now, to answer the question, we can modify it and ask, "When can we ignore y?" The answer is clearly that we can ignore y and have $I(q)$ if and only if the first order condition is independent of y, which gives us the same FOC as in our analysis without the y signal. This will be satisfied if $\frac{f_a(q,y|a)}{f(q,y|a)}$ is independent of y.

Definition 4.3: q is a *Sufficient Statistic for* (q,y) *with respect to* $a \in A$ if and only if the conditional density of (q,y) is multiplicative separable in y and a:

$$f(q,y|a) = g(q,y)\cdot h(q|a)\,.$$

We say that y is informative about a if q is not a sufficient statistic as defined above.

This is a statistical property which says that when we want to make an inference about the random variable. $\tilde{a}$, if q is a sufficient statistic as above then we can ignore y. (Again, remember that here a is *known* in equilibrium, but we have the same flavor of a statistical inference problem.) We thus have the following proposition:

Proposition 4.6: Assume that the first order approach is valid. y should be included in the incentive scheme if and only if y is informative about a.

Consider the case where q is a sufficient statistic for (q, y) as defined above. We can interpret this result as follows. Given a choice $a \in A$, we can think of q being a r.v. whose distribution is dependent on a, and then, once q is realized (but maybe not yet revealed,) then y is a r.v. whose distribution is dependent on q. This can be depicted using the following "causality" diagram,

$$a \overset{\widetilde{\varepsilon}}{\rightarrow} q \overset{\widetilde{\eta}}{\rightarrow} y\,,$$

that is, given a, some random shock $\widetilde{\varepsilon}$ determines q, and given the realized value of q, some random shock $\widetilde{\eta}$, which is independent of a, determines y. Thus, y is a noisier signal of a compared to q, or we say that y is a *garbling* of q.

Corollary 4.3: If q is a sufficient statistic then, if the principal is restricted to contract on one signal then $I(q)$ is better than $I(y)$ for the principal.

Corollary 4.4: Random compensation schemes are not optimal (for the separable utility case.)

The second corollary follows since a random incentive scheme $\widetilde{I}(x)$ is a payment based on a r.v. y which is independent of a, making q a sufficient statistic. (If the agent's utility isn't separable in q and a, then the agent's risk attitude depends on a, and randomizations of the incentive scheme may be beneficial for the principal.)

5.6 Incentives in Teams: Group Production

We now explore the situation in which several agents together produce some output. The following model is based on section 2 in Holmstrom (1982a):

- Consider a group of n agents, each choosing an action $a_i \in A_i \subset \Re$, for $i \in \{1, 2, ..., n\}$.
- Output is given by $x(a_1, a_2, ..., a_n, \varepsilon) \in \Re$, where ε is some random noise. For now we will perform the analysis with no noise, that is, set $\varepsilon = 0$. We assume that $\frac{\partial x}{\partial a_i} > 0$ $\forall i$, and $\forall a_i$, that is, output is increasing in each agent's action (effort.) Finally, assume that $x(\cdot)$ is concave so that we can restrict attention to interior solutions.

- Agent's utilities are given by $u_i(m_i, a_i) = m_i - g_i(a_i)$ where m_i denotes monetary income, and $g_i(a_i)$ is the private cost of effort, with $g_i' > 0$, and $g_i'' > 0$. (Note that agents are risk neutral in money.) Let $\overline{u}_i$ be outside option (which determines each agent's IR).

5.6.1 First Best: The Planner's Problem

Assume that there are no incentive problems, and that a planner can force agents to choose a particular action (so we are also ignoring IR constraints.) The first best solution maximizes total surplus, and solves,

$$\max_{a_1,...,a_n} x(a_1, ..., a_n) - \sum_{i=1}^{n} g_i(a_i)$$

which yields the FOCs,

$$\frac{\partial x(a)}{\partial a_i} = g_i'(a_i) \;\; \forall\, i = 1, ..., n\,. \tag{5.7}$$

That is, the marginal benefit from agent i's action equals the marginal private cost of agent i.

5.6.2 Second Best: The "Partnership" Problem

Consider the partnership problem where the agents jointly own the output. Assume that the actions are not contractible and the agents must resort to an incentive scheme $\{s_i(x)\}_{i=1}^{n}$ under the restriction of a balanced budget ("split-the-pie") rule:

$$\sum_{i=1}^{n} s_i(x) = x \;\; \forall\, x\,, \tag{5.8}$$

and we also impose a "limited liability" restriction,

$$s_i(x) \geq 0 \;\; \forall\, x\,. \tag{5.9}$$

(We will also assume that the $s_i(x)$ functions are differentiable in x. This is not necessary and we will comment on this later.)

We solve the partnership problem using Nash Equilibrium (NE), and it is easy to see that given an incentive scheme $\{s_i(x)\}_{i=1}^{n}$,any NE must satisfy,

$$s_i'(x) \cdot \frac{\partial x}{\partial a_i} = g_i'(a_i)$$

Question: Can the Partnership achieve FB.?

For the partnership to achieve FB efficiency, we must find an incentive scheme for which the NE coincides with the FB solution. That is, from the FOC of the FB problem, (5.7) above, we must have $s_i'(x) \equiv 1 \; \forall \, x$, and $\forall \, i$, which implies that

$$\sum_{i=1}^{n} s_i'(x) = n.$$

But from (5.8) we have

$$\sum_{i=1}^{n} s_i'(x) = 1 \; \forall \, x \, ,$$

which implies that a budget balanced partnership cannot achieve FB efficiency. The intuition is standard, and is related to the "Free riding" problem common to such problems of externalities.

Question: How can we solve this? (in the deterministic case)

One solution is to violate the budget balanced constraint, (5.8) above, by letting $\sum s_i(x) < x$ for some levels of output x.

Example 4.1 Consider the following incentive scheme:

$$s_i(x) = \begin{cases} s_i^* \text{ if } x = x_{FB}^* \\ 0 \text{ if } x \neq x_{FB}^* \end{cases}$$

where s_i^* is arbitrarily chosen to satisfy: $\sum_{i=1}^{n} s_i^* = x_{FB}^*$, and $s_i^* > g_i(a_i^*) \; \forall \, i$. (This can be done if we assume that $\overline{u}_i = 0 \;\; \forall i$.) It is easy to see that this scheme will yield the FB as a NE for the case where $g_i(0) = 0 \; \forall \, i$, $x(0, ..., 0) = 0$, and $\sum_{i=1}^{n} g_i'(0) < \sum_{i=1}^{n} \frac{\partial x(0,...,0)}{\partial a_i}$. (these are just Inada conditions that guarantee the solution.) □

1. One problem with the scheme above is that there are multiple NE. For example, $a_i = 0 \;\; \forall \, i$ is also a NE given the scheme above.
2. A second problem is that this scheme is not credible (or, more precisely, not renegotiation proof.) If one agent "slacks" and $x < x^*$ is realized, then all agents have an incentive to renegotiate. That is, in the scheme above, $\sum_{i=1}^{n} s_i < 0$ is not *ex-post efficient* off the equilibrium path. Thus, with renegotiation we cannot achieve FB efficiency for partnerships.

3. As mentioned above, we do not need $s_i(x)$ to be differentiable in x. See the appendix in Holmstrom 1982a. (That is, the inability to achieve FB is more general.)

5.6.3 A solution: Budget-Breaker Principal

We now introduce a new, $(n+1)^{th}$ agent to the partnership who will play the role of the "Budget-Breaker." This is a theoretical foundation for the famous paper by Alchian and Demsetz (1972). The idea in Alchian and Demsetz is that if we introduce a monitor to the partnership problem, then this monitor (or "principal") will make sure that the agents do not free ride. The question then is, who monitors the monitor? Alchian and Demsetz argue that if the monitor has residual claims, then there will be no need to monitor him. One should note, however, that there is *no monitoring* here in Holmstrom's model. This is just another case of an "unproductive" principal who helps to solve the partnership problem.

Example 4.2: Consider a modification of Example 4.1 where we add a principal (budget breaker), denoted by $n+1$, and modify the incentive scheme as follows:

$$s_i(x) = \begin{cases} s_i^* \text{ if } x \geq x_{FB}^* \\ 0 \text{ if } x < x_{FB}^* \end{cases} \quad \text{for } i = 1, ..., n;$$

$$s_{n+1}(x) = \begin{cases} x - x_{FB}^* & \text{if } x \geq x^* \\ x & \text{if } x < x^* \end{cases}$$

In equilibrium,

$$\sum_{i=1}^{n+1} s_i(x) = x_{FB}^*,$$

and $s_{n+1}(x_{FB}^*) = 0.$ □

Add Uncertainty: $\varepsilon \neq 0$

If agents are risk neutral, then it is easy to extend the previous analysis to see that a partnership cannot achieve FB efficiency in a NE. (As before, we will get under-provision of effort in any NE.) It turns out that adding a Budget-Breaker will help achieve FB efficiency as the following analysis shows. Denoted the principal by $n+1$, and consider the incentive scheme,

$$\begin{aligned} s_i(x) &= x-k \ \forall\, i=1,...,n \\ s_{n+1}(x) &= nk-(n-1)x \end{aligned}$$

where k is chosen so that

$$(n-1)\int_x xf(x,a^*_{FB})dx = nk\,,$$

which guarantees that at the FB solution, the $(n+1)^{th}$ agent breaks even in expectation. Under this scheme, each agent $i=1,...,n$ solves,

$$\max_{a_i} E[x(a_1,...,a_n)]-k-g(a_i)\,,$$

which yields the FOC,

$$\frac{\partial Ex(a_1,...,a_n)}{\partial a_i} = g_i'(a_i)\ ,$$

which is precisely the FB solution.

The intuition is simple: This is exactly like a *Groves Mechanism.* Each agent captures the full extent of the externality since the principal "sells" the *entire* firm to each agent for the price k.

1. If the principal is risk-averse we cannot achieve FB efficiency as demonstrated above in the case with uncertainty. This follows since the principal will be exposed to some risk, and since he breaks even in the sense of expected utility then there is a reduction in total expected social surplus.

2. A problem in this setup is *collusion*: Agent i can go to the principal and say: "If I choose a_i^* you get zero, so we can split x after I choose $a_i=0$." This is very different from the problem of renegotiation proofness in the partnership problem. For collusion to be a problem we must have "secret" side-contracts between the principal and some agent. These issues are dealt with in the collusion literature.

3. Legros-Matthews (1993) extended Holmstrom's results, and they established necessary and sufficient conditions for a Partnership (with no budget-breaker) to achieve FB. Some interesting cases in their analysis are:

   Case 1: A_i finite and $x(\cdot)$ being a generic function. In this case if only one agent deviates then it will be clear who it was, so we can achieve FB without then problem of renegotiation. For example, consider the incentive scheme (ignoring IR constraints,)

   $$s_i(x) = \begin{cases} \frac{x}{n} & \text{if } x = x^* \\ \frac{1}{n-1}(F + x) & \text{if } x \neq x^* \text{ and } j \neq i \text{ deviated} \\ -F & \text{if } x \neq x^* \text{ and } i \text{ deviated} \end{cases}$$

   (In fact, it is enough to know *who didn't deviate* to do something similar.)

   Case 2: $A_i = [\underline{a_i}, \overline{a_i}] \subset \Re$ is compact, and $a^* \in (\underline{a_i}, \overline{a_i})$ (i.e., an interior FB solution). In this case we can have one agent, say i, randomize between choosing a_i^* with high probability and some other action with low probability, and all other agents will a_j^* for sure. We can now achieve actions that are arbitrarily close to the FB solution with appropriate schemes. As $\Pr\{a_i = a_i^*\} \to 1$, we approach the FB solution. Note, however, that there are two criticisms to this case: First, as $\Pr\{a_i = a_i^*\} \to 1$, we need fines for "bad" outcomes that approach $-\infty$ to support the FB choice of actions, and second, do we really think that agents randomize?

5.6.4 Relative Performance Evaluations

This is the second part of Holmstrom (1982a). The model is setup as follows:

- Risk neutral principal

- n risk averse agents, each with utility over income/action as before, $u_i(m_i, a_i) = u_i(m_i) - g_i(a_i)$, with $u' > 0,\ u'' < 0,\ g' > 0$, and $g'' > 0$.

- $y = (y_1(a),\ y_2(a), ..., y_m(a))$ is a vector of random variables (e.g., outputs, or other signals) which is dependent on the vector of actions, $a = (a_1, a_2, ..., a_n)$.

- $E(x|y,a)$ denotes the principal's expected profit given the actions and the signals. (Thus, we can think of y as signals which are interdependent through the a_i's and some noise, and x maybe part of y.)

- $G(y,a)$ denotes the distribution of y as a function of a with $g(y,a)$ being the density.

The principal's problem is therefore,

$$\left\{\begin{array}{lll}
\max\limits_{a,s_1(y),...,s_n(y)} & \int\limits_y \left[E(x|y,a) - \sum\limits_{i=1}^n s_i(y)\right] dG(y,a) & \\
\text{s.t.} & \int\limits_y u_i(s_i(y))dG(y,a) - g_i(a_i) \geq \overline{u}_i \ \ \forall\ i & \text{(IR)} \\
 & a_i \in \arg\max\limits_{a_i' \in A_i} \int\limits_y u_i(s_i(y))dG(y,(a_i',a_{-i})) - g_i(a_i') \ \ \forall\, i & \text{(IC)}
\end{array}\right.$$

(Note that if x is part of y then $E(x|y,a) \equiv x$.)

Definition 4.4: A function $T_i(y)$ is a sufficient statistic for y with respect to a_i if there exist functions $h_i(\cdot) \geq 0$ and $p_i(\cdot) \geq 0$ such that,

$$g(y,a) = h_i(y,a_{-i})p_i(T_i(y),a) \ \ \forall\,(y,a) \in \text{support}(g(\cdot,\cdot))$$

The vector $T(y) = (T_1(y),...,T_n(y))$ is sufficient for y with respect to a if each $T_i(y)$ is sufficient for y with respect to a_i.

This is just an extended version of the sufficient statistic definition we saw for the case of one agent. For example, if each $T_i(y)$ is sufficient for y with respect to a_i, we can intuitively think of this situation as one where each a_i generates a random variable $T_i(y)$, and y is just a garbling of the vector of random variables, $T(y) = (T_1(y),...,T_n(y))$, or as the figure describes the process,

$$\left.\begin{array}{ccc}
a_1 & \overset{noise}{\longrightarrow} & T_1(y) \\
\vdots & & \vdots \\
a_n & \overset{noise}{\longrightarrow} & T_n(y)
\end{array}\right\} \overset{noise}{\longrightarrow} y$$

Proposition 4.7: (Holmstrom, Theorem 5) Assume $T(y) = (T_1(y), ..., T_n(y))$ is sufficient for y with respect to a. Then, given any collection of incentive schemes $\{s_i(y)\}_{i=1}^n$, there exists a set of schemes $\{\widetilde{s}_i(T_i(y))\}_{i=1}^n$, that weakly Pareto dominates $\{s_i(y)\}_{i=1}^n$

Proof: Let $\{s_i(y)\}_{i=1}^n$ implement the Nash equilibrium $(a_1, ..., a_n)$, and consider changing i's scheme from $s_i(y)$ to $\tilde{s}_i(T_i)$ as defined by:

$$\begin{aligned} u_i(\tilde{s}_i(T_i)) \equiv & \int\limits_{\{y:T_i(y)=T_i\}} u_i(s_i(y))\frac{1}{p_i(T_i, a)} \cdot g(y, a)dy \\ = & \int\limits_{\{y:T_i(y)=T_i\}} u_i(s_i(y))h_i(y, a_{-i})dy \end{aligned}$$

By definition, (IC_i) and (IR_i) are not changed since agent i's expected utility given his choice a_i is unchanged. Also,

$$\tilde{s}_i(T_i) \leq \int\limits_{\{y:T_i(y)=T_i} s_i(y) \cdot h_i(y, a_{-i})dy\,,$$

because $u'' < 0$, and T_i is constant whereas $s_i(y)$ is random *given* T_i. Integrating over T_i :

$$\int\limits_y \tilde{s}_i(T_i(y))g(y, a)dy \leq \int\limits_y s_i(y)g(y, a)dy\,.$$

This can be done for each $i = 1, ..., n$ while setting the actions of $j \neq i$ as given (to preserve the NE solution), and by offering $\{\tilde{s}_i(T_i)\}_{i=1}^n$ the principal will implement $(a_1, ..., a_n)$ at a (weakly) lower cost. *Q.E.D.*

The intuition is the same as for single agent in Holmstrom (1979): The collection $(T_1, ..., T_n)$ gives better information than y. Thus we can think of y as a garbling (or even a mean-preserving spread) of the vector $T(y)$.

Yet again, this looks like a statistical inference problem, but it is not; it is an equilibrium problem. It turns out that the "mechanics" of optimal incentives look like the mechanics of optimal inference. This is like "reverse engineering"; we use to the statistical inference properties that a would cause on $T(y)$, and use incentive based on these properties to make sure that the agents will choose the "correct" a.

Application: Yardstick Competition

We return to the simple case of $x = y$, so that the signal is profits, but consider the restricted case in which:

$$x(a,\theta) = \sum_{i=1}^{n} x_i(a_i,\theta_i).$$

That is, total profits equal the sum of individual profits generated by each agent individually, and each agent's profits are a function of his effort and some individual noise θ_i, where θ_i, θ_j may be correlated.

Proposition 4.8: If θ_i and θ_j are independent for all $i \neq j$, and x_i is increasing in θ_i, then $\{s_i(x_i)\}_{i=1}^{n}$ is the optimal form of the incentive schemes.

Proof: Let $f_i(x_i, a_i)$ be the density of x_i given a_i. Then define:

$$\begin{aligned} g(x,a) &= \prod_{j=i}^{n} f_j(x_1, a_1)\ , \\ p_i(x_i, a) &= f_i(x_i, a_i)\,, \\ h_i(x, a_{-i}) &= \prod_{j\neq i} f_j(x_j, a_j)\,, \end{aligned}$$

and apply Proposition 4.7. *Q.E.D.*

Now consider a different scenario: Let $x_i = a_i + \varepsilon_i + \eta$, where

$$\varepsilon_i \sim N(0, \frac{1}{\tau_i})\ \ \forall\, i = 1, ..., n$$

is an idiosyncratic noise that is independent across the agents, and

$$\eta \sim N(0, \frac{1}{\tau_0})$$

is some common shock. Therefore, we don't have independence as in the previous paragraph. (Note, we use the notion of *precision*, which is expressed by the τ_i's. These are the inverse of variance, $\tau_i \equiv \frac{1}{\sigma_i^2}$; the more precision a signal has, the less is its variance.)

Proposition 4.9: Let

$$\alpha_i = \frac{\tau_i}{\sum_{j=1}^n \tau_j}\,, \quad \forall\, i = 1, ..., n\,,$$

$$\overline{x} = \sum_{i=1}^n \alpha_i x_i$$

then, in this scenario, $s_i(x_i, \overline{x})$ is the optimal form of the incentive scheme.

Proof: We prove this using the sufficient statistic result. Since $x_i = a_i + \eta + \varepsilon_i$ then $\varepsilon_i = x_i - a_i - \eta$, and we can write:

$$F(\hat{x}_1, ..., \hat{x}_n, a) =$$

$$k \int\limits_{-\infty}^{\infty} \left[\underbrace{\int\limits_{-\infty}^{\hat{x}_1 - a_1 - \eta} e^{-\frac{1}{2}\tau_1\varepsilon_1^2} d\varepsilon_1}_{I_1} \cdot \underbrace{\int\limits_{-\infty}^{\hat{x}_2 - a_2 - \eta} e^{-\frac{1}{2}\tau_2\varepsilon_2^2} d\varepsilon_2 ...}_{I_2} \cdots \underbrace{\int\limits_{-\infty}^{\hat{x}_n - a_n - \eta} e^{1\frac{1}{2}\tau_n\varepsilon_n^2} d\varepsilon_n}_{I_n} \right] e^{-\frac{1}{2}\tau_0\eta^2} d\eta$$

each of the inner integrals, I_i, can be written as:

$$I_i = \int\limits_{-\infty}^{\hat{x}_i} e^{-\frac{1}{2}\tau_i(x_i - a_i - \eta)^2} dx_i$$

To obtain $f(\hat{x}_1, ..., \hat{x}_n, a)$ we need to partially differentiate $F(\cdot)$ with respect to $\hat{x}_i$, for $i = 1, ..., n$ sequentially, which yields:

$$f(x, a) = \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\sum_{j=1}^n \tau_j(x_j - a_j - \eta)^2 + \tau_0\eta^2\right]} d\eta$$

Let $\overline{\tau}_{-i} = \sum\limits_{j\neq i} \tau_j$ and $\overline{z}_{-i} = \sum\limits_{j\neq i} \frac{\tau_j}{\overline{\tau}_{-i}}(x_j - a_j)$ and note that:

$$\sum_{j=1}^{n} \tau_j (x_j - a_j - \eta)^2$$

$$= \sum_{j\neq i} \tau_j [\overbrace{(x_j - a_j - \overline{z}_{-i})}^{A_j} + \overbrace{(\overline{z}_{-i} - \eta)}^{B}]^2 + \tau_i (x_i - a_i - \eta)^2$$

$$= \sum_{j\neq i} \tau_j (x_j - a_j - \overline{z}_{-i})^2 + \sum_{j\neq i} \tau_j (\overline{z}_{-i} - \eta)^2 + \tau_i (x_i - a_i - \eta)^2 - \overbrace{\sum_{j\neq i} \tau_j 2 A_j B}^{=0}$$

where the last term is equal to zero because B is independent of j so, $\sum\limits_{j\neq i} \tau_j 2 A_j B = 2B \sum\limits_{j\neq i} \tau_j A_j$, and it is easy to check that $\sum\limits_{j\neq i} \tau_j A_j = 0$. So we have,

$$f(x,a) = \int\limits_{-\infty}^{\infty} e^{\frac{1}{2}[\sum\limits_{j\neq i}(x_j - a_j - \overline{z}_{-i})^2} \cdot e^{-\frac{1}{2}[\sum\limits_{j\neq i}(\overline{z}_{-i} - \eta)^2 + \tau_i (x_i - a_i - \eta)^2 + \tau_0 \eta^2]} d\eta$$

$$= \underbrace{e^{\frac{1}{2}[\sum\limits_{j\neq i}(x_j - a_j - \overline{z}_{-i})^2]}}_{h(x, a_{-i})} \cdot \underbrace{\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}[\sum\limits_{j\neq i}(\overline{z}_{-i} - \eta)^2 + \tau_i (x_i - a_i - \eta)^2 + \tau_0 \eta^2]} d\eta}_{\hat{p}_i(\overline{z}_{-i}, x_i, a_i, \eta)}$$

Letting $\overline{\tau} \equiv \sum\limits_{j=1}^{n} \tau_j$, we can write,

$$\overline{z}_{-i} = \frac{1}{\overline{\tau}_{-i}} \sum_{j\neq i} \tau_j x_j - \sum_{j\neq i} \frac{\tau_j a_j}{\overline{\tau}_{-j}}$$

$$= \frac{\overline{\tau}}{\overline{\tau}_{-i} \cdot \overline{\tau}} \left[\left(\sum_{j=1}^{n} \tau_j x_j \right) - \tau_i x_i \right] - \sum_{j\neq i} \frac{\tau_j a_j}{\overline{\tau}_{-i}}$$

$$= \frac{\overline{\tau}\, \overline{x} - \tau_i x_i}{\overline{\tau}_{-i}} - \sum_{j\neq i} \frac{\tau_j a_j}{\overline{\tau}_{-i}}$$

So we can rewrite $\hat{p}_i(\overline{z}_{-i}, x_i, a_i, \eta)$ as $p_i(\overline{x}, x_i, a, \eta)$ and we can now apply the sufficient statistic result from Proposition 4.7. *Q.E.D.*

(Note: having the ε_i's with mean is not crucial, we can have $\varepsilon_i = \mu_i + \varepsilon_i'$, where $\varepsilon_i' \sim N(0, \frac{1}{\tau_i})$ which gives us any mean we wish.)

>From the two scenarios we saw, we can conclude that competition (via rank-order tournaments, e.g., $s(x_i, \overline{x})$ as the incentive scheme) is not useful per-se but only as a way of getting more information. With independent shocks relative performance schedules only add noise and reduce the principal's profits. Thus, if we believe that Mutual fund managers face some common shock as well as an idiosyncratic shock, then we should see "yardstick competition." In a firm, however, if two divisions have unrelated technology (no common shock), then the division managers should not have relative evaluation schemes.

1. There is a literature on tournaments, e.g., Lazear-Rosen (1981), which showed that rank-order tournaments can increase effort [also, Green-Stokey (1983), Nalebuff-Stiglitz (1983)]. However, from Holmstrom we see that only under very restrictive conditions will rank-order tournaments be *optimal* (for this, ordinal rankings need to be a sufficient statistic.)

2. In all the multi-agent literature it is assumed that the principal chooses her "preferred" Nash Equilibrium if there are several. Mookherjee RES 1984 demonstrated that there may be a NE that is better for the agents. Consider the following example: $i \in \{1, 2\}$, $a_i \in \{a_L, a_m, a_H\}$, $g_i(a_L) < g_i(a_m) < g_i(a_H)$, and $x_i = a_i + \eta$. (This is an "extreme" case of a common shock with no individual shock.) Assume that (a_m, a_m) is FB optimal choice for the principal. The principal can implement the FB with a rank-order tournament:

$$s_i = \begin{cases} u_i^{-1}(g_i(a_m) + \overline{u}) & \text{if } x_i \geq x_j \\ u_i^{-1}(g_i(a_L) + \overline{u} - \delta) & \text{if } x_i < x_j \end{cases}$$

It is easy to see that (a_m, a_m) is a NE implemented at the FB cost (agent's IR is binding). Notice, however, that (a_L, a_L) is also a NE, and the agents strictly prefer this because then,

$$u_i = g_i(a_m) + \overline{u} - g_i(a_L) > \overline{u}\,.$$

(We can overcome this problem if we can find two payments, $s_{\text{win}} > s_{\text{tie}}$, such that

$$g_i(a_m) - g_i(a_L) \leq u_i^{-1}(s_{\text{win}}) - u_i^{-1}(s_{\text{tie}}) \leq g_i(a_H) - g(a_m)$$

and we can use the scheme $s_{\text{win}} > s_{\text{tie}} > s_{\text{lose}}$.)

3. Ma (1988) suggested a way of getting around the multiple NE problem: use an *indirect mechanism* and employ subgame-perfect-implementation (can also use Nash implementation with integer-games à la Maskin):

Stage 1	:	players choose (a_1, a_2) which is observable to them but not to the principal
Stage 2	:	player 1 can "protest" (p) or "not protest" (np)
Stage 3	:	x_i's realized, s_i's played to agents.

   where,

   $$s_i(x_1, x_2, \text{np}) = \begin{cases} u_i^{-1}(g_i(a_m) + \overline{u} + \gamma) & \text{if } x_i > x_j \\ u_i^{-1}(g_i(a_m) + \overline{u}) & \text{if } x_i = x_j \\ u_i^{-1}(g_i(a_L) + \overline{u} - \delta) & \text{if } x_i < x_j \end{cases}$$

   $$s_1(x_1, x_2, \text{p}) = s_1(x_1, x_2, \text{np}) + \overbrace{\alpha[x_2 - E(x_2|a_2 = a_m)]}^{>0 \text{ iff } a_2 = a_H}(\alpha \text{ small})$$
   $$s_2(x_1, x_2, \text{p}) = u_2^{-1}(g_2(a_L) + \overline{u} - \beta), \ (\beta \text{ large})$$

   So, player 1's best response (BR) to $a_2 \in \{a_L, a_m\}$ is "NP", and $a_1 = a_m$ would be the BR ex-ante. Look at Normal Form, and it is easy to see that $((a_m, \text{"}NP\text{"}), a_m)$ is the unique NE (also SPE).

   figure here

   Note: It is important that (a_1, a_2) are observable by (at least) agent 1. If not, we can't use the standard implementation approach and can't get FB (see Ma).

5.7 Dynamic Models

5.7.1 Long-Term Agency Relationship

In the Adverse-Selection (hidden information) models we saw that long-term commitment contracts will do better than short-term contracts (or Long-Term renegotiable contracts).

Stylized facts: Sequences of short-term contracts are common, e.g., piece-rates for laborers and sales commissions for sales representatives. That is, many schemes pay for per period performance.

Question: When will Short-Term contracts be as good as Long-Term?

This question is addressed by Fudenberg-Holmstrom-Milgrom (1990) (FHM hereafter.) We will outline a simplified version of FHM. Consider a 2-period relationship. In each period $t \in \{1, 2\}$ we have the sequence of events as described by the following time line:

Figure here

The *technology* is given by the following distributions: In period t, the agent exerts effort (action) a_t, and the output, x_t, is distributed according to $x_1 \sim F_1(x_1|a_1)$, and $x_2 \sim F_2(x_2|a_1, a_2, x_1, \sigma_1)$, where σ_1 is a signal observed by the agent. That is, second period technology can depend on all previous variables.

1. No discounting (no interest on money)
2. Agent has all the bargaining power. (Note that this will cause our program to look different with respect to individual rationality, but the essence of the moral hazard problem is unchanged, and we still need an incentive constraint for the agent. Now, however, the individual rationality constraint will be for the principal.)

Given the dynamic nature of the problem, let $a = (a_1, a_2(a_1, \sigma_1, x_1))$ be an *action plan* for the agent that has his second period action dependent on all first period observables. (These are observables to him, not necessarily to the principal.) Let $c = (c_1(\sigma_1, x_1), c_2(a_1, \sigma_1, \sigma_2, x_1, x_2))$ be a (contingent) *consumption plan* for the agent.

Assumption A1: x_t and s_t are observable and verifiable (contractible)

(This is also A1 in FHM.) Assumption A1 implies that a payment plan (incentive scheme) will take the form $s = (s_1(x_1), s_2(x_1, x_2))$.

We allow the agent's utility function to take on the most general form,

$$U(a_1, a_2, c_1, c_2, \sigma_1, \sigma_2),$$

and the principal's utility is given by (undiscounted) profits,

$$\pi = x_1 - s_1 + x_2 - s_2 \,.$$

A *Long-Term-Contract (LTC)* is a triplet, $\Delta = (a, c, s)$, where (a, c) are the agent's "suggested" plans, and s is the payment plan. The agent's expected utility, and the principal's expected profits *given* a LTC Δ are:

$$\begin{aligned} U(\Delta) &= E[U(a, c, \sigma)|a] \\ \pi(\Delta) &= E[x_1 - s_1(x_1) + x_2 - s_2(x_1, x_2)|a] \end{aligned}$$

Definition 4.5: We say that a LTC Δ is:

1. *Incentive Compatible* (IC) if $(a, c) \in \arg\max_{\hat{a},\hat{c}} E[U(\hat{a}, \hat{c}, \sigma)|\hat{a}]$
2. *Efficient* if it is (IC) and if there is no $\tilde{\Delta}$ such that $\pi(\tilde{\Delta}) \geq \pi(\Delta)$ and $U(\tilde{\Delta}) \geq U(\Delta)$ with at least one strict inequality.
3. *Sequentially Incentive Compatible* (SIC) if given any history of the first period, the continuation of Δ is IC in the second period.
4. *Sequentially Efficient* if it is (SIC), and, given any history of the first period, the continuation of Δ is efficient in the second period.

(*Note:* (3) and (4) in the definition above need to be formally defined with continuation utilities and profits. This is done in FHM, but since the idea is quite clear we will skip the formalities.)

We will now set up a series of assumptions that will guarantee that ST contracts (to be defined) will do as well as LTC's: (The numbering of the assumptions are as in FHM).

Assumption A3: The agent and the principal *have equal access to banks* between periods at the competitive market rate ($\delta = 1$).

Implication: the agent has a budget constraint:

$$c_1 + c_2 = s_1(x_1) + s_2(x_1, x_2).$$

Assumption A4: At the beginning of each period t, there is *common knowledge of technology.*

Implication: $F(x_2|a_1, a_2, x_1, \sigma_1) = F(x_2|a_2, x_1)$, or, (x_1, a_2) is a sufficient statistic for $(x_1, a_2, a_1, \sigma_1)$ with respect to x_2. That is, information provided by x_1 is sufficient to determine how a_2 affects x_2. A simple example of this assumption is the common time-separable case: $F_2(x_2|a_2)$, which implies that the periods are independent in technology.

Assumption A5: At the beginning of each period there is *common knowledge of the agent's preferences* over the continuation plans of any (a, c, s).

Implications:

1. $U(a, c, \sigma)$ cannot depend on σ (that is, there is no hidden information, or "types", in period 2.)

2. The continuation plan at $t = 2$ cannot depend on a_1.

3. The continuation plan at $t = 2$ cannot depend on c_1, if c_1 is not observable to the principal.

For simplicity we will assume that the agents utility function is given by (this satisfies A5):

$$U(\cdot, \cdot) = v_1(c_1) - g_1(a_1) + v_2(c_2) - g_2(a_2)$$

with the standard signs of the derivatives, $v_t^{'} > 0$, $v_t^{''} < 0$, $g_t^{'} > 0$, and $g_t^{''} > 0$. Thus, from now on we can ignore σ as if it did not exist (to satisfy A4 and A5.)

The importance of A4 and A5 is that they guarantee no "adverse selection" at the negotiation stage in period $t = 2$. (That is, at the re-contracting stage of any LTC.) We will later see examples of violations of these two assumptions, and the problems that are caused by these violations.

Given any continuation of a SIC contract, let $UPS(a_t, x_t)$ denote the *utility possibility set* given the history (a_t, x_t). (For a_0, x_0 this is not history dependent since there is no history before $t = 1$.) Let $\pi = UPF(u|a_1, x_1)$ denote the *frontier* of the UPS.

Assumption A6: For every (a_t, x_t), the function $UPF(u|a_t, x_t)$ is strictly decreasing in u.

Implication: The full set of incentives can be provided by efficient contracts (If the frontier were not strictly decreasing, we cannot keep the agent's utility fixed at u' and move to an efficient point.)

2**Figures here**

We now turn to the analysis of LTC's and STC's.

Definition 4.6: We say that a LTC Δ is *optimal* if it solves,

$$\begin{cases} \max\limits_{s,a,c} & E_{x_1,x_2}[v_1(c_1(x_1)) - g_1(a_1) + v_2(c_2(x_1,x_2)) - g_2(a_2)|a] \\ \text{s.t.} & (a,c) \in \arg\max\limits_{a_1,a_2,c_1,c_2} E_{x_1,x_2}[v_1(c_1)) - g_1(a_1) + v_2(c_2) - g_2(a_2)|a] \quad (\text{IC}_A) \\ & s_1(x_1) + s_2(x_1,x_2) = c_1(x_1) + c_2(x_1,x_2) \qquad (\text{BC}_A) \\ & E_{x_1x_2}[x_1 - s_1(x_1) + x_2 - s_2(x_1,x_2)|a] = 0 \qquad (\text{IR}_P) \end{cases}$$

Notes:

1. (IR_P) must bind because of A6 (downward sloping UPF)
2. A6 and (IR_P) binding imply that if the solution is optimal then it must be efficient.
3. We restrict attention to choices as functions only of the x_t's, since we ignore the σ's to satisfy A4 and A5.

Short-Term Contracts

A sequence of *short-term contracts* (STC) will specify one contract, $\Delta_1 = (a_1, c_1(x_1), s_1(x_1))$, at $t = 1$, and given a *realization* of x_1, the parties will specify a second contract, $\Delta_2 = (a_2, c_2(x_2), s_2(x_2))$, at $t = 2$.

Fact: Since Δ_2 depends on the realization of x_1, if the parties are rational, and A4-A5 are satisfied, then parties can *foresee the contract* Δ_2 for every realization of x_1.

This fact is straightforward, and it implies that for convenience, we can think of Δ as a complete contingent plan, $\Delta = (\Delta_1, \Delta_2)$, where,

$$\begin{aligned} \Delta_1 &= (a_1, c_1(x_1), s_1(x_1))\,, \\ \Delta_2 &= (a_2(x_1), c_2(x_1,x_2), s_2(x_1,x_2))\;. \end{aligned}$$

Question: What is the difference between such a complete contingent plan and a LTC?

The answer is that in a LTC, the agent *can commit* to Δ_2 ex-ante, whereas with STC's this is impossible. However, players can foresee Δ_2 that will arise in a sequence of STC's.

To solve for the *optimal sequence* of STC's we will work backward (i.e., use dynamic programming.) At $t = 2$, for every (x_1, s_1, c_1) the agent finds the optimal Δ_2^* by solving:

$$\begin{cases} \max\limits_{\{a_2(x_1),c_2(x_1,\cdot)s_2(x_1,\cdot)\}} & E_{x_2}[v_2(c_2(x_1,x_2)) - g_2(a_2(x_1))|a_2(x_1)] \\ \text{s.t.} & a_2(x_1), c_2(x_1,\cdot) \in \arg\max\limits_{a_2,c_2} E_{x_2}[v_2(c_2) - g_2(a_2))|a_2] \quad (\text{IC}_A^2) \\ & s_1 + s_2(x_1,\cdot) = c_1 + c_2(x_1,\cdot) \quad (\text{BC}_A) \\ & E_{x_2}[x_2 - s_2(x_1,x_2)|a_2(x_1)] = 0 \quad (\text{IR}_P^2) \end{cases}$$

Since the agent has perfect foresight at date $t = 1$, then anticipating $c_2(x_1,\cdot), a_2(x_1,\cdot)$ correctly he finds the optimal Δ_1^* by solving:

$$\begin{cases} \max\limits_{\Delta=(s,c,a)} & E_{x_1,x_2}[v_1(c_1(x_1)) - g_1(a_1) + v_2(c_2(x_1,x_2)) - g_2(a_2(x_1))|a] \\ \text{s.t.} & (a_1, c_1(x_1)) \in \arg\max\limits_{a_1,c_1} E_{x_1x_2}[v_1(c_1) - g_1(a_1) + v_2(c_2(x_1,x_2)) - g_2(a_2(x_1))|a] \quad (\text{IC}_A^1) \\ & E_{x_1}[x_1 - s_1(x_1)|a_1] = 0 \quad (\text{IR}_p^1) \end{cases}$$

Notes:

1. The agent's budget constraint, (BC_A), is only relevant at $t = 2$.
2. (IR_P^1) depends only on $x_1, s_1(x_1)$, since perfect foresight implies that $E[\pi_2] = 0$. (This need not hold for LTC's since the principal must break even over the whole relationship, and may have positive expected profits in one period for some histories, and negative expected profits in the other cases.)

3. Expectations about Δ_2^* are correct.

Proposition 4.10: (Theorem 2 in FHM) Under (A1) [verifiability], (A4) and (A5) [common knowledge of technology and preferences] and (A6) [decreasing UPF] any *efficient* LTC can be replaced by a *sequentially efficient* LTC which provides the same initial expected utility and profit levels.

Proof: Suppose Δ is an efficient LTC that is not sequentially efficient. $\Rightarrow \exists\ \hat{x}_1$ such that the continuation $a_2(\hat{x}_1), c_2(\hat{x}_1, x_2), s_2(\hat{x}_1, x_2)$ is not efficient. $\Rightarrow \exists\ \Delta_2'$ (a different continuation contract) which Pareto dominates Δ_2 after $\hat{x}_1$. Now, (A6) $\Rightarrow \exists\ \Delta_2'$ that gives the same expected continuation utility to the agent with a higher expected profit to the principal.

Figure Here

Construct a new LTC $\widetilde{\Delta}$ s.t. $\widetilde{\Delta}_1 = \Delta_1$, and,

$$\tilde{\Delta}_2(x_1) = \begin{cases} \Delta_2'(x_1) & \text{if } x_1 = \hat{x}_1 \\ \Delta_2(x_1) & \text{if } x_1 \neq x_1 \end{cases}$$

i.e., the same as Δ, but with continuation Δ' after $\hat{x}_1$. $\Rightarrow$ the agent's continuation utilities are unchanged, $\Rightarrow$(IC$_A$) is preserved, and this is common knowledge from (A4) and (A5). The principal strictly prefers the new continuation contract so (IR$_P$) is preserved since the principal is ex-ante weakly better off. (If she were strictly better off then Δ could not have been efficient.) So, $\tilde{\Delta}$ is sequentially efficient and gives the same U_0, π_0. *Q.E.D.*

Intuition: Just like complete Arrow-Debreu contingent markets.

Note: The reason the principal is only weakly better off (i.e., indifferent) is that the efficient LTC can be not sequentially efficient only for zero-probability histories. (e.g., continuous output space with a finite number of such histories, or histories that are off the equilibrium path.)

Question: What is the difference between an optimal sequentially efficient LTC and a series of optimal STC's?

In the series of optimal STC's the principal's rents are zero in *every period,* regardless of the history, whereas along the "play" of an optimal sequentially efficient LTC the principal may have different expected continuation profits for different histories at times $t \neq 1$.

Definition 4.7: A sequentially efficient LTC which gives the principal zero expected profits conditional on any history is called *sequentially optimal.*

Observation: If Δ is a sequentially optimal LTC, then its per-period continuation contracts constitute a sequence of optimal STC's.

Proposition 4.11: (Theorem 3 in FHM) Under the assumptions of Proposition 4.10, plus assumption A3 (equal access to banking,) then any optimal LTC can be replaced by a sequence of optimal STC's.

Proof: From Proposition 4.10 we know that there exists a sequentially efficient LTC that replaces the optimal LTC, so we are left to construct a sequentially efficient LTC with zero-expected profits for the principal at time 2, for any history of the first period. Given a sequentially efficient contract Δ, define:

$$\pi_2(x_1) = E_{x_2}[x_2 - s_2(x_1, x_2)|a_2(x_2)]$$

and assume that for some histories $\pi_2(x_1) \neq 0$ (if $\pi_2(x_1) = 0$ for all x_1, then we are done.) Using A3, we can define a new contract $\widehat{\Delta}$, such that

$$\begin{aligned} \hat{s}_1(x_1) &= s_1(x_1) - \pi_2(x_1), \\ \hat{s}_2(x_1, x_2) &= s_2(x_1, x_2) + \pi_2(x_1)\ \forall\,(x_1, x_2)\,, \end{aligned}$$

and leave $c_t(\cdot)$'s and $a_2(\cdot)$ unchanged. By construction we have the following four conditions:

1. (BC$_A$) is satisfied (both for the LTC $\widehat{\Delta}$ at $t = 1$, and for the resulting STC's at $t = 2$.)
2. For all x_1 :

$$\begin{aligned} E_{x_2}[x_2 - \hat{s}_2(x_1, x_2)|a_2(x_1)] &= E_{x_2}[x_2 - s_1(x_1, x_2) - \pi_2(x_1)|a_2(x_1)] \\ &= \pi_2(x_1) - \pi_2(x_1) \\ &= 0. \end{aligned}$$

3. The agent's incentives are unchanged since the $c_t(\cdot)$'s are unchanged. (And, they need not be changed since (BC$_A$) is satisfied.)

4. ex ante we have,

$$\begin{aligned} E_{x_1}[x_1 - \hat{s}_1(x_1)|a_1] &= E_{x_1}[x_1 - s_1(x_1) + \pi_2(x_1)|a_1] \\ &= E_{x_1x_2}[x_1 - x_1(x_1) + x_2 - s_2(x_1,x_2)|a_1, a_2(x_1)] \\ &= 0 \end{aligned}$$

   where the last equality follows from Δ being an optimal sequentially efficient LTC.

But notice that (1)-(4) above imply that $\widehat{\Delta}$ is sequentially optimal. *Q.E.D.*

Violating the Assumptions:

Case 1: Consumption not observable. This violates A5, that is, there is now no common knowledge of the agent's preferences at $t = 2$. This will cause a violation of Proposition 4.11, and an example is given in FHM, example 2.

Case 2: Agent cannot access bank. This is the case in Rogerson (1985). Restrict $s_t \equiv c_t$, so that the agent cannot borrow or save. In this case the agent would optimally like to "smooth" his consumption across time, so the principal is performing two tasks: First, she is giving the agent incentives in each period, and second, she is acting as a "bank" to smooth the agent's consumption. Rogerson looks at a stationary model and shows that under these assumptions optimal LTC's have $s_2(x_1, x_2)$ and not $s_2(x_2)$, that is, memory "matters". The intuition goes as follows: If for a larger x_1 the principal wants to give larger compensation, then both $s_1(x_1)$ and $s_2(x_1, \cdot)$ should rise. We need LTC's to *commit* to this incentives scheme, because with STC's the principal cannot commit to increase $s_2(\cdot)$ when x_1 is larger. Rogerson also looks at consumption paths given different conditions on the agent's utility function.

Case 3: No common knowledge of tecnology. That is,

$$x_2 \sim F_2(x_2|x_1, a_1, a_2)$$

in which case the agent's action in the first period affects the second period's technology. Consider the simple case where $x_1 = a_2 \equiv 0$, and $a_1 \in \{a_L, a_H\}$ with $g(a_H) > g(a_L)$.

Figure Here

Suppose that $a_1 = a_H$ is the optimal second-best choice the principal wants to implement, and at the time of negotiation 2 (or *renegotiation,*) the principal does not observe a_1. The optimal LTC is one in which $s(x_2)$ has the agent exposed to some risk, so that he has incentives to choose a_H. Note, however, that the optimal sequence of STC's is as follows: At $t = 1$, s_1 is a constant (there is nothing to condition on). At $t = 2$, s_2 must be a constant as well, which follows from the fact that the effort was *already taken,* and efficiency dictates that the principal should bear all the risk (*ex-post efficient* renegotiation.) This implies that if renegotiation is possible *after* effort has been exerted, but *before* outcome has been realized, then the *only ex-post credible* (RNP) contract has $s(x)$ being a constant, and no incentives are provided. This is exactly what Fudenberg and Tirole (1990) analyze. The optimal LTC of the standard second best scenario is not sequentially efficient, or, is not RNP. This implies that with renegotiation we cannot have the agent choosing a_H with probability 1. The solution is as follows: The agent chooses a_H with probability $p_H < 1$, and a_L with probability $1 - p_H$. At the renegotiation stage the principal plays the role of a monopolistic insurer (à la Stiglitz), and a menu of contracts is offered so that it is RNP ex-post (at the stage of negotiation 2 in the figure above). This is demonstrated by the following figure:

Figure Here

Other Results on Renegotiation in Agency

Hermalin and Katz (1991) look at following case (and more...):

Figure Here

The optimal LTC without renegotiation is the standard second-best $s(x)$. Any sequence of STC's has the same problem as Fudenberg and Tirole (1990). Hermalin and Katz consider a combination of ex-ante LTC in which renegotiation occurs *on* the equilibrium path. In the case where the agent's action is observable and verifiable (in contrast to some signal, y, of the action a) then FB is achieved with LTC and renegotiation. The procedure goes as follows:

1. At $t = 0$ the principal offers the agent a risky incentive scheme, $s(x)$, that implements a^*_{FB}. That is, the solution to the first stage in the Grossman-Hart decomposed process, which is *not* the SB standard contract but rather the lowest cost contract to implement a^*_{FB}.

2. Given *any* choice $a \in A$ that the agent actually chose, at $t = 2$ the principal offers the agent a constant $\hat{s}$ that is the certainty-equivalent of $s(x)$. That is, she offers $\hat{s} = E[v(s(x)|a]$, which implies that the agent's continuation utility is unchanged, so choosing a^*_{FB} after $t = 0$ is still optimal, and there is no ex-post risk. Thus, the FB is achieved.

Hermalin and Katz (1991) also looks at a signal $y \neq a$ being observable and not verifiable. If y is a sufficient statistic for x with respect to a, then we can implement $s(y)$ in the same fashion and improve upon $s(x)$.

Question: If renegotiation outside the contract helps, does this mean that the RNP principal is not applicable here?

The answer is no. We can think of renegotiation as a bargaining game where the principal makes take-it-or-leave-it offers. Then, this process can be written into the contract as follows: The principal offers $s(x)$ and the agent accepts/rejects, then the principal offers $\hat{s}$ and agent accepts/rejects. Thus, we put renegotiation into the contract. (Note that we can alternatively have *message-game* that replicates the process of renegotiation: after $s(x)$ has been offered, and a has been chosen, both the principal and agent announce $\hat{a}$, and if their announcements coincide then $\hat{s}$ is awarded. If the announcements are different, both parties are penalized.)

Other papers: **Ma** (1991)

Figure Here

can get $a_1 = a_H$ with probability 1, but this may not be optimal...

Segal-Tadelis (1996):

Figure Here

σ_1 is a signal that is always observed by agent, while the principal may choose to observe it at no cost, or commit not to observe it. In the optimal contract the principal may choose not to observe the signal even if x_2 is not a sufficient statistic. (Intuition: create endogenous asymmetric information at the renegotiation stage.)

Matthews (1995): Looks at renegotiation where the agent has all the bargaining power and the action is unobservable. Using a forward induction argument (note that agent makes the renegotiation offers, and we are thus in a signalling game) the unique RNP contract is one where the principal "sells" the firm to the agent.

5.7.2 Agency with lots of repetition

One might expect that as a relationship is repeated more often, then there is room to achieve efficiency. Indeed, Radner (1981) and Rubinstein & Yaari (....) show that if an agency relationship is repeated for a long time then we can arbitrarily approach the FB as $T \to \infty$. The idea goes as follows: Consider the agent's utility as:

$$U = \sum_{t=1}^{T} \delta^t [v(s_t(x^t)) - g(a_t)]$$

where $x^t = (x_1, x_2, ..., x_t)$, and assume that the model is stationary with i.i.d. shocks (this is the formulation in Radner(1981)). Then, as $T \to \infty$ and $\delta \to 1$ the principal can use a "counting scheme," i.e., count the outcomes and pay "well" if the distribution over x is commensurate with a^*_{FB}, and punish the agent otherwise. (The intuition is indeed simple, but the proof is quite complicated.)

1. From Rogerson (1985) we know that if the agent has access to a bank (lending and borrowing) then the FB is attainable *without* a principal; the agent can insure himself across time as follows: choose a^*_{FB} in each period, consume x^* in each period, lend (or borrow if negative) $\ell_t = x_t - x^*$ in each period, and

$$\lim_{T \to \infty} \frac{\ell_1 + \ell_2 + ... + \ell_T}{T} = 0 \text{ a.s.}$$

(We need to assure no bankruptcy; Yaari (....) deals with this issue.)

2. We need to ask ourselves if this kind of repetition, and the underlying conditions constitute a leading case? How interesting is this result (i.e., FB achieved with or without a principal.) The answer seems to be that these conditions are rather hard to justify which puts these results in some doubt.

5.7.3 Complex Environments and Simple Contracts

What have we learned so far from agency theory with moral hazard? We can roughly summarize it in two points: First, agency trades off risk with incentives. This is a "nice" result since it allows us to understand better what these relationships entail in terms of these trade-offs.. Second, we learned that the form of an optimal SB contract can be *anything.* For example, even a simple real-world observation like monotonicity is far from general. This lesson, that the optimal SB contract may look very strange, is less attractive as a result given its real world implications. Furthermore, in a very simple example which seems to be "well behaved", Mirrlees has shown that there is no SB solution (recall from Example 4.1 above.)

If we try and perform a *reality check,* it is quite clear that strange contracts (like that in Mirrlees's example) are not observed, whereas simple linear contracts seem to be common, where the linearity is applied to some aggregate measure of output. For example,

1. Piece rate per week (not per hour);

2. Sales commission per week/month/quarter;

3. Stocks or options with clauses like "can't sell before date t", which is similar to holding some percentage of the firm's value at date t.

Holmstrom and Milgrom (1987) (H-M hereafter) make the point that "strange" and complicated contracts are due to the simplistic nature of the models. In reality, things are much more complex, and the agent has room for manipulation. Thus, we would like to have incentive schemes that are robust to these complexities. As H-M show, it turns out that linear incentive schemes with respect to aggregates will be optimal in a complex environment that has some realistic flavors to it, and are robust to small changes in the environment.

Simplified version of Holmstrom and Milgrom (1987)

Consider a principal-agent relationship that lasts for T periods:

Figure Here

We make the following assumptions:

1. Let $x_t \sim f_t(x_t|a_t)$, which implies common knowledge of technology.

2. At time t, the agent observes past outcomes $(x_1, ..., x_{t-1})$, which implies that he can "adjust" effort a_t, so that the choice of effort is a history dependent strategy:

$$a_1, a_2(x_1), ..., a_t(x_1, x_2, ..., x_{t-1}), a_T(x_1, ..., x_{T-1})$$

3. Assume zero interest rate (no discounting). This implies that the agent cares only about total payment, $s(x_1, x_2, ..., x_T)$, since he can consume at the end of the relationship.

4. Utilities are given as follows: For the principal, profits (utility) are:

$$\pi = \sum_{t=1}^{T} x_t - s(x_1, ..., x_T)\,,$$

and for the agent, utility is given by:

$$U = -e^{-r[w+s-\sum\limits_{t=1}^{T} c(a_t)]}\,,$$

where $c(a_t)$ denotes the cost of effort level a_t,and w denotes the agent's initial wealth. (Note that this is a Constant Absolute Risk Aversion (CARA) utility function, which implies that both preferences and the optimal compensation are independent of w. Therefore, w.l.o.g. assume that $w = 0$.)

5. Agent has all the bargaining power. This is not how H-M '87 proceed (they have principal with all the bargaining power) but here we will use the F-H-M '90 results to simplify the analysis.

6. For simplicity, let $x_t \in \{0, 1\}$ for all t (e.g., "sale" or "no sale").

The following proposition is analogous to Theorem 5 in H-M.

Proposition 4.12: The optimal LTC takes on the form:

$$s(x_1, ..., x_T) = \alpha \sum_{t=1}^{T} x_t + \beta.$$

i.e., compensation is linear in aggregate output.

Proof: The assumptions of F-H-M (1990) are satisfied: (A1) x's are verifiable; (A3) Same access to bank with 0 interest rate; (A4) $x_t \sim f_t(x_t|a_t)$ implies common knowledge of technology; (A5) in each period t the agent's utility is:

$$-e^{-r[-\sum_{\tau=1}^{t-1} c(a_\tau)]} \cdot e^{-r[s-\sum_{\tau=t}^{T} c(a_\tau)]},$$

where the first term is a constant that is unknown to the principal, and the second term is known to the principal. Thus, the agent's preferences are common knowledge; (A6) we clearly have a downward sloping UPF. Now, using Proposition 4.11 (Theorem 3 in F-H-M), an optimal LTC can be implemented with a sequence of STC's:

$$\{s_1(x_1), s_2(x_1, x_2), ..., s_T(x_1, ..., x_T)\} ,$$

and we can just define $s \equiv \sum_{t=1}^{T} s_t(\cdot)$. To find the optimal sequence of STC's we solve backward: At $t = T$:

$$U = -\underbrace{e^{-r\left[\sum_{t=1}^{T-1}(s_t(\cdot)-c_t(a_t))\right]}}_{\text{constant}} \cdot e^{-r[s_T(\cdot)-c_T(a_T)]}$$

Thus, what happened in periods $t = 1, ..., T-1$ is irrelevant, and the agent solves the one-shot program:

$$\begin{cases} \max\limits_{s_T(\cdot),a_T} & E_{x_T}[-e^{-r[s_T(\cdot)-c(a_T)]}|a_T] \\ \text{s.t.} & a_T \in \arg\max\limits_{a} E_{xT}[-e^{-r(s_T(\cdot)-c(a)]}|a] \\ & E[x_T - s_T(\cdot)|a_T] = 0 \end{cases}$$

and the solution is: $\langle a^*, s^*(x_T = 0), s^*(x_T = 1)\rangle$. Now move back to period $T-1$, in which the agent's utility is given by

$$U = -e^{-r[\sum_{t=1}^{T-2}(s_t - c(a_t))]} \cdot e^{-r[s_{T-1}-c(a_{T-1})]} \cdot e^{-r[s_T - c(a_T)]} .$$

When we solve the program the first term is constant, and being forward-looking the agent maximizes:

$$\max_{s_{T-1},a_{T-1}} E_{x_{T-1},x_T}[-e^{-r[s_{T-1}-c(a_{T-1})]} \cdot e^{-r[s^*(x_T)-c(a^*)]} | a_{T-1}, a_T = a^*] .$$

But since period T does not depend on the history (due to the sequence of STC's), the objective function can be rewritten as:

$$\max_{s_{T-1},a_{T-1}} E_{x_{T-1}}[-e^{-r[s_{T-1}-c(a_{T-1})]} | a_{T-1}] \cdot \underbrace{E_{x_T}[e^{-r[s^*(x_T)-c(a^*)]} | a^*]}_{\text{constant}} ,$$

and again the agent solves a one-shot program, which is the same as for T. Thus, $a_{T-1} = a^*$, $s_{T-1} = s^*(x_{T-1})$. This procedure repeats itself for all t and we get:

$$\begin{aligned} s(x_1, ..., x_T) &= \sum_{t=1}^{T} s_t(x_t) \\ = \#\{t : x_t = 1\} \cdot s^*(1) &+ \#\{t : x_t = 0\} \cdot s^*(0) \\ = (\sum_{t=1}^{T} x_t) \cdot s^*(1) &+ (T - \sum_{t=1}^{T} x_t) s^*(0) \\ = \underbrace{[s^*(1) - s^*(0)]}_{\alpha} \sum_{t=1}^{T} x_t &+ \underbrace{T s^*(0)}_{\beta} \end{aligned}$$

Q.E.D.

Intuition of Linearity: Due to the CARA utility function of the agent, and the separability, each period is like the "one shot" single period problem and we have an optimal "slope" of the one-shot incentive scheme, $s^*(1) - s^*(0)$, which is constant and will give the agent incentives to choose a^* which is time/history independent.

1. It turns out that the principal need not observe $(x_1, ..., x_T)$, but it is enough for her to observe $\sum\limits_{t=1}^{T} x_t$ (and of course, this needs to be verifiable). In Holmstrom-Milgrom this is the assumption. (That is, the principal only observes aggregates, and these are verifiable.) Two common examples are first, a **Salesman,** whose compensation only depends on total sales (not when they occurred over the measurement period), and second, a **Laborer**, whose piece-rate depends on the number of items produced per day/week, etc. (not on when they were produced during the time period.)

2. With a non-linear scheme the following will occur. Since the agent observes the past performance (and the principal does not) then he can "calibrate" his next effort level:

   Figure Here

   If the past performance up to a certain time is relatively low, then the agent will work harder to compensate for that past. Similarly, if output is relatively high, then the agent will work less. However, this is not optimal since a^* in each period is optimal. (This is true assuming a concave $s(\cdot)$ as in the figure above. If it were convex then the reverse will happen.)

Question: What happens of we *fix time,* and let $T \to \infty$?

The idea goes as follows: Take a *fixed time horizon* and increase the number of periods while making the length of each period shorter, thus keeping total time fixed.

Fact: From the Central Limit Theorem, the average of many independent random variables will be normally distributed.

Notice, that by doing the above exercise, we are practically taking the average of more and more i.i.d. random variables. (This intuitively follows from the fact that each carries less "weight" since the time period for each is shorter.)

Question: If for the situation described above, $\lim\limits_{T\to\infty} \sum\limits_{t=1}^{T} x_t$ is normally distributed, are we in the "Mirrlees example" case? i.e., can we approximate the FB?

The answer is "yes" if the agent *cannot* observe $x_1, ..., x_T$ but only $\sum x_t$. If, however, the agent observes all the x_t's the theorem we proved implies a linear scheme and the answer is "no." Thus, we can conclude that letting the agent do "more" (namely, observe outcome and calibrate effort) gives us a simple scheme, and a well defined SB solution.

The Brownian-Motion Approximation

Consider our simple model, $x_t \in \{0, 1\}$, and $a_t \in A$ affects the probability distribution of x_t. Then, fix the time length of the relationship to the interval [0,1], and let $T \to \infty$, with the length of each period, $\frac{1}{T}$, going to zero. In the limit we will get a *Brownian Motion* (one dimensional) $\{x(t),\, t \in [0, 1]\}$, with

$$dx(t) = \mu(t)dt + \sigma dB(t)\,,$$

where $\mu(t)$ is the drift rate, σ is the standard deviation, and $B(t)$ is the *standard Brownian motion* (zero drift, unit variance) and $x(t)$ is the "sum" of all the changes up until time t.

Fact: If $\mu(t) = \mu$ for all t, then $x(1) \sim N(\mu, \sigma^2)$.

That is, if the drift is constant and equal to μ, then the value of x when $t = 1$ will be normally distributed with mean μ and variance σ^2. So, we can think of this as the continuous approximation of our earlier exercise, and if agent *does not* observe $x(t)$, for all $t < 1$ then we are "stuck" in a Mirrlees case of no SB solution. But, if $x(t)$ is observed by the agent then the Holmstrom-Milgrom results hold, and we get a nice linear scheme in $x(1)$, since this normal distribution is the result of a dynamic stochastic process. (The Mirrlees problem arises only in the static contract.)

Analysis of the Brownian-Motion Approximation

We assume that the Brownian Motion model is described as follows:

- The agent controls μ (the drift) by choice of $a \in A$ where "a" stands for the constant choice over the time length [0,1] at total cost $c(a)$. This implies that $x(1) \sim N(a, \sigma^2)$. From now on we will consider $x \equiv x(1)$.
- The principal offers the contracts (In this case the agent has CARA utility, so that having the principal get all the surplus does not change the optimal SB choice a^*; only the division of surplus is changed.)

First Best: The agent will be fully insured, and the principal wants to maximize the mean of x subject to the agent's (IR). Assuming that $\overline{u} = 0$ for the agent, the objective function is,

$$\max_a a - c(a)$$

and the FOC is, $c'(a) = 1$.

Second Best: We know from our previous analysis that the optimal scheme is a linear scheme: $s(x) = \alpha x + \beta$, where $x \sim N(a, \sigma^2)$. The agent maximizes:

$$\max_a \; Ex[e^{-r(\alpha x + \beta - c(a))}|a]$$

We can simplify this by having the agent maximize his *certainty equivalent* instead of maximizing expected utility, that is, maximize

$$CE = \underbrace{\alpha a + \beta}_{\text{mean}} - \underbrace{\frac{r}{2}\alpha^2\sigma^2}_{\text{risk premium}} - c(a)$$

(Note: CARA ⇒ the risk premium is independent of the agent's income.)

Assume that $c'(\cdot) > 0,$ and $c''(\cdot) > 0$, so that the first-order approach is valid, and since the agent maximizes,

$$\max_a \; \alpha a + \beta - \frac{r}{2}\alpha^2\sigma^2 - c(a)\,,$$

and the FOC is

$$c'(a) = \alpha\,.$$

The principal will set the agent's (IR) to bind, i.e., $CE = 0$:

$$\alpha a + \beta - \frac{r}{2}\alpha^2\sigma^2 - c(a) = 0$$

or,

$$\beta = c(a) + \frac{r}{2}\alpha^2\sigma^2 - \alpha a\,.$$

We can now substitute for β into the incentive scheme,

$$\begin{aligned} s(x) &= \alpha x + \beta \\ &= \alpha x + c(a) + \frac{r}{2}\alpha^2\sigma^2 - \alpha a, \end{aligned}$$

and the principal maximizes her expected profits, $E[x - s(x)|a]$, subject to the agent's (IC). Since $E[x] = a$, the principal's problem can be written as

$$\begin{cases} \max\limits_{a,\alpha} & a - c(a) - \frac{r}{2}\alpha^2\sigma^2 \\ \text{s.t.} & c'(a) = \alpha \end{cases}$$

By substituting $c'(a)$ for α in the objective function, the principals necessary (but not sufficient) FOC with respect to a becomes,

$$1 = c'(a) + rc'(a) \cdot c''(a)\sigma^2$$

Let's assume that the SOC is satisfied, and we have a unique maximizer, in which case we are done.

Example 4.2: A simple case is when $c(a) = \frac{k}{2}a^2$, and the principal's FOC is

$$ka + rk^2a\sigma^2 = 1$$

which yields,

$$a = \frac{1}{k + rk^2\sigma^2} \quad ; \quad \alpha = \frac{1}{1 + rk\sigma^2}$$

and we get a nice closed form solution with "realistic" results as follows:

1. $c'(a) < 1 \Rightarrow$ less effort in SB relative to FB.
2. $0 < \alpha < 1 \Rightarrow$ a sharing rule that "makes sense."
3. Appealing comparative statics: $\alpha \downarrow$ and $a \downarrow$ if either: (i) $r \uparrow$ (more risk aversion) or, (ii) $\sigma^2 \uparrow$ (more exogenous variance) That is, more risk implies less effort and a "flatter" (more insured) scheme.

Remark: A nice feature of the model is that if agent owns the firm he will choose $a = a^*_{FB}$ but he will be exposed to risk (follows from CARA).

1. Complicated environment $\Rightarrow$ simple optimal contracts
2. Nice tractable model, generalizes easily to $x = (x_1, ..., x_n)$ and $a = (a_1, ..., a_n)$ vectors.

5.8 Nonverifiable Performance

5.8.1 Relational Contracts (Levin)

5.8.2 Market Pressure (Holmstrom)

5.9 Multitask Model

Holmstrom-Milgrom (1991) analyze a model in which the agent has *multiple tasks*, for example, he produces some output using a machine, and at the same time needs to care for the machine's long-term quality. Using their model, H-M '91address the following questions:

1. Why are many incentive schemes "low powered?" (i.e., "flat" wages that do not depend on some measure of output.)
2. Why are certain verifiable signals left out of the contract? (assuming that the ones left in are not sufficient statistics.)
3. Should tasks be performed "in house" or rather purchased through the market?

Using the multitask model, H-M '91 show that incentive schemes not only create risk and incentives, but also allocate the agent's efforts among the various tasks he performs.

5.9.1 The Basic Model

- A risk averse agent chooses an effort vector $t = (t_1, ..., t_n) \geq 0$ at cost $c(t) \geq 0$, where $c(y)$ is strictly convex.
- Expected gross benefits to principal is $B(t)$. (The principal is risk neutral.)
- The agent's effort t also generates a vector of information signals: $x \in \Re^k$ given by,
$$x = \mu(t) + \varepsilon\,,$$
where, $\mu : \Re^n_+ \to \Re^k$ is concave, and the noise is multi-normal, $\varepsilon \sim N(0, \Sigma)$, $0 \in \Re^k$ is the vector of zeros, and Σ is a $k \times k$ covariance matrix.

- Given wage w and action t, the agent has exponential CARA utility given by,
$$u(w,t) = e^{-r[w-c(t)]}$$

- Following Holmstrom-Milgrom (1987) we assume that this is a final stage of a Brownian Motion model, so that the optimal scheme is linear, and given by,
$$w(x) = \alpha^T x + \beta\,,$$
$$\alpha^T x = \sum_{n=1}^{k} \alpha_n x_n\,,$$
and the agent's expected utility is equal to the certainty equivalent,
$$CE = \alpha^T \mu(t) + \beta - \frac{r^2}{2}\alpha^T \Sigma \alpha - c(t)$$

(where $\alpha^T \Sigma \alpha$ is the variance of $\alpha^T \varepsilon$.)

First Best: The principal ignores (IC) and only need to compensate the agent for his effort, so the principal's program is,
$$\max_t B(t) - c(t)\,.$$

Second Best: The principal maximizes,
$$\begin{cases} \max\limits_t & B(t) - \alpha^T \mu(t) - \beta \\ \text{s.t.} & t \in \arg\max \alpha^T \mu(t) - c(t) & \text{(IC)} \\ & \alpha^T \mu(t) + \beta - \frac{r^2}{2}\alpha^T \Sigma \alpha - c(t) \geq 0 & \text{(IR)} \end{cases}$$

As before (in H-M '87) (IR) binding gives β as a function of (α, t), so we can substitute this into the objective function, and get,
$$\begin{cases} \max\limits_t B(t) - c(t) - \frac{r^2}{2}\alpha^T \Sigma \alpha \\ \text{s.t.} \quad t \in \arg\max \alpha^T \mu(t) - c(t). \quad \text{(IC)} \end{cases}$$

(Note: $B(t)$ need not be part of x. For example, $B(t)$ can be a private benefit of the principal, or due to inaccurate accounting, we can have x being an inaccurate signal of true output.)

We can introduce the following simplification: $\mu(t) = t$. This implies that $x \in \Re^n$ (one interpretation is that there is one signal per task.) This is the case of *full dimensionality.* (Note that this is not really a special case: If $\mu(t) \in \Re^m$, $m > n$, then we can "reduce" the dimensionality by some combination of signals, and if $m < n$ then we can add signals with variance of infinity.)

>From this simplification:

$$CE = \alpha^T t + \beta - \frac{r}{2}\alpha^T \Sigma \alpha - c(t)$$

and the agent's FOC's are (we assume that we get an interior solution with $t >> 0$):

$$\alpha_i = c_i(t)\ \forall\, i = 1, ..., n$$

Following the first-order approach, we can substitute these FOC's into the principal's objective function where we use these FOCs as (IC) constraints. First, note that from the agent's FOCs we have (in vector notation):

$$\alpha(t) = \nabla c(t)$$

which implies that $\nabla \alpha(t) = [c_{ij}]$ which is the $n \times n$ matrix of the second derivatives of $c(t)$. Using the Inverse Function Theorem we get $\nabla t(\alpha) = [c_{ij}]^{-1}$ which we use later to perform comparative statics on $t(\cdot)$.

The principal maximizes,

$$\max_t\ B(t) - c(t) - \frac{r}{2}\alpha(t)^T \Sigma \alpha(t)$$

and since $\alpha(t) = \nabla c(t)$ from the agent's FOC, we can write $\alpha_i(t) = c_i(t)$ for each $i = 1, ...n$ (where $c_i(t) = \frac{\partial c}{\partial t_i}$,) and we get the principal's FOCs with respect to t,

$$B_i(t) = \alpha_i(t) + r \sum_{k=1}^{n} \sum_{j=1}^{n} \alpha_j(t) \delta_{jk} C_{ki}(t)\ \forall\, i = 1, ..., n$$

or in vector form:

$$\nabla B(t) = [I + r[c_{ij}]\Sigma]\alpha\,,$$

where I is the identity matrix, and thus we have,

$$\alpha = [I + r[c_{ij}]\Sigma]^{-1} \nabla B(t)\,. \tag{5.10}$$

Assuming that $\nabla c(t)^T \Sigma \nabla c(t)$ is a convex function of t will give sufficiency of the FOCs.

Benchmark: Stochastic and Technological Independence

To simplify we assume stochastic and technological independence which is given by,

- Σ is diagonal $\Rightarrow \sigma_{ij} = 0$ if $i \neq j$, (errors are stochastically independent.)
- $[c_{ij}]$ is diagonal $\Rightarrow c_{ij} = 0$ if $i \neq j$, (technologies of the different tasks are independent.)

under these assumptions the solution to (5.10) yields,

$$\alpha_i = \frac{B_i}{1 + rc_{ii}\sigma_i^2}\ \forall\, i = 1, ..., n \tag{5.11}$$

Observe that this solution implies:

1. "commissions" for the different tasks are independent (not surprising given the independence assumptions.)
2. α_i decreases in risk (higher r or higher σ^2)
3. α_i decreases in c_{ii} (if c_{ii} is larger, then the agent is less responsive to incentives for task i.)

A Special Case

Consider the case where $n = 2$, and only action t_1 can be measured:

$$x_i = t_i + \varepsilon_i\,,$$

where $var(\varepsilon_2) = \infty$, and $0 < var(\varepsilon_1) < \infty$, that is, $0 < \sigma_1^2 < \infty$, $\sigma_2^2 = \infty$, and $\sigma_{12} = \sigma_{21} = 0$.

>From the FOC (5.11) above we have that $\alpha_2 = 0$, and (assuming an interior solution $t_1, t_2 > 0$,)

$$\alpha_1 = \frac{B_1 - \frac{B_2 C_{12}}{C_{22}}}{1 + r\sigma_1^2 (C_{11} - \frac{C_{12}^2}{C_{22}})} \tag{5.12}$$

We can now ask what happens if t_1, t_2 are complements ($c_{12} < 0$) or substitutes ($c_{12} > 0$)? (i.e., via the agents cost function.) To answer this we can look at (5.12) above, and start with $c_{12} = 0$. As we change to $c_{12} > 0$ we see that α_1 decreases, and as we change to $c_{12} < 0$, α_1 increases.

Caveat: This is a *local argument* since C_{12} is a function of (t_1, t_2), which are in turn functions of α_1 through the incentives.

We can explain the intuition for the two directions above as follows:

1. Making C_{12} positive is making the tasks *substitutes in costs* for the agent. If we want both t_1 and t_2 to be performed, and we can only give incentives to t_1 through α_1, then increasing α_1 "kills" incentives for t_2 and increases incentives for t_1. This may be undesirable. (In fact, it is undesirable around $C_{12} = 0$.)

2. With C_{12} negative, the reverse happens: an increase in α increases $t_1$and reduces c_2 so that t_2 increases. Thus, α_1 gives incentives to both tasks. This result is actually global and does not depend on the local analysis performed above.

5.9.2 Application: Effort Allocation

Consider the case in which the agent's cost of efforts is a function of the *sum* of all efforts. This is a special case in which we assume that the efforts are extreme substitutes: The agent is indifferent between which tasks he performs, as long as his *total* effort is unchanged. We simplify by assuming that there are only two tasks, and further restrict attention t the following special case:

1. $c(t_1, t_2) = c(t_1 + t_2)$

2. There exists $\bar{t} > 0$ such that $c'(\bar{t}) = 0,\ c''(t) > 0\,\forall\, t.$

Figure Here

The idea behind assumption (2) above is that people will work $\bar{t}$ without incentives, not caring how $\bar{t}$ is allocated, as long as $t_1 + t_2 = \bar{t}$. However, providing incentives will affect the choice of effort. Note that this is a somewhat unorthodox assumption, however it is not extremely unrealistic. One way to think about this is that an agent will perform some minimal amount of performance either due to the prospect of getting fired, or due to the alternative of boredom..

Missing Incentives

Now assume that we cannot measure the effort for the first task. For example, the agent can be a contractor that is remodelling the principal's house, and t_1 can be courtesy, or attention to detail. On the other hand, we assume that t_2 is (imperfectly) measurable. Using the contractor example, t_2 can be time to completion, or how close the original plan was followed.

To formalize this assume that $\mu(t_1, t_2) = \mu(t_2) \in \Re$, a one dimensional effect of both tasks, and following the previous notation, let the measurable (verifiable) signal be given by,

$$x = \mu(t_2) + \varepsilon \ ,$$

and the linear compensation scheme be

$$s(x) = \alpha_2 x + \beta$$

Assumption: $B(t_1, t_2) > 0$ and increasing in both components, and $B(0, t_2) = 0 \; \forall \, t_2$ (t_1 is "essential" for the project to have value.)

Proposition 4.13: In the above set-up, $\alpha_2 = 0$ is optimal (even if the agent is risk neutral.)

Proof: With $\alpha_2 = 0$ ($\alpha_1 = 0$ since t_1 is not measurable) then principal maximizes: $B(t_1, \bar{t} - t_1)$ and due to his indifference, the agent will accommodate any solution. In this case there is no risk, and total surplus is

$$S^* = B(t_1^*, \bar{t} - t_1^*) - c(\bar{t}).$$

If $\alpha_2 > 0$, then the agent's choices are $t_1 = 0, \ t_2 = \hat{t}_2 \neq \bar{t}$, and social surplus is

$$\overbrace{B(0, \hat{t}_2)}^{0} - \overbrace{c(\hat{t})}^{>c(\bar{t})} - \overbrace{\frac{r}{2}\alpha_2^2\sigma^2}^{\geq 0} < S^*.$$

If $\alpha_2 < 0$, then $t_2 = 0$, and $t_1 < \bar{t}$ (since $c'(t_1) < 0 = c'(\bar{t})$), and total surplus is

$$\overbrace{B(t_1, 0)}^{<B(\bar{t},0)} - \overbrace{c(t_1)}^{>c(\bar{t})} - \overbrace{\frac{r}{2}\alpha^2\sigma^2}^{\geq 0} < S^*$$

Q.E.D.

1. It is important that $B(\cdot,\cdot)$ is a "private benefit,"or else principal can "sell" the project to the agent. The house contracting example is a very nice one, as is the example of a teacher having incentives to teach children both skills of succeeding in tests (measurable), and of creative thinking (not measurable) where the parents (or local government) have a private benefit from the children's' education.
2. This result is not "robust": it relies both on $B(0,t_2) = 0$ and $C(t_1,t_2) = C(t_1+t_2)$. But, the intuition is very appealing..

Low powered incentives in firms

Williamson (1985) observed that inside firms incentives are "low-powered" (e.g., wages to workers) compared to "high-powered" incentives offered to independent contractors. Also, employees work with the principal's assets while contractors work with their own assets. Using a variant of the previous set-up, this can be explained by multi-tasking.

Assume that,

$$B(t_1,t_2) = B(t_1) + v(t_2)\,,$$

with $B' > 0$, $v' > 0$, $B'' < 0$, $v'' < 0$ and $B(0) = v(0) = 0$. We can interpret $B(t_1)$ to be the current expected profit from activity t_1, e.g., effort in production, while $v(t_2)$ is the future value of the "assets" from activity t_2, e.g. preventive maintenance, etc.

Now let t_1 be measurable with the signal

$$x = \mu(t_1) + \varepsilon_x.$$

Let the change in the asset's value be $v(t_2) + \varepsilon_v$, and it is important to assume that the actual value accrues to the owner of the asset (for example, there can be some private benefit, imperfect markets, etc.) We finally assume that ε_x and ε_v are independent shocks.

As before, the incentive scheme will be linear, and given by

$$s(x) = \alpha x + \beta\,.$$

We now consider two alternatives for the relationship: Either the principal and agent enter a *contracting relationship*, in which case the agent owns asset, or they enter an *employment relationship*, in which case the principal owns the asset.

Define,

$$\begin{aligned} \pi^1 &= \max_{t_1} B(t_1) - C(t_1)\,, \\ \pi^2 &= \max_{t_2} v(t_2) - C(t_2)\,, \\ \pi^{12} &= \max_{t_1} B(t_1) + v(\bar{t} - t_1) - C(\bar{t})\,. \end{aligned}$$

That is, π^1 is maximal total surplus when only t_1 can vary, π^2 is maximal total surplus when only t_2 can vary, and π^{12} is maximal total surplus when we can choose t_1 subject to $t_1 + t_2 = \bar{t}$. We have the following proposition:

1. If $\pi^{12} > \max\{\pi^1, \pi^2\}$ then the optimal employment contract has $\alpha = 0$.
2. If contracting is optimal then $\alpha > 0$.
3. There exist $(r, \sigma_v^2, \sigma_\varepsilon^2)$ for which employment is optimal and other parameters for which contracting is optimal.
4. If employment is optimal for some $(r, \sigma_v^2, \sigma_\varepsilon^2)$, it is also optimal for larger values. The reverse for contracting.

1. If $v(t_2)$ accrues to the principal then $\alpha > 0 \Rightarrow t_2 = 0,\ c'(t_1) = \alpha$, and total surplus is

$$B(t_1) - c(t_1) - \frac{r}{2}\alpha^2\sigma_x^2 < \pi^1 \leq \pi^{12}$$

   whereas π^{12} can be achieved with $\alpha = 0$.
2. Idea: with no incentives, agent will not care about t_1 but rather only about t_2. With $\alpha > 0$, agent still chooses the same t_2 because he gets $v(t_2)$. $\Rightarrow \alpha > 0$ is optimal.
3. This is just to say that the π's can be ordered in any way depending on the parameters.
4. Intuition: if employment is optimal then more risk aversion implies that there is a stronger case for no risk in contract, so employment must still be optimal.

In chapter 5 we will discuss some theories of ownership, and this is an interesting model tat has implication to these issues.

5.9.3 Limits on outside activities

The following observation has been made by xxx: Some employees (usually high level) have more freedom to engage in "personal business" than others (e.g., private telephone conversations, undefined lunch breaks, etc.) This is another observation to which the multitask model adds some insight. To analyze this issue consider the following modifications to the model:

- **Tasks:** there are $k+1$ tasks, $(t, t_1, ..., t_k) \in \Re_+^{k+1}$, where only the first task, t, benefits the principal:

  $$B(t, t_1, ..., t_k) = p \cdot t$$

  (e.g., some market price for an output.)

- **Agent's costs:**

$$c(t, t_1, ..., t_k) = c(t + \sum_{i=1}^{k} t_i) - \sum_{i=1}^{k} v_i(t_i)$$

where $v_i(t_i)$ is the agent's private benefit from task i, with $v' > 0$, and $v'' < 0$. (e.g., having access to non pecuniary tasks like outside phone lines, long breaks, taking care of errands during work hours, etc.)

- Every personal task can be either *completely excluded* or *completely allowed* in the contract, but no personal task can be restricted to a level (i.e., "all-or-nothing".)
- $x = \mu(t) + \varepsilon$ is the signal, $s(x) = \alpha x + \beta$ is the incentive scheme.
- The principal can choose a contract that includes $A \subset \{1, ..., k\}$ of "allowable" personal tasks, and α, β for the incentive scheme.

To solve for the optimal contract we consider a *two stage solution process* which is similar in spirit to the two stage program of Grossman and Hart (1983): For every α, find $A(\alpha)$ that is optimal, then given $A(\alpha)$ choose α optimally.

Assume that an interior solution exists, so that given (α, A) the agent's FOC's yield:

$$\begin{aligned}\alpha &= c'(t + \sum_{i=1}^{k} t_i)\,, \\ \alpha &= v'(t_i)\ \forall\, i \in A\end{aligned}$$

Given α, use the following "cost-benefit" argument to determine the tasks that should not be excluded:

$$A = \{i : v_i(t_i(\alpha)) > p \cdot t_i(\alpha)\}$$

Figure Here

To understand the idea, look at figure above. For each task i, there exists some $\hat{t}_i$ such that for $t_i < \hat{t}_i$ the private benefit to the agent is larger than $p \cdot t_i$ and for $t_i > \hat{t}$ the reverse holds. From a social-surplus point of view, if $t_i(\alpha) < \hat{t}_i$, it is better to have the agent exert $t_i(\alpha)$ into his private benefit $v_i(t_i)$ rather than in the principal's private benefit $p \cdot t_i(\alpha)$. Thus, in figure, task 1 should be allowed and task 2 should not.

We can also see that an increase in α will (weakly) cause an increase in the set A. We also get:

Proposition 4.14: Assume $t(\alpha)$ is optimal then:

1. $\alpha = \frac{p}{1+r\sigma^2 dt/d\alpha}$,
2. If measurement is easier (σ^2 decreases), or if the agent is less risk averse (r decreases), then α and $A(\alpha)$ will be larger.
3. tasks that are excluded in the FB contract are also excluded in the SB contract, but for high $r\sigma^2$ some tasks that are included in a FB contract will be excluded in a SB contract.

(For a proof see the paper.)

The part of the proposition that is most interesting is part (2): The set A gets smaller (and incentives weaker) when we have measurement problems over t. A nice application is that without measurement problems, an outside sales force (independent contractors with no exclusions) is optimal, whereas with measurement problems an inside sales force is optimal (e.g., can't sell competitor's products, etc.)

5.9.4 Allocating tasks between two agents:

This is the last section of H-M '91. The results are:

1. It is never optimal to assign two agents to the same task.
2. The principal wants "information homogeneity": the hard-to-measure tasks go to one agent and the easy-to-measure tasks go to the other.

The intuition goes as follows: (1) we don't have "team" problem if agents are assigned to different tasks; (2) We avoid the multitask problems mentioned earlier; The agent with hard-to-measure tasks gets low incentives ("insider") and the other gets high incentives ("outsider").

Part IV

Theory of the Firm

5.10 Overview

- We discussed optimal contracts between agents and principals.
- We said nothing about whether these parties are in the same firm or two firms. Why do we have firms? What determines their boundaries?

5.10.1 Neoclassical Theory

The neoclassical theory of the firm vies the firm as a "black box" associated with a technology. That is, there is a clear relationship between inputs and outputs, say, according to some cost function.

Figure Here

The firms problem is given by the profit maximization program:

$$\max_{q} p \cdot q - C(q)$$

This program will (sometimes) give us the *minimum efficiency scale* (MES), that is, that quantity for which average costs are minimized. Also, we get a supply function, $q(p)$ from the solution to the profit maximization problem.

The standard assumptions in the neoclassical approach are that average costs decrease at first (due to some economies of scale, e.g., fixed costs,) but later, average costs increase (e.g., scarce managerial talent.) Thus, the firm's size is determined by either $q(p)$, or by the MES (the later is the prediction for the case of *free entry*).

There are, clearly, some straightforward problems with this simplistic approach. First, it says nothing about the internal organization of the firm, or about agency problems, incentives, etc. Second, consider any of the two predictions regarding the size of the firm, $q(p)$ or MES. The natural question that arises is, why can't two different firms belong to the same "outfit," i.e., why can a single firm have two plants? Taking this to the extreme, why can't all the production in the world be one giant, single entity owned firm? (e.g., like a planned economy.)

The most reasonable conclusion is that this seems like a *theory of plant size* given technology, and not a theory of the firm. That is, it does not determine the boundaries of the firm, as we understand firms to be defined (mostly through ownership.)

5.10.2 Agency theory

As mentioned earlier throughout the analysis of chapters 3 and 4, these theories analyze and describe how parties should contract, or set incentives, but not if these relationships are inside the firm or between separate firms. Thus, to some extent it can shed some light on the internal organization of firms, in the sense of optimal incentives, flows of information, and contracting between parties that interact together. Note, however, that these theories say nothing about when firms should merge, or when we should see vertical integration (e.g., the buyer "buys" the seller's firm). Thus, the same main criticism that applies to the neoclassical theory of the firm applies to most of agency theory as well. One important exception is the multitasking literature. Holmstrom and Milgrom (1994) take these ideas further to explore the question of when should we have vertical integration and when not depending on the characteristics of the multitasking arrangement. We will discuss this later in this chapter.

5.10.3 Transaction Cost Economics

This theory was first developed by Coase in his seminal paper (1937). Coase made the following two claims:

1. the "boundaries" of firms may depend on technology, but this alone cannot determine them. We should think about "transaction costs", defined by the costs of performing productive activities. (In the terms of our previous chapters, these costs can be considered as losses with respect to the FB outcome.)

2. Transaction costs may differ if the transactions are carried out *within* a firm or *across* firms. This comparison determines the boundaries of firms. Owners and market forces will impose the organizational design with the lowest transaction costs.

Sources of Transaction Costs

Bounded Rationality The ideas of Bounded Rationality are most commonly attributed to Simon. The main idea is that economic agents may not be able to fully understand the consequences of their actions, or may not have correct beliefs about the future, or that legal courts cannot "understand" the descriptions of contracts. The implications of such assumptions are that it will be hard to write contracts across firms (through the market,) and there will rise a need to do things "inside" a firm, with an authority structure that overcomes the costly decisions that would otherwise be needed using a price (market) mechanism. Simon (1951) analyzes a very simple and insightful model of such a problem.

A principal hires agent in the following setup:

- $x \in X$ is a decision to be made (e.g., a certain task to be performed, etc.)
- Agent's utility: $t - c(x)$ (t is the transfer, or wage from the principal)
- Principal's utility: $R(x) - t$
- FB: $\max\limits_{x \in X} R(x) - C(x)$.

It is interesting to note that the model, and analysis offered by Simon were presented before the idea of Arrow-Debreu contingent commodities. The interesting "limitations" that drive Simon's model are the following:

1. the FB $x \in X$ is *not known* in advance. Furthermore, there is no common set of "beliefs" so that we cannot perform expected-utility contacting as we did in the previous chapters. For example, X may be "hard" to describe in a contract ex ante. (This assumption would now be regarded as one of bounded rationality due to the Arrow-Debreu framework that we use now, and that was not an available tool when Simon wrote his model.)
2. Ex post bargaining may be very costly (over which $x \in X$ should be done).

The solution offered by Simon is the use of *authority* over a set $\overline{X} \subset X$ as follows: The agent gets a fixed wage t at date 1 (before the task is needed,) and at date 2 (when the decision needs to be made) the principal tells agent which $x \in \overline{X}$ to perform. If the principal asks the agent to perform some $x \in X \backslash \overline{X}$, then the agent can "quit." If $c(x)$ is not "too variable" over $x \in \overline{X}$ then the principal will choose

$$x \in \arg\max_{x \in \overline{X}} \ R(x)$$

which is “close” to FB.

According to Simon, a firm is defined by an authority (or, employment) relationship, in contrast to a market-price relationships, in which a certain task or action, $x \in \overline{X}$ is pre-specified, and a certain price is fixed.

The most common criticism of Simon’s approach is that this is indeed a nice theory of authority, but this does not “define” a firm. That is, given a certain set of authority based, and market based relationships, which subsets of agents will belong to the same firm? (This criticism is offered by Alchian and Demsetz (1972) and by Hart (1995).)

Monitoring Costs In chapters 3and 4 we developed and analyzed models of asymmetric information. If the principal could invest in “reducing” these asymmetries, we can think of such monitoring costs as the costs of asymmetric information compared to a FB world. Alchian and Demsetz (1972) claim that the task of managers is to monitor employees, and this monitoring produces incentives for employees to work. Alchian and Demsetz (A-D hereafter) claim that it must be harder to monitor agents that are outside the firm (e.g., independent contractor), and thus A-D conclude that the boundaries of the firm affect the incentives that can be provided.

A natural question that arises as a critique of this theory is, what is the “magic” solution that creating a firm offers, so that monitoring costs are lower in the firm? Furthermore, as for the case of the neoclassical firm, why don’t we see one huge firm? (which would be the organizational form with the lowest monitoring costs.) Thus, there is no foundation to this claim of A-D, and the predictions are not too appealing in the sense of what determines the boundaries of firms.

Bounded Rationality & Opportunism Williamson (1975, 1985) discussed the important consequences of three ingredients of bilateral relationships. The three ingredients are:

1. **Bounded rationality:** If agents are boundedly rational, then we cannot have complete contracts (in the sense of a complete set of contingencies specified in advance.)

2. **Opportunism:** Economic agents will do what is best for them selfishly.

3. **Relationship Specific Investments:** Both parties can invest individually, at some cost, to improve the social surplus from their relationship.

Williamson concludes that (1)+(2)+(3) ⇒ the *Hold-up* problem.

Example: Consider a buyer with valuation v, and a seller with cost c, so that total surplus from trade is given by

$$S = v - c$$

and this can be shared between the two parties. Assume that the two parties share any gains from trade according to a 50:50 equal sharing rule (e.g., Nash Bargaining with zero as the outside option.) Assume that the seller can incur a private fixed cost $a > 0$ to reduce his cost from c to $c' < c$, and assume that

$$a > \frac{c - c'}{2} .$$

The assumption of bounded rationality (in this case it is enough to assume that only courts are boundedly rational and the agents are completely rational) implies that the parties cannot contract on a. We can perform the analysis of this example: If the seller does not incur the investment a, he gets an ex-post utility of

$$u_0 = \frac{v - c}{2}.$$

which is also his ex-ante utility. If, however, he incurs a, he gets an ex-post utility of

$$u_a = \frac{v - c'}{2} ,$$

but his ex-ante utility will be $u_a - a$, and thus the difference between these options is,

$$\begin{aligned} \Delta u &= (u_a - a) - u_0 \\ &= \frac{v - c'}{2} - a - \frac{v - c}{2} \\ &= \frac{c - c'}{2} - a < 0 \end{aligned}$$

Thus, the socially efficient investment will not take place. ■

This is a simple example of the celebrated *hold up problem*: the inability to contract on ex ante investments, together with ex post opportunism, will cause under-investment by the parties involved, so that FB efficiency is not achieved.

Solution to Hold-up: Vertical Integration. That is, if the same owner gets both rents then efficient investments will be performed. So the hold up problem is alleviated inside the firm. This highlights the benefit to vertical integration. the *costs to vertical integration* are considered to be added bureaucratic (unproductive) activities, costs of information processing, etc. [Klein- Crawford-Alchian also look at this idea]. This trade-off will determine the boundaries of firms.

Criticism: We may have hold-up problems inside firms as well as across firms. Williamson is not clear on exactly how this hold-up problem is affected by integration.

5.10.4 The Property Rights Approach

Grossman-Hart (1986) and Hart-Moore (1990) define a firm by the non-human assets that it owns or controls (loosely speaking). Thus, a firm is defined using the notion of *residual control rights.* This definition of a firm is one of the important contributions of the G-H paper, which to a large extent sets the ideological foundations for the property rights approach. To begin, we differentiate between two different contracting assumptions:

1. The world of *complete contracts* is the one that we analyzed in chapters 2-4. (Hart (1995) refers to this as "comprehensive" contracts, in order to distinguish this setup of asymmetric information in which contracts are by definition "incomplete" due to lack of verifiability for some important variables, from the setup of first-best complete contracts.) In the complete contracts setting, the notion of "ownership" does not matter since we can specify what will be done with the assets, and how payoffs will be distributed, contingent on all possible states of the world. That is, there are no "gaps" or missing provisions (if something is missing, it must be unimportant). Also, we never need to renegotiate (following the RNP principle). In fact, the possibility to renegotiate is, as we saw, a burden rather than a benefit.

2. The world of *Incomplete contracts* is one in which we cannot right contracts on all possible contingencies or actions. Ownership will determine who has the *property rights* over the asset in contingencies that were not specified by the contract. That is, the ability to exclude others from using the asset and/or decide how it should be used. In such a case, the ability to renegotiate is crucial to achieve ex-post efficiency in these unspecified states.

We can start with an extreme case of contractual incompleteness: Imagine that we cannot describe the future states ex-ante, in which case the only "contracts" that can be written are those which determine the allocation of ownership. Now, add the following ingredients:

1. ex-post (after private investments and uncertainty) parties have symmetric information.
2. ex-post parties can "renegotiate" (decide how to use assets.)
3. Parties have "unlimited wealth" (so that any gains from trade can be realized.)
4. Parties can engage in relationship specific (private) investments after the allocation of ownership, but before uncertainty resolved.

This set of assumptions yields the following set of results:

A. Always get ex-post efficiency (follows from 1, 2, and 3 above)

B. We will generally get ex-ante inefficiency (which follows from the contractual incompleteness, and 1, 2, and 4 above.)

C. Ownership affects the ex-post "bargaining game", which implies that ownership affects the ex-ante inefficiencies.

These results are the important message of the property rights approach. They imply that we will choose the ownership structure (boundaries of the firms) so as to minimize ex-ante inefficiencies (deviations from FB.) Thus, there is a clear force that creates optimal boundaries of firms: residual rights of control and renegotiation to ex post optimal outcomes.

Note that this has a flavor of the "Fundamental Transformation" of Williamson; there is ex-ante competition, but ex-post there is bilateral monopoly because of relationship specific investments.

5.11 Incomplete Contracts, Hold Up, and Property Rights

In this chapter we will build a deeper understanding of the underlying model that is commonly used to explore the hold up problem, and form the basic building blocks for the property rights approach.

5.11.1 A Simple Formal Model

Consider two assets, a_1 and a_2 and two managers, M_1 and M_2. Only M_i can enhance his productivity using a_i and he can operate it ex-post. M_2 then has the option, using a_2 to supply a "widget" to M_1 who uses a_1 to transform the widget into a final product.

Figure Here

The timing is as follows:

Figure Here

We can simplify the model by assuming first that there is no uncertainty in costs or benefits but only in the "efficient" trade. That is, we can think of a parameter of trade, say the type of widget, that can vary and will determine the value from trade.

Assumption 5.1: Due to incomplete contracts we cannot specify in advance which widget should be traded, or how assets should be used.

The idea is simple: Assume that there are N states of nature, for each state a different "type" of widget should be supplied (for efficient trade) and all others are worthless. As $N \to \infty$, with $c > 0$ cost per "clause" in a contract, no contract is "optimal."

Assumption 5.2: Parties have rational expectations with respect to future payoffs in each state, i.e., they can do Dynamic Reprogramming.

Remark: This assumption has some "tension" with the assumption of incomplete contracts. It means that the "Bounded Rationality" is on the part of courts and the legal system. Maskin and Tirole (1998) show that if this is the case the parties can write "message-game" contracts and achieve FB.

Ownership Structures: The only "contracts" that we allow the parties to write are *ownership contracts*. Namely, at the beginning, the allocation of the assets can be determined. We denote by A the assets owned by M_1, and by B the assets owned by M_2, where $A, B \subseteq \{a_1, a_2\} \cup \emptyset$, and $A = \{\{a_1, a_2\} \cup \emptyset\} \backslash B$. We will distinguish between 3 possible ownership structures:

	A owned by M_1	B owned by $M_2(B)$
No integration	a_1	a_2
Type 1 integration	a_1, a_2	$\emptyset$
Type 2 integration	$\emptyset$	a_1, a_2

Due to incomplete contracts, the ex-ante ownership structure will be important in determining incentives for investments.

Investments: Each manager has an investment as follows:

- M_2 invests e to reduce production costs. We assume that the personal cost to M_2 of investing e is equal to e.
- M_1 invests i to enhance benefit from widget. We assume that the personal cost to M_1 of investing i is equal to i.

Trade: We assume that ex post there is no asymmetric information as follows:

Assumption 5.3: Investments i and e are ex-post observable but not verifiable. Also, all preferences are common knowledge.

An immediate implication is that, if M_1 and M_2 decide to trade ex post then they will trade *efficiently* due to the Coase Theorem being satisfied expost (this follows from the parties having symmetric information at the ex post bargaining stage.) If the parties trade efficiently at a price p then ex post utilities (ignoring the "sunk costs" of investments) are:

$$\begin{aligned} \pi_i &= R(i) - p \\ \pi_2 &= p - C(e) \end{aligned}$$

and total surplus is,

$$S = R(i) - C(e)\,.$$

We make the standard assumptions to get a "well behaved" problem, $R'(i) > 0$, $R''(i) \leqslant 0$, $C'(e) < 0$, $C''(e) > 0$.

Note: We can see in what way the renegotiation is important for this model; the parties could achieve ex-post efficiency *only if* they can renegotiate freely.

An important question is what determines p? To answer this question we need to specify a *bargaining process*. We will use Nash Bargaining, which requires us to specify the "disagreement point" which is a reference point for the renegotiation. For this we describe the case of "no trade" which is the natural disagreement point for this setup.

No Trade: If M_1 and M_2 do not trade they can both go to a "general" widget market and trade there. The price of the general widget is set at $\overline{p}$ and the utilities (profits) of each party transacting in the general market are:

$$\begin{aligned}\pi_1 &= r(i,A) - \overline{p} \\ \pi_2 &= \overline{p} - c(e,B)\end{aligned}$$

Note that utilities depend on investments and on which assets are owned. For this to be "consistent," the investment must be in some form of *human capital* since we assume that it has an effect on utilities even when $A = \emptyset$ for M_1, or when $B = \emptyset$ for M_2. (Similarly investments should have an effect when either $a_1 \notin A$, or $a_2 \notin B$.) The total ex-post surplus when the parties choose no trade is:

$$S = r(i,A) - c(e,B).$$

Assumption 5.4: Total surplus under trade is always higher than that under no trade:

$$R(i) - C(i) > r(i,A) - c(e,B) \forall A, B, i.e.$$

This assumption captures the ideas of Williamson that the investments i and e are *relationship specific*, i.e., worth more inside the relationship than outside. We will also assume relationship specificity in the marginal sense:

Assumption 5.5:

$$R'(i) > r'(i,\{a_1,a_2\}) \geq r'(i,\{a_1\}) \geq r'(i,\emptyset) \;\forall\, i\,,$$

$$|C'(i)| > |c'(e,\{a_1,a_2\})| \geq |c'(e,\{a_2\})| \geq |c'(e,\emptyset)| \;\forall\, e\,.$$

1. The idea of investment in human capital is seen by the first inequality being strict. If M_1 works with all the assets, but we are in a no trade situation (i.e., M_2 will not be the one who "operates" a_2,) then $R' > r'$. Otherwise we allow for weak inequalities.
2. If we only have absolute specificity, $R - C > r - c$, but we do not have marginal relationship specificity, then the results will be different. (See Baker, Gibbons and Murphy (1997) for more on this.)
3. It is important to remember that for ex post efficiency we are assuming that all the functions, R, r, C, and c, and both investments, i and e, are observable but not verifiable so that they cannot enter into a contract.
4. We need to add some technical assumption to guarantee a "nice" interior solution: $R'(0) > 2,\ R'(\infty) < 1,\ C'(0) < -2,\ C'''(\infty) > -1$.

First Best Investment Levels

In a FB world we allow the agents to contract on all variables, so that the optimal investments must maximize ex ante total surplus,

$$\max_{i,e}\ R(i) - C(e) - i - e$$

The FOCs are,

$$\begin{aligned} R'(i^*) &= 1\,, \\ C'(e^*) &= -1. \end{aligned}$$

(or, $|C'(e^*)| = 1$) and FB investments are denoted by i^* and e^*.

Second Best Investment Levels

Recall that symmetric information and ex-post renegotiation must imply that ex-post the parties will always choose to trade. The question is, therefore, what is then the role of $r(\cdot)$ and $c(\cdot)$? The answer is that these two values must be used to determine the *disagreement point.* If M_1 owns $A \subset \{a_1, a_2\} \cup \emptyset$ and M_2 owns $B = \{\{a_1, a_2\} \cup \emptyset\} \backslash A$, then Nash Bargaining implies that they will split the *gains from trade* compared to no-trade equally between themselves. Thus, profits are:

$$\begin{aligned}\pi_1 &= r(i,A) - \overline{p} + \frac{1}{2}[(R(i) - C(e)) - (r(i,A) - c(e,B))] - i \\ \pi_2 &= \overline{p} - c(e,B) + \frac{1}{2}[(R(i) - C(e)) - (r(i,A) - c(e,B))] - e\end{aligned}$$

Note: Earlier we described the profits from trade for M_1 to be $\pi_1 = R(i) - p$, which means that we can now get the expression for the trade price,

$$p = \overline{p} + \frac{1}{2}[(R - r) - (c - C)] \,.$$

Now, given π_1 and π_2 above, we will get M_1 maximizing π_1, and the FOC is:

$$\frac{1}{2}R'(i) + \frac{1}{2}r'(i,A) = 1 \,,$$

and similarly for M_2,

$$\frac{1}{2}|C'(e)| + \frac{1}{2}|c'(e,B)| = 1$$

Proposition 5.1: Under any ownership structure, second best investments are strictly less than FB investments i^*, e^*.

Proof: Follows immediately from assumptions on the marginal specificity:

$$R'(i) > \frac{1}{2}R'(i) + \frac{1}{2}r'(i,A) = 1 \; \forall A \;,$$

which together with $R'' < 0$ implies that $i < i^*$. The same is true for e. *Q.E.D.*

The intuition is simple: There is ex post expropriation of rents which leads to a "free rider" problem: costs are absorbed in full ex ante, but gains are split 50:50 ex post.

Lemma 5.1: Transferring assets from M_i to M_j, $j \neq i$, weakly increases the investment of M_j and reduces the investment of M_i.

Proof: Consider w.l.o.g. transferring an asset to M_1 from $M_{2.}$ $\Rightarrow A \subset \tilde{A}$ and $\tilde{B} \subset B$. Consider M_1, whose FOC implies that under assets A,

$$\frac{1}{2}R'(i) + \frac{1}{2}r'(i,A) = 1\,, \tag{5.13}$$

and under assets $\widetilde{A}$,

$$\frac{1}{2}R'(\widetilde{i}) + \frac{1}{2}r(\widetilde{i},\widetilde{A}) = 1\,. \tag{5.14}$$

>From Assumption 5.5,

$$\frac{1}{2}R'(i) + \frac{1}{2}r'(i,A) \leq \frac{1}{2}R'(i) + \frac{1}{2}r'(i,\widetilde{A}),$$

which together with (5.13) and (5.14) imply that

$$\frac{1}{2}R'(\widetilde{i}) + \frac{1}{2}r(\widetilde{i},\widetilde{A}) \leq \frac{1}{2}R'(i) + \frac{1}{2}r'(i,\widetilde{A})\,.$$

Since $R'' < 0$, and $r'' \leq 0$ then it must be that $i \leq \tilde{i}$. A similar argument works for e. *Q.E.D.*

The conclusion is clear: *Ownership matters.* It is important to note that ownership does not mater because it potentially affects the decision of trade or no trade, but rather because it affects the *disagreement point,* which in turn affects incentives through the ex post Nash Bargaining solution.

We can now answer the important question regarding the boundaries of firms, when forms are defined by the allocation of assets which are owned by the same person (or entity): How should ex-ante ownership be allocated? The answer is clearly to maximize ex-ante expected surplus. We have the following cases:

Case 1: *Inelastic investment decisions:* (Definition 1 in Hart (1995), p. 44.) The rough idea is that the marginal effect of investments by one party are "mostly" independent of the allocation of assets, or more formally,

$$R'(i) \cong r'(i,A)\ \forall\, i, A\,.$$

(This is the case where M_1 has inel;astic investments, so that M_1's investment decision is almost constant.) If this were the case then Type 2 integration is optimal. (A similar but reverse argument applies if M_2's investment is inelastic.)

Case 2: *Unimportant investment:* The rough idea is that if, for example, $C(e) + e$ is "very small" relative to $R(i) - i$ then,

$$\begin{aligned} S &= R(i) - C(e) - i - e \\ &= R(i) - i - (C(e) - e) \\ &\cong R(i) - i \end{aligned}$$

in which case Type 1 integration is optimal. Thus, we don't give assets to an "unimportant" party. (A similar but reverse argument applies if M_1's investment is unimportant.)(A similar but reverse argument applies if M_2's investment is inelastic.)

Case 3: *Independent Assets:* This is the case where

$$r'(i, \{a_1, a_2\}) \equiv r'(i, a_1)$$

and

$$c'(e, \{a_1, a_2\}) = c'(e, a_2)$$

and in this case No Integration is optimal. The intuition is as follows: Start from any type of integration, I_i, and change the allocation of assets so that we give a_j to M_j. This does not affect M_i, but it gives M_J better incentives.

Case 4: *Complimentary Assets:* (We consider the case of *strictly* complementary assets.) This is the case where either

$$r'(i, a_1) \equiv r'(i, \emptyset)\,, \tag{5.15}$$

in which case type 2 integration is optimal, or,

$$c'(e, a_2) \equiv c'(e, \emptyset)\,, \tag{5.16}$$

in which case type 1 integration is optimal. Again, to see the intuition start from No Integration, if (5.15) holds then M_1's incentives are not reduced if a_2 goes to M_2, and M_2's incentives are (weakly) increased when this transfer occurs. (similarly for a_1 to M_1.)

Case 5: *Essential Human Capital:* We say that M_1 is essential if

$$c'(e, \{a_1, a_2\}) \equiv c'(e, \emptyset)\,,$$

and we say that M_2 is essential if

$$r'(i, \{a_1, a_2\}) \equiv r'(i, \emptyset) .$$

The idea is that if a certain manager is essential, then the assets themselves don't enhance incentives, but only the "presence" of the essential party does. Thus, if M_1 is essential then Type 1 Integration is optimal. The intuition is that assets do not help M_2 without the presence of M_1, so M_1 might as well get all the assets. (Similarly for M_2.) A conclusion is that if both M_1 and M_2 are essential then all three ownership structures are equally good. This follows from the fact that if both are essential, neither's incentives are affected by asset ownership, but only by the presence of the other party.

(All the cases above are summarized in Proposition 2 in Hart (1995 p. 45).)

Corollary: Joint ownership is never optimal.

This follows from the argument of strict complementarity. If both parties own a_1, then neither party can use it independently if there is no trade (this assume that joint ownership takes the form of veto power over the use of the assets in case of a disagreement.) If we were able to "split" a_1 into two separate assets, that are worthless independently, then these two are strictly complementary parts, and from Case 4 above they should both be owned by either M_1 or M_2.

1. If both can use a_1 when there is no trade (e.g., a patent) then joint ownership may be optimal (See Aghion and Tirole (1994).)
2. In a repeated setting or "reputation" setting, joint ownership can cause severe punishments off the equilibrium path, so it may give better incentives on the equilibrium path (See Halonen (1997).)
3. We ignore stochastic ownership but this may actually be optimal (e.g., M_1 owns a_1 and a_2 is owned by M_1 with probability p, and by M_2 with probability $1 - p$.)

Taking the Theory to Reality

1. Strictly complementary assets are usually owned by one party. This is documented by Klein-Crawford-Alchian (1978) in the famous GM-Fisher Body case. Joskow (1985) demonstrates that for electricity plants and coal mines, a situation that would imply that there are geographic complementarities, there is a lot of vertical integration. (Or LT contracts.)

2. Independent assets are usually owned separately. Thus, we can think of the disintegration wave of the 1990's as a situation in which disintegration doesn't affect incentives, and we may reduce transaction costs and information processing costs by doing this.

Criticisms

1. Oversimplifies internal structure of the firm. First, most agents get their incentives not through ownership. Second, what is the role of management or hierarchy? (One can argue that managers are more "essential" but why?) Third, we do see joint ownership, why? (e.g., professional services.)

2. Lack of theoretical foundations: Why property rights and not something else? (Tirole (1997) (and Maskin and Tirole(1998)) discuss some sort of "inconsistency" in the theory.)

5.11.2 Incomplete Contracts and Hold-up

We saw that incomplete contracts leads to the hold-up problem. A question that is of theoretical interest is, how much "incompleteness" is necessary to get hold-up? That is, if people can specify some characteristics of trade in advance, even if this characterization is not optimal (say, it is not optimal with probability 1,) will the hold-up problem persist?

Noncontingent Contracts

(Based on Edlin and Reichelstein, AER 1996)

- A buyer (B) and seller (S) can transact ex-post and transfer any quantity $q \in [0, 1]$ from S to B.

- Both can invest in relationship specific investments, and ex post utilities for each party are given by,

$$u_B = R(i,\theta)\cdot q - t$$
$$u_S = t - c(e,\theta)\cdot q$$

where θ is some random variable which represents the ex post uncertainty, and as in the previous section the ex-ante cost of i is i, and of e is e. Given a choice of quantity q, and a realization of uncertainty θ, total ex post surplus is given by:

$$S = [R(i,\theta) - c(e,\theta)]\cdot q\,,$$

which implies that ex post either $q = 0$ or $q = 1$ is optimal, but ex ante we don't know which is since the uncertainty is not yet resolved..

Assumption: R, c, θ, i, and e are ex-post observable but not verifiable. (Like in the Grossman-Hart-Moore setting.)

We allow the parties to write "specific performance" contracts, i.e., contracts that specify $(\overline{q}, \overline{t})$ in advance, and each party can unilaterally ask for $(\overline{q}, \overline{t})$ to be enforced. That is, $(\overline{q}, \overline{t})$ is verifiable, and both parties are legally bound to follow the contract unless they choose to nullify it by renegotiating it to another contract. Note, however, that for any $\overline{q} \in (0,1)$, the probability of $(\overline{q}, \overline{t})$ being ex-post efficient is zero.

We allow the parties to renegotiate as follows:

- Given $(\overline{q}, \overline{t})$, parties can and will use unrestricted renegotiation to get to an ex-post efficient outcome.
- Gains from renegotiation are split 50:50 (Nash)

Thus, the role of $(\overline{q}, \overline{t})$ is to *set the disagreement point* for a reference point at the renegotiation stage. This is identical to the role of asset ownership in the Grossman-Hart-Moore setup.

Assumption: The parties' costs and benefits from trade are additively separable in investments and uncertainty:

$$\begin{aligned} R(i,\theta) &= r(i) + \widetilde{r}(\theta) \\ c(e,\theta) &= c(e) + \widetilde{c}(\theta) \end{aligned}$$

First Best To solve for the first best outcome, assume that i and e are contractible, and we solve,

$$\max_{i,e} E_\theta[\max\{[(r(i)+\widetilde{r}(\theta)-c(e)-\widetilde{c}(\theta)],0\}]-i-e\,,$$

and the FOC with respect to i is,

$$\underbrace{\frac{\partial}{\partial i}\left(E_\theta \max\{r(i)+\widetilde{r}(\theta)-c(e)-\widetilde{c}(\theta),0\}\right)}_{D}-1=0\,,$$

where $D=r'(i)$ if trade occurs, and $D=0$ if no trade occurs. (Recall that trade will occur if and only if it is efficient to trade.) Thus, we can rewrite the FOC as,

$$r'(i)\cdot\Pr\{r(i)-c(e)+\widetilde{r}(\theta)-\widetilde{c}(\theta)>0\}=1$$

Similarly, the FOC with respect to e is,

$$-c'(e)\cdot\Pr\{\text{trade}\}=1.$$

Second Best

Proposition : There exists $\overline{q}\in[0,1]$ such that a specific performance contract $(\overline{q},\overline{t})$ achieves FB investments

Proof:: Consider the buyer who maximizes, given $(\overline{q},\overline{t})$:

$$\begin{aligned}\max_i E_\theta\{[r(i)+\widetilde{r}(\theta)]\overline{q}-\overline{t}\\ +\frac{1}{2}[\max\{(r(i)+\widetilde{r}(\theta)-c(e)-\widetilde{c}(\theta)),0\}\\ -[r(i)+\widetilde{r}(\theta)-c(e)-\widetilde{c}(\theta)]\overline{q}]\}-i\end{aligned}$$

and the FOC is:

$$\frac{1}{2}r'(i)\cdot\overline{q}+\frac{1}{2}r'(i)\cdot\Pr\{r(i)+\widetilde{r}(\theta)-c(e)-\widetilde{c}(\theta)>0\}=1\,.$$

>From the FB. program we saw that the two FOC's in that program with respect to i and e will solve for FB investments i^*,e^*. Now let,

$$\overline{q}=\Pr\{r(i^*)+\widetilde{r}(\theta)-c(e^*)-\widetilde{c}(\theta)>0\}\,.$$

It is easy to see that i^* will be a solution to the buyer's FOC if he believes that seller chooses e^*. Similarly, we can easily show that the same $\overline{q}$ will cause the seller to choose e^* when he believes that the buyer chooses i^*. $\Rightarrow i^*, e^*$ is a Nash Equilibrium. *Q.E.D.*

The conclusion of the Edlin-Reichelstein model is that under the assumptions of their model, hold-up and specific performance have opposing forces on investments. The idea of their solution is given by:

- For values of θ such that $q^* = 0$, the buyer (seller) gets an "investment subsidy" ("tax") when $\overline{q} > 0$. i.e., this gives him a "positive" disagreement point for renegotiation and there is an over-incentive to invest.

- For values of θ such that $q^* = 1$ the reverse is true.

$\Rightarrow$ on average the FB. investment is attainable.

Relation to Literature

- Hart-Moore (EMA 1988): Similar framework but do not allow specific contracts or any contracts ("at-will contracts" maximum incompleteness). $\Rightarrow$ get underinvestment always.

- Aghion-Dewatripont-Rey (EMA 1994): Different approach to "playing" with ex-post renegotiation. Idea: We can design the renegotiation game using simple rules like (i) allocating all the bargaining power to one party, and (ii) specify default options in case that renegotiation falls (e.g., Financial hostages, per-diem penalties).

- Che-Hausch (AER forthcoming)...

- Segal-Whinston (1998 mimeo)...

Conclusion: observability and non-verifiability is not enough to explain hold-up.

Foundations for "maximum incompleteness:"

- In Edlin-Reichelstein, parties knew what good should be traded but not if trade should occur. (i.e., level)
- What does it mean not to know "what should be traded?"

1. **Rational Approach:** Parties have diffused priors over which type of many types of "widgets" will be the optimal one to trade.
2. **Courts have Bounded Rationality:** Parties can understand what will be optimal in every future state but cannot describe this to courts ex ante.
3. **Unforeseen Contingencies:** Parties have a form of bounded rationality; They cannot foresee the future states (suffer from informational "ignorance").

Rational Approach: (Based on Segal, RES forthcoming)

- N different types of widgets, ex-ante indistinguishable ex-post only one is the optimal-efficient one to trade.
- Buyer's valuation: $R(i)$ cost of investment: i

  Seller's cost: $c(e)$ cost of investment: e
- **Contract:** specify a fixed trade, e.g., widget #17, and a monetary transfer t.

This setup implies that the disagreement point is *independent* of the optimal widget.

Starting point: all widgets except the optimal one are worthless and costless - "cheap imitations" - and exhibit $R = c = 0$.

The Buyer's ex-ante problem is:

$$\max_i \; u_B = \frac{1}{N}[R(i) - t] + \frac{N-1}{N} \cdot \frac{1}{2} \cdot [(R(i) - c(e)) - 0]$$

where the first term follows from the fact that given a contract, there is a probability of $\frac{1}{N}$ that this ex-ante contract is optimal and there will be no renegotiation, and the second term follows from the fact that with probability

$\frac{N-1}{N}$ the contract will not be optimal, and we will have renegotiation in which the gains from renegotiation are split equally.. The FOC with respect to i is,

$$\left(\frac{1}{N}+\frac{N-1}{2N}\right)R'(i)\underset{N\to\infty}{\longrightarrow}\frac{1}{2}R'(i)\,.$$

That is, as $N \to \infty$, we converge to an "at-will trade" (no contract) outcome, i.e., complete hold-up due to "maximum incompleteness." The intuition is simple: The disagreement point is independent of the optimal widget, and does not depend on observable but not-verifiable information (unlike in the Edlin-Reichelstein model.) Thus, the conclusion is that we can improve over the no-contract case if the disagreement point will be made sensitive to non-verifiable information.

How can this be done? We can resort to a "message game" as follows:

- after investments and uncertainty, Buyer chooses which widget should be traded, i.e., give Buyer authority. This will cause the optimal widget to be chosen, and the FB outcome is achieved. (Full incentives ti the parties.)

What do we need to revive hold-up?

Assumption: Some of the $N-1$ non-optimal widgets are worthless as before, and some are "very expensive" regardless of investments: For some widgets, $R = C = A >> 0$ ("gold" widgets) $A > R(i) - t$.

In this case, if we use Buyer authority, the buyer will always choose a "gold" widget, and then they will renegotiate from there. Does this guarantee that we get holdup? Not necessarily:

Example: For any given N, let $g(N) \equiv \#$ of gold widgets, and this is known to the court. We can use the following message game: First, the Seller can "veto" $g(N)$ widgets. Second, the buyer chooses one remaining widget. This will lead to FB investments.

What is needed to get hold-up? Either (1) Courts don't know $g(N) \Rightarrow$ resort to general message games, or (2) Courts know $g(N)$ but only finitely many widgets can be described.

Related: Hart-Moore (1998), Maskin-Tirole (1998).