Dirk Bergemann
Department of Economics
Yale University

Information Economics

March 2009

Springer-Verlag
Berlin Heidelberg NewYork
London Paris Tokyo
Hong Kong Barcelona
Budapest
## Contents

### Part I. Introduction

1. **Game Theory** .................................................. 9  
   1.1 Game theory and parlor games - a brief history .......................... 9  
   1.2 Game theory in microeconomics ............................................. 10  

2. **Information Economics** ........................................... 11

3. **Akerlof’s Lemon Model** ........................................... 13  
   3.1 Basic Model .......................................................... 13  
   3.2 Extensions ............................................................ 14

4. **Wolinsky’s Price Signal’s Quality** ............................... 17  
   4.1 Conclusion ............................................................ 17  
   4.2 Reading ............................................................... 18

### Part II. Static Games of Incomplete Information

5. **Harsanyi’s insight** ................................................ 21

6. **Definition of a Bayesian Game** ................................... 23  
   6.1 Global Games ......................................................... 24  
   6.1.1 Incomplete Information ............................................. 25  
   6.2 Common Knowledge ................................................... 26

7. **Purification** ......................................................... 27

8. **Global Games** ....................................................... 29  
   8.0.1 Investment Game .................................................... 29

### Part III. Dynamic Games of Incomplete Information

9. **Job Market Signalling** ............................................. 33  
   9.1 Pure Strategy .......................................................... 33  
   9.2 Perfect Bayesian Equilibrium .......................................... 33  
   9.3 Equilibrium Domination ................................................ 36  
   9.4 Informed Principal .................................................... 38  
   9.4.1 Maskin and Tirole’s informed principal problem ..................... 38
Part IV. Moral Hazard

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5 Introduction and Basics</td>
<td>47</td>
</tr>
<tr>
<td>10.6 Binary Example</td>
<td>47</td>
</tr>
<tr>
<td>10.6.1 First Best</td>
<td>48</td>
</tr>
<tr>
<td>10.6.2 Second Best</td>
<td>48</td>
</tr>
<tr>
<td>10.7 General Model with finite outcomes and actions</td>
<td>49</td>
</tr>
<tr>
<td>10.7.1 Optimal Contract</td>
<td>50</td>
</tr>
<tr>
<td>10.7.2 Monotone Likelihood Ratio</td>
<td>51</td>
</tr>
<tr>
<td>10.7.3 Convexity</td>
<td>51</td>
</tr>
<tr>
<td>10.8 Information and Contract</td>
<td>53</td>
</tr>
<tr>
<td>10.8.1 Informativeness</td>
<td>53</td>
</tr>
<tr>
<td>10.8.2 Additional Signals</td>
<td>54</td>
</tr>
<tr>
<td>10.9 Linear contracts with normally distributed performance and exponential utility</td>
<td>54</td>
</tr>
<tr>
<td>10.9.1 Certainty Equivalent</td>
<td>55</td>
</tr>
<tr>
<td>10.9.2 Rewriting Incentive and Participation Constraints</td>
<td>55</td>
</tr>
<tr>
<td>10.10 Readings</td>
<td>56</td>
</tr>
</tbody>
</table>

Part V. Mechanism Design

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Introduction</td>
<td>59</td>
</tr>
<tr>
<td>12. Adverse selection: Mechanism Design with One Agent</td>
<td>61</td>
</tr>
<tr>
<td>12.1 Monopolistic Price Discrimination with Binary Types</td>
<td>61</td>
</tr>
<tr>
<td>12.1.1 First Best</td>
<td>62</td>
</tr>
<tr>
<td>12.1.2 Second Best: Asymmetric information</td>
<td>62</td>
</tr>
<tr>
<td>12.2 Continuous type model</td>
<td>64</td>
</tr>
<tr>
<td>12.2.1 Information Rent</td>
<td>64</td>
</tr>
<tr>
<td>12.2.2 Utilities</td>
<td>65</td>
</tr>
<tr>
<td>12.2.3 Incentive Compatibility</td>
<td>65</td>
</tr>
<tr>
<td>12.3 Optimal Contracts</td>
<td>67</td>
</tr>
<tr>
<td>12.3.1 Optimality Conditions</td>
<td>68</td>
</tr>
<tr>
<td>12.3.2 Pointwise Optimization</td>
<td>68</td>
</tr>
<tr>
<td>13. Mechanism Design Problem with Many Agents</td>
<td>71</td>
</tr>
<tr>
<td>13.1 Introduction</td>
<td>71</td>
</tr>
<tr>
<td>13.2 Model</td>
<td>71</td>
</tr>
<tr>
<td>13.3 Mechanism as a Game</td>
<td>72</td>
</tr>
<tr>
<td>13.4 Second Price Sealed Bid Auction</td>
<td>74</td>
</tr>
<tr>
<td>14. Dominant Strategy Equilibrium</td>
<td>77</td>
</tr>
</tbody>
</table>
15. Bayesian Equilibrium ................................................................. 79
   15.1 First Price Auctions ............................................................ 79
   15.2 Optimal Auctions ............................................................... 81
       15.2.1 Revenue Equivalence .................................................. 81
       15.2.2 Optimal Auction .......................................................... 82
   15.3 Additional Examples ........................................................... 84
       15.3.1 Procurement Bidding .................................................... 84
       15.3.2 Bilateral Trade ............................................................. 84
       15.3.3 Bilateral Trade: Myerson and Satterthwaite ..................... 84
       15.3.4 Readings .................................................................. 84

16. Efficiency ................................................................................. 85
   16.1 First Best ............................................................................ 85
   16.2 Second Best ......................................................................... 86

17. Social Choice ........................................................................... 87
   17.1 Social Welfare Functional ..................................................... 87
   17.2 Social Choice Function ......................................................... 87
Part I

Introduction
1. Game Theory

Game theory is the study of multi-person decision problems. The focus of game theory is interdependence, situations in which an entire group of people is affected by the choices made by every individual within that group. As such they appear frequently in economics. Models and situations of trading processes (auction, bargaining) involve game theory, labor and financial markets. There are multi-agent decision problems within an organization, many person may compete for a promotion, several divisions compete for investment capital. In international economics countries choose tariffs and trade policies, in macroeconomics, the FRB attempts to control prices.

Why game theory and economics? In competitive environments, large populations interact. However, the competitive assumption allows us to analyze that interaction without detailed analysis of strategic interaction. This gives us a very powerful theory and also lies behind the remarkable property that efficient allocations can be decentralized through markets.

In many economic settings, the competitive assumption does not makes sense and strategic issues must addressed directly. Rather than come up with a menu of different theories to deal with non-competitive economic environments, it is useful to come up with an encompassing theory of strategic interaction (game theory) and then see how various non-competitive economic environments fit into that theory. Thus this section of the course will provide a self-contained introduction to game theory that simultaneously introduces some key ideas from the theory of imperfect competition.

1. What will each individual guess about the other choices?
2. What action will each person take?
3. What is the outcome of these actions?

In addition we may ask

1. Does it make a difference if the group interacts more than once?
2. What if each individual is uncertain about the characteristics of the other players?

Three basic distinctions may be made at the outset

1. non-cooperative vs. cooperative games
2. strategic (or normal form) games and extensive (form) games
3. games with perfect or imperfect information

In all game theoretic models, the basic entity is a player. In noncooperative games the individual player and her actions are the primitives of the model, whereas in cooperative games coalition of players and their joint actions are the primitives.

1.1 Game theory and parlor games - a brief history

1. 20s and 30s: precursors
   a) E. Zermelo (1913) chess, the game has a solution, solution concept: backwards induction
1. Game Theory
   b) E. Borel (1913) mixed strategies, conjecture of non-existence
2. 40s and 50s: core conceptual development
   a) J. v. Neumann (1928) existence in of zero-sum games
   c) J. Nash (1950) Nonzero sum games
3. 60s: two crucial ingredients for future development: credibility (subgame perfection) and incomplete information
   a) R. Selten (1965,75) dynamic games, subgame perfect equilibrium
   b) J. Harsanyi (1967/68) games of incomplete information
4. 70s and 80s: first phase of applications of game theory (and information economics) in applied fields of economics
5. 90s and on: real integration of game theory insights in empirical work and institutional design
   • For more on the history of game theory, see Aumann’s entry on “Game Theory” in the New Palgrave Dictionary of Economics.

1.2 Game theory in microeconomics

1. decision theory (single agent)
2. game theory (few agents)
3. general equilibrium theory (many agents)
2. Information Economics

We shall start with two very simple, but classic models and results to demonstrate (i) how asymmetry of information changes classical efficiency results of markets and (ii) how asymmetry of information changes classical arguments about the role of prices and the equilibrium process.

Economics of information examines the role of information in economic relationship. It is therefore an investigation into the role of imperfect and incomplete information. In the presence of imperfect information, learning, Bayesian or not, becomes important, and in consequence dynamic models are prevalent. In this class we will focus mostly on incomplete or asymmetric information.

The nature of information economics as a field is perhaps best understood when contrasted with the standard general equilibrium theory. Information consists of a set of tools rather than a single methodology. Furthermore, the choice of tools is very issue driven. Frequently we will make use of the following tools:

- small number of participants
- institutions may be represented by constraints
- noncooperative (Bayesian) game theory
- simple assumptions on bargaining: Principal-Agent paradigm

We refer to the Principal-Agent paradigm as a setting where one agent, called the Principal, can make all the contract offers, and hence has (almost) all bargaining power and a second agent, called the Agent, can only choose whether to accept or reject the offer by the principal. With this general structure the interactive decision problem can often be simplified to a constrained optimization problem, where the Principal has an objective function, and the Agent simply represents constraints on the Principal’s objective function. Two recurrent constraints will be the participation constraint and the incentive constraint.

The following scheme organizes most of the models prevalent in information economics according to two criteria: (i) whether the informed or an uninformed agent has the initiative (makes the first move, offers the initial contract, arrangement; and (ii) whether the uninformed agent is uncertain about the action or the type (information) of the informed agent.

<table>
<thead>
<tr>
<th>hidden action</th>
<th>hidden information</th>
</tr>
</thead>
<tbody>
<tr>
<td>informed agent</td>
<td>signalling</td>
</tr>
<tr>
<td>uninformed agent</td>
<td>moral hazard</td>
</tr>
</tbody>
</table>
3. Akerlof’s Lemon Model

3.1 Basic Model

This section is based on ?. Suppose an object with value $v \sim \mathcal{U}[0, 1]$ is offered by the seller. The valuations are

$$u_s = \theta_s v$$

and

$$u_b = \theta_b v$$

with $\theta_b > \theta_s$, and hence trading is always pareto-optimal. But trade has to be voluntary. We then ask is there a price at which trade occurs. Suppose then at a price $p$ trade would occur. What properties would the price have to induce trade. The seller sells if

$$\theta_s v \leq p$$

and thus by selling the object he *signals* that

$$v \leq \frac{p}{\theta_s} \quad (3.1)$$

The buyer buys the object if

$$\theta_b \mathbb{E}[v] \geq p \quad (3.2)$$

and as he knows that (3.1) has to hold, he can form a conditional expectation, that

$$\theta_b \mathbb{E}[v | p] \geq p \Leftrightarrow \theta_b \frac{p}{2\theta_s} \geq p \quad (3.3)$$

Thus for the sale to occur,

$$\theta_b \geq \frac{p}{\theta_s} \quad (3.4)$$

Thus unless, the tastes differ substantially, the market breaks down completely:

- market mechanism in which a lower prices increases sales fails to work as lowering the price decreases the average quality, lower price is “bad news”.
- market may not disappear but display lower volume of transaction than socially optimal.
3.2 Extensions

In class, we made some informed guesses how the volume of trade may depend on the distribution of private information among the sellers. In particular, we ventured the claim that as the amount of private information held by the sellers decreases, the possibilities for trade should increase. We made a second observation, namely that in the example we studied, for a given constellation of preferences by buyer and seller, represented by $\theta_b$ and $\theta_s$, trade would either occur with positive probability for all prices or it would not occur at all. We then argued that this result may be due to the specific density in the example, but may not hold in general. We now address both issues with a uniform density with varying support and constant mean:

$$v \sim \mathcal{U} \left[ \frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon \right]$$

with $\varepsilon \in [0, \frac{1}{2}]$. We can formalize the amount of private information by the variance of $f (\cdot)$ and how it affects the efficiency of the trade.

We may redo the analysis above where the seller signals by selling the object at a fixed price $p$ that

$$v \leq \frac{p}{\theta_s}.$$  

The buyer buys the object if

$$\theta_b \mathbb{E} [v | p] \geq p,$$

The expected value is now given by

$$\mathbb{E} [v | p] = \frac{1}{2} - \varepsilon + \frac{p}{2\theta_s},$$

and hence for sale to occur

$$\theta_b \frac{1}{2} - \varepsilon + \frac{p}{2} \geq p$$

and inequality prevails if, provided that $\varepsilon \in [0, \frac{1}{2})$ and $2\theta_s > \theta_b$,

$$p \leq \frac{1}{2} \theta_b \theta_s \frac{1 - 2\varepsilon}{2\theta_s - \theta_b}.$$  

Consider next the efficiency issue. All types of sellers are willing to sell if

$$p = \theta_s \left( \frac{1}{2} + \varepsilon \right)$$

in which case the expected conditional value for the buyer is

$$\frac{1}{2} \theta_b \geq \theta_s \left( \frac{1}{2} + \varepsilon \right)$$

or equivalently

$$\theta_b \geq \theta_s (1 + 2\varepsilon)$$

Thus as the amount of private information, measured by $\varepsilon$, decreases, the inefficiencies in trade decrease. Notice that we derived a condition in terms of the preferences such that all sellers wish to sell. However
even if condition (3.7) is not met, we can in general support some trade. This is indicated already in the inequality (3.5):

\[ \theta_b \frac{1}{2} \frac{1}{2} - \varepsilon + \frac{p}{\theta_s} \geq p \]  

(3.8)

as the lhs increases slower in \( p \), then the rhs of the inequality above. Thus there might be lower prices, which will not induce all sellers to show up at the market, yet allow some trade. We show how this can arise next. For a positive mass of seller willing to sell the price may vary between \( p \in (\theta_s \left( \frac{1}{2} - \varepsilon \right), \theta_s \left( \frac{1}{2} + \varepsilon \right)) \). If we then insert the price as a function of \( x \in (-\varepsilon, \varepsilon] \), we find

\[ \theta_b \frac{1}{2} - \varepsilon + \frac{\theta_s(x + \frac{1}{2})}{\theta_s} \geq \theta_s \left( \frac{1}{2} + x \right) \]

or

\[ \theta_b \frac{1}{2} - \varepsilon + \frac{x}{\theta_s} \geq \theta_s \left( \frac{1}{2} + x \right) \]

For a given \( \theta_b, \theta_s \) and \( \varepsilon \), we may then solve the (in-)quality for \( x \), to find the (maximal) price \( p \) where the volume of trade is maximized, or:

\[ x = \frac{\theta_b (1 - \varepsilon) - \theta_s}{2 \theta_s - \theta_b} \]

Thus \( x \) is increasing in \( \theta_b \) and decreasing in \( \varepsilon \), confirming now in terms of the volume of trade our intuition about private information and efficiency. In fact, we can verify that for a given \( \varepsilon \in [0, \frac{1}{2}) \), the volume of trade is positive, but as \( \varepsilon \to \frac{1}{2} \) becomes arbitrarily small and converges as expected to zero for \( \theta_b < 2 \theta_s \).
4. Wolinsky’s Price Signal’s Quality

This section is based on ?. Suppose a continuum of identical consumers. They have preferences
\[ u = \theta v - p \]
and the monopolist can provide low and high quality: \( v = \{0, 1\} \) at cost \( 0 < c_0 < c_1 \). The monopolist
selects price and quality simultaneously. Assume that \( \theta > c_1 \) so that it is socially efficient to produce the
high quality good. Assume that the consumers do not observe quality before purchasing. It is clear that an
equilibrium in which the monopolist sells and provides high quality cannot exist.

Suppose now that some consumers are informed about the quality of the product, say a fraction \( \alpha \). Observe first that if the informed consumers are buying then the uniformed consumer are buying as well. When is the seller better off to sell to both segments of the market:
\[ p - c_1 \geq (1 - \alpha) (p - c_0) \]
or
\[ \alpha p \geq c_1 - (1 - \alpha) c_0. \]  \hspace{1cm} (4.1)

We can then make two observations:
- high quality is supplied only if price is sufficiently high, “high price can signal high quality”.
- a higher fraction, \( \alpha \), of informed consumers favors efficiency as it prevents the monopolist from cutting
  quality
- the informational externality favors government intervention as individuals only take private benefit and
  cost into account.

4.1 Conclusion

We considered “hidden information” or “hidden action” models and suggested how asymmetry in information
may reduce or end trade completely. In contrast to the canonical model of goods, which we may call
“search goods”, where we can assert the quality by inspection, we considered “experience goods” (?, ?),
where the quality can only be ascertained after the purchase. The situation is only further acerbated with
“credence goods” (?).

In both models, there was room for a third party, government or other institution, to induce pareto
improvement. In either case, and improvement in the symmetry of information lead to an improvement
in the efficiency of the resulting allocation. This suggest that we may look for optimal or equilibrium
arrangements to reduce the asymmetry in information, either through:
- costly signalling
- optimal contracting to avoid moral hazard, or
- optimal information extraction through a menu of contract (i.e. mechanism design).
4. Wolinsky’s Price Signal’s Quality

4.2 Reading

The lecture is based on Chapter 1 in ?, Chapter 2 in ?, and Chapter 5 in ?.
Part II

Static Games of Incomplete Information
5. Harsanyi’s insight

Harsanyi’s insights is illustrated by the following example.

Example: Suppose payoffs of a two player two action game are either:

\[ \begin{array}{ccc}
    & H & T \\
    H & 1,1 & 0,0 \\
    T & 0,1 & 1,0 \\
\end{array} \]

or

\[ \begin{array}{ccc}
    & H & T \\
    H & 1,0 & 0,1 \\
    T & 0,0 & 1,1 \\
\end{array} \]

i.e. either player II has dominant strategy to play \( H \) or a dominant strategy to play \( T \). Suppose that II knows his own payoffs but player I thinks there is probability \( \alpha \) that payoffs are given by the first matrix, probability \( 1 - \alpha \) that they are given by the second matrix. Say that player II is of type 1 if payoffs are given by the first matrix, type 2 if payoffs are given by the second matrix. Clearly equilibrium must have: II plays \( H \) if type 1, \( T \) if type 2; I plays \( H \) if \( \alpha > \frac{1}{2} \), \( T \) if \( \alpha < \frac{1}{2} \). But how to analyze this problem in general?

All payoff uncertainty can be captured by a single move of “nature” at the beginning of the game. See figure 14.

As with repeated games, we are embedding simple normal form games in a bigger game.
6. Definition of a Bayesian Game

A (finite) static incomplete information game consists of
- Players 1,...,I
- Actions sets $A_1,...,A_I$
- Sets of types $T_1,...,T_I$
- A probability distribution over types $p \in \Delta (T)$, where $T = T_1 \times ... \times T_I$
- Payoff functions $g_1,...,g_I$, each $g_i : A \times T \rightarrow \mathbb{R}$

Interpretation: Nature chooses a profile of players types, $t \equiv (t_1,...,t_I) \in T$ according to probability distribution $p(.)$. Each player $i$ observes his own type $t_i$ and chooses an action $a_i$. Now player $i$ receives payoffs $g_i (a,t)$.

A strategy is a mapping $s_i : T_i \rightarrow A_i$. Write $S_i$ for the set of such strategies, and let $S = S_1 \times ... \times S_I$. Player $i$’s payoff function of the incomplete information game, $u_i : S \rightarrow \mathbb{R}$, is

$$u_i(s) = \sum_{t \in T} p(t) g_i(s(t),t)$$

where $s = (s_1,...,s_I)$ and $s(t) = (s_1(t_1),...,s_I(t_I))$.

(Old) Definition: Strategy profile $s^*$ is a pure strategy Nash equilibrium if

$$u_i(s_i^*,s_{-i}^*) \geq u_i(s_i,s_{-i}^*) \text{ for all } s_i \in S_i \text{ and } i = 1,...,I$$

This can be re-written as:

$$\sum_{t \in T} p(t) g_i ((s_i^*(t_i),s_{-i}^*(t_{-i})),t) \geq \sum_{t \in T} p(t) g_i ((s_i(t_i),s_{-i}^*(t_{-i})),t)$$

for all $s_i \in S_i$ and $i = 1,...,I$

Writing $p(t_i) = \sum_{t_{-i}} p(t_i,t_{-i})$ and $p(t_{-i}|t_i) \equiv \frac{p(t_i,t_{-i})}{p(t_i)}$, this can be re-written as:

$$\sum_{t_{-i} \in T_{-i}} p(t_{-i}|t_i) g_i ((s_i^*(t_i),s_{-i}^*(t_{-i})),t) \geq \sum_{t_{-i} \in T_{-i}} p(t_{-i}|t_i) g_i ((a_i,s_{-i}^*(t_{-i})),t)$$

for all $t_i \in T_i$, $a_i \in A_i$ and $i = 1,...,I$

Example: Duopoly

\[
\begin{align*}
\pi_2(q_1,q_2;c_L) &= [(a - q_1 - q_2) - c_L] q_2 \\
\pi_2(q_1,q_2;c_H) &= [(a - q_1 - q_2) - c_H] q_2 \\
\pi_1(q_1,q_2) &= [(a - q_1 - q_2) - c] q_1 \\
\end{align*}
\]

$T_1 = \{ c \}$, $T_2 = \{ c_L,c_H \}$, $p(c_H) = \theta$. A strategy for player 1 is a quantity $q_1^*$. A strategy for player 2 is a function $q_2^* : T_2 \rightarrow \mathbb{R}$, i.e. we must solve for three numbers $q_1^*, q_2^*(c_H), q_2^*(c_L)$. 


6. Definition of a Bayesian Game

Assume interior solution. We must have:

\[ q_2^* (c_H) = \arg \max_{q_2} \left[ (a - q_1^* - q_2) - c_H \right] q_2 \]

i.e. \[ q_2^* (c_H) = \frac{1}{2} (a - q_1^* - c_H) \]

Similarly \[ q_2^* (c_L) = \frac{1}{2} (a - q_1^* - c_L) \]

We must also have:

\[ q_1^* = \arg \max_{q_1} \left[ (a - q_1 - q_2^* (c_H)) - c \right] q_1 + (1 - \theta) \left[ (a - q_1 - q_2^* (c_L)) - c \right] q_1 \]

i.e. \[ q_1^* = \arg \max_{q_1} \left[ (a - q_1 - \theta q_2^* (c_H) - (1 - \theta) q_2^* (c_L)) - c \right] q_1 \]

i.e. \[ q_1^* = \frac{1}{2} (a - c - \theta q_2^* (c_H) - (1 - \theta) q_2^* (c_L)) \]

Solution:

\[ q_2^* (c_H) = \frac{a - 2c_H + c}{3} + \frac{1 - \theta}{6} (c_H - c_L) \]
\[ q_2^* (c_L) = \frac{a - 2c_L + c}{3} - \frac{\theta}{6} (c_H - c_L) \]
\[ q_1^* = \frac{a - 2c + \theta c_H + (1 - \theta) c_L}{3} \]

6.1 Global Games

Consider a two player game, each agent must choose an action, "invest" or "not invest". The state \( \theta \) represents "fundamentals" of the economy and the payoff is given by

\[
\begin{array}{cccc}
I & N \\
I & \theta, \theta & \theta - 1, 0 \\
N & 0, \theta - 1 & 0, 0 \\
\end{array}
\]

**Complete Information.** Complete information: there is "common knowledge" of payoffs, including \( \theta \):

1. If \( 0 \leq \theta \leq 1 \), the game has multiple Nash equilibria:
   a) An equilibrium where all players invest (expecting all others to invest)
   b) An equilibrium where no one invests (since they expect no one else to invest)
2. If \( \theta \) is strictly greater than 1, everyone has a dominant strategy to invest
3. If \( \theta \) is strictly less than 0, everyone has a dominant strategy to not invest
6.1 Global Games

6.1.1 Incomplete Information

Suppose there is not common knowledge of payoff, suppose $\theta$ is not common knowledge. We have

$$\theta \sim U[-M, M]$$

for large $M > 0$, but each player only observe a signal $x_i$

$$x_i = \theta + \frac{1}{z} \varepsilon_i, \quad \varepsilon_i \sim U[-\frac{1}{2}, \frac{1}{2}]$$

and so $z$ represent the accuracy of the signal $x_i$.

Suppose the agents pursue a cut-off strategy:

$$s_i(x_i) = \begin{cases} I & \text{if } x_i > \overline{x} \\ N & \text{if } x_i < \overline{x} \end{cases}$$

Then the best response condition for investment is

$$E[\theta | x_i] Pr (s_j = I | x_i) + (1 - Pr (s_j = I | x_i)) E[\theta - 1 | x_i] \geq 0$$

(6.1)

Posterior and Beliefs. We have

$$\theta = x_i - \frac{1}{z} \varepsilon_i,$$

and hence

$$g_i (\theta | x_i) = \begin{cases} 0 & \text{if } \theta_i < x_i - \frac{1}{2z} \\ z & \text{if } x_i - \frac{1}{2z} < \theta_i < x_i + \frac{1}{2z} \\ 0 & \text{if } \theta_i > x_i + \frac{1}{2z} \end{cases}$$

Hence, simple enough:

$$E[\theta | x_i] = x_i.$$ 

Now

$$Pr (s_j = I | x_i) = Pr (x_j \geq \overline{x} | x_i) = Pr \left( \theta + \frac{1}{z} \varepsilon_j \geq \overline{x} | x_i \right) = Pr \left( x_i - \frac{1}{z} \varepsilon_i + \frac{1}{z} \varepsilon_j \geq \overline{x} \right)$$

which we could compute easily by the uniformity of the error term $\varepsilon_i$ and $\varepsilon_j$:

$$Pr \left( \frac{1}{z} (\varepsilon_j - \varepsilon_i) \geq \overline{x} - x_i \right).$$

(6.2)

But to compute the critical threshold, we do not have to do this, since we only need to evaluate at $x_i = \overline{x}:

$$x_i Pr (s_j = I | x_i) + (1 - Pr (s_j = I | x_i)) (x_i - 1) = 0 \iff$$

$$x_i \left( \frac{1}{z} (\varepsilon_j - \varepsilon_i) \geq \overline{x} - x_i \right) + \left( 1 - Pr \left( \frac{1}{z} (\varepsilon_j - \varepsilon_i) \geq \overline{x} - x_i \right) \right) (x_i - 1) = 0 \iff$$

$$\overline{x} \left( \frac{1}{z} (\varepsilon_j - \varepsilon_i) \geq \overline{x} - \overline{x} \right) + \left( 1 - Pr \left( \frac{1}{z} (\varepsilon_j - \varepsilon_i) \geq \overline{x} - \overline{x} \right) \right) (\overline{x} - 1) = 0 \iff$$

$$\overline{x} \left( \frac{1}{z} (\varepsilon_j - \varepsilon_i) \geq 0 \right) + \left( 1 - Pr \left( \frac{1}{z} (\varepsilon_j - \varepsilon_i) \geq 0 \right) \right) (\overline{x} - 1) = 0 \iff$$

$$\overline{x} \left( \frac{1}{z} + \frac{1}{2} (\overline{x} - 1) \right) = 0 \iff$$

$$\overline{x} = \frac{1}{2},$$

and thus we have a complete description of the equilibrium, and in fact it is now a unique equilibrium.
6.2 Common Knowledge

A state space $\Omega$ with $\omega \in \Omega$.

Knowledge is represented by a partition of $\Omega$, possibly different across players $h_i$ and $h_i(\omega)$ is the element of the partition that contains $\omega$.

The common prior is $p(\omega)$.

The posterior is

$$p(\omega | h_i) = \frac{p(\omega)}{\sum_{\omega' \in h_i} p(\omega')}$$

We say that player $i$ knows event $E$ at $\omega$ if

$$h_i(\omega) \subseteq E$$

and hence "player $i$ knows $E$"

$$K_i(E) = \{ \omega | h_i(\omega) \subseteq E \}$$

The event "everyone knows $E$" is

$$K_T(E) = \left\{ \omega \left| \bigcup_{i \in I} h_i(\omega) \subseteq E \right. \right\}$$

and hence

$$K^2_T(E) = \left\{ \omega \left| \bigcup_{i \in I} h_i(\omega) \subseteq K_T(E) \right. \right\},$$

and so on:

$$K^n_T(E) = \left\{ \omega \left| \bigcup_{i \in I} h_i(\omega) \subseteq K^{n-1}_T(E) \right. \right\}.$$

**Definition 6.2.1.** Event $E$ is common knowledge at $\omega$ if $\omega \in K^n_T(E)$. 
7. Purification

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1,-1</td>
<td>0,0</td>
</tr>
<tr>
<td>S</td>
<td>0,0</td>
<td>1,-1</td>
</tr>
</tbody>
</table>

Suppose I has some private value to going N, $\varepsilon \sim U[0,x]$; II has some private value to going N, $\eta \sim U[0,x]$

New game:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>$1+\varepsilon, -1+\eta$</td>
<td>$\varepsilon, 0$</td>
</tr>
<tr>
<td>S</td>
<td>$0, \eta$</td>
<td>$1,-1$</td>
</tr>
</tbody>
</table>

Best Responses: If I attaches probability $q$ to II going N, then I’s best response is N if $\varepsilon + q \geq 1 - q$ i.e. $\varepsilon \geq 1 - 2q$. Similarly, if II attaches prob. $p$ to I going N, then II’s best response is N if $\eta - p \geq - (1 - p)$ i.e. $\eta \geq 2p - 1$.

Thus players 1 and 2 must each have a threshold value of $\varepsilon$ and $\eta$, respectively, above which they choose N. We may label the thresholds so that the ex ante probability of choosing N is $p$ and $q$, respectively:

$$s_1(\varepsilon) = \begin{cases} N, & \text{if } \varepsilon \geq (1-p)x \\ S, & \text{if } \varepsilon < (1-p)x \end{cases}$$

$$s_2(\eta) = \begin{cases} N, & \text{if } \eta \geq (1-q)x \\ S, & \text{if } \eta < (1-q)x \end{cases}$$

Now $s_1$ is a best response to $s_2$ if and only if $(1-p)x = 1 - 2q$; $s_2$ is a best response to $s_1$ if and only if $(1-q)x = 2p - 1$. Thus

$$\begin{pmatrix} -x & 2 \\ -2 & x \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 1-x \\ 1+x \end{pmatrix} .$$

So

$$\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} -x & 2 \\ -2 & x \end{pmatrix}^{-1} \begin{pmatrix} 1-x \\ 1+x \end{pmatrix}$$

$$= \frac{1}{-x^2-4} \begin{pmatrix} x & -2 \\ -2 & -x \end{pmatrix} \begin{pmatrix} 1-x \\ 1+x \end{pmatrix}$$

$$= \frac{-1}{x^2 + 4} \begin{pmatrix} x & -2 \\ -2 & -x \end{pmatrix} \begin{pmatrix} 1-x \\ 1+x \end{pmatrix}$$

$$= \frac{2+x+x^2}{4+x^2} \begin{pmatrix} 2-2x \\ 2 \end{pmatrix}$$

$$\rightarrow \left( \frac{3}{2} \right), \text{ as } x \rightarrow 0$$
We say that the mixed strategy equilibrium of the complete information game is “purified” by adding the incomplete information.

Parameterize a complete information game by the payoff functions, \( g = (g_1, ..., g_I) \).

Let player \( i \) have a type \( \theta_i = (\theta^n_i)_{a \in \mathcal{A}} \in \mathbb{R}^{\mathcal{A}} \); let player \( i \)'s type be independently drawn according to smooth density \( f_i(\cdot) \) with bounded support, say \([-1, 1]^{\mathcal{A}}\). Let player \( i \)'s payoff in the \( \varepsilon \)-perturbed game (this is an incomplete information game) be

\[
\bar{g}_i(a, \theta) = g_i(a) + \varepsilon \theta^n_i
\]

Theorem (Harsanyi): Fix a generic complete information game \( g \) and any Nash equilibrium of \( g \). Then there exists a sequence of pure strategy equilibria of the \( \varepsilon \)-perturbed with the property that the probability distribution over actions of each player is converging to his strategy in the original Nash equilibrium.
8. Global Games

8.0.1 Investment Game

The state $\theta$ is drawn uniformly from the interval $[M,M]$ for large $M > 0$. Each player observes a signal $s_i = q + 1 \times #i$, where $#i \sim U(1, 2)$, and $x$ measures the accuracy of the signal. Stephen
Part III

Dynamic Games of Incomplete Information
9. Job Market Signalling

Consider the following game first analyzed by ?. Nature chooses a workers type (productivity) \( a \in \{1, 2\} \). The worker has productivity \( a \) with probability \( p(a) \). For notational convenience, we define

\[
p \triangleq p(a = 2).
\]

The worker can choose an educational level \( e \in \mathbb{R}_+ \). The worker chooses the firm The workers utility is

\[
u(w, e, a) = w - \frac{e}{a}.
\]

The objective function of the firm is given by

\[
\min_{a \in \{1, 2\}} \sum_{a \in \{1, 2\}} (a - w)^2 p(a | e)
\]

We observe that

\[
\frac{\partial^2 u(w, e, a)}{\partial e \partial a} = \frac{1}{a^2} > 0
\]

and thus type and strategic variable are complements (supermodular).

9.1 Pure Strategy

A pure strategy for the worker is a function

\[
\hat{e} : \{1, 2\} \to \mathbb{R}_+
\]

A pure strategy for the firm is a function

\[
w : \mathbb{R}_+ \to \mathbb{R}_+
\]

where \( w(e) \) is the wage offered to a worker of educational level \( e \).

9.2 Perfect Bayesian Equilibrium

The novelty in signalling games is that the uninformed party gets a chance to update its prior belief on the basis of a signal sent by the informed party. The updated prior belief is the posterior belief, depending on \( e \) and denoted by \( p(a | e) \). The posterior belief is a mapping

\[
\hat{p} : \mathbb{R}_+ \to [0, 1]
\]
In sequential or extensive form games we required that the strategy are sequential rational, or time consistent. We now impose a similar consistency requirement on the posterior beliefs by imposing the Bayes’ rule whenever possible. Since \( p(a | e) \) is posterior belief, and hence a probability function, it is required that:

\[
\forall e, \exists p(a | e), \text{ s.th. } p(a | e) \geq 0, \text{ and } \sum_{a \in \{1, 2\}} p(a | e) = 1. \tag{9.1}
\]

Moreover, when the firm can apply Bayes law, it does so, or

\[
\text{if } \exists a \text{ s.th. at } e^*(a) = e, \text{ then } p(a | e) = \frac{p(a)}{\sum_{a' | e^*(a) = e} p(a')}. \tag{9.2}
\]

We refer to educational choice which are selected by some worker-types in equilibrium, or if \( \exists a \text{ s.th. at } e^*(a) = e \), as “on-the-equilibrium path” and educational choices s.th. \( \exists a \text{ with } e^*(a) = e \) as “off-the-equilibrium-path”.

As before it will be sometimes easier to refer to \( p(e) \), \( p(a = 2 | e) \) and hence

\[
p(1 | e) = 1 - p(e).
\]

**Definition 9.2.1 (PBE).** A pure strategy Perfect Bayesian Equilibrium is a set of strategies \( \{e^*(a), w^*(e)\} \) and posterior beliefs \( p(e) \) such that:

1. \( \forall e, \exists p^*(a | e), \text{ s.th. } p^*(a | e) \geq 0, \text{ and } \sum_{a \in \{1, 2\}} p^*(a | e) = 1; \)
2. \( \forall i, w_i^*(e) = \sum_a p^*(a | e) a; \)
3. \( \forall a, e^*(a) \in \arg \max \{w^*(e) - \frac{w_i(e)}{a}\} \)
4. \( \text{if } \exists a \text{ s.th. at } e^*(a) = e, \text{ then:} \)

\[
p(a | e) = \frac{p(a)}{\sum_{a' | e^*(a) = e} p(a')}.
\]

**Definition 9.2.2 (Separating PBE).** A pure strategy PBE is a separating equilibrium if

\( a \neq a' \Rightarrow e(a) \neq e(a') \).

**Definition 9.2.3 (Pooling PBE).** A pure strategy PBE is a pooling equilibrium if

\( \forall a, a' \Rightarrow e(a) = e(a') \).

[Lecture 3]

**Theorem 9.2.1.**

1. A pooling equilibrium exists for all education levels \( e = e(1) = e(2) \in [0, p] \).
2. A separating equilibrium exists for all \( e(1) = 0 \) and \( e(2) \in [1, 2] \).

**Proof.** (1) We first construct a pooling equilibrium. For a pooling equilibrium to exist it must satisfy the following incentive compatibility constraints

\[
\forall e, 1 + p(e^*) - e^* \geq 1 + p(e) - e \tag{9.3}
\]

and
\[
\forall e, \ 1 + p(e^*) - \frac{e^*}{2} \geq 1 + p(e) - \frac{e}{2} \quad (9.4)
\]

Consider first downward deviations, i.e. \( e < e^* \), then (9.3) requires that
\[
p(e^*) - p(e) \geq e^* - e. \quad (9.5)
\]

Then consider upward deviations, i.e. \( e > e^* \), then (9.4) requires that
\[
p(e) - p(e^*) \leq \frac{1}{2} (e - e^*). \quad (9.6)
\]

We can then ask for what levels can both inequalities be satisfied. Clearly both inequalities are easiest to satisfy if
\[
\begin{align*}
  e < e^* & \Rightarrow p(e) = 0 \\
  e > e^* & \Rightarrow p(e) = 0,
\end{align*}
\]

which leaves us with
\[
\begin{align*}
  e < e^* & \Rightarrow p \geq e^* - e \\
  e > e^* & \Rightarrow -p \leq \frac{1}{2} (e - e^*). \quad (9.7)
\end{align*}
\]

We may then rewrite the inequalities in (9.7) as
\[
\begin{align*}
  e < e^* & \Rightarrow e^* \leq p + e \\
  e > e^* & \Rightarrow e^* \leq p + 2e. \quad (9.8)
\end{align*}
\]

As the inequality has to hold for all \( e \), the asserting that \( 0 \leq p \leq e^* \) holds, follows immediately.

(2) Consider then a separating equilibrium. It must satisfy the following incentive compatibility constraints
\[
\forall e, \ 1 + p(e^*_1) - e^*_1 \geq 1 + p(e) - e
\]

and
\[
\forall e, \ 1 + p(e^*_2) - \frac{e^*_2}{2} \geq 1 + p(e) - \frac{e}{2}. \quad (9.9)
\]

As along the equilibrium path, the firms must apply Bayes law, we can rewrite the equations as
\[
\forall e, \ 1 - e^*_1 \geq 1 + p(e) - e
\]

and
\[
\forall e, \ 2 - \frac{e^*_2}{2} \geq 1 + p(e) - \frac{e}{2}.
\]

Consider first the low productivity type. It must be that \( e^*_1 = 0 \). This leaves us with
\[
p(e) \leq e. \quad (9.10)
\]

But if \( e = e^*_2 \) is supposed to be part of a separating equilibrium, then \( p(e^*_2) = 1 \). Thus it follows further from (9.10) that \( e^*_2 \geq 1 \), for otherwise we would not satisfy \( 1 = p(e^*_2) \leq e^*_2 \). Finally, we want to determine an upper bound for \( e^*_2 \). AS we cannot require the high ability worker to produce too much effort otherwise he would mimic the lower ability, we can rewrite (9.9) to obtain:
which is easiest to satisfy if
\[ p(e) = 0, \]
and hence
\[ \forall e, 2 + e \geq e_2^* \]
which implies that:
\[ e_2^* \leq 2. \]
which completes the proof.

The theorem above only defines the range of educational choices which can be supported as an equilibrium but is not a complete equilibrium description as we have not specified the beliefs in detail. The job market signalling model suggests a series of further questions and issues. There were multiple equilibria, reducing the predictive ability of the model and we may look at different approaches to reduced the multiplicity:

- refined equilibrium notion
- different model: informed principal

The signal was modelled as a costly action. We may then ask for conditions:

- when do costly signals matter for Pareto improvements (or simply separation): Spence-Mirrlees single crossing conditions
- when do costless signals matter: cheap-talk games.

In the model, education could also be interpreted as an act of disclosure of information through the verification of the private information by a third party. It may therefore be of some interest to analyze the possibilities of voluntary information disclosure.

### 9.3 Equilibrium Domination

The construction of some of the equilibria relied on rather arbitrary assumptions about the beliefs of the firms for educational choice levels off the equilibrium path. Next we refine our argument to obtain some logical restrictions on the beliefs. The restrictions will not be based on Bayes law, but on the plausibility of deviations as strategic choices by the agents. The notions to be defined follow below.

**Definition 9.3.1 (Equilibrium Domination).** Given a PBE, the message \( e \) is equilibrium-dominated for type \( a \), if for all possible assessments \( p(a|e) \) (or simply \( p(e) \)):

\[
w^*(e^*(a)) - \frac{e^*(a)}{a} > w(e) - \frac{e}{a}
\]
or equivalently

\[
1 + p^*(e^*(a)) - \frac{e^*(a)}{a} > 1 + p(e) - \frac{e}{a}
\]
Definition 9.3.2 (Intuitive Criterion). If the information set following \( e \) is off the equilibrium path and \( e \) is equilibrium dominated for type \( a \), then, if possible,

\[
p(a|e) = 0. \tag{9.11}
\]

Remark 9.3.1. The qualifier, if possible, in the definition above refers to the fact that \( p(a|e) \) has to be a well-defined probability, and thus if the effort \( e \) is equilibrium dominated for all \( a \), then the requirement imposed by (9.11) is vacuous.

Definition 9.3.3. A PBE where all beliefs off the equilibrium path satisfy the intuitive criterion is said to satisfy the intuitive criterion.

Theorem 9.3.1 (Uniqueness). The unique Perfect Bayesian equilibrium outcome which satisfies the intuitive criterion is given by

\[
\{e_*^1 = 0, e_*^2 = 1, w^*(0) = 1, w^*(1) = 2\}. \tag{9.12}
\]

The beliefs are required to satisfy

\[
p^*(e) = \begin{cases} 
0, & \text{for } e = 0 \\
\in [0,e], & \text{for } 0 < e < 1 \\
1, & \text{for } e \geq 1
\end{cases}
\]

Remark 9.3.2. The equilibrium outcome is referred to as the least-cost separating equilibrium (or ? equilibrium)

Proof. We first show that there can’t be any pooling equilibria satisfying the intuitive criterion, and then proceed to show that only one of the separating equilibria survives the test of the intuitive criterion.

Suppose \( 0 \leq e^* \leq p \). Consider \( e \in (e^* + (1 - p), e^* + 2(1 - p)) \). Any message \( e \) in the interval is equilibrium dominated for the low productivity worker as

\[1 + p - e^* > 2 - e,\]

but any message in the interval is not equilibrium dominated for the high productivity worker, as

\[1 + p - \frac{e^*}{2} < 2 - \frac{e}{2}\]

and thus for \( e = e^* + (1 - p) \), we have

\[p < 1 - \frac{(1 - p)}{2} \Leftrightarrow \frac{1}{2}p < \frac{1}{2},\]

which certainly holds. Thus if \((e, p(e))\) satisfies the intuitive criterion, we must have \( p(e) = 1 \) for \( e \in (e^* + (1 - p), e^* + 2(1 - p)) \), but then pooling is not an equilibrium anymore as the high productivity worker has a profitable deviation with any \( e \in (e^* + (1 - p), e^* + 2(1 - p)) \).

Consider now the separating equilibria. For \( e_*^2 > 1 \), any \( e \in (1, e_*^2) \) is equilibrium dominated for the low ability worker as

\[1 > 2 - e,\]

but is not equilibrium dominated for the high ability worker, as

\[2 - \frac{e_*^2}{2} < 2 - \frac{e}{2}.\]

It follows that \( p(e) = 1 \) for all \( e \in (1, e_*^2) \). But then \( e_*^2 > 1 \), cannot be supported as an equilibrium as the high ability worker has a profitable deviation by lowering his educational level to some \( e \in (1, e_*^2) \), and still receive his full productivity in terms of wages. It remains \( e_*^2 = 1 \) as the unique PBE satisfying the intuitive criterion.
9. Job Market Signalling

Criticism. Suppose \( p \to 1 \).

Remark 9.3.3. Add the Kreps critique of equilibrium domination.

[Lecture 4]

9.4 Informed Principal

As plausible as the Cho-Kreps intuitive criterion may be, it does seem to predict implausible equilibrium outcomes in some situations. For example, suppose that, \( p \), the prior probability that a worker of type \( a = 2 \) is present, is arbitrarily large, \(( p \to 1)\). In that case, it seems a rather high cost to pay to incur an education cost of

\[ c(e = 1) = \frac{1}{2} \]

just to be able to raise the wage by a commensurately small amount \( \Delta w = 2 - (1 + p) \to 0 \) as \( p \to 1 \). Indeed, in that case the pooling equilibrium where no education costs are incurred seems a more plausible outcome. This particular case should serve as a useful warning not to rely too blindly on selection criteria (such as Cho-Kreps' intuitive criterion) to single out particular PBEs.

9.4.1 Maskin and Tirole’s informed principal problem

Interestingly, much of the problem of multiplicity of PBEs disappears when the timing of the game is changed to letting agent and principal sign a contract before the choice of signal. This is one important lesson to be drawn from ?.

To see this, consider the model of education as specified above and invert the stages of contracting and education choice. That is, now the worker signs a contract with his/her employer before undertaking education. This contract then specifies a wage schedule contingent on the level of education chosen by the worker after signing the contract.

Let \( \{w(e)\} \) denote the contingent wage schedule specified by the contract.

Consider the problem for the high productivity worker. Suppose he would like to make an offer by which he can separate himself from a low ability worker

\[
\max_{\{e,w(e)\}} \left\{ w(e) - \frac{e}{2} \right\}
\]

subject to

\[ w(e_1) - e_1 \geq w(e_2) - e_2 \quad (IC_1) \]

and

\[ w(e_2) - \frac{e_2}{2} \geq w(e_1) - \frac{e_1}{2} \quad (IC_2) \]

and

\[ a_1 - w(e_1) \geq 0 \quad (IR_1) \]

\[ a_2 - w(e_2) \geq 0 \quad (IR_2) \]

Thus to make incentive compatibility as easy as possible he suggests \( e_1 = 0 \) and \( w(e_1 = 0) = 1 \). As \( w(e_2) = 2 \), it follows that after setting \( e_2 = 1 \), he indeed maximizes his payoff.
Suppose instead he would like to offer a pooling contract. Then he would suggest

$$\max_{e,w} \left\{ w - \frac{e}{2} \right\}$$

$$1 + p - w \geq 0$$

which would yield $w = 1 + p$ for all $e$.

This suggest that there are two different cases to consider:

1. $1 + p \leq 2 - \frac{1}{2}$: a high productivity worker is better off in the “least cost” separating equilibrium than in the efficient pooling equilibrium.
2. $1 + p > 2 - \frac{1}{2}$: a high productivity worker is better off in the efficient pooling equilibrium.

It is easy to verify that in the case where $p \leq \frac{1}{2}$, the high productivity worker cannot do better than offering the separating contract, nor can the low productivity worker. More precisely, the high productivity worker strictly prefers this contract over any contract resulting in pooling or any contract with more costly separation. As for the low productivity worker, he has everything to lose by offering another contract which would identify himself.

In the alternative case where $p > \frac{1}{2}$, the unique equilibrium contract is the one where:

$$w^*(e) = 1 + p \text{ for all } e \geq 0.$$

Again, if the firm accepts this contract, both types of workers choose an education level of zero. Thus, on average the firm breaks even by accepting this contract, provided that it is as (or more) likely to originate from a high productivity worker than a low productivity worker. Now, a high productivity worker strictly prefers to offer this contract over any other separating contract. Similarly, a low productivity worker has everything to lose from offering another contract and thus identifying himself. Thus, in this case again this is the unique contract offer made by the workers.
10. Spence-Mirrlees Single Crossing Condition

10.0.2 Separating Condition

Consider now a general model with a continuum of types:

$$A \subseteq \mathbb{R}_+$$

and an arbitrary number of signals

$$E \subseteq \mathbb{R}_+$$

with a general quasilinear utility function

$$u(t, e, a) = t + v(e, a)$$

where we recall that the utility function used in Spence model was given by:

$$u(t, e, a) = t - \frac{e}{a}$$

We now want ask when is it possible in general to sustain a separating equilibrium for all \(n\) agents, such that

$$a \neq a' \Rightarrow e \neq e'$$

Suppose we can support a separating equilibrium for all types, then we must be able to satisfy for \(a' > a\) and without loss of generality \(e' > e\):

$$t + v(e, a) \geq t' + v(e', a) \Leftrightarrow t - t' \geq v(e', a) - v(e, a) \quad (10.1)$$

and

$$t' + v(e', a') \geq t + v(e, a') \Leftrightarrow t - t' \leq v(e', a') - v(e, a'), \quad (10.2)$$

where \(t \triangleq t(e)\) and \(t' \triangleq t(e')\), and by combining the two inequalities, we find

$$v(e', a) - v(e, a) \leq v(e', a') - v(e, a'), \quad (10.3)$$

or similarly, again recall that \(e < e'\)

$$v(e, a') - v(e, a) \leq v(e', a') - v(e', a). \quad (10.4)$$

We then would like to know what are sufficient conditions on \(v(\cdot, \cdot)\) such that for every \(a\) there exists \(e\) such that (10.3) can hold.
10.1 Supermodular

Definition 10.1.1. A function \( v : E \times A \to \mathbb{R} \) has increasing differences in \((e, a)\) if for any \( a' > a \), \( v(e, a') - v(e, a) \) is nondecreasing in \( e \).

Note that if \( v \) has increasing differences in \((e, a)\), it has increasing differences in \((a, e)\). Alternatively, we say the function \( v \) is supermodular in \((e, a)\).

If \( v \) is sufficiently smooth, then \( v \) is supermodular in \((e, a)\) if and only if \( \partial^2 v / \partial e \partial a \geq 0 \).

As the transfers \( t, t' \) could either be determined exogenously, as in the Spence signalling model through market clearing conditions, or endogenously, as in optimally chosen by the mechanism designer, we want to ask when we can make a separation incentive compatible. We shall not consider here additional constraints such as a participation constraints. We shall merely assume that \( t(a) \), i.e. the transfer that the agent of type \( a \) gets, \( t(a) \), conditional on the sorting allocation \( \{a, e(a), t(a)\} \) to be incentive compatible, is continuously differentiable and strictly increasing.

Theorem 10.1.1. A necessary and sufficient condition for sorting ((10.1) and (10.2)) to be incentive compatible for all \( t(a) \) is that

1. \( v(e, a) \) is strictly supermodular
2. \( \frac{\partial v}{\partial e} < 0 \) everywhere

Proof. We shall also suppose that \( v(\cdot, \cdot) \) is twice continuously differentiable. The utility function of the agent is given by

\[
t(\hat{a}) + v(e(\hat{a}), a),
\]

where \( a \) is the true type of agent \( a \) and \( e(\hat{a}) \) is the signal the informed agent sends to make the uninformed agent believe he is of type \( \hat{a} \).

(Sufficiency) For \( e(a) \) to be incentive compatible, \( \hat{a} \) must locally solve for the first order conditions of the agent at \( \hat{a} = a \), namely the true type:

\[
t'(\hat{a}) + \frac{\partial v(e(\hat{a}), a)}{\partial e} \frac{de}{d\hat{a}} = 0, \text{ at } \hat{a} = a
\]

But Fix the payments \( t(a) \) as a function of the type \( a \) to be \( t(a) \) and suppose without loss of generality that \( t(a) \) is continuously differentiable. The equality (10.5) defines a differential equation for the separating effort level

\[
\frac{de}{da} = \frac{t'(a)}{\frac{\partial v(e(a), a)}{\partial e}}
\]

for which a unique solution exists given an initial condition, say \( t(0) = 0 \). The essential element is that the first order conditions is only an optimal condition for agent with type \( a \) and nobody else, but this follows from strict supermodularity. As

\[
t'(\hat{a})
\]

and

\[
\frac{de}{d\hat{a}}
\]

are independent of the true type, \( a \), it follows that if

\[
\frac{\partial^2 v(e, a)}{\partial e \partial a} > 0,
\]
for all $e$ and $a$, then (10.5) can only identify truthtelling for agent $a$.

(Necessity) For any particular transfer policy $t(a)$, we may not need to impose the supermodularity condition everywhere, and it might often by sufficient to only impose it locally, where it is however necessary to guarantee local truthtelling, i.e.

$$\frac{\partial v(e(a), a)}{\partial e a} > 0$$

at $e = e(a)$. However, as we are required to consider all possible transfers $t(a)$ with arbitrary positive slopes, we can guarantee that for every $a$ and every $e$ there is some transfer problem $t(a)$ such that $e(a) = e$ by (10.6), and hence under the global condition on $t(a)$, supermodularity becomes also a necessary condition for sorting.

10.2 Supermodular and Single Crossing

In the discussion above we restricted our attention to quasilinear utility function. We can generalize all the notions to more general utility functions. We follow the notation in ?. Let

$$U(x, y; t) : \mathbb{R}^3 \to \mathbb{R}$$

where $x$ is in general the signal (or allocation), $y$ an additional instrument, such as money transfer and $t$ is the type of the agent.

**Definition 10.2.1 (Spence-Mirrlees).** The function $U$ is said to satisfy the (strict) Spence-Mirrlees condition if

1. the ratio

$$\frac{U_x}{U_y}$$

is (increasing) nondecreasing in $t$, and $U_y \neq 0$, and

2. the ratio has the same sign for every $(x, y, t)$.

**Definition 10.2.2 (Single-Crossing).** The function $U$ is said to satisfy the single crossing property in $(x, y; t)$ if for all $(x', y') \geq (x, y)$

1. whenever $U(x', y'; t) \geq U(x, y; t)$, then $U(x', y'; t') \geq U(x, y; t')$ for all $t' > t$;

2. whenever $U(x', y'; t) > U(x, y; t)$, then $U(x', y'; t') > U(x, y; t')$ for all $t' > t$;

? show that the notions of single crossing and the Spence Mirrlees conditions are equivalent.

Thus the supermodularity condition is also often referred to as the single-crossing or Spence-Mirrlees condition, where Mirrlees used the differential form for the first time in ?. The notions of supermodularity and single-crossing are exactly formulated in ? and for some corrections ? Some applications to supermodular games are considered by ? and ?. The mathematical theory is, inter alia, due to ? and ?.

10.3 Signalling versus Disclosure

Disclosure as private but certifiable information. Signalling referred to situations where private information is neither observable nor verifiable. That is to say, we have considered private information about such things as individual preferences, tastes, ideas, intentions, quality of projects, effort costs, etc., which cannot really be measured objectively by a third party. But there are other forms of private information, such as an
individual's health, the servicing and accident history of a car, potential and actual liabilities of a firm, earned income, etc., that can be certified or authenticated once disclosed. For these types of information the main problem is to get the party who has the information to disclose it. This is a simpler problem than the one we have considered so far, since the informed party cannot report false information. It can only choose not to report some piece of information it has available. The main results and ideas of this literature is based on ? and ?.

10.4 Reading

The material of this lecture is covered by Section 4 of ? and for more detail on the Spence-Mirrlees conditions, see ?.
Part IV

Moral Hazard
Today we are discussing the optimal contractual arrangement in a moral hazard setting.

10.5 Introduction and Basics

In contractual arrangements in which the principal offers the contract, we distinguish between

- hidden information - ‘adverse selection’

and

- hidden action - ‘moral hazard’

The main trade-off in adverse selection is between efficient allocation and informational rent. In moral hazard settings it is between risk sharing and work incentives. Today we are going to discuss the basic moral hazard setting. As the principal tries to infer from the output of the agent about the effort choice, the principal engages in statistical inference problems. Today we are going to develop basic insights into the optimal contract and use this as an occasion to introduce three important informational notions:

1. monotone likelihood ratio
2. garbling in the sense of Blackwell
3. sufficient statistic

The basic model goes as follows. An agent takes action that affect utility of principal and agent. Principal observes “outcome” \( x \) and possible some “signal” \( s \) but not “action” \( a \) of agent. Agent action in absence of contract is no effort. Influence agents action by offering transfer contingent on outcome.

**Example 10.5.1.** Employee employer effect (non observable), Property insurance (fire, theft)

- **FB:** agents action is observable (risk sharing)
- **SB:** agents action is unobservable (risk sharing-incentive)

10.6 Binary Example

The agent can take action \( a \in \{a_l, a_h\} \) at cost \( c \in \{c_l, c_h\} \). The outcomes \( x \in \{x_l, x_h\} \) occur randomly, where the probabilities are governed by the action as follows.

\[
p_l = \Pr(x_h | a_l) < p_h = \Pr(x_h | a_h)
\]

The principal can offer a wage, contingent on the outcome \( w \in \{w_l, w_h\} \) to the agent and the utility of the agent is

\[
u(w_i) - c_i
\]

and of the principal it is

\[
v(x_i - w_i)
\]

where we assume that \( u \) and \( v \) are strictly increasing and weakly concave.
10.6.1 First Best

We consider initially the optimal allocation of risk between the agents in the presence of the risk and with observable actions. If the actions are observable, then the principal can induce the agent to choose the preferred action \( a^* \) by

\[
 w(x_i, a) = -\infty
\]

if \( a \neq a^* \) for all \( x_i \). As the outcome is random and the agents have risk-averse preference, the optimal allocation will involve some risk sharing. The optimal solution is characterized by

\[
 \max_{\{w_h, w_l\}} \{ p_i v(x_h - w_h) + (1 - p_i) v(x_l - w_l) \}
\]

subject to

\[
 p_i u(w_h) + (1 - p_i) u(w_l) - c_i \geq U_i(\lambda)
\]

which is the individual rational constraint. Here we define

\[
 w(x_i, a^*) \triangleq w_i
\]

This is a constrained optimization problem, and the first order conditions from the Lagrangian

\[
 L(w_h, w_l, \lambda) = p_i v(x_h - w_h) + (1 - p_i) v(x_l - w_l) + \lambda (p_i u(w_h) + (1 - p_i) u(w_l) - c_i)
\]

are given by

\[
 \frac{V'(x_i - w_l)}{U'(w_l)} = \lambda,
\]

which is Borch’s rule.

10.6.2 Second Best

Consider now the case in which the action is unobservable and therefore

\[
 w(x_i, a) = w(x_i)
\]

for all \( a \). Suppose the principal wants to induce high effort, then the incentive constraint is:

\[
 p_h u(w_h) + (1 - p_h) u(w_l) - c_h \geq p_l u(w_h) + (1 - p_l) u(w_l) - c_l
\]

or

\[
 (p_h - p_l) (u(w_h) - u(w_l)) \geq c_h - c_l \tag{10.7}
\]

as \( p_l \to p_h \) \( w_h - w_l \) must increase, incentives become more high-powered. The principal also has to respect a participation constraint (or individual rationality constraint)

\[
 p_h u(w_h) + (1 - p_h) u(w_l) - c_h \geq \bar{U} \tag{10.8}
\]

We first show that both constraints will be binding if the principal maximizes

\[
 \max_{\{w_h, w_l\}} \left[ p_h (x_h - w_h) + (1 - p_h) (x_l - w_l) \right]
\]
subject to (10.7) and (10.8). For the participation constraint, principal could lower both payments and be better off. For the incentive constraint, subtract
\[
\frac{(1 - p_h) \varepsilon}{u'(w_h)}
\]
from \( w_h \) and add
\[
\frac{p_h \varepsilon}{u'(w_l)}
\]
to \( w_l \). Then the incentive constraint would still hold for \( \varepsilon \) sufficiently small. Consider then participation constraint. We would subtract utility
\[
(1 - p_h) \varepsilon
\]
from \( u(w_h) \) and add
\[
p_h \varepsilon
\]
to \( u(w_l) \), so that the expected utility remains constant. But the wage bill would be reduced for the principal by
\[
\varepsilon p_h (1 - p_h) \left( \frac{1}{u'(w_h)} - \frac{1}{u'(w_l)} \right) > 0,
\]
since \( w_h > w_l \) by concavity. The solution is :
\[
u(w_l) = U - \frac{c_h p_l - p_h c_l}{p_h - p_l}
\]
and
\[
u(w_h) = U - \frac{c_h p_l - p_h c_l}{p_h - p_l} + \frac{c_h - c_l}{p_h - p_l}
\]

10.7 General Model with finite outcomes and actions

Suppose
\[
x_i \in \{x_1, ..., x_I\}
\]
and
\[
a_j \in \{a_1, ..., a_J\}
\]
and the probability
\[
p_{ij} = \Pr(x_i | a_j)
\]
the utility is \( u(w) - a \) for agent and \( x - w \) for the principal.
10.7.1 Optimal Contract

The principals problem is then

\[
\max_{\{w_i\}_{i=1}^J} \left\{ \sum_{i=1}^I (x_i - w_i) p_{ij} \right\}
\]

given the wage bill the agent selects \( a_j \) if and only if

\[
\sum_{i=1}^I u(w_i) p_{ij} - a_j \geq \sum_{i=1}^I u(w_i) p_{ik} - a_k \quad (\mu_k)
\]

and

\[
\sum_{i=1}^I u(w_i) p_{ij} - a_j \geq U(\lambda)
\]

Fix \( a_j \), then the Lagrangian is

\[
L(w_{ij}, \lambda, \mu) = \left\{ \sum_{i=1}^I (x_i - w_i) p_{ij} \right\} + \sum_{k \neq j} \left\{ \sum_{i=1}^I u(w_i) (p_{ij} - p_{ik}) - (a_j - a_k) \right\} + \lambda \left\{ \sum_{i=1}^I u(w_i) p_{ij} - a_j \right\}
\]

Differentiating with respect to \( w_{ij} \) yields

\[
\frac{1}{u'(w_i)} = \lambda + \sum_{k \neq j} \mu_k \left( 1 - \frac{p_{ik}}{p_{ij}} \right), \quad \forall i
\] (10.9)

With a risk-avers principal the condition (10.9) would simply by modified to:

\[
\frac{v'(x_i - w_i)}{u'(w_i)} = \lambda + \sum_{k \neq j} \mu_k \left( 1 - \frac{p_{ik}}{p_{ij}} \right), \quad \forall i
\] (10.10)

In the absence of an incentive problem, or \( \mu_k = 0 \), (10.10) states the Borch rule of optimal risk sharing:

\[
\frac{v'(x_i - w_i)}{u'(w_i)} = \lambda, \quad \forall i,
\]

which states that the ratio of marginal utilities is equalized across all states \( i \).

We now ask for sufficient conditions so that higher outcomes are rewarded with higher wages. As \( w_i \) increases as the right side increases, we might ask for conditions when the rhs is increasing in \( i \). To do this we may distinguish between the ‘downward’ binding constraints \( (\mu_k, k < j) \) and the ‘upward’ binding constraints \( (\mu_k, k > j) \). Suppose first then that there were only ‘downward’ binding constraints, i.e. either \( j = J \), or \( \mu_k = 0 \) for \( k > j \). Then a sufficient for monotonicity in \( i \) would clearly be that

\[
\frac{p_{ik}}{p_{ij}}
\]

is decreasing in \( i \) for all \( k < j \).
10.7.2 Monotone Likelihood Ratio

The ratio
\[
\frac{p_{ik}}{p_{ij}}
\]
are called likelihood ratio. We shall assume a monotone likelihood ratio condition, or \( \forall i < i', \forall k < j : \)
\[
\frac{p_{ik}}{p_{ij}} > \frac{p'_{ik}}{p'_{ij}}
\]
or reversing the ratio
\[
\frac{p_{ij}}{p_{ik}} < \frac{p'_{ij}}{p'_{ik}}
\]
so that the likelihood ratio increases with the outcome. In words, for a higher outcome \( x_{i'} \), it becomes more likely, the action profile was high \( (j) \) rather than low \( (k) \). By modifying the ratio to:
\[
\frac{p'_{jk}}{p'_{ij}} < \frac{p'_{jk}}{p'_{ik}}
\]
we get yet a different interpretation. An increase in the action increases the relative probability of a high outcome versus a low outcome.

The appearance of the likelihood ratio can be interpreted as the question of estimating a ‘parameter’ \( a \) from the observation of the ‘sample’ \( x_i \) as the following two statements are equivalent:

\( a_j \) is the maximum likelihood estimator of \( a \) given \( x_i \), \( \forall k, \frac{p_{ik}}{p_{ij}} \leq 1 \).

As is explained in Hart-Holmström [1986]:

The agent is punished for outcomes that revise belief about \( a_j \) down, while he is rewarded for outcomes that revise beliefs up. The agency problem is not an inference problem in a strict statistical sense; conceptually, the principal is not inferring anything about the agent’s action from \( x_i \) because he already knows what action is being implemented. Yet, the optimal sharing rule reflects precisely the pricing of inference.

\[
\left( 1 - \frac{p_{ik}}{p_{ij}} \right) > 0 \text{ if } p_{ik} < p_{ij}
\]
\[
\left( 1 - \frac{p_{ik}}{p_{ij}} \right) < 0 \text{ if } p_{ik} > p_{ij}
\]

The monotonicity in the optimal contract is then established if we can show that the downward binding constraints receive more weight than the upward binding constraints. In fact, we next give conditions which will guarantee that \( \mu_k = 0 \) for all \( k > j \).

10.7.3 Convexity

The cumulative distribution function of the outcome is convex in \( a \): for \( a_j < a_k < a_l \) and \( \lambda \in [0, 1] \) such that for
\[
a_k = \lambda a_j + (1 - \lambda) a_l
\]
we have
$P_{ik} \leq \lambda P_{ij} + (1 - \lambda) P_{il}$

The convexity condition then states that the return from action are stochastically decreasing. Next we show in two steps that monotone likelihood and convexity are sufficient to establish monotonicity. The argument is in two steps and let $a_j$ be the optimal action.

First, we know that for some $j < k$, $\mu_j > 0$. Suppose not, then the optimal choice of $a_k$ would be same if we were to consider $A$ or $\{a_k, ..., a_J\}$. But relative to $\{a_k, ..., a_J\}$ we know that $a_k$ is the least cost action which can be implemented with a constant wage schedule. But extending the argument to $A$, we know that a constant wage schedule can only support the least cost action $a_1$.

Second, consider $\{a_1, ..., a_k\}$. Then $a_k$ is the most costliest action and by MLRP $w_i$ is increasing in $i$. We show that $a_k$ remains the optimal choice for the agent when we extend his action to $A$. The proof is again by contradiction. Suppose there is $l > k$ s.th.

$$\sum_{i=1}^{l} u(w_i)p_{ik} - a_k < \sum_{i=1}^{l} u(w_i)p_{il} - a_l,$$

and let $j < k$ be an action where $\mu_k > 0$ and hence

$$\sum_{i=1}^{l} u(w_i)p_{ik} - a_k = \sum_{i=1}^{l} u(w_i)p_{ij} - a_j.$$

Then there exists $\lambda \in [0, 1]$ such that

$$a_k = \lambda a_j + (1 - \lambda) a_l$$

and we can apply convexity, by first rewriting $p_{ik}$ and using the cumulative distribution function

$$\sum_{i=1}^{l} \left( u(w_i) - u(w_{i+1}) \right) P_{ik} + u(w_I) - a_k$$

By rewriting the expected value also for $a_j < a_k < a_l$, we obtain

$$\sum_{i=1}^{l} \left( u(w_i) - u(w_{i+1}) \right) P_{ik} + u(w_I) - a_k \geq$$

$$\lambda \left( \sum_{i=1}^{l} \left( u(w_i) - u(w_{i+1}) \right) P_{kj} + u(w_I) - a_j \right)$$

$$+ (1 - \lambda) \left( \sum_{i=1}^{l} \left( u(w_i) - u(w_{i+1}) \right) P_{il} + u(w_I) - a_l \right)$$

The later is of course equivalent to

$$\lambda \left( \sum_{i=1}^{l} u(w_i)p_{ij} - a_j \right) + (1 - \lambda) \left( \sum_{i=1}^{l} u(w_i)p_{il} - a_l \right)$$

The inequality follows from the convexity and the fact that we could assume $w_i$ to be monotone, so that $u(w_i) - u(w_{i+1}) \leq 0$.

A final comment on the monotonicity of the transfer function: Perhaps a good reason why the function $w_i$ must be monotonic is that the agent may be able to artificially reduce performance. In the above example, he would then artificially lower it from $x_j$ to $x_i$ whenever the outcome is $x_j$. 
10.8 Information and Contract

10.8.1 Informativeness

We now ask how does the informativeness of a signal structure affect the payoff of the agent. Consider a stochastic matrix $Q$ which modifies the probabilities $p_{ij}$ such that

$$
\hat{p}_{kj} = \sum_{i=1}^{I} q_{ki} p_{ij},
$$

such that $q_{ki} \geq 0$ and

$$
\sum_{k=1}^{K} q_{ki} = 1
$$

for all $i$. The matrix is called stochastic as all its entries are nonnegative and every column adds up to one. Condition (10.11) is a generalization of the following idea. Consider the information structure given by $p_{ij}$. Each time the signal $x_i$ is observed, it is garbled by a stochastic mechanism that is independent of the action $a_j$, which may be interpreted here as a state. It is transformed into a vector of signals $\hat{x}_k$. The term $q_{ki}$ can be interpreted as a conditional probability of $\hat{x}_k$ given $x_i$. Clearly $\hat{p}_{kj}$ are also probabilities.

Suppose the outcomes undergo a similar transformation so that

$$
\sum_{k=1}^{K} \hat{x}_k \hat{p}_{kj} = \sum_{i=1}^{I} x_i p_{ij},
$$

for all $j$. Thus the expected surplus stays the same for every action. It can be achieved (in matrix language) by assuming that $\hat{x} = Q^{-1} x$.

In statistical terms, inferences drawn on $a_j$ after observing $\hat{x}$ with probabilities $\hat{p}$ will be less precise than the information drawn from observing $x$ with probability $p$.

Consider now in the $(\hat{p}, \hat{x})$ model a wage schedule $\hat{w}_i$ which implements $a_j$. Then we find a new wage schedule for the $(p, x)$ problem based on $\hat{w}$ such that $a_j$ is also implementable in $(p, x)$ problem, yet we will see that it involves less risk imposed on the agent and hence is better for the principal. Let

$$
u(w_i) = \sum_{k=1}^{K} q_{ki} u(\hat{w}_k)
$$

(10.12)

We can then write

$$
\sum_{i=1}^{I} p_{ij} u(w_i) = \sum_{i=1}^{I} p_{ij} \left( \sum_{k=1}^{K} q_{ki} u(\hat{w}_k) \right)
$$

$$
= \sum_{k=1}^{K} \hat{p}_{kj} u(\hat{w}_k)
$$

But the implementation in $(p, x)$ is less costly for the principal than the one in $(\hat{p}, \hat{x})$ which proves the result. But (10.12) immediately indicates that it is less costly for the principal as the agent has a concave utility function. This shows in particular that the optimal contract is only a second best solution as the first best contains no additional noise.
10.8.2 Additional Signals

Suppose besides the outcome, the principal can observe some other signal, say \( y \in Y = \{ y_1, ..., y_l, ..., y_L \} \), which could inform him about the performance of the agent. The form of the optimal contract can then be verified to be

\[
\frac{1}{w^i(w_i)} = \lambda + \sum_{k \neq j} \mu_k \left( 1 - \frac{p_{ik}}{p_{ij}} \right).
\]

Thus the contract should integrate the additional signal \( y \) if there exists some \( x_i \) and \( y_l \) such that for \( a_j \) and \( a_k \)

\[
\frac{p_{ik}^l}{p_{ij}^l} \neq \frac{p_{ik}}{p_{ij}},
\]

but the inequality simply says that

\( x \)

is not a sufficient statistic for \((x, y)\) as it would be if we could write the conditional probability as follows

\[
f(x_i, y_l | a_j) = h(x_i, y_l) g(x_i | a_j).
\]

Thus, Hart-Holmström [1986] write:

The additional signal \( s \) will necessarily enter an optimal contract if and only if it affects the posterior assessment of what the agent did; or perhaps more accurately if and only if \( s \) influences the likelihood ratio.

10.9 Linear contracts with normally distributed performance and exponential utility

This constitutes another “natural” simple case. Performance is assumed to satisfy \( x = a + \epsilon \), where \( \epsilon \) is normally distributed with zero mean and variance \( \sigma^2 \). The principal is assumed to be risk neutral, while the agent has a utility function:

\[
U(w, a) = -e^{-r(w-c(a))}
\]

where \( r \) is the (constant) degree of absolute risk aversion \( (r = -U''/U') \), and \( c(a) = \frac{1}{2}ca^2 \).

We restrict attention to linear contracts:

\[
w = \alpha x + \beta.
\]

A principal trying to maximize his expected payoff will solve:

\[
\max_{a, \alpha, \beta} E(x - w)
\]

subject to:
\[ E(-e^{-r(w-c(a))}) \geq U(\bar{w}) \]

and

\[ a \in \arg\max_a E(-e^{-r(w-c(a))}) \]

where \( U(\bar{w}) \) is the default utility level of the agent, and \( \bar{w} \) is thus its certain monetary equivalent.

### 10.9.1 Certainty Equivalent

The certainty equivalent \( w \) of random variable \( x \) is defined as follows

\[ u(w) = E[u(x)] \]

The certainty equivalent of a normally distributed random variable \( x \) under CARA preferences, hence \( w \) which solves

\[ -e^{-rw} = E(-e^{-rx}) \]

has a particularly simple form, namely

\[ w = \mathbb{E}[x] - \frac{1}{2}r\sigma^2 \quad (10.13) \]

The difference between the mean of random variable and its certain equivalent is referred to as the risk premium:

\[ \frac{1}{2}r\sigma^2 = \mathbb{E}[x] - w \]

### 10.9.2 Rewriting Incentive and Participation Constraints

Maximizing expected utility with respect to \( a \) is equivalent to maximizing the certainty equivalent wealth \( \tilde{w}(a) \) with respect to \( a \), where \( \tilde{w}(a) \) is defined by

\[ -e^{-r\tilde{w}(a)} = E(-e^{-r(w-c(a))}) \]

Hence, the optimization problem of the agent is equivalent to:

\[ a \in \arg\max \{ \tilde{w}(a) \} = \]

\[ \in \arg\max \left\{ \alpha a + \beta - \frac{1}{2}ca^2 - \frac{r}{2}a^2\sigma^2 \right\} \]

which yields

\[ a^* = \frac{\alpha}{c} \]

Inserting \( a^* \) into the participation constraint

\[ \alpha \frac{\alpha}{c} + \beta - \frac{1}{2}c \left( \frac{\alpha}{c} \right)^2 - \frac{r}{2}a^2\sigma^2 = \bar{w} \]
yields an expression for $\beta$,

$$
\beta = \bar{w} + \frac{r}{2} \alpha^2 \sigma^2 - \frac{1}{2} \frac{\alpha^2}{c}
$$

This gives us the agent’s effort for any performance incentive $\alpha$. The principal then solves:

$$
\max_{\alpha} \left\{ \frac{\alpha}{c} - (\bar{w} + \frac{r}{2} \alpha^2 \sigma^2 + \frac{1}{2} \frac{\alpha^2}{c}) \right\}
$$

The first order conditions are

$$
\frac{1}{c} - (r \sigma^2 + \frac{\alpha}{c}) = 0,
$$

which yields:

$$
\alpha^* = \frac{1}{1 + r \sigma^2}
$$

Effort and the variable compensation component thus go down when $c$ (cost of effort), $r$ (degree of risk aversion), and $\sigma^2$ (randomness of performance) go up, which is intuitive. The constant part of the compensation will eventually decrease as well as $r$, $c$ or $\sigma^2$ become large, as

$$
\beta = \bar{w} + \left( \frac{1}{2} \frac{r \sigma^2 - \frac{1}{c}}{(1 + r \sigma^2)^2} \right).
$$

10.10 Readings

The following are classic papers in the literature on moral hazard: ?, ?, ?, ?, ?, ?, ??.

[LECTURE 7]
Part V

Mechanism Design
11. Introduction

Mechanism design. In this lecture and for the remainder of this term we will look at a special class of games of incomplete information, namely games of mechanism design. Examples of these games include: (i) monopolistic price discrimination, (ii) optimal taxation, (iii) the design of auctions, and (iv) the provision of public goods.

The adverse selection model when extended to many agent will ultimately form the theoretical centerpiece of the lectures. In the later case we refer to it as the

many informed agents:
mechanism design
(social choice)

“mechanism design” problem which encompasses auctions, bilateral trade, public good provision, taxation and many other general asymmetric information problems. The general mechanism design problem can be represented in a commuting diagram:

The upper part of the diagram represents the social choice problem, where the principal has to decide on the allocation contingent on the types of the agent. The lower part represents the implementation problem, where the principals attempts to elicit the information by announcing an outcome function which maps the message (information) he receives from the privately informed agents into allocation. From the point of view of the privately informed agents, this then represents a Bayesian game in which they can influence the outcome through their message.

Information extraction, truth telling, and incentives. All these examples have in common there is a “principal” (social planner, monopolist, etc.) who would like to condition her action on some information that is privately known by the other players, called “agents”. The principal could simply ask the agents for their information, but they will not report truthfully unless the principal gives them an incentive to do so, either by monetary payments or with some other instruments that she controls. Since providing these incentives is costly, the principal faces a trade-off that often results in an efficient allocation.

Principal’s objective. The distinguishing characteristic of the mechanism design approach is that the principal is assumed to choose the mechanism that maximizes her expected utility, as opposed to using a particular mechanism for historical or institutional reasons. (Caveat: Since the objective function of the
principal could just be the social welfare of the agents, the range of problem which can be studies is rather large.)

**Single agent.** Many applications of mechanism design consider games with a single agent (or equivalent with a continuum of infinitesimal small agents). For example, in a second degree price discrimination by a monopolist, the monopolist has incomplete information about the consumer’s willingness to pay for the good. By offering a price quantity schedules he attempts to extract the surplus from the buyer.

**Many agents.** Mechanism design can also be applied to games with several agents. The auctions we studied are such an example. In a public good problem, the government has to decide whether to supply a public good or not. But as it has incomplete information about how much the good is valued by the consumers, it has to design a scheme which determines the provision of the public good as well as transfer to be paid as a function of the announced willingness to pay for the public good.

**Three steps.** Mechanism design is typically studied as three-step game of incomplete information, where the agents' type - e.g. their willingness to pay - are private information. In step 1, the principal designs a “mechanism”, “contract”, or “incentive scheme”. A mechanism is a game where the agents send some costless message to the principal, and as a function of that message the principal selects an allocation. In step 2, the agents simultaneously decide whether to accept or reject the mechanism. An agent who rejects the mechanism gets some exogenously specified “reservation utility”. In step 3, the agents who accepted the mechanism play the game specified by the mechanism.

**Maximization subject to constraints.** The study of mechanism design problems is therefore formally a maximization problem for the principal subject to two classes of constraints. The first class is called the “participation” or “individual rationality” constraint, which insures the participation of the agent. The second class are the constraints related to truth-telling, what we will call “incentive compatibility” constraints.

**Efficiency and distortions.** An important focus in mechanism design will be how these two set of constraints in their interaction can prevent efficient outcomes to arise: (i) which allocation y can be implemented, i.e. is incentive compatible?, (ii) what is the optimal choice among incentive compatible mechanisms?, where optimal could be efficient, revenue maximizing.
12. Adverse selection: Mechanism Design with One Agent

Often also called self-selection, or “screening”. In insurance economics, if a insurance company offers a
tariff tailored to the average population, the tariff will only be accepted by those with higher than average
risk.

12.1 Monopolistic Price Discrimination with Binary Types

A simple model of wine merchant and wine buyer, who could either have a coarse or a sophisticated taste,
which is unobservable to the merchant. What qualities should the merchant offer and at what price?
The model is given by the utility function of the buyer, which is

\[ v(\theta_i, q_i, t_i) = u(\theta_i, q_i) - t_i = \theta_i q_i - t_i, \quad i \in \{l, h\} \]  

(12.1)

where \( \theta_i \) represent the marginal willingness to pay for quality \( q_i \) and \( t_i \) is the transfer (price) buyer \( i \) has
to pay for the quality \( q_i \). The taste parameters \( \theta_i \) satisfies

\[ 0 < \theta_l < \theta_h < \infty. \]  

(12.2)

The cost of producing quality \( q \geq 0 \) is given by

\[ c(q) \geq 0, c'(q) > 0, c''(q) > 0. \]  

(12.3)

The ex-ante (prior) probability that the buyer has a high willingness to pay is given by

\[ p = \Pr(\theta_i = \theta_h) \]

We also observe that the difference in utility for the high and low valuation buyer for any given quality \( q \)

\[ u(\theta_h, q) - u(\theta_l, q) \]

is increasing in \( q \). (This is know as the Spence-Mirrlees sorting condition.) If the taste parameter \( \theta_i \) were
a continuous variable, the sorting condition could be written in terms of the second cross derivative:

\[ \frac{\partial^2 u(\theta, q)}{\partial \theta \partial q} > 0, \]

which states that taste \( \theta \) and quality \( q \) are complements. The profit for the seller from a bundle \( (q, t) \) is
given by

\[ \pi(t, q) = t - c(q) \]
12. Adverse selection: Mechanism Design with One Agent

12.1.1 First Best

Consider first the nature of the socially optimal solution. As different types have different preferences, they should consume different qualities. The social surplus for each type can be maximized separately by solving

$$\max_{q_i} \{\theta_i q_i - c(q_i)\}$$

and the first order conditions yield:

$$q_i = q_i^* \iff c'(q_i^*) = \theta_i \Rightarrow q_i^* < q_h^*.$$  

The efficient solution is the equilibrium outcome if either the monopolist can perfectly discriminate between the types (first degree price discrimination) or if there is perfect competition. The two outcomes differ only in terms of the distribution of the social surplus. With a perfectly discriminating monopolist, the monopolist sets

$$t_i = \theta_i q_i$$  \hspace{1cm} (12.4)

and then solves for each type separately:

$$\max_{\{t_i, q_i\}} \pi(t_i, q_i) \iff \max_{\{t_i, q_i\}} \{\theta_i q_i - c(q_i)\},$$

using (12.4). Likewise with perfect competition, the sellers will break even, get zero profit and set prices at

$$t_i = c(q_i^*)$$

in which case the buyer will get all the surplus.

12.1.2 Second Best: Asymmetric information

Consider next the situation under asymmetric information. It is verified immediately that perfect discrimination is now impossible as

$$\theta_h q_h^* - t_h^* = (\theta_h - \theta_l) q_h^* > 0 = \theta_l q_l^* - t_l$$ \hspace{1cm} (12.5)

but sorting is possible. The problem for the monopolist is now

$$\max_{\{t_l, q_l, t_h, q_h\}} (1 - \pi) t_l - c(q_l) + \pi (t_h - c(q_h)))$$

subject to the individual rationality constraint for every type

$$\theta_i q_i - t_i \geq 0 \hspace{1cm} (IR_i)$$ \hspace{1cm} (12.7)

and the incentive compatibility constraint

$$\theta_i q_i - t_i \geq \theta_j q_j - t_j \hspace{1cm} (IC_i)$$ \hspace{1cm} (12.8)

The question is then how to separate. We will show that the binding constraint are $IR_l$ and $IC_h$, whereas the remaining constraints are not binding. We then solve for $t_l$ and $t_h$, which in turn allows us to solve for $q_h$, and leaves us with an unconstrained problem for $q_l$. Thus we want to show

$$(i) \hspace{0.1cm} IR_l \hspace{0.1cm} \text{binding,} \hspace{0.1cm} (ii) \hspace{0.1cm} IC_h \hspace{0.1cm} \text{binding,} \hspace{0.1cm} (iii) \hspace{0.1cm} \hat{q}_h \geq \hat{q}_l \hspace{0.1cm} (iv) \hspace{0.1cm} \hat{q}_h = q_h^*$$ \hspace{1cm} (12.9)
Consider (i). We argue by contradiction. As
\[
\theta_h q_h - t_h \geq \max_{l \in C_h} \theta_h q_l - t_l \geq \theta_l q_l - t_l
\]  
(12.10)
suppose that \(\theta_h q_h - t_h > 0\), then we could increase \(t_h, t_l\) by an equal amount, satisfy all the constraints and increase the profits of the seller. Contradiction.

Consider (ii) Suppose not, then as
\[
\theta_h q_h - t_h > \max_{l \in C_h} \theta_h q_l - t_l \geq \theta_l q_l - t_l = 0
\]  
(12.11)
and thus \(t_h\) could be increased, again increasing the profit of the seller.

(iii) Adding up the incentive constraints gives us \((IC_l) + (IC_l)\)
\[
\theta_h (q_h - q_l) \geq \theta_l (q_h - q_l)
\]  
(12.12)
and since:
\[
\theta_h > \theta_l \Rightarrow \hat{q}_h - \hat{q}_l \geq 0.
\]  
(12.13)
Next we show that \(IC_l\) can be neglected as
\[
t_h - t_l = \theta_h (q_h - q_l) \geq \theta_l (q_h - q_l). \tag{12.14}
\]
This allows to say that the equilibrium transfers are going to be
\[
t_l = \theta_l q_l
\]  
(12.15)
and
\[
t_h - t_L = \theta_h (q_h - q_L) \Rightarrow t_h = \theta_h (q_h - q_L) + \theta_l q_l.
\]
Using the transfer, it is immediate that
\[
\hat{q}_h = q_h^*
\]
and we can solve for the last remaining variable, \(\hat{q}_l\). 
\[
\max_{q_l} \{(1 - p) (\theta_l q_l - c(q_l)) + p (\theta_h (q_h^* - q_L) + \theta_l q_l - c(q_h^*)))\}
\]
but as \(q_h^*\) is just as constant, the optimal solution is independent of constant terms and we can simplify the expression to:
\[
\max_{q_l} \{(1 - p) (\theta_l q_l - c(q_l)) - p (\theta_h - \theta_l) q_l\}
\]
Dividing by \((1 - p)\) we get
\[
\max_{q_l} \left\{ \theta_l q_l - c(q_l) - \frac{p}{1 - p} (\theta_h - \theta_l) q_l \right\}
\]
for which the first order conditions are
\[
\theta_l - c'(q_l) - \frac{p}{1 - p} (\theta_h - \theta_l) q_l = 0
\]
This immediately implies that the solution \(\hat{q}_l\):
12. Adverse selection: Mechanism Design with One Agent

\[ c'(q_l) < \theta_1 \iff q_l < q^*_l \]

and the quality supply to the low valuation buyer is inefficiently low (with the possibility of complete exclusion).

Consider next the information rent for the high valuation buyer, it is

\[ I(q_l) = (\theta_h - \theta_l) q_l \]

and therefore the rent is increasing in \( q_l \) which is the motivation for the seller to depress the quality supply to the low end of the market.

The material is explicated in \(^7\), Chapter 2.

[ Lecture 8 ]

12.2 Continuous type model

In this lecture, we consider the adverse selection model with a continuous type model and a single agent. This will set the state for the multiple agent model.

12.2.1 Information Rent

Consider the situation in the binary type model. The allocation problem for the agent and principal can well be represented in a simple diagram.

The information rent is then

\[ I(\theta_h) = (\theta_h - \theta_l) q_l \]

and it also represents the difference between the equilibrium utility of low and high type agent, or

\[ I(\theta_h) = U(\theta_h) - U(\theta_l) \]

If we extend the model and think of more than two types, then we can think of the information rent as resulting from two adjacent types, say \( \theta_k \) and \( \theta_{k-1} \). The information rent is then given by

\[ I(\theta_k) = (\theta_k - \theta_{k-1}) q_{k-1} \]

As the information rent is also the difference between the agent’s net utility, we have

\[ I(\theta_k) = U(\theta_k) - U(\theta_{k-1}) \]

well as we see not quite, but more precisely:

\[ I(\theta_k) = U(\theta_k | \theta_k) - U(\theta_{k-1} | \theta_k) \]

where

\[ U(\theta_k | \theta_l) \]

denotes in general the utility of an agent of (true) type \( \theta_l \), when he announces and pretends to be of type \( \theta_k \). As we then extend the analysis to a continuous type model, we could ask how the net utility of the agent of type \( \theta_k \) evolves as a function of \( k \). In other words as \( \theta_{k-1} \to \theta_k \):

\[ \frac{U(\theta_k | \theta_k) - U(\theta_{k-1} | \theta_k)}{\theta_k - \theta_{k-1}} = \frac{(\theta_k - \theta_{k-1})}{(\theta_k - \theta_{k-1})} q_{k-1} \]
and assuming continuity:
\[ q_{k-1} \rightarrow q_k, \]
we get
\[ U' (\theta) = q (\theta) \]
which is identical using the specific preferences of our model to
\[ U' (\theta) = \frac{\partial u}{\partial \theta} (q (\theta), \theta). \]
Thus we have a description of the equilibrium utility as a function of \( q(\theta) \) alone rather than \((q(\theta), t(\theta))\).
The pursuit of the adverse selection problem along this line is often referred to as “Mirrlees’ trick”.

12.2.2 Utilities

The preferences of the agent are described by
\[ U(x, \theta, t) = u(x, \theta) - t \]
and the principals
\[ V(x, \theta, t) = v(x, \theta) + t \]
and \( u, v \in C^2 \). The type space \( \theta \in \Theta \subset \mathbb{R}_+ \).
The social surplus is given by
\[ S(x, \theta) = u(x, \theta) + v(x, \theta). \]
The uncertainty about the type of the agent is given by:
\[ f(\theta), F(\theta). \]
We shall assume the (strict) Spence-Mirrlees conditions
\[ \frac{\partial u}{\partial \theta} > 0, \quad \frac{\partial^2 u}{\partial \theta^2} > 0. \]

12.2.3 Incentive Compatibility

**Definition 12.2.1.** An allocation is a mapping
\[ \theta \rightarrow y(\theta) = (x(\theta), t(\theta)). \]

**Definition 12.2.2.** An allocation is implementable if \( y = (x, t) \) satisfies truthtelling:
\[ u(x(\theta), \theta) - t(\theta) \geq u\left(x\left(\hat{\theta}\right), \theta\right) - t\left(\hat{\theta}\right), \quad \forall \theta, \hat{\theta} \in \Theta. \]

**Definition 12.2.3.** The truthtelling net utility is denoted by
\[ U(\theta) \triangleq U(\theta | \theta) = u(x(\theta), \theta) - t(\theta), \quad (12.16) \]
and the net utility for agent \( \theta \), misreporting by telling \( \hat{\theta} \) is denoted by
\[ U\left(\hat{\theta} | \theta\right) = u\left(x\left(\hat{\theta}\right), \theta\right) - t\left(\hat{\theta}\right). \]
We are interested in (i) describing which contracts can be implemented and (ii) in describing optimal contracts.

**Theorem 12.2.1.** The direct mechanism $y(\theta) = (x(\theta), t(\theta))$ is incentive compatible if and only if:

1. The truthtelling utility is described by:
   \[
   U(\theta) - U(0) = \int_0^\theta u(x(s), s) \, ds,
   \]  
   (12.17)

2. $x(s)$ is nondecreasing.

**Remark 12.2.1.** The condition (12.17) can be restated in terms of first order conditions:

\[
\frac{dU}{d\theta} = \frac{\partial u}{\partial y} \frac{dy}{d\theta} + \frac{\partial u}{\partial \theta} = 0.
\]  
(12.18)

**Proof.** Necessity. Suppose truthtelling holds, or

\[
U(\theta) \geq U\left(\hat{\theta} | \theta\right), \forall \theta, \hat{\theta},
\]

which is:

\[
U(\theta) \geq U\left(\hat{\theta} | \theta\right) \triangleq U\left(\hat{\theta}\right) + u(x(\theta), \theta) - u(x(\hat{\theta}), \hat{\theta}),
\]

and thus

\[
U(\theta) - U\left(\hat{\theta}\right) \geq u(x(\theta), \theta) - u(x(\hat{\theta}), \hat{\theta}).
\]  
(12.19)

A symmetric condition gives us

\[
U\left(\hat{\theta}\right) - U(\theta) \geq u\left(x(\hat{\theta}), \theta\right) - u\left(x(\theta), \hat{\theta}\right).
\]  
(12.20)

Combining (12.19) and (12.20), we get:

\[
u(x(\theta), \theta) - u\left(x(\hat{\theta}), \theta\right) \geq U(\theta) - U\left(\hat{\theta}\right) \geq u\left(x(\theta), \hat{\theta}\right) - u\left(x(\hat{\theta}), \theta\right).
\]

Suppose without loss of generality that $\theta > \hat{\theta}$, then monotonicity of $x(\theta)$ is immediate from

\[
\frac{\partial^2 u}{\partial \theta \partial x}.
\]

Dividing by $\theta - \hat{\theta}$, and taking the limit as

\[
\theta \to \hat{\theta}
\]

at all points of continuity of $x(\theta)$ yields

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

and thus we have the representation of $U$. As $x(\theta)$ is nondecreasing, it can only have a countable number of discontinuities, which have Lebesgue measure zero, and hence the integral representation is valid, independent of continuity properties of $x(\theta)$, which is in contrast with the representation of incentive compatibility by the first order condition (12.18).
Sufficiency. Suppose not, then there exists \( \theta \) and \( \hat{\theta} \) s.th.

\[
U(\hat{\theta} | \theta) \geq U(\theta).
\] (12.21)

Suppose without loss of generality that \( \theta \geq \hat{\theta} \). The inequality (12.21) implies that the inequality in (12.19) is reversed:

\[
u(x(\hat{\theta}), \theta) - u(x(\hat{\theta}), \hat{\theta}) > U(\theta) - U(\hat{\theta})
\]

integrating and using (12.17) we get

\[
\int_{\theta}^{\hat{\theta}} u_s(x(\theta), s) \, ds > \int_{\theta}^{\hat{\theta}} u_s(x(s), s) \, ds
\]

and rearranging

\[
\int_{\theta}^{\hat{\theta}} \left[ u_s(x(\theta), s) - u_s(x(s), s) \right] \, ds > 0
\] (12.22)

but since

\[
\frac{\partial^2 u}{\partial \theta \partial x} > 0
\]

and the monotonicity condition implied that this is not possible, and hence we have the desired contradiction.

[Lecture 9]

### 12.3 Optimal Contracts

We can now describe the problem for the principal.

\[
\max_{y(\theta)} \mathbb{E}_\theta [v(x(\theta), \theta) + t(\theta)]
\]

subject to

\[
u(x(\theta), \theta) - t(\theta) \geq u(x(\hat{\theta}), \theta) - t(\hat{\theta}), \quad \forall \theta, \hat{\theta}
\]

and

\[
u(x(\theta), \theta) - t(\theta) \geq \bar{u}.
\]

The central idea is that we restate the incentive conditions for the agents by the truth-telling utility as derived in Theorem 12.2.1.
Adverse selection: Mechanism Design with One Agent

12.3.1 Optimality Conditions

In this manner, we can omit the transfer payments from the control problem and concentrate on the optimal choice of $x$ as follows:

$$\max_{x(\theta)} \mathbb{E}_{\theta} [S(x(\theta), \theta) - U(\theta)]$$

(12.23)

subject to

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

(12.24)

and

$x(\theta)$ nondecreasing

(12.25)

and

$U(\theta) \geq \bar{u}$,

(12.26)

which is the individual rationality constraint. Next we simplify the problem by using integration by parts. Recall, that integration by parts, used in the following form

$$\int dU(1 - F) = U(1 - F) + \int Uf$$

and hence

$$\int Uf = -U(1 - F) + \int dU(1 - F)$$

As

$$\mathbb{E}_{\theta} [U(\theta)] = \int_0^1 U(\theta) f(\theta) d\theta = -[U(\theta)(1 - F(\theta))]_0^1 + \int_0^1 \frac{dU(\theta)}{d\theta} \frac{1 - F(\theta)}{f(\theta)} f(\theta) d\theta$$

(12.27)

we can rewrite (12.23) and using (12.24), we get

$$\max_{x(\theta)} \mathbb{E}_{\theta} \left[ S(x(\theta), \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(x(\theta), \theta) - U(0) \right]$$

subject to (12.25) and (12.26).and thus we removed the transfers out of the program.

12.3.2 Pointwise Optimization

Consider the optimization problem for the principal and omitting the monotonicity condition, we get

$$\Phi(x, \theta) = S(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(x, \theta).$$

(12.29)

Assumption 12.3.1 $\Phi(x, \theta)$ is quasiconcave and has a unique interior maximum in $x \in \mathbb{R}_+$ for all $\theta \in \Theta$.

Theorem 12.3.2. Suppose that Assumption 12.3.1 holds, then the relaxed problem is solved at $y = y(\theta)$ if:
1. $\Phi_x(x(\theta), \theta) = 0$.
2. The transfer $t(\theta)$ satisfy

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

Then $y = (x, t)$ solve the relaxed program.

The idea of the proof is simple. By integration by parts we removed the transfers. Then we can choose $x$ to maximize the objective function and later choose $t$ to solve the differential equation. Given that $\partial u/\partial \theta > 0$, the IR contract can be related as

$$U(0) \geq \bar{u}.$$ 

It remains to show that $x$ is nondecreasing.

**Assumption 12.3.3** $\partial^2 \Phi(x, \theta)/\partial x \partial \theta \geq 0$.

**Theorem 12.3.4.** Suppose that assumptions 12.3.1 and 12.3.3 are satisfied. Then the solution to the relaxed program satisfies the original solution in the unrelaxed program.

**Proof.** Differentiating

$$\Phi_x(x(\theta), \theta) = 0 \text{ for all } \theta$$

and hence we obtain

$$\frac{dx(\theta)}{d\theta} = \frac{\Phi_{xx}}{\Phi_{xx}}$$

and as

$$\Phi_{xx} \leq 0$$

has to be satisfied locally, we know that $x(\theta)$ has to be increasing, which concludes the proof.

Next we interpret the first order conditions and obtain first results:

$$S_x(x(\theta), \theta) = \frac{1 - F(\theta)}{f(\theta)} u_{xx}(x(\theta), \theta) \geq 0$$

implies efficient provision for $\theta = 1$ and underprovision of the contracted activity for all but the highest type $\theta = 1$. To see that $\hat{x}(\theta) < x^*(\theta)$, we argue by contradiction. Suppose not, or $\hat{x}(\theta) \geq x^*(\theta)$. Observe first that for all $x$ and $\theta < 1$

$$\frac{\partial \Phi(x, \theta)}{\partial x} < \frac{\partial S(x, \theta)}{\partial x}$$

by:

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

Thus at $x^*(\theta)$, by definition

$$\frac{\partial S(x^*(\theta), \theta)}{\partial x} = 0$$
and using (12.30)
\[
\frac{\partial \Phi (x^* (\theta), \theta)}{\partial x} < \frac{\partial S (x^* (\theta), \theta)}{\partial x} = 0
\]

But as \( \Phi (x, \theta) \) is quasiconcave the derivative with respect to \( x \) satisfies the (downward single crossing condition), which implies that for
\[
\frac{\partial \Phi (x (\theta), \theta)}{\partial x} = 0
\]
to hold \( x (\theta) < x^* (\theta) \).

The first order condition can be written as to represent the trade-off:
\[
f (\theta) S_x (x (\theta), \theta) = [1 - F (\theta)] u_{x \theta} (x (\theta), \theta).
\]

Consider then the \textit{monopoly interpretation}, where:
\[
\max_p (1 - F (p)) (p - c)
\]
and the associated first order condition is given by:
\[
f (p) (p - c) = 1 - F (p)
\]

The material in this lectures is bases on ?.

[\textsc{Lecture 10}]
13. Mechanism Design Problem with Many Agents

13.1 Introduction

Mechanism design explores the means of implementing a given allocation of available resources when the information is dispensed in the economy. In general two questions arise:

- can \( y(\theta) \) be implemented incentive compatible?
- what is the optimal choice among incentive-compatible mechanisms?

Today we shall introduce several general concepts:

- revelation principle
- efficiency
- quasi-linear utility

We emphasize that we now look at it as a static game of the agents and hence the designer is acting non-strategically. The optimal design is a two stage problem. Mechanism design takes a different perspective on the game and the design of games.

13.2 Model

Consider a setting with \( I \) agents, \( \mathcal{I} = \{1, ..., I\} \). The principal, mechanism designer, center, is endowed with subscript 0. They must make a collective choice among a set of possible allocations \( Y \).

Each agent observes privately a signal \( \theta_i \in \Theta_i \), his type, that determines his preferences over \( Y \), described by a utility function \( u_i(y, \theta) \) for all \( i \in \mathcal{I} \). The prior distribution over types \( P(\theta) \) is assumed to be common knowledge. In general the type could contain any information agent \( i \) possesses about his preferences, but it could also contain information about his neighbor, competitors, and alike.

**Definition 13.2.1.** A social choice function is a mapping:

\[
f: \Theta_1 \times ... \times \Theta_I \rightarrow Y.
\]

The problem is that \( \theta = (\theta_1, ..., \theta_I) \) is not publicly observable when the allocation \( y \) is to be decided. However each agent can send a message to the principal \( m_i \in M_i \). The message space can be arbitrary, with

\[
M = \prod_{i=1}^{I} M_i.
\]

When the agents transmit a message \( m \), the center will choose a social allocation \( y \in Y \) as a function of the message received.

**Definition 13.2.2.** An outcome function is a mapping

\[
g: M_1 \times ... \times M_I \rightarrow Y.
\]
Each agent transmits the message independent and simultaneous with all other agents. Thus a mechanism defines a game with incomplete information for which must choose an equilibrium concept, denoted by $c$. A strategy for agent $i$ is a function

$$m_i : \Theta_i \rightarrow M_i.$$ 

**Definition 13.2.3.** A mechanism $\Gamma = (M_1, \ldots, M_I, g(\cdot))$ is a collection of strategy sets $(M_1, \ldots, M_I)$ and an outcome function

$$g : M_1 \times \ldots \times M_I \rightarrow Y.$$ 

With a slight abuse of notation, we use the same notation, $M_i$, for the set of strategies and their range for a particular agent $i$. Graphically, we have the following commuting diagram:

\[
\begin{array}{ccc}
\Theta_1 \times \ldots \times \Theta_I & \xrightarrow{f(\cdot)} & Y \\
\downarrow m_{\theta,c}(\cdot) & & \\
M_1 \times \ldots \times M_I & \xrightarrow{g(\cdot)} & \\
\end{array}
\]

where $m_{\theta,c}$ is the mapping that associates to every $I$-tuple of true characteristics $\theta$, the equilibrium messages of this game for the equilibrium concept $c$. The strategy space of each individual agent is often called the message space.

### 13.3 Mechanism as a Game

The central question we would like to ask whether a particular objective function, or social choice function $f(\cdot)$ can be realized by the game which is the mechanism.

**Definition 13.3.1.** A mechanism $\Gamma = (M_1, \ldots, M_I, g(\cdot))$ implements the social choice function $f(\cdot)$ if there is an equilibrium profile

$$(m^*_1(\theta_1), \ldots, m^*_I(\theta_I))$$

of the game induced by $\Gamma$ such that

$$g(m^*_1(\theta_1), \ldots, m^*_I(\theta_I)) = f(\theta_1, \ldots, \theta_I).$$

The identification of implementable social choice function is at first glance a complex problem because we have to consider all possible mechanism $g(\cdot)$ on all possible domains of strategies $M$. However a celebrated result (valid for all of the implementation versions above), the revelation principle, simplifies the task.

**Definition 13.3.2.** A mechanism is direct if $M_i = \Theta_i$ and $g(\cdot) = f(\cdot)$ for all $i$.

**Definition 13.3.3.** A direct mechanism is said to be revealing if $m_i(\theta_i) = \theta_i$ for each $\theta \in \Theta$.

**Definition 13.3.4.** The social choice function $f(\cdot)$ is truthfully implementable (or incentive compatible) if the direct revelation mechanism

$$\Gamma = (\Theta, f(\cdot))$$

has an equilibrium $(m^*_1(\theta_1), \ldots, m^*_I(\theta_I))$ in which $m^*_i(\theta_i) = \theta_i$ for all $\theta_i \in \Theta_i$, for all $i$. 

We shall test these notions with our example of adverse selection with a single buyer. Suppose the message space is

\[ M \]

and the outcome function is

\[ g : M \rightarrow X \times T \]

The best response strategy of the agent is a mapping

\[ m^* : \Theta \rightarrow M \]

such that

\[ m^*(\theta) \in \arg \max_{m \in M} \{ u(g(m), \theta) \} \]

and he obtains the allocation

\[ g(m^*(\theta)) \] (13.1)

Therefore we say that \( \Gamma = (g(\cdot), M) \) implement \( f \), where \( f \) satisfies

\[ g(m^*(\theta)) = f(\theta) \] (13.2)

for every \( \theta \). Thus based on the indirect mechanism \( \Gamma = (g(\cdot), M) \) we can define a direct mechanism as suggested by (13.2):

\[ \Gamma_d = (f(\cdot), \Theta) . \]

**Theorem 13.3.1 (Revelation Principle).**

If the social choice function \( f(\cdot) \) can be implemented through some mechanism, then it can be implemented through a direct revelation (truthful) mechanism: \( \Gamma_d = (f(\cdot), \Theta) \).

**Proof.** Let \( (g, M) \) be a mechanism that implements \( f(\cdot) \) through \( m^*(\cdot) \) be the equilibrium message:

\[ f(\theta) = g(m^*(\theta)) \]

Consider the direct mechanism

\[ \Gamma_d = (f(\cdot), \Theta) . \]

If it were not truthful, then an agent would prefer to announce \( \theta' \) rather than \( \theta \), and he would get

\[ u(f(\theta), \theta) < u(f(\theta'), \theta) \]

but by definition of implementation, and more precisely by (13.2), these would imply that

\[ u(g(m^*(\theta)), \theta) < u(g(m^*(\theta'), \theta)) \]

which is a contradiction to the hypothesis that \( m^*(\cdot) \) would be an equilibrium of the game generated by the mechanism

\[ \Gamma(g(\cdot), M) . \]
The revelation principle states that it is sufficient to restrict attention to mechanisms which offer a menu of contracts: agent announces $\theta$, and will get

$$y(\theta) = (x(\theta), t(\theta))$$

(13.3)

but many mechanisms in the real world are indirect in the sense that goods are provided in different qualities and/or quantities and the transfer is then a function of the choice variable $x$.

We can use the taxation principle which says that every direct mechanism is equivalent to a nonlinear tariff. How? Offering a menu of allocations $\{x(\theta)\}_{\theta \in \Theta}$ and let

$$t(x) \triangleq t(x(\theta)) = t(\theta)$$

be the corresponding nonlinear tariff, so that the menu is

$$\{x, t(x)\}_{x \in X}$$

The proof that the nonlinear tariff is implementing the same allocation as the direct mechanism is easy. Suppose there are $\theta, \theta'$ such that $x(\theta) = x(\theta')$. If $t(\theta) \neq t(\theta')$, then the agent would have an interest to misrepresent his state of the world, and hence the direct mechanism couldn’t be truth-telling. It follows that $t(\theta) = t(\theta')$. In consequence, the transfer function can be uniquely defined as

$$i \text{f } x = x(\theta) \Rightarrow t(x) = t(\theta),$$

which leads to the nonlinear tariff. The notion of a direct mechanism may then represent a normative approach, whereas the indirect mechanism represents a positive approach.

### 13.4 Second Price Sealed Bid Auction

Consider next the by now familiar second price auction and let us try to translate it into the language of mechanism design. Consider the seller with two bidders and $\theta_i \in [0, 1]$. The set of allocations is $Y = X \times T$, where $y_i = (x_i, t_i)$, where $x_i \in \{0, 1\}$ denotes the assignment of the object to bidder $i$: no if 0, yes if 1. The utility function of agent $i$ is given by:

$$u_i(y_i, \theta_i) = \theta_i x_i - t_i$$

The second price sealed bid auction implements the following social choice function

$$f(\theta) = (x_0(\theta), x_1(\theta), x_2(\theta); t_0(\theta), t_1(\theta), t_2(\theta))$$

with

$x_1(\theta) = 1$, if $\theta_1 \geq \theta_2$; $= 0$ if $\theta_1 < \theta_2$.

$x_2(\theta) = 1$, if $\theta_1 < \theta_2$; $= 0$ if $\theta_1 \geq \theta_2$.

$x_0(\theta) = 0$, for all $\theta$;

and

$$t_1(\theta) = \theta_2 x_1(\theta)$$

$$t_2(\theta) = \theta_1 x_2(\theta)$$

$$t_0(\theta) = t_1(\theta) x_1(\theta) + t_2(\theta) x_2(\theta)$$

The message $m_i$ is the bid for the object:
The outcome function is
\[
g : M \rightarrow Y
\]
with
\[
g = \begin{cases} 
  x_1 = 1, t_1 = m_2, & \text{if } m_1 \geq m_2 \\
  x_2 = 0, t_2 = 0, & \text{if } m_1 < m_2 
\end{cases}
\]

A variety of other examples can be given:

Example 13.4.1 (Income tax). \(x\) is the agent’s income and \(t\) is the amount of tax paid by the agent; \(\theta\) is the agent’s ability to earn money.

Example 13.4.2 (Public Good). \(x\) is the amount of public good supplied, and \(t_i\) is the consumer \(i\)’s monetary contribution to finance it; \(\theta_i\) indexes consumer surplus from the public good.

The example of the second price sealed bid auction illustrates that as a general matter we need not only to consider to directly implementing the social choice function by asking agents to reveal their type, but also indirect implementation through the design of institutions by which agents interact. The formal representation of such institutions is known as a mechanism.
14. Dominant Strategy Equilibrium

Example 14.0.3 (Voting Game). Condorcet paradox

Next we consider implementation in dominant strategies. We show that in the absence of prior restrictions on the characteristics, implementation in dominant equilibria is essentially impossible.

**General Environments.** Next, we prove the revelation principle for the dominant equilibrium concept. Equivalent results can be proven for Nash and Bayesian Nash equilibria.

**Definition 14.0.1.** The strategy profile $m^* = (m^*_1(\theta_1), ..., m^*_I(\theta_I))$ is a dominant strategy equilibrium of mechanism $\Gamma = (M, g(\cdot))$ if for all $i$ and all $\theta_i \in \Theta_i$:

$$u_i(g(m^*_1(\theta_1), m_{-i}), \theta_i) \geq u_i(g(m'_i, m_{-i}), \theta_i)$$

for all $m'_i \neq m_i$, $\forall m_{-i} \in M_{-i}$.

**Theorem 14.0.1 (Revelation Principle).** Let $\Gamma = \{M, g(\cdot)\}$ be a mechanism that implements the social choice function $f(\cdot)$ for the dominant equilibrium concept, then there exists a direct mechanism $\Gamma' = \{\Theta, f(\cdot)\}$ that implements $f(\cdot)$ by revelation.

**Proof.** (Due to \cite{...}). Let $m^*(\theta) = (..., m^*_i(\theta_i), ...)$ be an $I$-tuple of dominant messages for $(M, g(\cdot))$. Define $g^*$ to be the composition of $g$ and $m^*$.

$$g^* \triangleq g \circ m^*$$

or

$$g^* (\theta) \triangleq g (m^*(\theta)). \quad (14.1)$$

By definition

$$g^* (\theta) = f (\theta).$$

In fact

$$\Gamma^* = (\Theta, g^*(\cdot))$$

is a direct mechanism. Next we want to show that the mechanism implements $f(\cdot)$ as an equilibrium in dominant strategies. The proof is by contradiction. Suppose not, that is the transmission of his true characteristics is not a dominant message for agent $i$. Then there exists $(\hat{\theta}_i, \theta_{-i})$ such that

$$u_i(g^*(\hat{\theta}_i, \theta_{-i}), \theta_i) > u_i(g^*(\theta), \theta_i).$$

But by the definition (14.1) above, the inequality can be rewritten as

$$u_i(g(m^*(\hat{\theta}_i, \theta_{-i})), \theta_i) > u_i(g(m^*(\theta)), \theta_i).$$

But this contradicts our initial hypothesis that $\Gamma = (M, g(\cdot))$ implements $f(\cdot)$ in a dominant strategy equilibrium, as $m^*$ is obviously not an equilibrium strategy for agent $i$, which completes the proof.
Dominant Strategy Equilibrium

The notion of implementation can then be refined in Definition ?? in a suitable way. This is implementation in a very robust way, in terms of strategies and in informational requirements as the designer doesn’t need to know \( P(\cdot) \) for the successful implementation. But can we always implement in dominant strategies:

**Definition 14.0.2.** The social choice function \( f(\cdot) \) is dictatorial if there is an agent \( i \) such that for all \( \theta \in \Theta \),

\[
f(\theta) \in \{ x \in X : u_i(x, \theta_i) \geq u_i(y, \theta_i) \text{ for all } y \in X \}.
\]

Next, we can state the celebrated result from ? and ?.

**Theorem 14.0.2 (Gibbard-Satterthwaite).** Suppose that \( X \) contains at least three elements, and that \( R_i = \mathcal{P} \) for all \( i \), and that \( f(\Theta) = X \). Then the social choice function is truthfully implementable if and only if it is dictatorial.

**Quasi-Linear Environments.** The quasilinear environment is described by

\[
v_i(x, t, \theta_i) = u_i(x, \theta_i) - t_i.
\]

Denote the efficient allocation by \( x^*(\theta) \). Then the generalization of the Vickrey auctions states: ?, ?, ?.

**Theorem 14.0.3 (Vickrey-Clark-Groves).** The social choice function \( f(\cdot) = (x^*, t_1(\cdot), \ldots, t_I(\cdot)) \) is truthfully implementable in dominant strategies if for all \( i \):

\[
t_i(\theta_i) = \left[ \sum_{j \neq i} u_j(x^*(\theta), \theta_j) \right] - \left[ \sum_{j \neq i} u_j(x^*(\theta_{-i}), \theta_j) \right].
\]  

(14.2)

**Proof.** If truth is not a dominant strategy for some agent \( i \), then there exist \( \theta_i, \hat{\theta}_i, \text{ and } \theta_{-i} \) such that

\[
u_i(x^*(\hat{\theta}_i, \theta_{-i}), \theta_i) + t_i(\hat{\theta}_i, \theta_{-i}) > v_i(x^*(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i})\]  

(14.3)

Substituting from (14.2) for \( t_i(\hat{\theta}_i, \theta_{-i}) \) and \( t_i(\theta_i, \theta_{-i}) \), this implies that

\[
\sum_{j=1}^{I} u_j(x^*(\hat{\theta}_i, \theta_{-i}), \theta_j) > \sum_{j=1}^{I} u_j(x^*(\theta), \theta_j),
\]  

(14.4)

which contradicts \( x^*(\cdot) \) being an optimal policy. Thus, \( f(\cdot) \) must be truthfully implementable in dominant strategies.

This is often referred to as the pivotal mechanism. It is *ex-post efficient* but it may not satisfy *budget balance*:

\[
\sum_{i \in I} t_i(\theta_i) = 0.
\]

We can now weaken the implementation requirement in terms of the equilibrium notion.
15. Bayesian Equilibrium

In many cases, implementation in dominant equilibria is too demanding. We therefore concentrate next on
the concept of Bayesian implementation. Note however that with a single agent, these two concepts are
equivalent.

Definition 15.0.3. The strategy profile \( s^* = (s^*_1(v_1), ..., s^*_I(v_I)) \) is a Bayesian Nash equilibrium of mech-
anism \( \Gamma = (S_1, ..., S_I, g (\cdot)) \) if for all \( i \) and all \( v \in \Theta \):

\[
\mathbb{E}_{\nu_{-i}} [u_i (g (s^*_i (v_i), s^*_{-i} (v_{-i})), v_i) | v_i] \geq \mathbb{E}_{\nu_{-i}} [u_i (g (s_i, s^*_{-i} (v_{-i})), v_i) | v_i]
\]

for all \( s_i \neq s^*_i (v), \forall s^*_{-i} (v) \).

15.1 First Price Auctions

We replace the type notation \( v \) by the vector of private valuations

\( v = (v_1, ..., v_I) \)

and denote the distribution over types by symmetric

\( f (v_i), F (v_i) \).

For agent \( i \) to win with value \( v \), he should have a higher value than all the other agents, which happens
with probability

\( G (v) \triangleq (F (v))^{I-1} \)

and the associated density is

\( g (v) = (I - 1) f (v) (F (v))^{I-2}. \)

This is often referred to as the first order statistic of the sample of size \( I - 1 \).

Theorem 15.1.1. The unique symmetric Bayes-Nash equilibrium in the first price auction is given by

\[
b^* (v) = \int_0^v y \left[ \frac{g (y)}{G (v)} \right] dy
\]

Remark 15.1.1. The expression in the integral is thus the expected value of highest valuation among the
remaining \( I - 1 \) agents, provided that the highest valuation is is below \( v \). (Thus the bidder acts if he were
to match the expected value of his competitor.)
Proof: We start the proof by making some “inspired guesses,” and then go back to rigorously prove the guesses. This technique is how the first economist to study auctions might have originally derived the equilibrium. That economist might have made the following good guesses about the equilibrium $b^* (\cdot)$ being sought:

(i) All types of bidder submit bids: $b (v) \neq \emptyset$ for all $v \geq 0$.

(ii) Bidders with greater values bid higher: $b (v) > b (z)$ for all $v > z$.

(iii) The function $b (\cdot)$ is differentiable.

Obviously, many functions other than $b^* (\cdot)$ satisfy Guesses (i) – (iii). We now show that the only possible equilibrium satisfying them is $b^* (\cdot)$. From (iii), $b (\cdot)$ has an inverse function, which we denote as $\phi (\cdot)$, such that $b (\phi (\hat{b})) = \hat{b}$ for all numbers $\hat{b}$ in the range of $b (\cdot)$. Value $v = \Phi (\hat{b})$ is the value a bidder must have in order to bid $\hat{b}$ when using strategy $b (\cdot)$. Because $b (\cdot)$ strictly increases, the probability that bidder $i$ wins if he bids $b$ is

$$Q (\hat{b}) = F (\phi (\hat{b}))^{n-1} = G (\phi (b)).$$

(15.1)

Why? Well, observe that since $b (\cdot)$ is strictly increasing, the other bidders’ values are continuously distributed, we can ignore ties: another bidder $j$ bids the same $\hat{b}$ only if his value is precisely $v_j = \phi (\hat{b})$, which occurs with probability zero. So the probability of bidder $i$ winning with bid $\hat{b}$ is equal to the probability of the other bidders bidding no more than $\hat{b}$. Because $b (\cdot)$ strictly increases, this is the probability that each other bidder’s value is not more than $\phi (\hat{b})$. This probability is $G (\phi (b)) = F (\phi (\hat{b}))^{n-1}$, the probability that the maximum of the bidders’ value is no more than $\phi (b)$. Thus, the expected profit of a type $v$ bidder who bids $b$ when the others use $b (\cdot)$ is

$$\pi (v, \hat{b}) = (v - \hat{b}) G (\phi (\hat{b})).$$

(15.2)

Because of (iii), the inverse function $\phi (\cdot)$ is differentiable. Letting $\phi' (\hat{b})$ be its derivative at $\hat{b}$, the partial derivative of $\pi (v, \hat{b})$ with respect to $\hat{b}$ is

$$\pi_b (v, b) = -G (\phi (b)) + (v - b) g (\phi (b)) \Phi' (b).$$

(15.3)

Since $b (\cdot)$ is an equilibrium, $b (v)$ is an optimal bid for a type $v$ bidder, i.e. $b = b (v)$ maximizes $\pi (v, b)$. The first order condition $\pi_b (v, b (v)) = 0$ holds:

$$-G (\phi (b (v))) + (v - b (v)) g (\phi (b (v))) \Phi' (b (v)) = 0.$$ 

(15.4)

Use $\Phi (b (v)) = v$ and $\Phi' (b (v)) = 1/b' (v)$ to write this as

$$-G (v) + \frac{(v - b (v)) g (v)}{b' (v)} = 0.$$ 

(15.5)

This is a differential equation that almost completely characterizes the exact nature of $b (\cdot)$. It is easy to solve. Rewrite it as

$$G (v) b' (v) + g (v) = vg (v).$$

(15.6)

Since $g (v) = G' (v)$, the left side is the derivative of $G (v) b (v)$. So we can integrate both sides from any $v_0$ to any $v$ (using “y” to denote the dummy integration variable):

$$G (v) b (v) - G (v_0) b (v_0) = \int_{v_0}^{v} yg (y) dy$$

(15.7)
>From Guess A, all types bid, so we can take $v_0 \to 0$. We know $b'(0) \geq 0$, as $r = 0$. Hence, $G(v_0)b'(v_0) \to 0$ as $v_0 \to 0$. Take this limit in (6.5) and divide by $G(v)$ to obtain

$$b(v) = \int_0^v y \left( \frac{g(y)}{G(v)} \right) dy = b^*(v).$$

(15.8)

This is the desired result.

Finally, we can rewrite (15.5) using integration by parts:

$$\int u'v = uv - \int uv'$$

as

$$b^*(v) = v - \int_0^v \left( \frac{F'(y)}{F(v)} \right)^{t-1} dy$$

### 15.2 Optimal Auctions

#### 15.2.1 Revenue Equivalence

We are now describing the optimal auction and the revenue equivalence result on the basis of the result we obtained earlier. Recall the following theorem which describes the incentive compatible equilibrium utility of agent $i$:

**Theorem 15.2.1.** The direct mechanism $y(v) = (q(v), t(v))$ is incentive compatible if and only if:

1. the truth-telling utility is described by:
   $$U_i(v_i) - U_i(0) = \int_{s_i}^{v_i} u_{v_i}(q_i(s_i), s_i) ds_i,$$
   (15.9)

2. $q_i(s_i)$ is nondecreasing.

The utility function of agent $i$ in the single unit auction is

$$u(v_i, q_i) = v_i q_i$$

where

$$q_i = \Pr(x_i = 1)$$

Thus we can rewrite the theorem in our context as follows:

**Theorem 15.2.2.** The direct mechanism to sell a single object $y(v) = (q(v), t(v))$ is incentive compatible if and only if:

1. the truth-telling utility is described by:
   $$U_i(v_i) - U_i(0) = \int_{s_i}^{v_i} q_i(s_i) ds_i,$$  
   (15.10)
Bayesian Equilibrium

2. \( q_i(s_i) \) is nondecreasing.

We can now state the celebrated revenue equivalence theorem:

**Theorem 15.2.3 (Revenue Equivalence).**
Given any two auctions mechanisms (direct or not) which agree
1. on the assignment probabilities \( q_i(v) \) for every \( v \)
2. on the equilibrium utility \( U_i(0) \) for all

lead to identical revenues.

In fact, we have the remarkable revenue equivalence theorem. The argument relies on the important (and more general) revelation principle. To see which social choice rules are implementable by some game, it is enough to focus attention on direct revelation games where players truthfully report their types and an allocation is chosen (perhaps randomly) as a function of their reported types.

\[ v_i - b_i = \int_0^{v_i} q_i(s_i) \, ds_i \]

## 15.2.2 Optimal Auction

We described the optimization problem by the principal with the single agent as follows:

\[
\max \mathbb{E}_v \left[ S(q(v), v) - \frac{1 - F(v)}{f(v)} u_v(q(v), v) - U(0) \right]
\]  
(15.11)

and the extension to many agents is trivial:

\[
\max \mathbb{E}_{q(v)} \left[ S(q(v), v) - \sum_{i=1}^I \left( \frac{1 - F_i(v_i)}{f_i(v_i)} u_{iv}(q_i(v), v_i) - U_i(0) \right) \right]
\]  
(15.12)

and making use of the auction environment we get

\[
\max_{q(v)} \int_{v_i=0}^1 \cdots \int_{v_I=0}^1 \left( \sum_{i=1}^I q_i(v) \left[ v_i - \left[ \frac{1 - F_i(v_i)}{f_i(v_i)} \right] \right] \right) \times \left[ \prod_{i=1}^I f_i(v_i) \right] \, dv_1 \cdots dv_I
\]

where

\[ q = (q_1, \ldots, q_I) \]  
(15.13)

and

\[ q_i \geq 0, \sum_{i=1}^I q_i \leq 1. \]

But the optimization problem can be solved pointwise for every \( v = (v_1, \ldots, v_I) \):

\[
\max_q \sum_{i=1}^I q_i(v) \left[ v_i - \left[ \frac{1 - F_i(v_i)}{f_i(v_i)} \right] \right]
\]
Let us denote, following Myerson, by $\hat{v}_i$ the virtual utility of agent $i$

$$\hat{v}_i \triangleq v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}$$

It is then optimal to set the reserve price for every agent such that $r_i = \hat{v}_i$ satisfies:

$$\hat{v}_i - \frac{1 - F_i(v_i)}{f_i(v_i)} = 0$$
**Theorem 15.2.4 (Optimal Auction).** The optimal policy $q^* = (q_1^*, ..., q_I^*)$ is then given by:

1. $q_i^* = 0$ if all $\tilde{v}_i < 0$
2. $\sum_i q_i^* = 1$ if $\exists \tilde{v}_i \geq 0$
3. $q_i^* > 0 \Rightarrow \tilde{v}_i = \max \{\tilde{v}_1, ..., \tilde{v}_I\}$.

### 15.3 Additional Examples

#### 15.3.1 Procurement Bidding

A single buyer and many sellers competing to sell the service or product. The private information is now regarding the cost of providing the service.

The virtual utility is now

$$c_i + \frac{F_i(c_i)}{f_i(c_i)}$$

Information rent with privately informed buyers is due to the fact that buyers wish to **understate** their valuation.

Information rent with privately informed sellers is due to the fact that sellers wish to overstate their cost.

#### 15.3.2 Bilateral Trade

Consider a seller (with valuation $v_1$) and a buyer (with valuation $v_2$) who wishes to engage in a trade:

The virtual utility now becomes identical to:

$$\int \int \left( v_2 - \frac{1 - F_2(v_2)}{f_2(v_2)} \right) - \left( v_1 + \frac{F_1(v_1)}{f_1(v_1)} \right)$$

Point wise optimization occurs when

$$\left( v_2 - \frac{1 - F_2(v_2)}{f_2(v_2)} \right) - \left( v_1 + \frac{F_1(v_1)}{f_1(v_1)} \right) > 0$$

But for

$$v_2 = v_1$$

The difference of the virtual utilities is negative

$$\left( v_2 - \frac{1 - F_2(v_2)}{f_2(v_2)} \right) - \left( v_1 + \frac{F_1(v_1)}{f_1(v_1)} \right) < 0$$

And thus in general inefficient trade with two sided private information

#### 15.3.3 Bilateral Trade: Myerson and Sattherthwaite

#### 15.3.4 Readings

The surveys by $?$ and $?$ are highly recommended. The notion of incentive compatible and the basic structure of a collective decision rule preserving some kind of informational decentralization is due to $?$. The various notions of efficiency are discussed in $?$.  

]
16. Efficiency

A game theorist or a mediator who analyzes the Pareto efficiency of behavior in a game with incomplete information must use the perspective of an outsider, so he cannot base his analysis on the players’ private information. An outsider may be able to say how the outcome will depend on the players’ types. That is, he can know the mechanisms but not its outcome. Thus, Holmstrom and Myerson (1983) argued that the concept of efficiency for games with incomplete information should be applied to the mechanisms, rather than to outcomes, and the criteria for determining whether a particular mechanism $\mu$ is efficient should depend only on the commonly known structure of the game, not on the privately own types of the individual players.

16.1 First Best

A definition of Pareto efficiency in a Bayesian collective-choice problem is: A mechanism is efficient if and only if no other feasible mechanism can be found that might make some other individuals better off and would certainly not make other individuals worse off. However, this definition is ambiguous in several ways. In particular, we must specify what information is to be considered when determining whether an individual is “better off” or “worse off.”

A mechanism $\gamma$ can now be thought of as contingent allocation plan

$$\gamma : \Theta \rightarrow \Delta(Y)$$

and consequently

$$\gamma (y | \theta)$$

expresses the conditional probability that $y$ is realized given a type profile $\theta$.

One possibility is to say that an individual is made worse off by a change that decreases his expected utility payoff as would be computed before his own type or any other individuals’ type are specified. This standard is called the \textit{ex ante welfare criterion}. Thus, we say that a mechanism $\gamma$ is \textit{ex ante} Pareto superior to another mechanism $\mu$ if and only if

$$\sum_{\theta \in \Theta} \sum_{y \in Y} p(\theta) \gamma (y | \theta) u_i (y, \theta) \geq \sum_{\theta \in \Theta} \sum_{y \in Y} p(\theta) \mu (y | \theta) u_i (y, \theta), \forall i \in I,$$

and this inequality is strict for at least one player in $I$. Next define

$$U_i (\mu | \theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{y \in Y} p_i (\theta_{-i} | \theta_i) \mu (y | \theta) u_i (y, \theta)$$

and notice that
Another possibility is to say that an individual is made worse off by a change that decreases his conditionally expected utility, given his own type (but to given the type of any other individuals). An outside observer, who does not know any individual’s type, would then say that a player $i$ “would certainly not be made worse off (by some change of mechanism)” in this sense if this conditionally expected utility will not be decreased (by the change) for any possible type of player $i$. This standard is called the **interim welfare criterion**, because it evaluates each player’s welfare after he learns his own type but before he learns any other player’s type. Thus, we say that a mechanism $\gamma$ is *interim Pareto superior* to another mechanism $\mu$ if and only if

$$U_i(\gamma | \theta_i) \geq U_i(\mu | \theta_i), \forall i \in I, \forall \theta_i \in \Theta_i,$$

and this inequality is strict for at least one type of one player in $I$.

Yet another possibility is to say that an individual is made worse off by a change that decreases his conditionally expected utility, given the types of all individuals. An outsider observer would then say that a player “would certainly not be made worse off” in this sense if his conditionally expected utility would not be decreased for any possible combination of types for all the players. This standard is called *ex post welfare criterion*, because it uses the information that would be available after all individuals have revealed their types. Thus, we say that a mechanism $\gamma$ is *ex post pareto superior* to another mechanism $\mu$ if and only if

$$\sum_{y \in Y} \gamma(y | \theta) u_i(y, \theta) \geq \sum_{y \in Y} \mu(y | \theta) u_i(y, \theta), \forall i \in I, \forall \theta \in \Theta,$$

and this inequality is strict for at least one player in $N$ and at least one possible combination of types $\theta$ in $\Theta$ such that $p(\theta) > 0$.

Given any concept of feasibility, the three welfare criteria (ex ante, interim, ex post) give rise to three different concepts of efficiency. For any set of mechanisms $\Phi$ (to be interpreted as the set of “feasible” mechanisms in some sense), we say that a mechanism $\mu$ is *ex ante efficient* in the set $\Phi$ if and only if $\mu$ is in $\Phi$ and there exists no other mechanism $v$ that is in $\Phi$ and is ex ante Pareto superior to $\mu$. Similarly $\mu$ is *interim efficient* in $\Phi$ if and only if $\mu$ is in $\Phi$ and there exists no other mechanism $v$ that is in $\Phi$ and is interim Pareto superior to $\mu$; and $\mu$ is *ex post efficient* in $\Phi$ if and only if $\mu$ is in $\Phi$ and there exists no other mechanism $v$ that is in $\Phi$ and is ex post Pareto superior to $\mu$.

### 16.2 Second Best

Similar efficiency notions can be defined in a second best environment. The notions will be unaltered, the change is in the set of mechanisms which we are considered to be feasible. The set may be restricted, e.g., to the set of incentive feasible mechanism, that is all mechanisms which satisfy the incentive compatibility condition.
17. Social Choice

Can individual preferences be aggregated into social preferences, or more directly into social decisions. Denote by $X$ the set of alternatives. Each individual $i \in I$ has a rational preference relation $\succeq_i$. Denote the set of preferences by $\mathcal{R}$ and $\mathcal{P}$.

17.1 Social Welfare Functional

Definition 17.1.1. A social welfare functional is a rule

$$F : \mathcal{R}^I \to \mathcal{R}. $$

The strict preference relation derived from $F$ is denoted by $F_p$.

Definition 17.1.2. $F$ is paretnian if for any pair $(x, y) \in X$ and for all $\succeq_I \in A$,

$$\forall i, x \succ_i y \Rightarrow xF_p y.$$  

Example 17.1.1. Borda count: $x \succ_i y, x \succ_i z \Rightarrow c_i (x) = 1, c_i (y) = 2, c_i (z) = 3$.

Definition 17.1.3. $F$ satisfies independence of irrelevant alternatives if for $\succeq_I$ and $\succeq_I'$ with the property that $\forall i, \forall (x, y)$:

$$\succeq_i \setminus \{x, y\} = \succeq_i' \setminus \{x, y\} \Rightarrow F(\{x, y\}) = F'(\{x, y\}).$$

Example 17.1.2. Borda count doesn’t satisfy IIA for as we change from $x \succ_1 z \succ_1 y$, and $y \succ_2 x \succ_2 z$, to $x \succ_i y \succ_i z$, and $y \succ_j z \succ_j x$, the relative position remains unchanged yet the Borda count changes.

Example 17.1.3. Condorcet paradox with majority voting: $x \succ_1 y \succ_1 z, z \succ_2 x \succ_2 y, y \succ_3 z \succ_3 x$ yields cyclic and non-transitive preferences.

Definition 17.1.4. $F$ is dictatorial if $\exists i$ such that $\forall (x, y) \in X, \forall \succeq_I \in A$,

$$x \succ_i y \Rightarrow xF_p y.$$  

Theorem 17.1.1. Suppose $|X| \geq 3$ and $A = \mathcal{R}$. Then every $F$ which is paretnian and satisfies independence of irrelevant alternatives is dictatorial.

Proof. See ?.

The result demonstrates the relevance of procedures and rules for social aggregation. To escape the impossibility results, one may either (i) restrict the domain of the set of preferences or (ii) weaken the requirement of rational preferences.

$$F : A \to \mathcal{R}$$

(i) (ii)

17.2 Social Choice Function

Definition 17.2.1. A social choice function $f : A \to X$ assigns $f(\succeq_I) \in X$ for every $\succeq_I \in A$. 