A (Very) Brief Introducion to Mechanism Design

Arunava Sen

Indian Statistical Institute, New Delhi

Static and Dynamic Mechanism Design Workshop
August 1, 2015

The Problem

The Problem

- Mechanism design theory is concerned with resource allocation (understood very broadly) in multi-agent environments.

The Problem

- Mechanism design theory is concerned with resource allocation (understood very broadly) in multi-agent environments.
- Key feature: determination of an “optimal” allocation depends on information which agents possess *privately*.

The Problem

- Mechanism design theory is concerned with resource allocation (understood very broadly) in multi-agent environments.
- Key feature: determination of an “optimal” allocation depends on information which agents possess *privately*.
- This private information must therefore be *elicited* from the agents.

The Problem

- Mechanism design theory is concerned with resource allocation (understood very broadly) in multi-agent environments.
- Key feature: determination of an "optimal" allocation depends on information which agents possess *privately*.
- This private information must therefore be *elicited* from the agents.
- Agents are sophisticated - they recognize that they may (depending on beliefs that they have about the information revealed by the other agents) be served better by lying rather than by telling the truth.

The Problem

- Mechanism design theory is concerned with resource allocation (understood very broadly) in multi-agent environments.
- Key feature: determination of an “optimal” allocation depends on information which agents possess *privately*.
- This private information must therefore be *elicited* from the agents.
- Agents are sophisticated - they recognize that they may (depending on beliefs that they have about the information revealed by the other agents) be served better by lying rather than by telling the truth.
- Computing the optimal allocation from incorrect information may entail serious errors;

Problem contd.

- Challenge is to devise a *mechanism* or a procedure for communicating the information of agents such that the outcome is an optimal allocation even when these agents behave strategically.

Problem contd.

- Challenge is to devise a *mechanism* or a procedure for communicating the information of agents such that the outcome is an optimal allocation even when these agents behave strategically.
- Mechanism Design theory can therefore be thought of as a theory of the design of *institutions* or the design of the rules of interactions amongst fully strategic agents in order to achieve desirable outcomes.

Problem contd.

- Challenge is to devise a *mechanism* or a procedure for communicating the information of agents such that the outcome is an optimal allocation even when these agents behave strategically.
- Mechanism Design theory can therefore be thought of as a theory of the design of *institutions* or the design of the rules of interactions amongst fully strategic agents in order to achieve desirable outcomes.
- We consider some motivating examples.

Motivating Example: Voting

- n voters, n odd.

Motivating Example: Voting

- n voters, n odd.
- Voters have to collectively select one of the two proposals a or b.

Motivating Example: Voting

- n voters, n odd.
- Voters have to collectively select one of the two proposals a or b.
- For cach voter i, either a is "better than" b or b is better than a. This information is known *only* to i - revealed by voting.

Motivating Example: Voting

- n voters, n odd.
- Voters have to collectively select one of the two proposals a or b.
- For cach voter i, either a is “better than” b or b is better than a. This information is known *only* to i - revealed by voting.
- Consider majority voting: all voters vote either a or b and the proposal which gets the highest aggregate number of votes is selected.

Motivating Example: Voting

- n voters, n odd.
- Voters have to collectively select one of the two proposals a or b.
- For cach voter i, either a is “better than” b or b is better than a. This information is known *only* to i - revealed by voting.
- Consider majority voting: all voters vote either a or b and the proposal which gets the highest aggregate number of votes is selected.
- Voters realize that they are playing a game. They can vote either a or b (their strategy sets) and the outcome and payoff depends not only on how they vote but also on how *everyone else* votes.

Voting contd.

Voting contd.

- How will voters vote?

Voting contd.

- How will voters vote?
- Their vote does not matter unless the other voters are exactly divided in their opinion on a and b. In this case a voter gets to choose the proposal she wants. She will clearly hurt herself by misrepresenting her preferences.

Voting contd.

- How will voters vote?
- Their vote does not matter unless the other voters are exactly divided in their opinion on a and b. In this case a voter gets to choose the proposal she wants. She will clearly hurt herself by misrepresenting her preferences.
- In the language of game theory, truth-telling is a weakly dominiant strategy.

Voting contd.

- Three proposals or candidates a, b and c?

Voting contd.

- Three proposals or candidates *a*, *b* and *c*?
- Consider a generalization of the rule proposed above. Each voter votes for her best proposal. Select the proposal which is best for the largest number number of voters. If no such proposal exists, select *a* (which can be thought of as a *status quo* proposal).

Voting contd.

- Three proposals or candidates *a*, *b* and *c*?
- Consider a generalization of the rule proposed above. Each voter votes for her best proposal. Select the proposal which is best for the largest number number of voters. If no such proposal exists, select *a* (which can be thought of as a *status quo* proposal).
- What behaviour does this rule induce? Is truth-telling a dominant strategy once again?

Voting contd.

Voting contd.

- NO!

Voting contd.

- NO!
1	2	3
c	b	a
b	a	b
a	c	c

  Table : Voter Preferences

Voting contd.

- NO!

1	2	3
c	b	a
b	a	b
a	c	c

Table : Voter Preferences

- Suppose voter 1's true preference is c better than b than a while she believes that voters 2 and 3 are going to vote for b and a respectively. Then voting truthfully will yield a while lying and voting for b will get b which is better than a according to her *true* preferences.

Voting contd.

- NO!
1	2	3
c	b	a
b	a	b
a	c	c

  Table : Voter Preferences
- Suppose voter 1's true preference is c better than b than a while she believes that voters 2 and 3 are going to vote for b and a respectively. Then voting truthfully will yield a while lying and voting for b will get b which is better than a according to her *true* preferences.
- Are there voting rules which will induce voters to reveal their true preferences?

Example: Bilateral Trading

- Two agents, a seller S and a buyer B.

Example: Bilateral Trading

- Two agents, a seller S and a buyer B.
- Seller has a single object which the buyer is potentially interested in buying.

Example: Bilateral Trading

- Two agents, a seller *S* and a buyer *B*.
- Seller has a single object which the buyer is potentially interested in buying.
- The seller and buyer have valuations $v_s, v_b \in \Re_+$, known privately. Assume that they are iid random variables - uniformly distributed on $[0, 1]$.

Example: Bilateral Trading

- Two agents, a seller S and a buyer B.
- Seller has a single object which the buyer is potentially interested in buying.
- The seller and buyer have valuations $v_s, v_b \in \Re_+$, known privately. Assume that they are iid random variables - uniformly distributed on $[0, 1]$.
- Consider the following trading rule proposed by Chatterjee and Samuelson. Seller and buyer announce “bids” x_s and x_b. Trade takes place only if $x_b > x_s$. If trade occurs, it does so at price $\frac{x_b+x_s}{2}$. If no trade occurs both agents get 0; if it occurs, then payoffs for the buyer and seller are $v_b - \frac{x_b+x_s}{2}$ and $\frac{x_b+x_s}{2} - v_s$ respectively.

Bilateral Trading contd.

- Game of incomplete information.

Bilateral Trading contd.

- Game of incomplete information.
- A linear Bayes-Nash equilibrium of the game exists where $x_b = \frac{2}{3}v_b + \frac{1}{12}$ and $x_s = \frac{2}{3}v_s + \frac{1}{4}$. Trade takes place only if $v_b - v_s > \frac{1}{4}$.

Bilateral Trading contd.

- Game of incomplete information.
- A linear Bayes-Nash equilibrium of the game exists where $x_b = \frac{2}{3}v_b + \frac{1}{12}$ and $x_s = \frac{2}{3}v_s + \frac{1}{4}$. Trade takes place only if $v_b - v_s > \frac{1}{4}$.
- Efficiency would require trade to take place whenever $v_b > v_s$. There are realizations of v_b, v_s where there is no trade in equilibrium where it would be efficient to have it.

Bilateral Trading contd.

- Game of incomplete information.
- A linear Bayes-Nash equilibrium of the game exists where $x_b = \frac{2}{3}v_b + \frac{1}{12}$ and $x_s = \frac{2}{3}v_s + \frac{1}{4}$. Trade takes place only if $v_b - v_s > \frac{1}{4}$.
- Efficiency would require trade to take place whenever $v_b > v_s$. There are realizations of v_b, v_s where there is no trade in equilibrium where it would be efficient to have it.
- Are there other trading rules where agents participate voluntarily and equilibrium outcomes are always efficient?

A General Formulation

- n agents
- Set of feasible alternatives/outcomes/ allocations A.

A General Formulation

- *n* agents
- Set of feasible alternatives/outcomes/ allocations *A*.
- Each agent *i* has some private information $\theta_i \in \Theta_i$, θ_i is *i*'s *type*.

A General Formulation

- n agents
- Set of feasible alternatives/outcomes/ allocations A.
- Each agent i has some private information $\theta_i \in \Theta_i$, θ_i is i's *type*.
- Agent i has a payoff function $v_i : \Theta_i \times A \to \Re$. Every realization of θ_i determines a payoff function for i (private values model).

A General Formulation

- n agents
- Set of feasible alternatives/outcomes/ allocations A.
- Each agent i has some private information $\theta_i \in \Theta_i$, θ_i is i's *type*.
- Agent i has a payoff function $v_i : \Theta_i \times A \to \Re$. Every realization of θ_i determines a payoff function for i (private values model).
- A profile $\theta \equiv (\theta_1, .., \theta_n)$ is an n tuple which describes the "state of the world". Notation (θ_i', θ_{-i}) refers to profile where the i^{th} component of the profile θ is replaced by θ_i'.

A General Formulation

- n agents
- Set of feasible alternatives/outcomes/ allocations A.
- Each agent i has some private information $\theta_i \in \Theta_i$, θ_i is i's *type*.
- Agent i has a payoff function $v_i : \Theta_i \times A \to \Re$. Every realization of θ_i determines a payoff function for i (private values model).
- A profile $\theta \equiv (\theta_1, .., \theta_n)$ is an n tuple which describes the "state of the world". Notation (θ_i', θ_{-i}) refers to profile where the i^{th} component of the profile θ is replaced by θ_i'.
- A Social Choice Function (SCF) is a mapping $f : \Theta_1 \times \Theta_2 \times ... \times \Theta_n \to A$.

Incentive Compatibility - Dominant Strategy

- A SCF represents the collective goals of the agents/ the objectives of a Principal/Designer. Mechanism design theory investigates the class of SCFs that are “attainable” or incentive compatible when agents are aware of their strategic opportunities.

Incentive Compatibility - Dominant Strategy

- A SCF represents the collective goals of the agents/ the objectives of a Principal/Designer. Mechanism design theory investigates the class of SCFs that are "attainable" or incentive compatible when agents are aware of their strategic opportunities.
- A SCF f is *strategy-proof* if

$$v_i(f(\theta), \theta_i) \geq v_i(f(\theta'_i, \theta_{-i}), \theta_i)$$

holds for all θ_i, θ'_i, θ_{-i} and i.

Incentive Compatibility - Dominant Strategy

- A SCF represents the collective goals of the agents/ the objectives of a Principal/Designer. Mechanism design theory investigates the class of SCFs that are "attainable" or incentive compatible when agents are aware of their strategic opportunities.
- A SCF f is *strategy-proof* if
$$v_i(f(\theta), \theta_i) \geq v_i(f(\theta_i', \theta_{-i}), \theta_i)$$
holds for all θ_i, θ_i', θ_{-i} and i.
- If a SCF is strategy-proof, then truth-telling is a dominant strategy for each agent. Strategy-proofness is dominant-strategy incentive-compatibility.

Bayes-Nash incentive-compatibility

Bayes-Nash incentive-compatibility

- Truth-telling gives a higher expected utility than lying for each agent when these expectations are computed with respect to beliefs regarding the types of other agents and assuming that other agents are telling the truth.

Bayes-Nash incentive-compatibility

- Truth-telling gives a higher expected utility than lying for each agent when these expectations are computed with respect to beliefs regarding the types of other agents and assuming that other agents are telling the truth.
- Assume that $\mu_i : \Theta_1 \times ... \times \Theta_n \to [0, 1]$ denotes the beliefs of agent *i* i.e $\mu(\theta) \geq 0$ and $\int_\theta d\mu_i(\theta) = 1$. Let $\mu_i(.|\theta_i)$ denote agent *i*'s beliefs over the types of other agents conditional on her type being θ_i.

Bayes-Nash incentive-compatibility

- Truth-telling gives a higher expected utility than lying for each agent when these expectations are computed with respect to beliefs regarding the types of other agents and assuming that other agents are telling the truth.
- Assume that $\mu_i : \Theta_1 \times ... \times \Theta_n \rightarrow [0,1]$ denotes the beliefs of agent i i.e $\mu(\theta) \geq 0$ and $\int_\theta d\mu_i(\theta) = 1$. Let $\mu_i(.|\theta_i)$ denote agent i's beliefs over the types of other agents conditional on her type being θ_i.
- A SCF f is Bayesian incentive-compatible if

$$\int_{\theta_{-i}} v_i(f(\theta),\theta_i) d\mu_i(\theta_{-i}|\theta_i) \geq \int_{\theta_{-i}} v_i(f(\theta_i',\theta_{-i}),\theta_i) d\mu_i(\theta_{-i}|\theta_i)$$

for all θ_i, i.

BIC contd.

- BIC is with respect to a given prior.

BIC contd.

- BIC is with respect to a given prior.
- A SCF is strategy-proof $\Rightarrow$ it is BIC.

BIC contd.

- BIC is with respect to a given prior.
- A SCF is strategy-proof $\Rightarrow$ it is BIC.
- A SCF is BIC with respect to *all* priors $\Rightarrow$ it is strategy-proof.

BIC contd.

- BIC is with respect to a given prior.
- A SCF is strategy-proof $\Rightarrow$ it is BIC.
- A SCF is BIC with respect to *all* priors $\Rightarrow$ it is strategy-proof.
- A strategy-proof SCF is robust with respect to beliefs. However may not exist.

Goals of Mechanism Design Theory

Goals of Mechanism Design Theory

- What are SCFs that are strategy-proof or BIC?

Goals of Mechanism Design Theory

- What are SCFs that are strategy-proof or BIC?
- A more general goal is to identify the “best” or optimal SCF within the class of incentive-compatible SCFs. For instance, we might wish to design an auction which maximizes expected revenue to the seller and so on.

Goals of Mechanism Design Theory

- What are SCFs that are strategy-proof or BIC?
- A more general goal is to identify the “best” or optimal SCF within the class of incentive-compatible SCFs. For instance, we might wish to design an auction which maximizes expected revenue to the seller and so on.
- Very Important: the domain of preferences - the structure of the set A, the sets Θ_i and the function v_i.

Goals of Mechanism Design Theory

- What are SCFs that are strategy-proof or BIC?
- A more general goal is to identify the “best” or optimal SCF within the class of incentive-compatible SCFs. For instance, we might wish to design an auction which maximizes expected revenue to the seller and so on.
- Very Important: the domain of preferences - the structure of the set A, the sets Θ_i and the function v_i.
- Examples: Single-peaked domains, quasi-linear preferences, indifference, randomisation...

The Revelation Principle

The Revelation Principle

- Why do we care about truth-telling? Maybe "good" outcomes can arise when everyone lies?

The Revelation Principle

- Why do we care about truth-telling? Maybe “good” outcomes can arise when everyone lies?
- Answer: it is without loss of generality - Revelation Principle.

The Revelation Principle

- Why do we care about truth-telling? Maybe “good” outcomes can arise when everyone lies?
- Answer: it is without loss of generality - Revelation Principle.
- Let $f : \Theta_1 \times \ldots \times \Theta_n \rightarrow A$ be a scf.

The Revelation Principle

- Why do we care about truth-telling? Maybe “good” outcomes can arise when everyone lies?
- Answer: it is without loss of generality - Revelation Principle.
- Let $f : \Theta_1 \times \ldots \times \Theta_n \to A$ be a scf.
- A *mechanism* is an $n+1$ tuple, $M_1, M_2, \ldots, M_n$ are *message spaces* for each agent and $g : M_1 \times M_2 \ldots \times M_n \to A$ is a *strategic outcome function*.

The Revelation Principle

- Why do we care about truth-telling? Maybe “good” outcomes can arise when everyone lies?
- Answer: it is without loss of generality - Revelation Principle.
- Let $f : \Theta_1 \times \ldots \times \Theta_n \to A$ be a scf.
- A *mechanism* is an $n + 1$ tuple, $M_1, M_2, \ldots, M_n$ are *message spaces* for each agent and $g : M_1 \times M_2 \ldots \times M_n \to A$ is a *strategic outcome function*.
- The message are arbitrary - no notion of truth-telling.

Revelation Principle contd.

Revelation Principle contd.

- Let $\theta_i \in \Theta_i$ be a type for i.

Revelation Principle contd.

- Let $\theta_i \in \Theta_i$ be a type for i.
  - $m_i^*(\theta_i) \in M_i$ is a weakly dominant dominant strategy at θ_i for i if $v(g(m_i^*(\theta_i), m_{-i}), \theta_i)) \geq v(m_i, m_{-i}), \theta_i))$ for all $m_i \in M_i$ and $m_{-i} \in M_{-i}$.

Revelation Principle contd.

- Let $\theta_i \in \Theta_i$ be a type for i.
  - $m_i^*(\theta_i) \in M_i$ is a weakly dominant dominant strategy at θ_i for i if $v(g(m_i^*(\theta_i), m_{-i}), \theta_i)) \geq v(m_i, m_{-i}), \theta_i))$ for all $m_i \in M_i$ and $m_{-i} \in M_{-i}$.
- The mechanism $(M_1, \ldots, M_n; g)$ *implements* the scf f if, for all $\theta \in \Theta_1 \times \ldots \times \Theta_n$, and $i \in I$, there exists $m_i^*(\theta_i)$ such that:
  1. $m_i^*(\theta)$ is weakly dominant for i at θ_i.
  2. $g(m_1^*(\theta_1) \ldots m_n^*(\theta_n)) = f(\theta)$.

Revelation Principle contd.

- Let $\theta_i \in \Theta_i$ be a type for i.
  - $m_i^*(\theta_i) \in M_i$ is a weakly dominant dominant strategy at θ_i
$$\text{for } i \text{ if } v(g(m_i^*(\theta_i), m_{-i}), \theta_i)) \geq v(m_i, m_{-i}), \theta_i))$$
$$\text{for all } m_i \in M_i \text{ and } m_{-i} \in M_{-i}.$$
- The mechanism $(M_1, \ldots, M_n; g)$ *implements* the scf f if, for all $\theta \in \Theta_1 \times \ldots \times \Theta_n$, and $i \in I$, there exists $m_i^*(\theta_i)$ such that:
  1. $m_i^*(\theta)$ is weakly dominant for i at θ_i.
  2. $g(m_1^*(\theta_1) \ldots m_n^*(\theta_n)) = f(\theta)$.
- f is implementable if there exists a mechanism that implements it.

Revelation Principle contd.

- Revelation Principle (for dominant strategies). f is implementable $\Rightarrow$ f is strategy-proof.
- Proof: Suppose $(M_1, \ldots, M_n; g)$ implements f. Pick $\theta_i, \theta_i', \theta_{-i}$. Then,

$$\begin{aligned} v(f(\theta_i, \theta_{-i}), \theta_i) &= v(g(m_i^*(\theta_i), m_{-i}^*(\theta_{-i}))\theta_i) \\ &\geq v(g(m_i^*(\theta_i'), m_{-i}^*(\theta_{-i}), \theta_i) \\ &= v(f(\theta_i', \theta_{-i}), \theta_i). \end{aligned}$$

Revelation Principle contd.

Revelation Principle contd.

- Another way of expressing this is the following: if f is implementable, then it must be the case that truth-telling is an equilibrium in the direct mechanism, i.e. when f is used as a mechanism.

Revelation Principle contd.

- Another way of expressing this is the following: if f is implementable, then it must be the case that truth-telling is an equilibrium in the direct mechanism, i.e. when f is used as a mechanism.
- Subtle point: The direct mechanism need not implement f even though f is implementable. This is because the direct mechanism may pick up additional non-optimal equilibria.

Revelation Principle contd.

- Another way of expressing this is the following: if f is implementable, then it must be the case that truth-telling is an equilibrium in the direct mechanism, i.e. when f is used as a mechanism.
- Subtle point: The direct mechanism need not implement f even though f is implementable. This is because the direct mechanism may pick up additional non-optimal equilibria.
- A similar Revelation Principle hold for Bayes-Nash equilibria.

Revelation Principle contd.

- Another way of expressing this is the following: if f is implementable, then it must be the case that truth-telling is an equilibrium in the direct mechanism, i.e. when f is used as a mechanism.
- Subtle point: The direct mechanism need not implement f even though f is implementable. This is because the direct mechanism may pick up additional non-optimal equilibria.
- A similar Revelation Principle hold for Bayes-Nash equilibria.
- RP may not hold for some solution concepts....