Introduction to Mechanism Design

Mathijs de Weerdt
Slides loosely based on EASSS tutorial with Paul Harrenstein

Faculty of Electrical Engineering, Mathematics, and Computer Science
Delft University of Technology

December 8, 2009

Relation to Multiagent Systems

- In case agents are self-interested and rational: social choice theory (first part of tutorial)
  - acting on behalf of people
  - acting on behalf of companies
  - "smart" virtual characters
- Specific results in case agents have private information (second part of tutorial), with a focus on strategic behavior
- Mostly relatively "old" game theory (50s-80s) that gives a background on what's (not) possible when designing multiagent systems

TUDelft

Outline

TUDelft

Relation to Game Theory

- Mechanism Design is inverse of Game Theory, and a subfield of Economics
- Develop games to deal with multiple rational players when these players may have different preferences (non-cooperative)
- How to make “good” decisions? (e.g., socially, environmentally, or from government point-of-view)

A Simple Example

Goal: Design a game (mechanism) that generates a desired outcome for all (relevant) preference profiles

Example Applications

- Deciding how to allocate radio spectrums
- Setting up a supply chain
- Influencing route choice behavior
- Allocating and planning transportation tasks

Assumptions

- Every player has preferences over outcomes
  - that do not change during mechanism (preferences nor outcomes)
  - every player knows its own preferences
- Only outcome counts (ignore/include long-term relationships, altruism)
- Full rationality (e.g., ultimatum game)
- Communication is possible (but not necessarily truthful)
- Mostly ignore forming of coalitions

Outline

What Is Social Choice Theory Trying to Accomplish?

- Goal: aggregate preferences of agents (e.g. voting)

In this tutorial

- Desirable properties
- Impossibility results (theorems)
- (Unique) solutions (mechanisms/rules)
- Abstract from domain of actions/strategies

Social Choice Rules

Definition

Let N be a set of players, X a set of alternatives, and Θ a class of types, in this case, preference relations over X. We define

social choice function (scf)	$f : \Theta^N \rightarrow X$	one social choice
social choice correspondence (scc)	$F : \Theta^N \rightarrow 2^X$	multiple choices
social welfare function (swf)	$\phi : \Theta^N \rightarrow \Theta$	social preference
social welfare correspondence (swc)	$\Phi : \Theta^N \rightarrow 2^\Theta$	multiple pref.'s

For simplicity we assume that $\Theta \subseteq L(X)$, i.e., the class of *linear preference orders*.

Social Choice Example with Two Alternatives

Given alternatives $X = \{a, b\}$, and 10 players with the following preferences: 3 players with $\theta_1 : a \succ b$ and 7 players with $\theta_2 : b \succ a$.

3	7
a	b
b	a

Question: Which alternative (a or b) is preferred?

Question: Formulate a social choice function $f : \Theta^N \rightarrow X$ for all situations with two alternatives.

Majority Voting with Two Alternatives

Definition

Majority voting is defined as follows.

- Given two alternatives a and b, three possible preference relations:

$$a \succ b \qquad a \sim b \qquad b \succ a$$

- Order candidates proportional to number of "votes" they obtain.
  - Social choice function f selects (one) candidate with the most votes.
  - Social choice correspondence F selects a subset of candidates that have the most votes.
  - Social welfare function ϕ defines the social order proportionally to the number of votes.

TUDelft

Anonymity and Neutrality

For voting rules we may require some "nice" properties, such as

Definitions

- *Anonymity*: The names of the players do not matter: if two players exchange preference relations, the outcome is not affected.
- *Neutrality*: The names of the alternatives do not matter: if we exchange a and b in the preference profile of each agent, then the outcome is affected accordingly.

Choice on Two Alternatives: May's Theorem

Definition

A social welfare function ϕ on two alternatives a and b is *positive responsive* if for all players i it holds that if a wins or it is a tie with b, and i strictly increases a in its ordering *ceteris paribus*, then alternative a wins.

Theorem

If $|X| = 2$, majority voting defines the only social welfare function satisfying anonymity, neutrality and positive responsiveness. *(May, 1952)*

Majority Rule on More than Two Alternatives

3	5	7	6
a	a	b	c
b	c	d	b
c	b	c	d
d	d	a	a

Question: Who should be the winner according to the majority rule?

Condorcet Winner

An alternative a is a *Condorcet (18th century) winner* if against any other alternative b there is a majority preferring a to b.

3	5	7	6
a	a	b	c
b	c	d	b
c	b	c	d
d	d	a	a

Question: Who is the Condorcet winner?

TU Delft

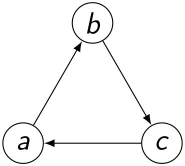
Examples of Social Choice Methods

- Majority Voting
- Condorcet Consistent Method
- Many, many others
  - e.g. Borda protocol (counting)
  - see also elections
- Why are there so many?

Condorcet Paradox

The Condorcet Paradox: A Condorcet winner does not always exist.

1	1	1
a	c	b
b	a	c
c	b	a



A Trivial Impossibility Result

Theorem

There is no anonymous and neutral social choice function (for more than 2 alternatives).

(Proof follows from Condorcet example.)

Properties of Social Welfare Functions

Definition

Properties of Social Welfare Functions

- *Pareto optimality*: If alternative *a* is unanimously preferred to alternative *b*, *b* should not be elected.
- *Non-dictatorship*: There is no player whose *preference profile* determines the strict preferences of the social welfare function.
- *Unrestricted domain*: The social welfare function should define a social preference order for any given set of preference profiles.
- *Independence of irrelevant alternatives (IIA)*: The social preference of two alternatives only depends on the *relative ordering* of these two alternatives in the individual preference relations.

Remark: IIA captures a consistency property of social choice rules. Lack of such consistency enables strategic manipulation.

Arrow's Impossibility Theorem

Theorem

Let $|X| \geq 3$ and $|N| \geq 2$, then, any social welfare function with unrestricted domain satisfying the Pareto property (or unanimity) and independence of irrelevant alternatives is dictatorial. (Arrow, 1951)

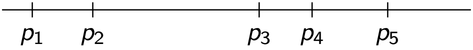
TUDelft

Circumventing the Impossibility Theorems

Argue against one of the axioms

- Independence of irrelevant alternatives
- Pareto property
- Unrestricted domain
    - quasi-linear utility functions
    - single-peaked preferences

Single-peaked Preferences



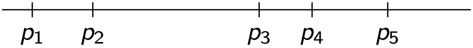
Question: What time to go to the bar?

The median is p_3.

- Not dictatorial: no single player determines outcome
- Incentive compatible (or strategy-proof), see next section
  - in contrast to taking the average of all peaks

TUDelft

Single-peaked Preferences



Question: What time to go to the bar?

The median is p_3.

- Not dictatorial: no single player determines outcome
- Incentive compatible (or strategy-proof), see next section
  - in contrast to taking the average of all peaks

Mechanism Design

TU Delft

Strategic Behavior and the Importance of Truthfulness

"Principles of voting make an election more of a game of skill than a real test of the wishes of the electors. My own opinion is that it is better for elections to be decided according to the wish of the majority than of those who happen to have most skill at the game."

(C.L. Dodgson a.k.a. Lewis Carroll, 1832-1898)

Remark: Some of the information to make a social choice is private. The outcome thus can be manipulated.

Incentive Compatibility of Social Choice Functions

Definition

A social choice function f is *dominant strategy incentive compatible* or *strategy-proof* or *truthful* w.r.t. a set Θ of type profiles, if for all $\theta^* \in \Theta$, all $\theta, \theta' \in \Theta$ and all $i \in N$:

$$f(\theta_1, \ldots, \theta_i^*, \ldots, \theta_n) \succcurlyeq_{\theta_i^*} f(\theta_1, \ldots, \theta_i', \ldots, \theta_n)$$

Intuition: A social choice function f is *strategy-proof* if for no player there are situations in which telling the truth can hurt.

Remark: Strategy-proof implementation is important!

Example: Solomon's Verdict (1/3)

He sent for a sword, and when it was brought, he said, "Cut the living child in two and give each woman half of it". The real mother, her heart full of love for her son, said to the king, "Please, Your Majesty, don't kill the child! Give it to her!" But the other woman said, "Don't give it to either of us; go on and cut it in two". Then Solomon said, "Don't kill the child! Give it to the first woman, she is its real mother."
(1 Kings 3: 16-28)

Example: Solomon's Verdict (2/3)

Three outcomes:

- a: First woman gets the baby.
- b: Second woman gets the baby.
- c: The baby is bisected.

For each woman, two possible types (good mother, bad mother):

- $\theta_1 : \quad a \succ b \succ c \qquad \theta_2 : \quad b \succ a \succ c$
- $\theta_1' : \quad a \succ c \succ b \qquad \theta_2' : \quad b \succ c \succ a$

	θ_2	θ_2'
θ_1	c	a
θ_1'	b	c

	θ_2	θ_2'
θ_1	c	a
θ_1'	b	c

TU Delft

Example: Solomon's Verdict (3/3)

- $\theta_1 : \quad a \succ b \succ c$ $\qquad \theta_2 : \quad b \succ a \succ c$
- $\theta_1' : \quad a \succ c \succ b$ $\qquad \theta_2' : \quad b \succ c \succ a$

	θ_2	θ_2'
θ_1	c	a
θ_1'	b	c

Question: Suppose that (θ_1, θ_2') is the case. Is it a dominant strategy for the second woman (the bad mother) to reveal her true preferences?

Question: Does there exist another mechanism that is truthful, for every decision Solomon needs to make?

Example: Solomon's Verdict (3/3)

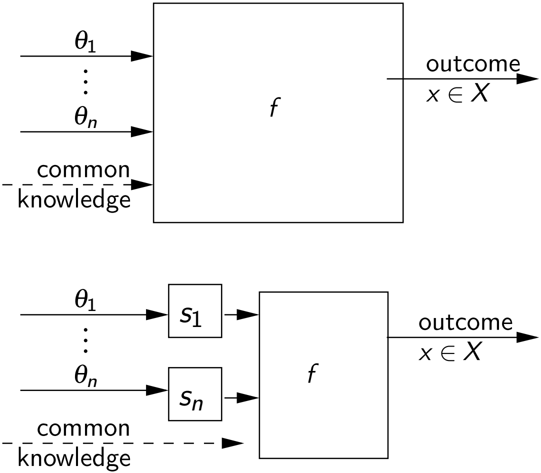
- $\theta_1: \quad a \succ b \succ c \qquad \theta_2: \quad b \succ a \succ c$
- $\theta_1': \quad a \succ c \succ b \qquad \theta_2': \quad b \succ c \succ a$

	θ_2	θ_2'
θ_1	c	a
θ_1'	b	c

Question: Suppose that (θ_1, θ_2') is the case. Is it a dominant strategy for the second woman (the bad mother) to reveal her true preferences?

Question: Does there exist another mechanism that is truthful, for every decision Solomon needs to make?

Direct vs. Indirect Mechanisms



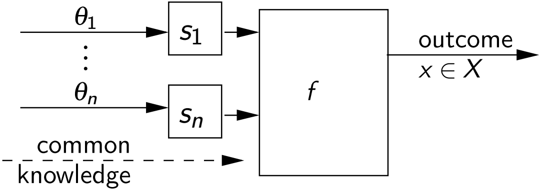
Remark: A type can contain more than just preferences.

The Revelation Principle

Theorem

If there exists any mechanism that implements an scf f in dominant strategies then there exists a direct mechanism that implements f in dominant strategies.

Intuition: For implementation in dominant strategy equilibria we can restrict ourselves to direct mechanisms.

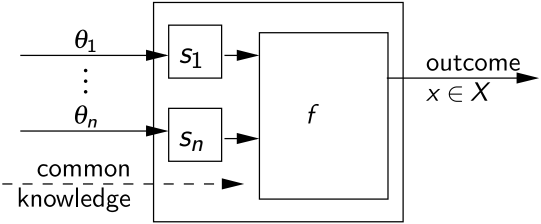


The Revelation Principle

Theorem

If there exists any mechanism that implements an scf f in dominant strategies then there exists a direct mechanism that implements f in dominant strategies.

Intuition: For implementation in dominant strategy equilibria we can restrict ourselves to direct mechanisms.



Properties of Social Choice Functions

Recall. . .

Definitions

- A social choice function f is *dictatorial* if there is a player i such that for all preference profiles θ and all outcomes x in X: $f(\theta) \succsim_{\theta_i} x$.
- A social choice function f is *strategy-proof* if if for all $\theta^* \in \Theta$, all $\theta, \theta' \in \Theta$ and all $i \in N$:

$$f(\theta_1, \ldots, \theta_i^*, \ldots, \theta_n) \succsim_{\theta_i^*} f\left(\theta_1, \ldots, \theta_i', \ldots, \theta_n\right)$$

TUDelft

Gibbard-Satterthwaite

Theorem

Let $|X| \geq 3$ and $|N| \geq 2$. Let f be a social choice function such that for all $x \in X$, there is some type profile θ with $f(\theta) = x$. Then, f is strategy-proof with respect to the set of all type profiles (unrestricted domain) if and only if f is dictatorial. (Gibbard, 1973 and Satterthwaite, 1975)

Intuition: For all non-trivial (i.e., non-dictatorial) social choice functions there are circumstances (i.e., type profiles) in which it is profitable for some player to lie.

Around Gibbard-Satterthwaite

Ways to circumvent the Gibbard-Satterthwaite theorem

- Restrict domain of preferences (e.g., add money, or single-peaked preferences)
- Weaken solution concept (e.g., not strategy-proof in dominant strategies, but in *Nash equilibrium*, i.e., no incentive to deviate assuming no one does)
- Probabilistic social choice functions (e.g., lottery-dictatorship)

TUDelft

eBay auction

TUDelft

William Vickrey (1914–1996)

- Nobel Prize in Economic Sciences at October 8, 1996
- Studied road transportation in Washington, D.C. in 1959

The second-price auction

Definition

In a *second-price auction* the winner is the highest bidder and its payment is the second-highest bid.

A strategy followed by players is *dominant* if players cannot obtain an outcome they prefer more by any other strategy.

A mechanism is *strategy-proof* if the strategy based on their true preferences is a dominant strategy.

Intuition: Players have no need to strategize in such a mechanism.

Why is the second-price auction strategy-proof?

Theorem

The second-price auction is strategy-proof (Vickrey, 1961).

Intuition: You are never worse off by bidding exactly your true valuation, because the price is independent of your own bid.

- Your valuation: v_i
- Your bid: b_i
- The maximum bid of all others: b
- Your utility if you win: $v_i - b$

Consider bidding higher, lower, or exactly your valuation…

Bid higher than your true valuation, i.e. $b_i > v_i$

Other bidders bid higher than you:
$b > b_i > v_i$

$\Downarrow$

You don't get the item.

Other bidders bid lower than you, but higher than your true valuation:
$b_i > b > v_i$

$\Downarrow$

You win the item, but have net utility:
$v_i - b < 0$.

Other bidders bid lower than your true valuation:
$b_i > v_i > b$

$\Downarrow$

You win the item and have net utility:
$v_i - b > 0$.

TUDelft

Bid lower than your true valuation, i.e. $v_i > b_i$

Other bidders bid higher than you and your true valuation:
$b > v_i > b_i$

$\Downarrow$

You don't get the item.

Other bidders bid higher than you, but lower than your true valuation:
$v_i > b > b_i$

$\Downarrow$

You lose the item, while you could have positive net utility.

Other bidders bid lower than you:
$v_i > b_i > b$

$\Downarrow$

You win the item and have net utility:
$v_i - b > 0$.

TUDelft

Bid exactly your true valuation, i.e. $b_i = v_i$

Other bidders bid higher than you:
$b > v_i$

$\Downarrow$

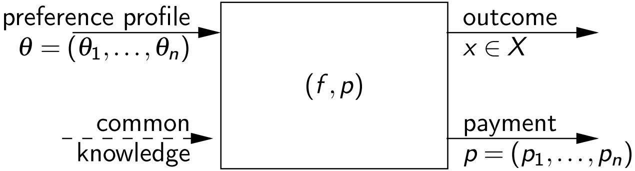
You don't get the item.

Other bidders bid lower than you:
$v_i > b$

$\Downarrow$

You win the item and have net utility: $v_i - b > 0$.

Introducing payments



- Mechanism also includes a payment function $p : \Theta^N \rightarrow \mathbb{R}^N$ that defines the payment for each agent.
- Utility of each agent i depends linearly on payment: $u_i : X \times \Theta^N \rightarrow \mathbb{R}$, i.e. $u_i(x, \theta) = v_i(x, \theta) - p_i(\theta)$ (and is therefore called a quasi-linear utility function)
- NB: The (preference) profile θ can contain other (private) information besides the valuation functions v_i of the agents.

Groves Mechanisms (1/2)

Definition

A mechanism (f, p) is called a *Groves mechanism* if

- $f(\theta) \in \arg\max_{x \in X} \sum_{i \in N} v_i(x, \theta)$, i.e. f maximizes social welfare;
- for some function $h_{-i} : \Theta_{-i}^N \rightarrow \mathbb{R}$, we have that for all valuations θ it holds that $p_i(\theta) = h_i(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n) - \sum_{j \neq i} v_j(f(\theta), \theta)$.

Groves Mechanisms (2/2)

Theorem

Every VCG mechanism for agents with quasi-linear preferences is strategy-proof and efficient.

Intuition: The payment function aligns each agent's utility with the social welfare and is independent of the agent's declared type.

Proof.

The utility of agent i is $v_i(f(\theta),\theta) + \sum_{j \neq i} v_j(f(\theta),\theta) - h_i(\theta_{-i})$.
Ignore h_i, because does not depend on θ_i.
$f(\theta) \in \arg\max_{x \in X} \sum_{i \in N} v_i(x,\theta)$, so efficient by definition, and strategy-proofness follows, because other types $\theta_i' \neq \theta_i$ cannot improve the utility of agent i. □

TU Delft

Individual Rationality

Question: When do agents want to take part in such a Groves mechanism (f, p)?

Individual Rationality

Definition

If every agent $i \in N$ has a positive utility, i.e. if $v_i(f(\theta_{-i}, \theta_i), \theta) - p_i(\theta_{-i}, \theta_i) \geq 0$, we say that it is (ex post) *individually rational (IR)* for the agents to participate.

Clarke pivot rule

Definition

The *Clarke pivot rule* is the payment function where $h_i(\theta) = \max_{x \in X} \sum_{j \neq i} v_i(x, \theta)$.

Theorem

The Vickrey-Clarke-Groves (VCG) mechanisms (i.e., Groves mechanism with the Clarke pivot payment) is (ex-post) individually rational.

Intuition: The utility of agent i with the Clarke pivot payment is $v_i(f(\theta), \theta) + \sum_{j \neq i} v_j(f(\theta), \theta) - \max_{x \in X} \sum_{j \neq i} v_i(x, \theta)$. This is its marginal contribution to the social welfare.

Are Groves Mechanisms the Way to Go?

Theorem

Groves mechanisms are the only efficient and strategy-proof mechanisms for agents with quasi-linear preferences and general valuation functions, amongs all direct-revelation mechanisms. (Green and Laffont, 1977)

Remark: This has also been extended to hold for Bayesian-Nash mechanisms by Kirshna and Perry (1998) and Williams (1999).

Shortcomings of VCG

- What if players form coalitions?
- What to do with the collected payments?
- What if we cannot compute the optimal solution and the payments exactly?

Budget-balancedness (1/3)

Definitions

- A mechanism is (ex post) *budget-balanced* if $\sum_i p_i = 0$.
- A mechanism is (ex post) *weakly budget-balanced* if $\sum_i p_i \geq 0$.

Theorem

The VCG mechanism can be made (ex post) weakly budget-balanced.

Budget-balancedness (2/3)

Theorem

No strategy-proof mechanism can implement an efficient (ex post) budget-balanced scf. *(Hurwicz, 1975 and Green-Laffont, 1979)*

However, sometimes part of the money can be *redistributed*.

Budget-balancedness (3/3)

Example

Bilateral trading

- one buyer, one seller, one good
- valuations v_1 and v_2

Theorem

Even in bilateral trading no mechanism can implement an efficient (ex-post) weak budget-balanced and (interim) IR scf, even in Bayes-Nash equilibrium. *(Myerson-Satterthwaite, 1983)*

Intuition: Not efficient if there is a chance that the buyer truly has a lower valuation than the seller.

Computational properties

Theorem

VCG requires exact optimization (approximation loses IC) in general. (Nisan and Ronen, 2007)

However

- IC is possible for some problems where approximation solution is *maximal in range (MIR)*, i.e. max and not depending on agents' types (eg multi-unit auctions).
- MIR for cost-minimization allocation problems can give solutions arbitrarily far from optimal.

Circumvent the shortcomings of VCG

- In which situations can all/most of the payments be redistributed?
- Under which conditions can the optimal solution be computed efficiently?
- For which problems can we find MIR algorithms?

Summary Social Choice Theory

Social choice theory: making social decisions

- With two alternatives: majority rule *(May).*
- With more alternatives:
  - Voting rules not neutral and anonymous (not always a Condorcet winner).
  - If IIA, Pareto, and unrestricted domain then dictatorship *(Arrow).*
- Single-peaked preferences: median voter rule.

Summary Mechanism Design

Mechanism design: type includes all relevant information.

- For every strategy-proof mechanism there is also a direct mechanism (revelation principle).
- If unrestricted range and domain and strategy-proof then dictatorship *(Gibbard-Satterthwaite).*
- Groves mechanisms are the only direct, efficient, strategy-proof mechanisms with payments *(Green-Laffont).*
- No strategy-proof mechanism can be efficient and budget-balanced *(Hurwicz, and Green-Laffont).*
- Difficulties: optimal solution required, not budget-balanced.

TU Delft

Current / Future Work

- Online mechanisms (see PhD thesis Rugierro Cavallo, 2008)
- Distributed mechanisms (see chapter by David Parkes, 2008)
- Other effects of the mechanism process and/or outcome (e.g., building/destroying relationships)
- Dynamically changing preferences and/or outcomes (convincing)
- Monitoring and enforcing outcomes (including payments), two-stage mechanisms (Mezzetti, 2004)

Recommended reading

R. J. Aumann.
Game theory.
In J. Eatwell, M. Milgate, and P. Newman, editors, *Game Theory*, The New Palgrave, pages 1–54. Macmillan, London and Basingstoke, 1987.

Y. Shoham and K. Leyton-Brown.
Multiagent Systems: Algorithmic, Game-theoretic, and Logical Foundations.
Cambridge University Press, 2008.

J. Moore.
Implementation, contracts, renegotiations in environments with complete information.
In J. Laffont, editor, *Advances in Economic Theory*, chapter 5, pages 182–282. Cambridge University Press, 1992.

N. Nisan.
Introduction to mechanism design (for computer scientists).
In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, chapter 9, pages 209–242. Cambridge University Press, 2007.