Intelligent Autonomous Agents:

Lecture 12: Mechanism Design

Ralf Möller
Hamburg University of Technology

Acknowledgement

Material from CS 886

Advanced Topics in AI

Electronic Market Design

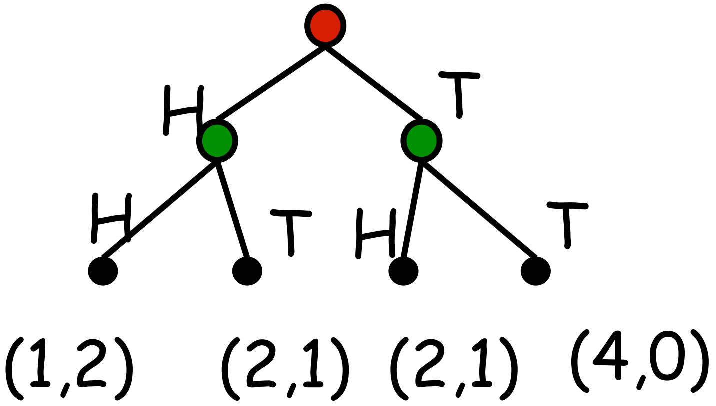
Kate Larson

Waterloo Univ.

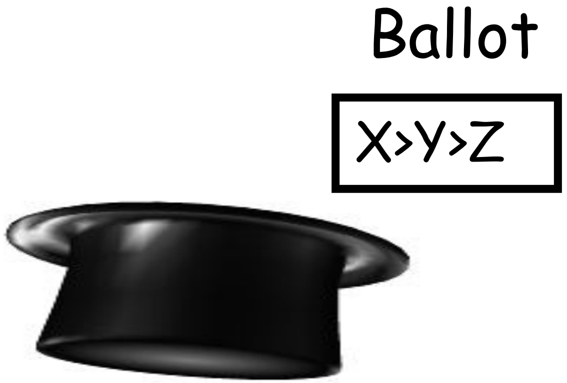
Introduction

So far we have looked at

- Game Theory
  - Given a game we are able to analyze the strategies agents will follow



- Social Choice Theory
  - Given a set of agents' preferences we can choose some outcome



Introduction

- Now: Mechanism Design
  - Game Theory + Social Choice
- Goal of Mechanism Design is to
  - Obtain some outcome (function of agents' preferences)
  - But agents are rational
    - They may lie about their preferences
- Goal: Define the rules of a game so that in equilibrium the agents do what we want

Fundamentals

- Set of possible outcomes, O
- Agents $i \in I$, $|I|=n$, each agent i has type $\theta_i \in \Theta_i$
  - Type captures all private information that is relevant to agent's decision making
- Utility $u_i(o, \theta_i)$, over outcome $o \in O$
- Recall: goal is to implement some system-wide solution
  - Captured by a social choice function (SCF)

$$f: \Theta_1 \times \ldots \times \Theta_n \rightarrow O$$

$f(\theta_1,\ldots\theta_n)=o$ is a collective choice

Examples of social choice functions

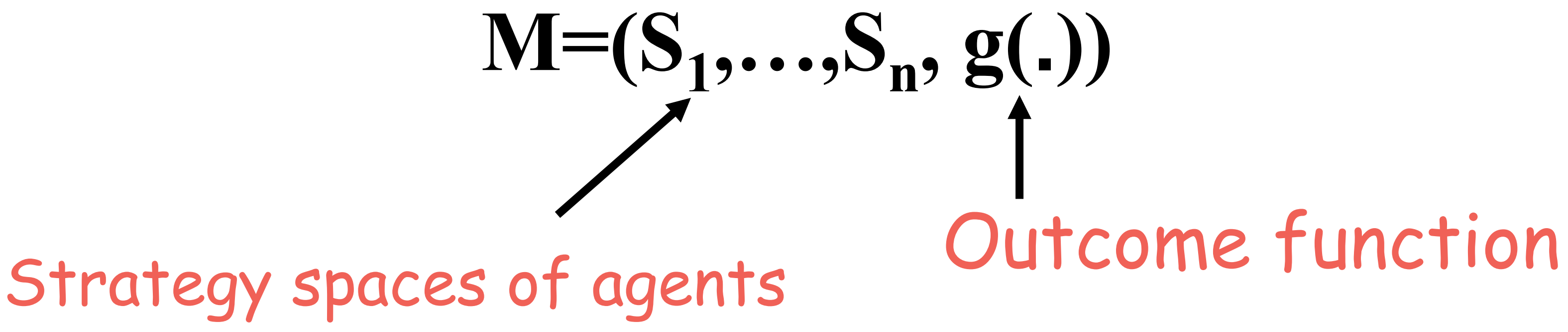
- Voting: choose a candidate among a group
- Public project: decide whether to build a swimming pool whose cost must be funded by the agents themselves
- Allocation: allocate a single, indivisible item to one agent in a group

Mechanisms

- Recall: We want to implement a social choice function
  - Need to know agents' preferences
  - They may not reveal them to us truthfully
- Example:
  - 1 item to allocate, and want to give it to the agent who values it the most
  - If we just ask agents to tell us their preferences, they may lie

Mechanism Design Problem

- By having agents interact through an institution we might be able to solve the problem
- Mechanism:



$$M=(S_1,\ldots,S_n, g(.))$$

Strategy spaces of agents

Outcome function

$$g:S_1 x\ldots x\ S_n \rightarrow O$$

Implementation

- A mechanism $\mathbf{M=(S_1,\ldots,S_n,g(.))}$ implements social choice function $\mathbf{f(\theta)}$ if there is an equilibrium strategy profile $\mathbf{s^*(.)=(s^*_1(.),\ldots,s^*_n(.))}$ of the game induced by M such that

$$\mathbf{g(s_1^*(\theta_1),\ldots,s_n^*(\theta_n))=f(\theta_1,\ldots,\theta_n)}$$

for all

$$\mathbf{(\theta_1,\ldots,\theta_n) \in \Theta_1 x \ldots x\, \Theta_n}$$

Implementation

- We did not specify the type of equilibrium in the definition
- Nash

$u_i(s_i^*(\theta_i),s^*_{-i}(\theta_{-i}),\theta_i)$ ¸ $u_i(s_i'(\theta_i),s^*_{-i}(\theta_{-i}),\theta_i)$, $\forall\ i$, $\forall\ \theta$, $\forall\ s_i' \neq s_i^*$

- Bayes-Nash

$E[u_i(s_i^*(\theta_i),s^*_{-i}(\theta_{-i}),\theta_i)]$ ¸ $E[u_i(s_i'(\theta_i),s^*_{-i}(\theta_{-i}),\theta_i)]$, $\forall\ i$, $\forall\ \theta$, $\forall\ s_i' \neq s_i^*$

- Dominant

$u_i(s_i^*(\theta_i),s_{-i}(\theta_i),\theta_i)$ ¸ $u_i(s_i'(\theta_i),s_{-i}(\theta_{-i}),\theta_i)$, $\forall\ i$, $\forall\ \theta$, $\forall\ s_i' \neq s_i^*$, $\forall\ s_{-i}$

Direct Mechanisms

- Recall that a mechanism specifies the strategy sets of the agents
  - These sets can contain complex strategies
- **Direct mechanisms:**
  - Mechanism in which $S_i=\Theta_i$ for all i, and $g(\theta)=f(\theta)$ for all $\theta\in\mathbf{\Theta}_1\mathbf{x}\ldots\mathbf{x}\Theta_n$
- **Incentive-compatible:**
  - A direct mechanism is incentive-compatible if it has an equilibrium s^* where $s^*_i(\theta_i)=\theta_i$ for all $\theta_i\in\mathbf{\Theta}_i$ and all i
  - (truth telling by all agents is an equilibrium)
  - Strategy-proof if dominant-strategy equilibrium

Dominant Strategy Implementation

- Is a certain social choice function implementable in dominant strategies?
  - In principle we would need to consider all possible mechanisms

- **Revelation Principle** (for Dom Strategies)
  - Suppose there exists a mechanism M=(S_1,…,S_n,g(.)) that implements social choice function f() in dominant strategies. Then there is a direct strategy-proof mechanism, M', which also implements f().

Revelation Principle

"the computations that go on within the mind of any bidder in the nondirect mechanism are shifted to become part of the mechanism in the direct mechanism" [McAfee&McMillian 87]

- Consider the incentive-compatible direct-revelation implementation of an English auction

Revelation Principle: Proof

- $M=(S_1,\ldots,S_n,g())$ implements SCF f() in dom str.
  - Construct direct mechanism $M'=(\Theta^n,f(\theta))$
  - By contradiction, assume

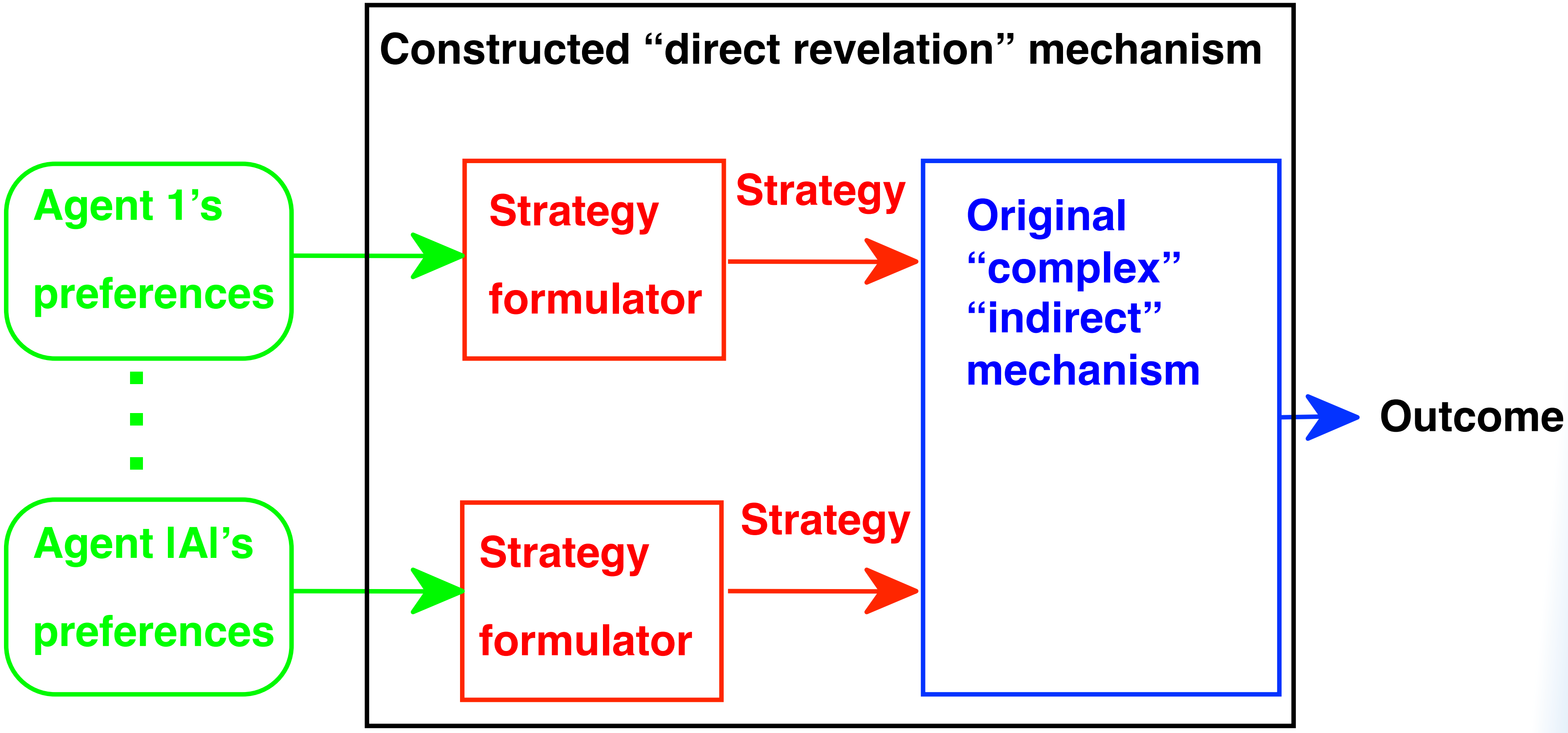
    $\exists\ \theta_i'\neq\theta_i$ s.t. $u_i(f(\theta_i',\theta_{-i}),\theta_i)>u_i(f(\theta_i,\theta_{-i}),\theta_i)$

    for some $\theta_i'\neq\theta_i$, some θ_{-i}.
  - But, because $f(\theta)=g(s^*(\theta))$, this implies

    $u_i(g(s_i^*(\theta_i'),s_{-i}^*(\theta_{-i})),\theta_i)>u_i(g(s^*(\theta_i),s^*(\theta_{-i})),\theta_i)$

    Which contradicts the strategy-proofness of s^* in M

Revelation Principle: Intuition



Theoretical Implications

- Literal interpretation: Need only study direct mechanisms
    - This is a smaller space of mechanisms
  - Negative results: If no direct mechanism can implement SCF f() then no mechanism can do it

  - Analysis tool:
    - Best direct mechanism gives us an upper bound on what we can achieve with an indirect mechanism
    - Analyze all direct mechanisms and choose the best one

Practical Implications

- Incentive-compatibility is “free” from an implementation perspective
- **BUT!!!**
  - A lot of mechanisms used in practice are not direct and incentive-compatible
  - Maybe there are some issues that are being ignored here

Quick review

- We now know
  - What a mechanism is
  - What is means for a SCF to be dominant strategy implementable
  - If a SCF is implementable in dominant strategies then it can be implemented by a direct incentive-compatible mechanism
- We do not know
  - What types of SCF are dominant strategy implementable

Gibbard-Satterthwaite Thm

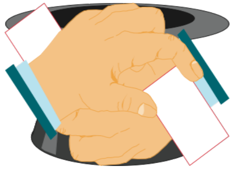
- Assume
  - O is finite and $|O| \geq 3$
  - Each $o \in \mathbf{O}$ can be achieved by social choice function f() for some θ

Then:

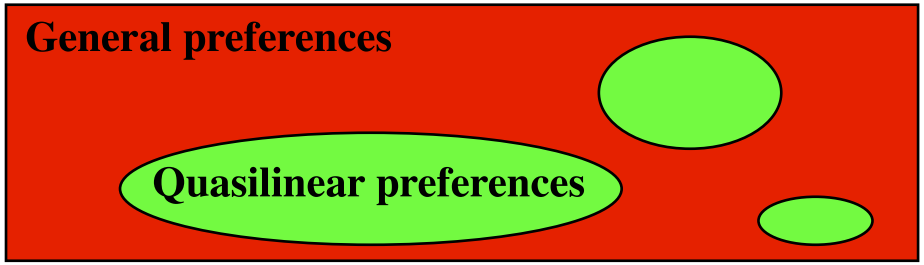
f() is truthfully implementable in dominant strategies if and only if f() is dictatorial

Circumventing G-S

- Use a weaker equilibrium concept
  - Nash, Bayes-Nash
- Design mechanisms where computing a beneficial manipulation is hard
  - Many voting mechanisms are NP-hard to manipulate (or can be made NP-hard with small "tweaks")
    [Bartholdi, Tovey, Trick 89] [Conitzer, Sandholm 03]
- Randomization
- Agents' preferences have special structure



Almost need this much



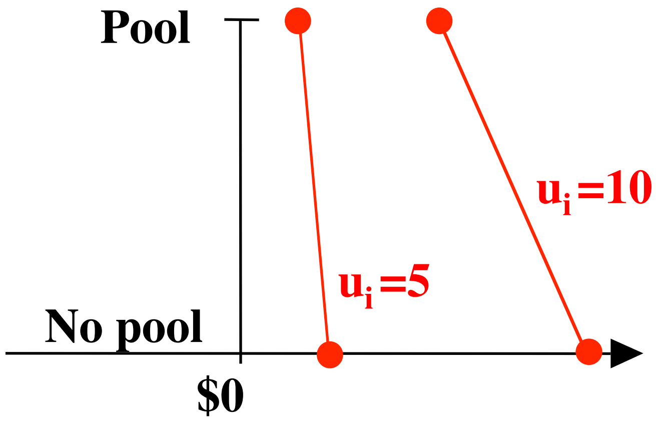
General preferences

Quasilinear preferences

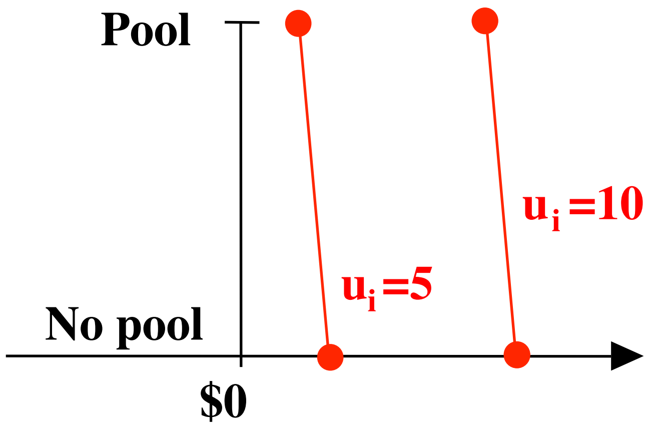
Quasi-Linear Preferences

- Example: x=”joint pool built” or “not”, m_i = \$
  - E.g. equal sharing of construction cost: $-c$ / |A|, so $v_i(x) = w_i(x) - c / |A|$
  - So, $u_i = v_i\ (x) + m_i$

General preferences



Quasilinear preferences



Quasi-Linear Preferences

- Outcome $o=(x,t_1,\ldots,t_n)$
  - x is a "project choice" and $t_i \in \mathbf{R}$ are transfers (money)
- Utility function of agent i
  - $u_i(o,\theta_i)=u_i((x,t_1,\ldots,t_n),\theta_i)=v_i(x,\theta_i)-t_i$
- Quasi-linear mechanism: $M=(S_1,\ldots,S_n,g(.))$ where $g(.)=(x(.),t_1(.),\ldots,t_n(.))$

Social choice functions and quasi-linear settings

- SCF is efficient if for all types $\theta=(\theta_1,\ldots,\theta_n)$
  - $\sum^n_{i=1} v_i(x(\theta),\theta_i) \geq \sum^n_{i=1} v_i(x'(\theta),\theta_i) \quad \forall\ x'(\theta)$
  - Aka social welfare maximizing

- SCF is budget-balanced (BB) if
  - $\sum^n_{i=1} t_i(\theta)=0$

  - Weakly budget-balanced if

    $\sum^n_{i=1} t_i(\theta)\geq 0$

Groves Mechanisms

[Groves 1973]

- A **Groves mechanism**, $M=(S_1,\ldots,S_n, (x,t_1,\ldots,t_n))$ is defined by
  - Choice rule $x^*(\theta')=\text{argmax}_x \sum_i v_i(x,\theta_i')$
  - Transfer rules
    - $t_i(\theta')=h_i(\theta_{-i}')-\sum_{j\neq i} v_j(x^*(\theta'),\theta'_j)$

where $h_i(.)$ is an (arbitrary) function that does not depend on the reported type θ_i' of agent i

Groves Mechanisms

- **Thm:** Groves mechanisms are strategy-proof and efficient (We have gotten around Gibbard-Satterthwaite!)

  Proof:

  Agent i's utility for strategy θ_i', given θ_{-i}' from agents j≠i is

  $U_i(\theta_i')=v_i(x^*(\theta'),\theta_i)-t_i(\theta')$

  $=v_i(x^*(\theta^i),\theta_i)+\sum_{j\neq i}v_j(x^*(\theta'),\theta'_j)-h_i(\theta'_{-i})$

  Ignore $h_i(\theta_{-i})$. Notice that

  $x^*(\theta')=\text{argmax} \sum_i v_i(x,\theta'_i)$

  i.e. it maximizes the sum of reported values.

  Therefore, agent i should announce $\theta_i'=\theta_i$ to maximize its own payoff

- **Thm**: Groves mechanisms are unique (up to $h_i(\theta_{-i})$)

VCG Mechanism

(aka Clarke tax mechanism aka Pivotal mechanism)

- Def: Implement efficient outcome,

$$x^{*}=\text{argmax}_{x}\sum_{i} v_{i}(x,\theta_{i}')$$

Compute transfers

$$t_{i}(\theta')=\sum_{j\neq i} v_{j}(x^{-i},\theta'_{j}) -\sum_{j\neq i}v_{j}(x^{*}, \theta_{i}')$$

Where $x^{-i}=\text{argmax}_{x} \sum_{j\neq i}v_{j}(x,\theta_{j}')$

VCGs are efficient and strategy-proof

Agent's equilibrium utility is:

$$u_{i}(x^{*},t_{i},\theta_{i})=v_{i}(x^{*},\theta_{i})-[\sum_{j\neq i} v_{j}(x^{-i},\theta_{j}) -\sum_{j\neq i}v_{j}(x^{*},\theta_{j})]$$

$$= \sum_{j} v_{j}(x^{*},\theta_{j}) - \sum_{j \neq i} v_{j}(x^{-i},\theta_{j})$$

= marginal contribution to the welfare of the system

Example: Building a pool

- The cost of building the pool is $300
- If together all agents value the pool more than $300 then it will be built
- Clarke Mechanism:
    - Each agent announces their value, v_i
    - If $\sum v_i \geq 300$ then it is built
    - Payments $t_i(\theta_i') = \sum_{j\neq i} v_j(x^{-i}, \theta'_j) - \sum_{j\neq i} v_j(x^*, \theta_i')$ if built, 0 otherwise

v1=50, v2=50, v3=250

Pool should be built

t_1=(250+50)-(250+50)=0

t_2=(250+50)-(250+50)=0

t_3=(0)-(100)=-100

Not budget balanced

Vickrey Auction

- Highest bidder gets item, and pays second highest amount
- Also a VCG mechanism
  - Allocation rule: get item if $b_i = \max_i[b_j]$
  - Every agent pays

$$t_i(\theta_i') = \sum_{j\neq i} v_j(x^{-i}, \theta'_j) - \sum_{j\neq i} v_j(x^*, \theta_i')$$

$\max_{j\neq i}[b_j]$

$\max_{j\neq i}[b_j]$ if i is not the highest bidder,

0 if it is

London Bus System

(as of April 2004)

- 5 million passengers each day
- 7500 buses
- 700 routes

- The system has been privatized since 1997 by using competitive tendering
- Idea: Run an auction to allocate routes to companies

The Generalized Vickrey Auction (VCG mechanism)

- Let *G* be set of all routes, *I* be set of bidders
- Agent *i* submits bids $v_i^*(S)$ for all bundles $S \subseteq G$
- Compute allocation S* to maximize sum of reported bids

$$V^*(I)=\max_{(S1,\ldots,SI)}\sum_i v_i^*(S_i)$$

- Compute best allocation without each agent *i*:

$$V^*(I\backslash i)=\max_{(S1,\ldots,SI)}\sum_{j\neq i} v_i^*(S_i)$$

- Allocate Si* for each agent, each agent pays

$$P(i)=v_i^*(S_i^*)-[V^*(I)-V^*(I\backslash i)]$$

Clarke tax mechanism…

- Pros
  - Social welfare maximizing outcome
  - Truth-telling is a dominant strategy
  - Feasible in that it does not need a benefactor ($\sum_i \; m_i \leq 0$)

Clarke tax mechanism…

- Cons
- Budget balance not maintained (in pool example, generally $\sum_i m_i < 0$)
  - Have to burn the excess money that is collected
  - Thrm. [Green & Laffont 1979]. Let the agents have quasilinear preferences $u_i(x, m) = m_i + v_i(x)$ where $v_i(x)$ are arbitrary functions. No social choice function that is (ex post) welfare maximizing (taking into account money burning as a loss) is implementable in dominant strategies
- Vulnerable to collusion
  - Even by coalitions of just 2 agents

Implementation in Bayes-Nash equilibrium

- Goal is to design the rules of the game (aka mechanism) so that in **Bayes-Nash** equilibrium (s_1, …, s_n), the outcome of the game is $f(\theta_1, \ldots, \theta_n)$

- Weaker requirement than dominant strategy implementation
  - An agent's best response strategy may depend on others' strategies
    - Agents may benefit from counterspeculating each others'
      - Preferences, rationality, endowments, capabilities…

  - Can accomplish more than under dominant strategy implementation
    - E.g., budget balance & Pareto efficiency (social welfare maximization) under quasilinear preferences …

Expected externality mechanism

[d'Aspremont & Gerard-Varet 79; Arrow 79]

- Like Groves mechanism, but sidepayment is computed based on agent's revelation v_i , averaging over possible true types of the others v_{-i} *
- Outcome $(x, t_1, t_2, \ldots, t_n)$
- *Quasilinear* preferences: $u_i(x, t_i) = v_i(x) - t_i$
- *Utilitarian* setting: Social welfare maximizing choice
  - Outcome $x(v_1, v_2, \ldots, v_n) = \text{argmax}_x \sum_i v_i(x)$
    - Others' expected welfare when agent i announces v_i is

$$\xi(v_i) = \int_{v_{-i}} p(v_{-i}) \sum_{j \neq i} v_j(x(v_i , v_{-i}))$$

  - Measures change in expected externality as agent i changes her revelation

* *Assume that an agent's type is its value function*

Expected externality mechanism

[d'Aspremont & Gerard-Varet 79; Arrow 79]

- **Thrm.** Assume quasilinear preferences and statistically independent valuation functions v_i. A utilitarian social choice function f: v -> (x(v), t(v)) can be implemented in Bayes-Nash equilibrium if $t_i(v_i) = \xi(v_i) + h_i(v_{-i})$ for arbitrary function h
- Unlike in dominant strategy implementation, budget balance is achievable
  - Intuitively, have each agent contribute an equal share of others' payments
  - Formally, set $h_i(v_{-i}) = -[1 / (n-1)] \sum_{j \neq i} \xi(v_j)$
- Does not satisfy participation constraints (aka individual rationality constraints) in general
  - Agent might get higher expected utility by not participating

Participation Constraints

- Agents cannot be forced to participate in a mechanism
  - It must be in their own best interest

- A mechanism is **individually rational** (IR) if an agent's (expected) utility from participating is (weakly) better than what it could get by not participating

Participation Constraints

- Let $u_i^*(\theta_i)$ be an agent’s utility if it does not participate and has type θ_i
- Ex ante IR: An agent must decide to participate before it knows its own type
  - $E_{\theta 2\Theta}[u_i(f(\theta),\theta_i)]$¸ $E_{\theta_i 2\Theta_i}[u_i^*(\theta_i)]$
- Interim IR: An agent decides whether to participate once it knows its own type, but no other agent’s type
  - $E_{\theta_{-i} 2\Theta_{-i}}[u_i(f(\theta_i,\theta_{-i}),\theta_i)]$¸ $u_i^*(\theta_i)$
- Ex post IR: An agent decides whether to participate after it knows everyone’s types (after the mechanism has completed)
  - $u_i(f(\theta),\theta_i)$¸ $u_i^*(\theta_i)$

Quick Review

- Gibbard-Satterthwaite
  - Impossible to get non-dictatorial mechanisms if using dominant strategy implementation and general preferences
- Groves
  - Possible to get dominant strategy implementation with quasi-linear utilities
    - Efficient
- Clarke (or VCG)
  - Possible to get dominant strat implementation with quasi-linear utilities
    - Efficient, interim IR
- D'AGVA
  - Possible to get Bayesian-Nash implementation with quasi-linear utilities
    - Efficient, budget balanced, ex ante IR

Other mechanisms

- We know what to do with
  - Voting
  - Auctions
  - Public projects
- Are there any other “markets” that are interesting?

Bilateral Trade (e.g., B2B)

- Heart of any exchange
- 2 agents (one buyer, one seller), quasi-linear utilities
- Each agent knows its own value, but not the other’s
- Probability distributions are common knowledge

- Want a mechanism that is
  - Ex post budget balanced
  - Ex post Pareto efficient: exchange to occur if v_b, v_s
  - (Interim) IR: Higher expected utility from participating than by not participating

Myerson–Satterthwaite Thm

- **Thm**: In the bilateral trading problem, no mechanism can implement an ex-post BB, ex post efficient, and interim IR social choice function (even in Bayes–Nash equilibrium).

Proof

- Seller's valuation is s_L w.p. α and s_H w.p. $(1-\alpha)$
- Buyer's valuation is b_L w.p. β and b_H w.p. $(1-\beta)$. Say $b_H > s_H > b_L > s_L$
- By revelation principle, can focus on truthful direct revelation mechanisms
- p(b,s) = probability that car changes hands given revelations b and s
  - Ex post efficiency requires: $p(b,s) = 0$ if ($b = b_L$ and $s = s_H$), otherwise $p(b,s) = 1$
  - Thus, $E[p|b=b_H] = 1$ and $E[p|b = b_L] = \alpha$
  - $E[p|s = s_H] = 1-\beta$ and $E[p|s = s_L] = 1$
- m(b,s) = expected price buyer pays to seller given revelations b and s
  - Since parties are risk neutral, equivalently m(b,s) = actual price buyer pays to seller
  - Since buyer pays what seller gets paid, this maintains budget balance ex post
  - $E[m|b] = (1-\alpha)\ m(b, s_H) + \alpha\ m(b, s_L)$
  - $E[m|s] = (1-\beta)\ m(b_H, s) + \beta\ m(b_L, s)$

Proof

- Individual rationality (IR) requires
  - $b\ E[p|b] - E[m|b] \geq 0$ for $b = b_L, b_H$
  - $E[m|s] - s\ E[p|s] \geq 0$ for $s = s_L, s_H$
- Bayes-Nash incentive compatibility (IC) requires
  - $b\ E[p|b] - E[m|b] \geq b\ E[p|b'] - E[m|b']$ for all b, b'
  - $E[m|s] - s\ E[m|s] \geq E[m|s'] - s\ E[m|s']$ for all s, s'
- Suppose $\alpha=\beta= ½$, $s_L=0$, $s_H=y$, $b_L=x$, $b_H=x+y$, where $0 < 3x < y$. Now,
- $IR(b_L)$: $½\ x - [\ ½\ m(b_L,s_H) + ½\ m(b_L,s_L)] \geq 0$
- $IR(s_H)$: $[½\ m(b_H,s_H) + ½\ m(b_L,s_H)] - ½\ y \geq 0$
- Summing gives $m(b_H,s_H) - m(b_L,s_L) \geq y-x$
- Also, $IC(s_L)$: $[½\ m(b_H,s_L) + ½\ m(b_L,s_L)] \geq [½\ m(b_H,s_H) + ½\ m(b_L,s_H)]$
  - I.e., $m(b_H,s_L) - m(b_L,s_H) \geq m(b_H,s_H) - m(b_L,s_L)$
- $IC(b_H)$: $(x+y) - [½\ m(b_H,s_H) + ½\ m(b_H,s_L)] \geq ½\ (x+y) - [½\ m(b_L,s_H) + ½\ m(b_L,s_L)]$
  - I.e., $x+y \geq m(b_H,s_H) - m(b_L,s_L) + m(b_H,s_L) - m(b_L,s_H)$
  - So, $x+y \geq 2\ [m(b_H,s_H) - m(b_L,s_L)] \geq 2(y-x)$. So, $3x \geq y$, contradiction. QED

Does market design matter?

- You often here “The market will take care of “it”, if allowed to.”
- Myerson–Satterthwaite shows that under reasonable assumptions, the market will **NOT** take care of efficient allocation

- For example, if we introduced a disinterested 3rd party (auctioneer), we could get an efficient allocation