CPS 173
Mechanism design

Vincent Conitzer
conitzer@cs.duke.edu

Mechanism design: setting

- The center has a set of outcomes O that she can choose from
  - Allocations of tasks/resources, joint plans, …
- Each agent i draws a type θ_i from Θ_i
  - usually, but not necessarily, according to some probability distribution
- Each agent has a (commonly known) valuation function v_i: $\Theta_i \times O \rightarrow \mathfrak{R}$
  - Note: depends on θ_i, which is not commonly known
- The center has some objective function g: $\Theta \times O \rightarrow \mathfrak{R}$
  - $\Theta = \Theta_1 \times ... \times \Theta_n$
  - E.g., efficiency ($\Sigma_i v_i(\theta_i, o)$)
  - May also depend on payments (more on those later)
  - The center does not know the types

What should the center do?

- She would like to know the agents' types to make the best decision
- Why not just ask them for their types?
- Problem: agents might lie
- E.g., an agent that slightly prefers outcome 1 may say that outcome 1 will give him a value of 1,000,000 and everything else will give him a value of 0, to force the decision in his favor
- But maybe, if the center is clever about choosing outcomes and/or requires the agents to make some payments depending on the types they report, the incentive to lie disappears…

Quasilinear utility functions

- For the purposes of mechanism design, we will assume that an agent's utility for
  - his type being θ_i,
  - outcome o being chosen,
  - and having to pay π_i,

  can be written as $v_i(\theta_i, o) - \pi_i$
- Such utility functions are called quasilinear
- Some of the results that we will see can be generalized beyond such utility functions, but we will not do so

Definition of a (direct-revelation) mechanism

- A deterministic mechanism without payments is a mapping $o: \Theta \rightarrow O$
- A randomized mechanism without payments is a mapping $o: \Theta \rightarrow \Delta(O)$
  - $\Delta(O)$ is the set of all probability distributions over O
- Mechanisms with payments additionally specify, for each agent i, a payment function $\pi_i : \Theta \rightarrow \Re$ (specifying the payment that that agent must make)
- Each mechanism specifies a Bayesian game for the agents, where i's set of actions $A_i = \Theta_i$
  - We would like agents to use the truth-telling strategy defined by $s(\theta_i) = \theta_i$

The Clarke (aka. VCG) mechanism [Clarke 71]

- The Clarke mechanism chooses some outcome o that maximizes $\Sigma_i v_i(\theta_i', o)$
  - θ_i' = the type that i reports
- To determine the payment that agent j must make:
  - Pretend j does not exist, and choose o_{-j} that maximizes $\Sigma_{i \neq j} v_i(\theta_i', o_{-j})$
  - j pays $\Sigma_{i \neq j} v_i(\theta_i', o_{-j}) - \Sigma_{i \neq j} v_i(\theta_i', o) = \Sigma_{i \neq j} (v_i(\theta_i', o_{-j}) - v_i(\theta_i', o))$
- We say that each agent pays the externality that she imposes on the other agents
- (VCG = Vickrey, Clarke, Groves)

Incentive compatibility

- Incentive compatibility (aka. truthfulness) = there is never an incentive to lie about one's type
- A mechanism is dominant-strategies incentive compatible (aka. strategy-proof) if for any i, for any type vector $\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n$, and for any alternative type θ_i', we have

  $v_i(\theta_i, o(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n)) - \pi_i(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n) \geq$

  $v_i(\theta_i, o(\theta_1, \theta_2, \ldots, \theta_i', \ldots, \theta_n)) - \pi_i(\theta_1, \theta_2, \ldots, \theta_i', \ldots, \theta_n)$
- A mechanism is Bayes-Nash equilibrium (BNE) incentive compatible if telling the truth is a BNE, that is, for any i, for any types θ_i, θ_i',

  $\Sigma_{\theta_{-i}} P(\theta_{-i}) [v_i(\theta_i, o(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n)) - \pi_i(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n)] \geq$

  $\Sigma_{\theta_{-i}} P(\theta_{-i}) [v_i(\theta_i, o(\theta_1, \theta_2, \ldots, \theta_i', \ldots, \theta_n)) - \pi_i(\theta_1, \theta_2, \ldots, \theta_i', \ldots, \theta_n)]$

The Clarke mechanism is strategy-proof

- Total utility for agent j is

  $v_j(\theta_j, o) - \Sigma_{i \neq j} (v_i(\theta_i', o_{-j}) - v_i(\theta_i', o)) =$

  $v_j(\theta_j, o) + \Sigma_{i \neq j} v_i(\theta_i', o) - \Sigma_{i \neq j} v_i(\theta_i', o_{-j})$
- But agent j cannot affect the choice of o_{-j}
- Hence, j can focus on maximizing $v_j(\theta_j, o) + \Sigma_{i \neq j} v_i(\theta_i', o)$
- But mechanism chooses o to maximize $\Sigma_i v_i(\theta_i', o)$
- Hence, if $\theta_j' = \theta_j$, j's utility will be maximized!

- Extension of idea: add any term to agent j's payment that does not depend on j's reported type
- This is the family of Groves mechanisms [Groves 73]

Individual rationality

- A selfish center: "All agents must give me all their money." – but the agents would simply not participate
  - If an agent would not participate, we say that the mechanism is not individually rational
- A mechanism is ex-post individually rational if for any i, for any type vector $\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n$, we have

  $v_i(\theta_i, o(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n)) - \pi_i(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n) \geq 0$
- A mechanism is ex-interim individually rational if for any i, for any type θ_i,

  $\Sigma_{\theta_{-i}} P(\theta_{-i}) [v_i(\theta_i, o(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n)) - \pi_i(\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_n)] \geq 0$
  - i.e., an agent will want to participate given that he is uncertain about others' types (not used as often)

Additional nice properties of the Clarke mechanism

- Ex-post individually rational, assuming:
  - An agent's presence never makes it impossible to choose an outcome that could have been chosen if the agent had not been present, and
  - No agent ever has a negative value for an outcome that would be selected if that agent were not present
- Weakly budget balanced - that is, the sum of the payments is always nonnegative - assuming:
  - If an agent leaves, this never makes the combined welfare of the other agents (not considering payments) smaller

Generalized Vickrey Auction (GVA) (= VCG applied to combinatorial auctions)

- Example:
  - Bidder 1 bids ({A, B}, 5)
  - Bidder 2 bids ({B, C}, 7)
  - Bidder 3 bids ({C}, 3)
- Bidders 1 and 3 win, total value is 8
- Without bidder 1, bidder 2 would have won
  - Bidder 1 pays 7 - 3 = 4
- Without bidder 3, bidder 2 would have won
  - Bidder 3 pays 7 - 5 = 2
- Strategy-proof, ex-post IR, weakly budget balanced
- Vulnerable to collusion (more so than 1-item Vickrey auction)
  - E.g., add two bidders ({B}, 100), ({A, C}, 100)
  - What happens?
  - More on collusion in GVA in [Ausubel & Milgrom 06, Conitzer & Sandholm 06]

Clarke mechanism is not perfect

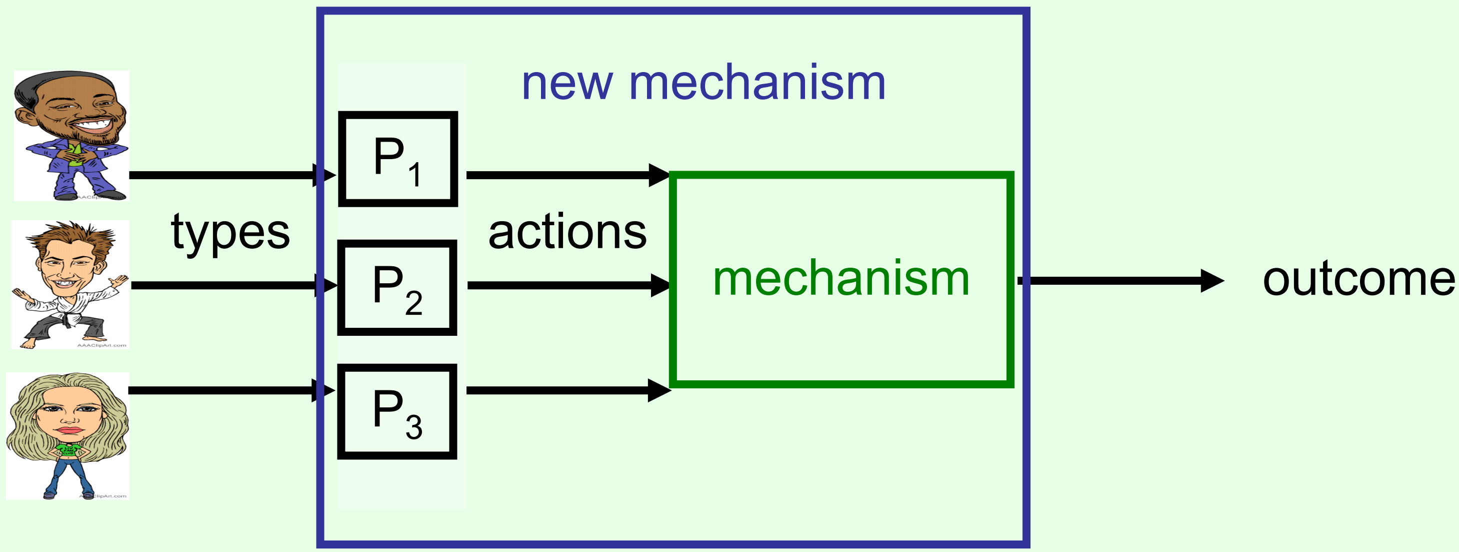
- Requires payments + quasilinear utility functions
- In general money needs to flow away from the system
  - Strong budget balance = payments sum to 0
  - In general, this is impossible to obtain in addition to the other nice properties [Green & Laffont 77]
- Vulnerable to collusion
  - E.g., suppose two agents both declare a ridiculously large value (say, $1,000,000) for some outcome, and 0 for everything else. What will happen?
- Maximizes sum of agents' utilities (if we do not count payments), but sometimes the center is not interested in this
  - E.g., sometimes the center wants to maximize revenue

Why restrict attention to truthful direct-revelation mechanisms?

- Bob has an incredibly complicated mechanism in which agents do not report types, but do all sorts of other strange things
- E.g.: Bob: “In my mechanism, first agents 1 and 2 play a round of rock-paper-scissors. If agent 1 wins, she gets to choose the outcome. Otherwise, agents 2, 3 and 4 vote over the other outcomes using the Borda rule. If there is a tie, everyone pays $100, and…”
- Bob: “The equilibria of my mechanism produce better results than any truthful direct revelation mechanism.”
- Could Bob be right?

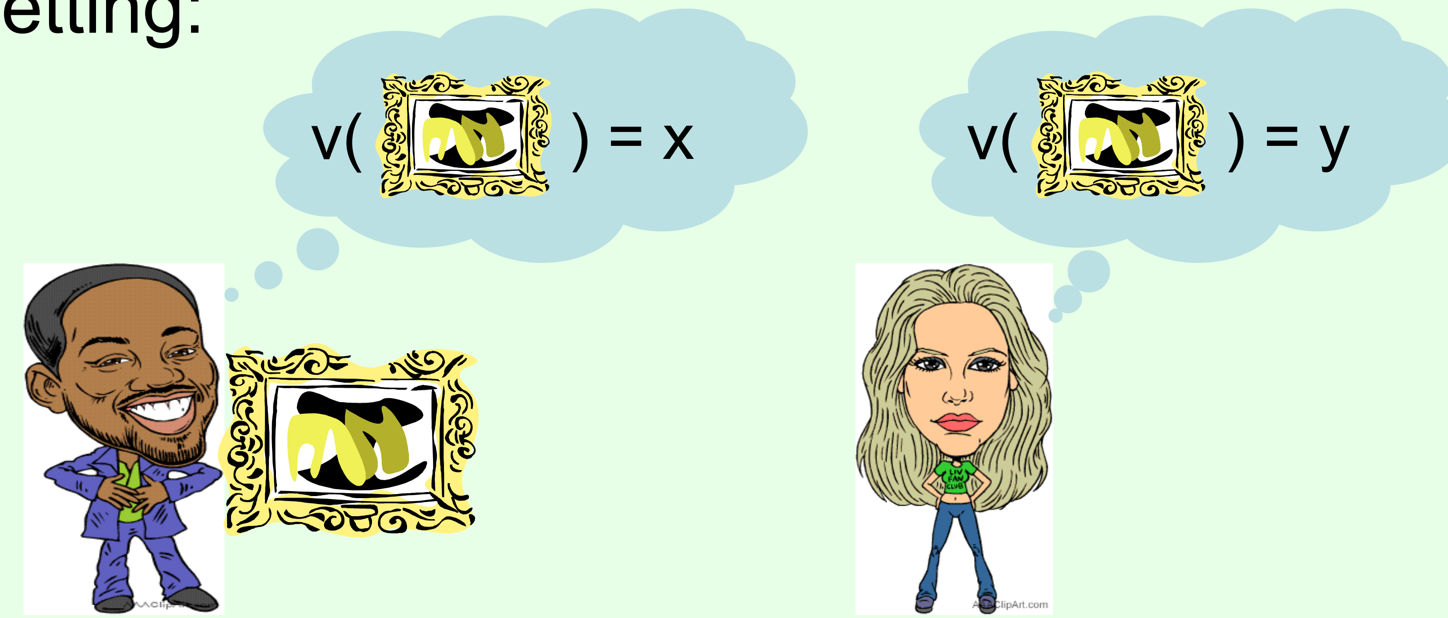
The revelation principle

- For any (complex, strange) mechanism that produces certain outcomes under strategic behavior (dominant strategies, BNE)…
- … there exists a (dominant-strategies, BNE) incentive compatible direct revelation mechanism that produces the same outcomes!



Myerson-Satterthwaite impossibility [1983]

- Simple setting:



- We would like a mechanism that:
  - is efficient (trade if and only if y > x),
  - is budget-balanced (seller receives what buyer pays),
  - is BNE incentive compatible, and
  - is ex-interim individually rational
- This is impossible!

A few computational issues in mechanism design

- Algorithmic mechanism design
  - Sometimes standard mechanisms are too hard to execute computationally (e.g., Clarke requires computing optimal outcome)
  - Try to find mechanisms that are easy to execute computationally (and nice in other ways), together with algorithms for executing them
- Automated mechanism design
  - Given the specific setting (agents, outcomes, types, priors over types, …) and the objective, have a computer solve for the best mechanism for this particular setting
- When agents have computational limitations, they will not necessarily play in a game-theoretically optimal way
  - Revelation principle can collapse; need to look at nontruthful mechanisms
- Many other things (computing the outcomes in a distributed manner; what if the agents come in over time (online setting); …)