Harvard CS 1360, Spring 2025
Instructor: Ariel Procaccia
Scribe: Taig Singh

The VCG Mechanism

—

Lecture 8

We will begin this lecture by defining what mechanisms are, and then we will introduce the VCG mechanism.

Definition 1 (Mechanisms with Payments)**.** A mechanism is defined as the following:

- A set of players $N = \{1, \ldots, n\}$ and a set of alternatives A.
- Each player $i \in N$ has a valuation function $v_i : A \to \mathbb{R}$.
- Players have quasi-linear utilities: for $x \in A$ and payment $p_i \in \mathbb{R}$,

$$u_i(x, p_i) = v_i(x) - p_i.$$

- A (direct revelation) mechanism $M = (f, p)$ takes as input a valuation profile $\mathbf{v} = (v_1, \ldots, v_n)$ and returns an alternative $f(\mathbf{v})$ and payments $p(\mathbf{v})$, where $p_i(\mathbf{v})$ is the payment of player i.

Mechanisms and voting rules are different. The input of voting rules are preference profiles (σ where σ_i is the ranking of alternatives for voter i), whereas the input of mechanisms are valuation profiles (v where v_i is the valuation function for agent i that maps alternatives to values in $\mathbb{R}$). Additionally, while both voting rules and mechanisms output an alternative ($f(\sigma)$ or $f(v)$), mechanisms also output a payment $p_i(v)$ for each agent i.

Example 1 (Auctioning a single item)**.** There is one item and n players with bids $b_1, \ldots, b_n$. The highest bidder gets the item. $A = \{a_1, \ldots, a_n\}$, where $i \in N$ gets the item in alternative a_i. Valuation functions for each agent are defined as

$$v_i(x) = \begin{cases} b_i, & x = a_i, \\ 0, & x \neq a_i. \end{cases}$$

We now define the mechanism as follows:

$$f(\mathbf{v}) = a_i \text{ such that } b_i = \max_j b_j.$$

In a *first-price auction,*

$$p_i(\mathbf{v}) = b_i, \quad p_j(\mathbf{v}) = 0 \text{ for } j \neq i,$$

whereas in a *Vickrey auction*,

$$p_i(\mathbf{v}) = \max_{j \neq i} b_j, \quad p_j(\mathbf{v}) = 0 \text{ for } j \neq i.$$

We will now extend our definition of strategyproofness to mechanisms.

Definition 2 (Strategyproofness)**.** A mechanism $M = (f, p)$ is strategyproof if for all valuation profiles $\mathbf{v}$, for all $i \in N$ and for all v_i',

$$u_i(f(\mathbf{v}), p_i(\mathbf{v})) \geq u_i(f(v_i', \mathbf{v}_{-i}), p_i(v_i', \mathbf{v}_{-i})).$$

Here, a mechanism is strategyproof if truthful reporting of agents' valuation functions is a dominant strategy. This is an analogous definition of strategyproofness as given for voting rules because instead of reporting rankings of alternatives, they are now just reporting valuation functions.

Note that the first-price auction is not strategyproof, because the highest bidder can benefit by decreasing their bid to match (or be slightly above) the second highest bid. By contrast:

Theorem 1. *The Vickrey auction is strategyproof.*

Proof. For a given agent, let v be their valuation, let b be their bid, and let b' be the highest bid that is not their bid. We will proceed with casework. We wish to show that deviating from $b = v$ to $b \neq v$ is not a useful deviation, implying that $b = v$ is a dominant strategy.

Consider the case where the agent deviates to $b < v$. Then, if $b' > v$, the agent loses in either case so the agent's utility is 0 if $b = v$ or $b < v$. If $b < b' < v$, the agent loses and receives a utility of 0 but if $b = v$ then they would win with utility $v - b' > 0$. If $b' < b$, then the agent wins in either case and the agent's utility is $v - b'$ if $b = v$ or $b < v$. Thus, by deviating from $b = v$ to $b < v$, the agent's utility either stays the same or decreases and so this is not a useful deviation.

Now consider the case where the agent deviates to $b > v$. Then, if $b' > b$, the agent loses in either case so the agent's utility is 0 if $b = v$ or $b > v$. If $v < b' < b$, the agent wins and receives a utility of $v - b' < 0$ but if $b = v$ the agent would lose and receive a utility of 0. If $b' < v$, then the agent wins in either case and the agent's utility is $v - b'$ if $b = v$ or $b > v$. Thus, by deviating from $b = v$ to $b > v$, the agent's utility either stays the same or decreases and so this is not a useful deviation.

Thus, $b = v$ is the dominant strategy for all agents, and so the Vickrey auction is strategyproof. □

William Vickrey (1914 - 1996) was a professor of economics at Columbia. He was known for receiving the Nobel Prize posthumously.

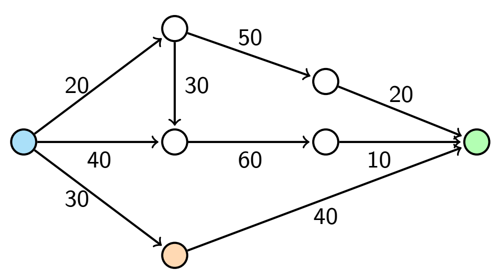
Definition 3 (The VCG Mechanism)**.** The Vickrey-Clarke-Groves (VCG) mechanism is defined by:

- A welfare-maximizing choice rule:
$$f(\mathbf{v}) \in \arg\max_{x \in A} \sum_{i \in N} v_i(x).$$
- A payment rule p, where A^{-i} is the set of alternatives that are available when i is not present:
$$p_i(\mathbf{v}) = \max_{x \in A^{-i}} \sum_{j \neq i} v_j(x) - \sum_{j \neq i} v_j(f(\mathbf{v})).$$

This is a generalization of the second-price auction described before. Here, the payment rule states that each player has to pay the marginal harm they cause to the other agents in the auction due to their bid. The first term is the other bidders' social welfare in the event that bidder i doesn't exist, and the second term is the other bidders' social welfare when bidder i does exist.

Example 2 (The VCG Mechanism 1)**.** Player 1 has value 7 for the item, and player 2 has value 3. Under VCG, player 1 gets the item and pays the second-highest bid of 3, while player 2 pays nothing. This follows because if player 1 didn't exist, player 2 would receive the item, and so $p_1(\mathbf{v}) = 3 - 0 = 0$, and if player 2 didn't exist, player 1 would still get the item so $p_2(\mathbf{v}) = 7 - 7 = 0$.

Example 3 (The VCG Mechanism 2)**.** Consider the following graph, where alternatives are represented by paths from the blue node to the green node and players are represented by directed edges with their associated costs. The valuation of an edge is minus its cost if it is selected.



Note here that A^{-i} are paths that don't include the edge associated with player i. What is the payment of the blue-orange agent? Without this edge, the minimum weighted path from blue to green (the alternative that incurs the least social cost) has cost $20 + 50 + 20 = 90$. With the blue-orange edge, the cost incurred by *other* agents is just 40 (because the minimum weighted path in this case is the blue → orange → yellow path, which has a social cost of 70). Thus, the payment made by the blue-orange agent is $(-90) - (-40) = -50$ implying that this agent gets paid \$50.

Theorem 2. *The VCG mechanism is strategyproof.*

Proof. Recall that:

$$f(\mathbf{v}) \in \arg\max_{x \in A} \sum_{i \in N} v_i(x),$$

and the utility of player i is:

$$\begin{aligned} v_i(f(\mathbf{v})) - p_i(\mathbf{v}) &= v_i(f(\mathbf{v})) - \left[\max_{x' \in A^{-i}} \sum_{j \neq i} v_j(x') - \sum_{j \neq i} v_j(f(\mathbf{v})) \right] \\ &= \max_{x \in A} \sum_{j \in N} v_j(x) - \max_{x' \in A^{-i}} \sum_{j \neq i} v_j(x') \end{aligned}$$

Now, consider a misreport to v_i'. Then, player i's utility becomes

$$\begin{aligned} v_i(f(v_i', \mathbf{v}_{-i})) - p_i(v_i', \mathbf{v}_{-i}) &= \sum_{j \in N} v_j(f(v_i', \mathbf{v}_{-i})) - \max_{x' \in A^{-i}} \sum_{j \neq i} v_j(x') \\ &\leq \max_{x \in A} \sum_{j \in N} v_j(x) - \max_{x' \in A^{-i}} \sum_{j \neq i} v_j(x') \end{aligned}$$

Implying that agent i has no useful misreports. Thus, the VCG mechanism is strategyproof. Note that player i has no control over the second term in this utility equation. □

Definition 4 (Individually Rational Mechanisms)**.** For a mechanism $M = (f, p)$, denote by $f(\mathbf{v}_{-i}) \in A_{-i}$ the outcome of the mechanism when i isn't present. M is individually rational if:

$$u_i(f(\mathbf{v}), p(\mathbf{v})) \geq u_i(f(\mathbf{v}_{-i}), 0).$$

In other words, M is individually rational if for each player i, player i is at least as well off as they would be without participating in the mechanism at all.

Theorem 3. *The VCG mechanism is individually rational.*

Proof. The difference $u_i(f(\mathbf{v}), p(\mathbf{v})) - u_i(f(\mathbf{v}_{-i}), 0)$ is

$$\begin{aligned} & v_i(f(\mathbf{v})) - \left[\sum_{j \neq i} v_j(f(\mathbf{v}_{-i})) - \sum_{j \neq i} v_j(f(\mathbf{v})) \right] - v_i(f(\mathbf{v}_{-i})) \\ &= \max_{x \in A} \sum_{j \in N} v_j(x) - \sum_{j \in N} v_j(f(\mathbf{v}_{-i})) \\ &\geq 0, \end{aligned}$$

where the inequality follows from the fact that the maximum social welfare must be at least as high as the social welfare of some alternative $\mathbf{v}_{-i}$. □