TTIC 31260 - Algorithmic Game Theory (Winter 2026)

Lecturer: Avrim Blum — Lecture 9: February 9, 2026

Mechanism Design I: Auctions, Examples, and VCG

Readings: Chapter 9.3, 9.5

1 Overview

Today we're going to talk about *mechanism design*: designing the rules for some joint interaction so that people acting in their own best interests leads to some overall desirable outcome. We will begin with the illustrative case of a single-item auction and then broaden our scope to consider mechanism design problems more broadly. In contrast to our social choice discussion from last time, we will now allow for schemes with *payments*.

2 Introduction and the Vickrey Auction

Suppose that TTIC has a spare printer, and would like to get it to the person who values it the most. Specifically, suppose there are n people, and each person has some private value $v_i \in \mathbb{R}^+$ for receiving the printer and value 0 for not getting it. We'd like the printer to go to $\arg\max_i v_i$. One potential mechanism is to just ask everyone to report their v_i's and give it to the person who reports the highest value. Unfortunately, some person might say that they liked the printer enough to pay $\$1,000,000$, even if they only liked it enough to pay \$10, because they could win with that bid and wouldn't have to pay anything.

To fix this problem, we will allow for schemes with *payments* and we will assume that people's utilities are *linear in money*, that is, the utility u_i for person i if they get the printer and pay p is $v_i - p$.

Given that we allow payments, a second attempt would be ask each person how much they would be willing to pay, and to sell it to the highest bidder at their bid. In this case, there is a different problem. The highest bidder i would rather bid something just very slightly more than the second-highest bidder, so they can pay less than v_i and have positive utility for the outcome of the "auction". In particular, bidding v_i is a bit silly since then you have guaranteed your utility will be 0 no matter what.

Finally, there are two very related mechanisms that don't have either of these problems. One is an ascending auction, where the bidding starts at \$1 and all bidders start with their hands raised. The price rises in small increments and bidders lower their hands whenever they feel like dropping out of the auction; finally, once there is only one bidder left, the auction ends and the bidder remaining gets the printer at that price. In this auction, there is no reason for any bidder to lower their hand *before* their value v_i is reached (that would just ensure a utility of 0) and no reason for any bidder to keep their hand rasied *after* their value v_i has been passed (that would just ensure a negative utility); so each bidder will lower their hand exactly when their v_i is reached and the printer will go to the bidder of highest value. Second, consider a second-price, or Vickrey auction, where everyone submits *sealed* bids, and the printer goes to the i with the highest bid, at the second-highest price. To give some notation, give the printer to $i = argmax_j v_j$ and charge price $p = max(v_{-i})$ (where v_{-i} denotes all the bids except the bid of player i). It turns out that in a Vickrey auction, it is a *dominant strategy* to give your true value as a bid. That is, regardless of what every other bidder does, any particular bidder is at least as well-off with the result if they tell the truth.

Claim 1 *The Vickrey auction is dominant-strategy truthful, or incentive compatible. For any* i*, any* v_{-i}*,* $u_i(Vic(v_i, v_{-i})) \geq u_i(Vic(v_i', v_{-i}))$*, where* $Vic(\vec{v})$ *is the outcome of running the Vickrey auction on vector of bids* $\vec{v}$*.*

There are several proofs of this claim. Here's one.

Proof: Consider bidder i and let p be the maximum bid out of all the *other* bidders.

If $v_i > p$, then announcing v_i gives a utility of $v_i - p$. Announcing any other value will either produce the same outcome (getting the item and paying p) or will result in someone else getting the item, giving a utility of 0. So, bidder$i$ is best off saying v_i.

If $v_i = p$, there is zero utility whether or not you win the auction. So you may as well tell the truth.

If $v_i < p$, then if you bid truthfully, you don't win and get utility 0. Any other bid will either produce the same outcome or else will result in getting the item at a cost of $p > v_i$, yielding negative utility. ■

Another way to think of it: Vickrey is like allowing everyone to go last, giving each person a price p they can buy at or not that *does not depend on what they announced.* E.g., based on everyone else, an offer is proposed to the person, and only then does the system look at their announced value and act on their behalf.

Notice that the ascending auction also gives the item to the highest bidder at a price equal to the second-highest value. So, Vickrey is very much like having the system act on your behalf in an ascending auction.

A scheme is said to be *economically efficient* if it maximizes social welfare. Social welfare is defined as the sum of everyone's utilities, including the auctioneer who has zero value on the item (but values money like everyone else). So, payments cancel out in computing social welfare, and social welfare is maximized by having the item go to the agent who values it the most. So, the Vickrey auction and ascending auction are incentive-compatible and economically efficient.

Claim 2 *Vickrey auctions maximize social welfare: that is, the winner of the auction is the person who values the item the most.*

Maximizing social welfare is a somewhat intuitive notion. It makes sense that we want to get our good to whoever wants it the most. But why do we care about incentive compatibility? One reason is that from the bidder's perspective, playing the game has now become easy. They can do less strategizing and just tell the truth. It's like the difference between buying groceries and buying a car. When you buy groceries, you see the prices, and can decide what to buy based on just what *you* want. When buying a car, you haggle, and have to think about how much the seller might be willing to go down, etc. Another is that from the auctioneer's perspective, there is more predictability in how the mechanism behaves: the auctioneer doesn't need to worry about what bidders might believe about other bidders, for instance.

Now, suppose that TTIC has two identical printers. There are two natural mechnisms one might try here:

1. Sell printer #1 to the highest bidder at the second-highest bid; sell printer #2 to the second-highest bidder at the third highest bid.
2. Sell them to the top 2 bidders at the third-highest bid.

It turns out the first one isn't incentive compatible (see homework). The second one is.

What if the two printers are not idential? Suppose one printer is much nicer, and the other one is an older printer that often jams. What if there are also ink cartridges that are worth more to a bidder if they also get the printer? The amazing thing is the Vickrey auction can be generalized to essentially any setting where you have payments and the players have "quasi-linear utilities". This will be the VCG mechanism.

3 Mechanism design, general setup

We assume we have n players, and a set of "alternatives" A (we will also call them outcomes or allocations) such as who gets all the various items. Each player i has a valuation function $v_i : A \to \mathbb{R}$. These can be arbitrary (e.g., you can have items worth more together than separately like a printer and ink, and you can even have higher value on allocations that give more items to your friends and lower value on allocations that give items to your enemies). The one assumption we will make is that players' utilities are *quasilinear*: The utility for player i of allocation a, paying p_i is $u_i = v_i(a) - p_i$.

Definition 1 *A* direct revelation mechanism $(f, p) : V \to A$ *takes in a sequence v of valuation functions $v = (v_1, v_2, \ldots, v_n)$, selects some allocation a and some payments $p = (p_1, p_2, \ldots, p_n)$. Specifically, $f(v) = a$ and $p(v)$ is the vector of payments with player i making a payment of $p_i(v)$. Note that payments could be negative, say for cases like a procurement auction.*

Definition 2 *A mechanism (f, p) is* incentive-compatible *(dominant strategy truthful) if for all $v = (v_1, ..., v_n)$ and all i, for all v'_i, we have:*

$$v_i(f(v)) - p_i(v) \geq v_i(f(v'_i, v_{-i})) - p_i(v'_i, v_{-i}).$$

In other words, misreporting can never help.

Definition 3 *The* social welfare *of some allocation a is $\sum_j v_j(a)$. A mechanism is* efficient *if it maximizes social welfare.*

Here is the amazing thing: there exists a mechanism, called VCG (Vickrey-Clarke-Groves) that in this very general setting is both incentive-compatible and produces the allocation that maximizes social welfare.

The key ideas is that we should somehow set the payments so that, when individuals maximize their personal utility, this coincides with them maximizing the overall social welfare. The VCG mechanism does just that. We will give several versions, starting with one that's simplest to analyze and then modifying it to havs additional important properties.

4 VCG, version 1

Given a vector v of reported valuation functions,

- Let $f(v) = argmax_{a \in A} \sum_j v_j(a)$ be the allocation that maximizes social welfare according to the reported valuations.
- Let $p_i(v) = -\sum_{j \neq i} v_j(f(v))$. I.e., we're going to pay each player i an amount equal to the sum of everone else's reported valuations for the allocation chosen.

What is the utility of player i if i truthfully reports v_i?

$$u_i(f(v), p) \;=\; v_i(f(v)) + \sum_{j \neq i} v_j(f(v)) \;=\; \sum_j v_j(f(v)).$$

Notice that this is the social welfare of the allocation chosen accoding to the reported valuations. Notice also that we chose the $f(v)$ that maximizes social welfare according to these valuations. Now, what if player i instead reports some $v_i' \neq v_i$? Then, their utility is

$$u_i(f(v_i', v_{-i}), p) \;=\; v_i(f(v_i', v_{-i})) + \sum_{j \neq i} v_j(f(v_i', v_{-i})) \;=\; \sum_j v_j(f(v_i', v_{-i}))$$

Notice that we have $v_i(...)$ not $v_i'(...)$ because v_i is their true valuation function. So, the utility is the social welfare of the new allocation $f(v_i', v_{-i})$ according to the same original vector v. This can't be higher than the social welfare of the previous allocation *because that one was chosen to be social-welfare maximizing* according to v. So, bidder i maximizes their own utility by reporting v_i rather than any other v_i'. This proves that VCG version 1 is incentive-compatible and (by design) produces the allocation that maximizes social welfare.

Now, one problem with VCG version 1 is that if you think of this as an auction, it requires the auctioneer to give money to the bidders! For example, in the case of auctioning a printer, the winner gets the printer for free ($\sum_{j \neq i} v_j(f(v)) = 0$) and everybody else gets paid the amount that the top bidder valued the printer. E.g., if the top bidder valued the printer at \$100, then that bidder gets the printer and everyone else gets \$100 cash.

However, notice that if we add to each $p_i(v)$ some amount that depends on v_{-i} only and is not influenced at all by v_i, then that is just a constant as far as player i is concerned and doesn't change the incentive compatibility of the mechanism. This leads us to our next algorithm

5 VCG - general version

For each $i = 1, ..., n$, let h_i be some function over v_{-i}. Given a vector v of valuation functions,

- Let $f(v) = argmax_{a \in A} \sum_j v_j(a)$ be the allocation that maximizes social welfare.
- Let $p_i(v) = h_i(v_{-i}) - \sum_{j \neq i} v_j(f(v))$

As we just argued, this is incentive-compatible too.

Now, there is a specific set of h_i's that have the nice properties:

- The auctioneer never pays the bidder
- Assuming the v_i's themselves are non-negative, no player ever gets negative utility. This is called *ex-post individual rationality.* For example, in an auction of goods where people's valuations depend only on what they get, then among other things this implies that people who don't get anything don't have to pay anything.

This set of h_i's is called the *Clarke pivot rule*:

$$h_i(v_{-i}) = \max_a \left[\sum_{j \neq i} v_j(a) \right] = \max_a \left[v_{-i}(a) \right]$$

This now gives us the standard version of VCG.

6 VCG - standard version

Given a vector v of valuation functions,

- Let $f(v) = argmax_{a \in A} \sum_j v_j(a)$ be the allocation that maximizes social welfare.
- Let $p_i(v) = \max_a \left[v_{-i}(a) \right] - \sum_{j \neq i} v_j(f(v))$.

In other words, you charge each player i an amount equal to how much less happy they make everyone else (how much they reduce everyone else's total social welfare) by causing $f(v)$ to be chosen rather than $f(v_{-i})$. This is often called *charging them their externality.*

Why does this satisfy $p_i(v) \geq 0$? This is because the first term is the max.

Why does this satisfy individual rationality assuming the v_i functions are non-negative? That's because:

$$\begin{aligned} u_i(f(v), p) &= v_i(f(v)) + \sum_{j \neq i} v_j(f(v)) - \max_a \left[v_{-i}(a) \right] \\ &= \max_a [v(a)] - max_a \left[v_{-i}(a) \right]. \end{aligned}$$

This can't be negative because one option for the first "a" is to use the second "a", and v_i itself is non-negative.

You should convince yourself that in the case of selling a single item, this becomes exactly the Vickrey auction.