1 Mechanism Design (incentive-aware algorithms, inverse game theory)

- How to give away a printer
- The Vickrey Auction
- Social Welfare, incentive-compatibility
- The VCG (Vickrey-Clarke-Groves) mechanism

Today we're going to talk about an area called *mechanism design*, also sometimes called *incentive-aware algorithms* or *inverse game theory*. But rather than start with the big picture, let's start with an example.

1.1 How to give away a printer

Say TTIC has a spare printer, and wants to give it to whoever can make the most use of it. To make this formal, let's say there are n people, and assume each person i has some value $v_i \geq 0$ (called their *private value*) on getting the printer, and 0 on not getting it. We'll assume everyone knows their own v_i.

What we want to do is to give the printer to the person with the highest v_i. So, one option is we ask each person for their own v_i and we then compute the argmax (the i for which v_i is maximum) and give it to that person.

Can anyone see any potential problems with this? The problem is people might lie (*misreport* is the formal term) because they want the printer.

So, let's assume one more thing, which is that we (TTIC) have the ability to charge people money, and that the utility of person i for getting the item and paying p is $v_i - p$. (The definition of *utility* for our purposes is: the thing that people/players want to maximize. If there are probabilities, then we assume they want to maximize expected utility. But everything today will be deterministic.) Let's use $u_i(x)$ to denote the utility of player i for outcome x. For instance, $u_i(x) = 0$ for the outcome "get nothing, pay nothing".

To be clear: while we are giving us the ability to charge money, it's goal isn't to make money but just to use this to help in getting the printer to the right person.

What about asking people how much they would pay (ask each person i to write down a bid b_i) and then give the printer to the person who bids the highest, charging them that amount. Any potential problems with this? Would you write down the amount that you value the printer as your bid? No. Because even if you win, you'd get zero utility.

Here is something interesting we can do called a Vickrey auction.

1.2 The Vickrey auction

The Vickrey auction:

- Ask everyone each person i to report the value of the printer to them (let's call this their "bid" b_i).
- Give the printer to $arg \max_i b_i$ (the person of highest reported value).
- Charge that person the *second-highest* bid.

Claim: Vickrey is *dominant-strategy truthful*, aka *incentive-compatible*. Specifically, for any valuations $v_1, \ldots, v_n$, for any player i, for any vector of bids of the other players (call this b_{-i}), we have:

$$u_i(\text{Vickrey}(v_i, b_{-i})) \;\geq\; u_i(\text{Vickrey}(v_i', b_{-i})).$$

Notation: given a vector v, will write v_{-i} as the vector removing the ith component, and "(x, v_{-i})" as the vector v with the ith component replaced by x.

In other words: even if you knew what all the other bids were, and got to choose what to bid based on those, you would still be best off (have highest utility) bidding your true value on the printer.

Can anyone see why?

Proof: Consider player i and let p be the highest bid among everyone else.

Case 1: $v_i > p$. In this case, if player i announces truthfully then it gets the item and pays p and has positive utility. Any other v_i' either will produce the same outcome or will result in someone else getting the item, for a utility of 0.

Case 2: $v_i = p$. Then it doesn't matter. Utility is 0 no matter what.

Case 3: $v_i < p$. In this case, announcing truthfully, player i doesn't get the item and has utility 0. Any other v_i' will either have the same outcome or else (if $v_i' > p$) will give him the item at a cost of p, yielding negative utility.

■

Another way to think of it: Vickrey is like a system that bids for you, up to a maximum bid of whatever you tell it, in an ascending auction where prices go up by tiny epsilons (i.e., just like ebay). Initially everyone is in the game and then they drop out as their maximum bids are reached. In this game, you would want to give v_i as your maximum bid (there is no advantage to dropping out early, and no reason to continue past v_i).

So, Vickrey is incentive-compatible and (assuming everyone bids their valuations, which they should because of IC) gives the printer to the person of highest value for it. This is called "maximizing social welfare".[1]

[1] Economists will call this "efficient". E.g., an "efficient market" is one that gets goods to the people who value them the most. We won't use that terminology because it clashes with the "runs quickly" meaning of "efficient".

1.3 What about two printers?

What if TTIC has two (equal quality) printers?

One option is you could do one Vickrey auction after another. Equivalently, take in the bids, give one printer to the highest bidder at a price equal to the 2nd-highest bid, and then give the other printer to the 2nd-highest bidder at a price equal to the 3rd highest bid. Does this work (is it incentive-compatible)? No. Why not?

How about giving the printers to the top 2 bidders at the 3rd-highest price? Does this work? Yes! Why?

> If you don't get a printer, would you regret your decision of bidding your true value v_i and want to raise your bid? No. If you do get a printer, you have no way of lowering the price you paid, and are happier than (or at least as happy as) if you lowered your bid below the 3rd highest and didn't get the printer.

So, this procedure (a) is incentive-compatible and (b) gives the printers to the two people who value them the most—i.e., it maximizes social welfare—if everyone bids truthfully (which they might as well do, by (a)).

So, you can think of this as inverse game theory, because we are designing the rules of the game so that if people act in their own interest, an outcome that we want will occur.

1.4 More general scenarios: the formal setup

What if TTIC has two printers but one is nicer than the other? Or maybe some things that go together (like bagels and cream cheese) or even cubicle space where you might care not only about the space you get but maybe you also would prefer a space near to someone else with the same advisor? The amazing thing is the Vickrey auction can be generalized to essentially any setting where you have payments and the players have what are called "quasi-linear utilities". This will be the Vickrey-Clarke-Groves, or VCG, mechanism. Let's now define the general setup formally.

We have n players and a set of alternatives A (will also call them "allocations"), such as who gets the printers or what the assignment of students to cubicles is. It can be arbitrarily complicated (we're not going to be worried about running time here). Each player i has a valuation *function* $v_i : A \to \mathbb{R}$.

We assume **quasi-linear utilities**: The utility for alternative $a \in A$ and paying a payment p is $v_i(a) - p$. It's called "quasi-linear" because it is linear in money, even if it might be some weird function over the alternatives. E.g., if we have multiple items we are allocating, the utility does not have to be additive over the items you get (maybe you need several together to build a product or maybe two printers isn't much better than one, and it even can depend on what other people get!) but you are assumed to be linear in money.

The **social welfare** of an allocation a is $SW(a) = \sum_i v_i(a)$. Notice that it involves the values, *not the utilities.* However, we can think of it as the sum of utilities if we also put the utility of the center (TTIC in this case) in the picture, so the money cancels out.

A **direct revelation mechanism** is a function that takes in a sequence $v = (v_1, ..., v_n)$ of valuation functions, and selects an alternative $a \in A$, along with a vector p of payments. It will be convenient to split it into two functions: $f(v) = a$ and $p(v) =$ the vector of payments. We will use $p_i(v)$ to denote the payment of player i.

A direct revelation mechanism (f, p) is **incentive-compatible** if for every $v = (v_1, ..., v_n)$, every i, every v_i', we have:

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

I.e., misreporting can never help.

Here is the amazing thing: there exists a mechanism, called VCG, that in this very general setting is both (a) incentive compatible and (b) produces the alternative that maximizes social welfare if everyone reports truthfully (which they should, due to (a)).

1.5 The Vickrey-Clarke-Groves (VCG) mechanism

The basic idea is to design payments so that everyone wants to optimize what *we* want to optimize, namely social welfare. There are a couple versions. Let's start with the simplest to analyze:

VCG version 1: Given a vector of reported valuation functions v,

- Let $f(v)$ be the allocation that maximizes social welfare with respect to v.
  I.e., $f(v) = arg\max_{a \in A} \sum_j v_j(a)$.
- Pay each player i an amount equal to the sum of everyone else's reported valuations.
  I.e., $p_i(v) = -\sum_{j \neq i} v_j(f(v))$.

Analysis: Suppose player i reports truthfully. Then its utility will be

$$v_i(f(v)) + \sum_{j \neq i} v_j(f(v)) \quad = \quad \sum_j v_j(f(v)) \; = \; \max_a \sum_j v_j(a) \quad \text{(by the definition of } f(v))$$

Suppose instead player i reports v_i'. Call the resulting vector v'. Then player i's utility will be:

$$v_i(f(v')) + \sum_{j \neq i} v_j(f(v')) \quad = \quad \sum_j v_j(f(v')) \; \leq \; \max_a \sum_j v_j(a).$$

So, misreporting can only hurt. This means that the mechanism is incentive-compatible, and by design it maximizes social welfare when everyone reports truthfully.

Problems with version 1: If you think of this as an auction, a big problem is this requires the auctioneer to give money to the bidders! E.g, in the case of the printer, it corresponds to giving the top guy the printer for free, and pay everyone else the amount the top guy valued it. That way everyone gets a utility equal to what the top guy got.

However, notice that if we add to each $p_i(v)$ something that depends on v_{-i} only (and not influenced at all by v_i) then it is just a constant as far as player i is concerned and so still incentive-compatible. This suggests the following generalization:

VCG - general version: Let h_i be any function over v_{-i} for each $i = 1 \ldots n$. Now, given a vector of reported valuation functions v,

- Let $f(v)$ be the allocation that maximizes social welfare with respect to v.
  I.e., $f(v) = arg\max_{a \in A} \sum_j v_j(a)$.
- 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 that (a) the center is never paying the bidders/players, and (b) on the other hand, assuming the v_i's themselves are non-negative, no player gets negative utility: this is called "individual rationality" (e.g., if we're allocating goods and 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": $h_i(v_{-i}) = \max_a \sum_{j \neq i} v_j(a)$. This gives us the following:

VCG - standard version: Given a vector of reported valuation functions v,

- Let $f(v)$ be the allocation that maximizes social welfare with respect to v.
  I.e., $f(v) = arg\max_{a \in A} \sum_j v_j(a)$.
- Let $p_i(v) = \max_a \sum_{j \neq i} v_j(a) - \sum_{j \neq i} v_j(f(v))$.

In other words, you charge each player i an amount equal to how much less happy they make everyone else by participating in the maximization. This is often called "charging them their externality".

Why does this satisfy $p_i(v) \geq 0$? Answer: because the 1st term is a max.

Why does this satisfy individual rationality? Think of it this way: if you got value 3 but hurt everyone else's total value by more than 3, then this couldn't have been the maximum social welfare allocation since a better allocation would have been to use the the optimal allocation without you and give you nothing.

What does this look like for the case of the single printer? Everyone who doesn't get the printer pays nothing (both terms are equal to the maximum guy's value). The person who gets the printer pays the second-highest valuation (since the sum of everyone else's valuations went from the second-highest valuation down to zero.) So it reduces to the Vickrey auction in that case.