Algorithmic Game Theory, Summer 2025		Lecture 24 (8 pages)
	Voting	
Thomas Kesselheim		Last Update: June 18, 2025

Today, we will turn to a mechanism-design problem, in which you once again don't want to have money involved: Voting. Say, you want to design presidential elections. How would you design the rules? If the choice will be only among two candidates, this is easy.

Example 24.1. *If there are only two candidates A and B, let each voter vote for one of the candidates. The winner is whoever receives the majority of votes. (Ignore the issue of tie-breaking here.)*

This "voting rule" fulfills almost every design goal that we could ask for. For example, there is no other "better" outcome. Only a minority of voters would prefer the other candidate. All votes count the same, so the identities of voters do not matter. While this may sound boring, just consider the US presidential elections. There, frequently candidates get elected that get fewer votes than another candidates.

Another advantage is that participating as a voter is easy. (We always assume that you know your preferences.) You simply vote for the candidate that you prefer. No voter can influence the outcome in his favor by not voting for his favorite candidate.

If we have three or more candidates, things are not nearly as clear, clean, and simple. There are different voting and ranking rules and some of them fulfill more desirable properties than others. As we will see today, there is a plethora of possible voting rules, some of which are better than others, but none of them actually fulfills all the desired properties at once. Indeed, we will also prove formally that this is just not possible.

1 Formal Problem Statement

Again we have n agents N. Furthermore, there are m candidates Γ. Each of the agents has a complete preference list over the candidates without ties. That is, agent i has some preference relation $\succ_i$, which is a total order over Γ. We write $a \succ_i b \succ_i c$ if agent i strictly prefers candidate a over candidate b and b over c. Note that this also implies $a \succ_i c$ (transitivity).

There are two problems of interest:

- In the *voting problem*, given preferences $\succ_1, \ldots, \succ_n$, we have to come up with one winner from Γ.
- In the *ranking problem*, given preferences $\succ_1, \ldots, \succ_n$, we have to come up with a *society ranking* $>$ of all candidates. That is, besides finding a winner, we also have to choose a second, a third, and so on.

So, our goal is to design a *voting rule* or a *ranking rule*. In the case of two candidates, these two notions coincide.

These problems are different from most of what we have seen so far. We do not have an "objective function". There are different ways to evaluate the outcome and naturally there are trade-offs. Furthermore, the outcome affects the entire society. In all other cases, if we simply took the union of two sets of agents, we could still have treated them as separate groups.

2 Examples of Voting and Ranking Rules

Let us see possible examples of voting and ranking rules with three or more candidates and let us find out their weaknesses.

2.1 Plurality Voting

In plurality voting, every voter can vote for one candidate. The candidate who got the most votes wins the election. This can be translated into a voting rule within our framework by soliciting the entire preferences but only counting how often one candidate is ranked in the top-most position by an agent. We can easily also turn this voting rule into a ranking rule by ordering the candidates by decreasing number of votes.

We all know examples in which plurality voting fails to give a "natural outcome". Suppose there are two candidates A and B from one party (so their policy is very similar) and another candidate C from a very different party.

60 % of the voters prefer the party of A and B, 40 % prefer C's party. In terms of preferences this may look as follows:

$$35\,\% : A \succ_i B \succ_i C$$
$$25\,\% : B \succ_i A \succ_i C$$
$$40\,\% : C \succ_i A \succ_i B$$

In this case, C wins the election despite the fact that most voters would be in favor of the party of A and B.

In the light of strategic behavior this also means that we cannot expect truthful reports. The agents in the $B \succ_i A \succ_i C$ group could decide to all report $A \succ_i B \succ_i C$ instead. Then candidate A wins, who they prefer to C.

2.2 Runoff Voting and Instant-Runoff Voting

A very common way to deal with the deficiencies of plurality voting is to run a second round in which the voters can choose between the two candidates that got the highest number of votes in the first round. This is called *runoff voting*. In our framework it can be captured by first counting two first preferences just as in plurality voting. Among the top two candidates, call them X and Y, one counts for how many i one has $X \succ_i Y$ in comparison to $Y \succ_i X$. Whoever gets the majority wins the vote.

If one has the full preference lists of all agents, one can also extend this procedure to multiple stages, called *instant-runoff voting*. Again, we solicit the preferences of all agents and count how often a candidate is ranked first. The candidate who receives the smallest number of first-rank votes is then eliminated (that is, he gets removed from all preferences) and the procedure is re-iterated until only a single candidate is left. For three candidates, this is equal to runoff voting, for more the result is potentially different.

Obviously, the outcome on the above example is different now. First, candidate B gets eliminated. Then, candidate C is eliminated because only 40 % prefer him to A.

However, sometimes it may still be interesting for agents to misreport. For example, consider the following preferences.

$$30\,\% : A \succ_i B \succ_i C$$
$$45\,\% : B \succ_i C \succ_i A$$
$$25\,\% : C \succ_i A \succ_i B$$

In this case, candidate A wins. The $B \succ_i C \succ_i A$ group could instead report $C \succ_i A \succ_i B$, then candidate C would win, who they prefer to B.

2.3 Dictatorship

An obvious feasible choice of a ranking rule is to take agent i's preferences (or the ones of any other fixed agent). While this is clearly not what we have in mind when speaking of elections in a democracy, it does have a positive property: No agent can gain anything from misreporting the preferences.

3 Who should win elections?

What should ideal voting or ranking rules fulfill? Let us introduce four examples.

- Anonymity: The identities of the agents do not matter. That is, the outcome on preference profile $(\succ_1, \ldots, \succ_n)$ is the same as on $(\succ_{\pi(1)}, \ldots, \succ_{\pi(n)})$ for any permutation $\pi\colon N \to N$.
- Monotonicity: If one agent moves a candidate up in the ranking, keeping everything else fixed, the candidate does not move down in the combined society's ranking.
- Condorcet winner criterion: If, in pairwise comparisons, one candidate is always preferred to any other candidate in the majority of all preference profiles, it is the winner of the election. That is, if there is A such that for all B there are more i with $A \succ_i B$ than $B \succ_i A$, then A is the winner of the election.
- Condorcet loser criterion: If for some A for all B there are more i with $B \succ_i A$ than $A \succ_i B$, then A does not win the election.

3.1 Plurality Voting

Let us come back to our initial example of plurality voting. We choose the candidate who is ranked highest in the plurality of agents' preferences. This voting rule is anonymous and monotone.

Recall the following example, in which C is the winner.

$$35\,\% : A \succ_i B \succ_i C$$
$$25\,\% : B \succ_i A \succ_i C$$
$$40\,\% : C \succ_i A \succ_i B$$

Observe that A is the Condorcet winner. In A vs. B, he gets $75\,\%$ of the votes; in A vs. C, he gets $60\,\%$ of the votes. So, the Condorcet winner criterion is not fulfilled.

Furthermore, C is the Condorcet loser in this case. We observe that in C vs. A and also in C vs. B he gets $40\,\%$ of the votes. That is, the Condorcet loser criterion is not fulfilled either.

Instant-runoff voting does not fulfill monotonicity or the Condorcet winner criterion (see page 221 in the Karlin/Peres book). It does fulfill the Condorcet loser criterion because a Condorcet loser would never survive the final round.

3.2 Borda Count and Positional Voting

A very general voting/ranking approach works as follows. Depending on the position in the preference list, each candidate gets a score from every agent. The social ranking is determined by ordering the candidates by sum of scores. Every vector of $a_1 \geq \ldots \geq a_m$ is such a positional voting rule. Plurality voting is recovered by $a_1 = 1, a_2 = \ldots, a_m = 0$. Again, any of these rules is anonymous and monotone.

Another common rule is the *Borda count*, in which one sets $a_1 = m, a_2 = m-1, \ldots, a_m = 1$. Let us consider our example from above:

$$35\,\% : A \succ_i B \succ_i C$$
$$25\,\% : B \succ_i A \succ_i C$$
$$40\,\% : C \succ_i A \succ_i B$$

If we have $n = 100$ voters, A gets a total score of $35 \times 3 + 25 \times 2 + 40 \times 2 = 235$, B gets a score of $35 \times 2 + 25 \times 3 + 40 \times 1 = 185$, C gets a score of $35 \times 1 + 25 \times 1 + 40 \times 3 = 180$.

Note that equivalently, we could perform pairwise votes between any two candidates and sum up the number of votes each candidate gets in his $m - 1$ pairwise comparisons.

Other voting rules follow the same scheme. For example, the Eurovision Song Contest uses $a_1 = 12, a_2 = 10, a_3, = 8, a_4 = 7, a_5 = 6, a_6 = 5, a_7 = 4, a_8 = 3, a_9 = 2, a_{10} = 1, a_{11} = \ldots = a_m = 0$, where agents correspond to countries participating in the vote.

While in our example above, indeed the Condorcet winner wins and the Condorcet loser loses, this is generally not true for any of these rules. Indeed, no such rule fulfills the Condorcet winner criterion. Consider the following example, in which B wins Borda count, despite the fact that A is the Condorcet winner.

$$60\,\% : A \succ_i B \succ_i C$$
$$40\,\% : B \succ_i C \succ_i A$$

We could also ask at this point whether the Condorcet winner criterion is actually desirable. In this particular example, A is polarizing: Some agents really dislike him, so B might be a reasonable compromise. The reason is that the criterion only makes a binary comparison. It does not matter how much an agent likes one candidate better than the other.

3.3 Copeland Rule

Yet another approach is to perform pairwise comparisons and to count how many a candidate wins. The *Copeland score* of a candidate is defined to be the difference of how many pairwise comparisons this candidate wins and how many he loses. So, it is an integer in the range of $-(m - 1)$ to $m - 1$. The candidates are then ranked by decreasing Copeland score.

This rule fulfills the Condorcet winner and loser criteria. A Condorcet winner always has the highest possible score of $m - 1$, a Condorcet loser always has score $-(m - 1)$, which is the lowest possible score.

4 Strategic Vulnerability

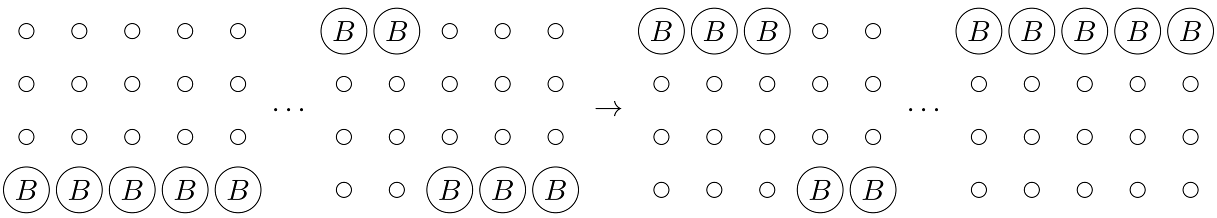
As we will show, it is not a coincidence that agents want to lie about their preferences in the above voting rules. Indeed, there are theorems by Arrow (1950) and Gibbard and Satterthwaite (1973/1975), which show that every ranking or voting rule has these deficiencies unless it is a dictatorship.

For the remainder of today, we will show Arrow's theorem that deals with *ranking rules*. That is, given the preferences $(\succ_1, \ldots, \succ_n)$, we have to come up with a combined single ranking $>$ expressing the preferences of society.

Definition 24.2. *A ranking rule is* strategically vulnerable *if there are an agent i, a preference profile $(\succ_1, \ldots, \succ_n)$, an alternative preference report $\succ_i'$, and two candidates A and B such that $A \succ_i B$ and $B > A$ but $A >' B$, where $>$ is the social ranking under $(\succ_1, \ldots, \succ_n)$ and $>'$ is the social ranking under $(\succ_1, \ldots, \succ_{i-1}, \succ_i', \succ_{i+1}, \ldots, \succ_n)$.*

The first observation will be that we can rephrase the property of strategic vulnerability using the following property.

Definition 24.3. *A ranking rule is* independent of irrelevant alternatives *(IIA) if for any two candidates A and B, the fact whether A is ranked before or after B in society's ranking only depends on how the agents rank A and B against each other (in pairwise comparisons). That is, given two preference profiles $(\succ_1, \ldots, \succ_n)$ and $(\succ_1', \ldots, \succ_n')$ such that $A \succ_i B$ whenever $A \succ_i' B$, then $A > B$ if and only if $A >' B$.*



B ranked last B ranked first

Figure 1: Candidate B is always ranked first or last in the society ranking under such preference profiles and IIA. If more and more agents change their preference, there has to be a transition from last to first. Therefore, a single agent can make the difference

This also seems be a natural property. Interestingly, any ranking rule that is not strategically vulnerable has to fulfill it.

Lemma 24.4. *If a ranking rule does not fulfill IIA, it is strategically vulnerable.*

Proof. If the ranking rule does not fulfill IIA, there have to be $(\succ_1, \ldots, \succ_n)$ and $(\succ'_1, \ldots, \succ'_n)$ such that $A \succ_i B$ whenever $A \succ'_i B$ but $A > B$ and $B >' A$.

We consider the sequence of $n+1$ preference profiles $(\succ_1, \ldots, \succ_t, \succ'_{t+1}, \ldots, \succ'_n)$. In one of the n steps in between, the relative ranks of A and B have to switch. Let this switch happen from i to $i+1$. That is, A is ranked before B on $(\succ_1, \ldots, \succ_{i-1}, \succ_i, \succ'_{i+1}, \ldots, \succ'_n)$ but after on $(\succ_1, \ldots, \succ_{i-1}, \succ'_i, \succ'_{i+1}, \ldots, \succ'_n)$.

If $B \succ_i A$, agent i can strategically misreport $\succ'_i$ instead of $\succ_i$. Otherwise, we have $A \succ'_i B$ and he can misreport $\succ_i$ instead of $\succ'_i$. $\square$

5 Arrow's Impossibility Theorem

Clearly, every dictatorship fulfills IIA because we simply use the preferences of one single agent. Another way of fulfilling IIA is to return an arbitrary order, which does not depend on the agents' preferences at all. As a matter of fact, these are essentially the only ranking rules that fulfill IIA. Every non-trivial ranking rule that you will come up with does not fulfill IIA. This is the statement of Arrow's theorem.

Definition 24.5. *A ranking rule is* unanimous *if the following condition holds. If candidate A is preferred to B by every agent i, then this candidate gets a better rank in the overall ranking. That is, if $A \succ_i B$ for all i, then $A > B$.*

Now, we are ready to state the theorem.

Theorem 24.6 (Arrow's theorem)**.** *A ranking rule for three or more candidates fulfills unanimity and IIA only if it is a dictatorship.*

The general idea of the proof is not too complicated. One considers preference profiles with the property that there is one candidate B who is ranked highest or lowest by each agent (but not consistently so). By IIA, candidate B will be either highest or lowest in the society ranking as well. Because of unanimity, there is a preference profile such that a change in the preference of one agent moves B from the bottom to the top.

This observation has an interesting consequence: The agent who decides whether B is ranked first or last also determines the relative ranking of all other candidates. This is due to the fact that we could move B between any two other candidates in his preference. Due to IIA, this would not change their relative ranking.

6 Outlook

As we have seen, there is a plethora of different voting rules. In each of them, agents sometimes have an incentive to misreport their preferences. Crucially, the counterexamples are often cyclic. That is, there are preferences of the forms $A \succ_i B \succ_i C$, $B \succ_i C \succ_i A$, and $C \succ_i A \succ_i B$. One can show, for example, that there are non-trivial strategy-proof voting rules if preferences have more of a structure. Most prominently, they could correspond to a left-right political spectrum and therefore be not cyclic.

Regarding other properties of voting rules, there are some that are less disputable than others. There is clearly no single best voting rule, simply because there are multiple ways to define who should be the winner.

Further Reading

- Chapter 13 in the Karlin/Peres book
- K. Arrow, Social Choice and Individual Values, 1951
- A. Gibbard, Manipulation of voting schemes: A general result, Econometrica, 1973
- M. Satterthwaite, Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions, Journal of Economic Theory, 1975

A Bonus: Proof of Arrow's Theorem

Definition 24.7. *A candidate B is* polarizing *with respect to a preference profile if each agent i ranks B highest or lowest.*

Lemma 24.8. *Consider a ranking rule that fulfills unanimity and IIA. If there is a polarizing candidate B in the strategy profile, then the ranking rule ranks B highest or lowest.*

Proof. We prove the claim by contradiction. Suppose, it is false. That is, given a preference profile $(\succ_1, \ldots, \succ_n)$ such that each agent i ranks B highest or lowest, the outcome of the ranking rule is such that there are A and C for which $A > B > C$. We now derive a preference profile $(\succ'_1, \ldots, \succ'_n)$ on which we will apply IIA.

For every i, we distinguish the relative rank of A, B, and C in $\succ_i$. There are the following four cases: (i) $B \succ_i A \succ_i C$, (ii) $B \succ_i C \succ_i A$, (iii) $A \succ_i C \succ_i B$, (iv) $C \succ_i A \succ_i B$.

In case (i), define $\succ'_i$ such that $B \succ'_i C \succ'_i A$; in case (iii), define $\succ'_i$ such that $C \succ'_i A \succ'_i B$; in the other cases, define $\succ'_i$ to be simply $\succ_i$.

What ranking $>'$ does the ranking rule decide on $(\succ'_1, \ldots, \succ'_n)$? By IIA, we have to have $A >' B$ and $B >' C$ because $(\succ'_1, \ldots, \succ'_n)$ keeps the positions of B relative to both A and C. Furthermore, by unanimity, $C >' A$ because every agent prefers C to A in the new preferences. This is a contradiction, so either A or C cannot exist. □

Definition 24.9. *Given a candidate B, we call an agent i B-pivotal if there is a preference profile $(\succ^*_1, \ldots, \succ^*_n)$ and an alternative preference $\succ^{**}_i$ such that B is polarizing and ranked lowest under $(\succ^*_1, \ldots, \succ^*_n)$ and polarizing and ranked highest under $(\succ^*_1, \ldots, \succ^*_{i-1}, \succ^{**}_i, \succ^*_{i+1}, \ldots, \succ^*_n)$.*

Lemma 24.10. *Consider a ranking rule that fulfills unanimity and IIA. For every candidate B, there is a B-pivotal agent.*

Proof. We consider a sequence of $n+1$ different preference profiles $(\succ_i^{(0)})_{i\in N}, \dots, (\succ_i^{(n)})_{i\in N}$. In $(\succ_i^{(0)})_{i\in N}$, all agents rank B lowest, in $(\succ_i^{(n)})_{i\in N}$, all agents rank B highest. From $(\succ_i^{(t-1)})_{i\in N}$ to $(\succ_i^{(t)})_{i\in N}$, we switch agent t's preferences from $\succ_t^{(0)}$ to $\succ_t^{(n)}$.

By Lemma 24.8, candidate B is ranked highest or lowest in the outcome on every $(\succ_i^{(t)})_{i\in N}$. By unanimity, he is ranked lowest in the outcome on $(\succ_i^{(0)})_{i\in N}$ and highest in the outcome on $(\succ_i^{(n)})_{i\in N}$. At some point t, there has to be a transition such that B is ranked lowest in $(\succ_i^{(t-1)})_{i\in N}$ and highest in $(\succ_i^{(t)})_{i\in N}$. These profiles fulfill the claim. □

Lemma 24.11. *Consider a ranking rule that fulfills unanimity and IIA. Let B be any candidate and let i be the agent with the property guaranteed by Lemma 24.10. Then i is a dictator on $\Gamma \setminus \{B\}$. That is, for $A, C \neq B$, we have $A > C$ if and only if $A \succ_i C$.*

Proof. It suffices to show that if $A \succ_i C$ then also $A > C$. We will construct a preference profile $(\succ'_1, \dots, \succ'_n)$ and apply IIA. Note that by IIA everything that matters are the relative positions of A, B, and C. In other words, we can assume without loss of generality that $\Gamma = \{A, B, C\}$.

We use the preferences $(\succ^*_1, \dots, \succ^*_n)$ and $\succ^{**}_i$ to construct $(\succ'_1, \dots, \succ'_n)$. For every $i' \neq i$, we know that B is either ranked first or last in $\succ^*_{i'}$. Construct $\succ'_{i'}$ by moving B to the same position as in $\succ_{i'}$ and keeping the relative order of A and C fixed. Furthermore, choose $\succ'_i$ as $A \succ'_i B \succ'_i C$.

First of all, by IIA considering candidates A and C, the relative order of A and C in the outcome on $(\succ'_1, \dots, \succ'_n)$ has to be the same as on $(\succ_1, \dots, \succ_n)$. Next, by IIA but this time regarding candidates A and B, A has to be ranked before B on $(\succ'_1, \dots, \succ'_n)$ because the relative order is just as in $(\succ^*_1, \dots, \succ^*_n)$. Finally, B has to be ranked before C on $(\succ'_1, \dots, \succ'_n)$ because their relative order is like in $(\succ^*_1, \dots, \succ^*_{i-1}, \succ^{**}_i, \succ^*_{i+1}, \dots, \succ^*_n)$. □

Proof of Theorem 24.6. Consider three distinct candidates A, B, and C. By Lemma 24.11, we know that there is a B-pivotal agent i, who is a dictator on $\Gamma \setminus \{B\}$. There is also a C-pivotal agent i', who is a dictator on $\Gamma \setminus \{C\}$. We claim that $i = i'$.

To see this, consider the preference profile $(\succ^*_1, \dots, \succ^*_n)$ and the preference $\succ^{**}_i$ that make i a B-pivotal agent. Suppose that $i \neq i'$. Then A would be ranked higher than B if and only if $A \succ^*_{i'} B$. However, what makes i B-pivotal is the fact that *only his* preference changes and makes B move from bottom to top, whereas the preferences of i' do not change. This is a contradiction. □

B Bonus: Gibbard-Satterthwaite Theorem

There is a similar statement regarding voting rules. In this case, the “truthfulness” property is called *strategy-proofness*. It is defined as follows.

Definition 24.12. *A voting rule is* strategy-proof *if for all preference profiles $\succ_1, \dots, \succ_n$, all agents i, and candidates A and B the following holds. If $A \succ_i B$ and B wins under $\succ_1, \dots, \succ_n$, then A does not win under any false report $\succ'_i$ of agent i.*

So, in words, this definition requires that no agent can enforce an outcome that he likes better by lying. This, once again, is conceptually the same as truthfulness in mechanism design with money.

The following theorem was discovered by Allan Gibbard (1973) and Mark Satterthwaite (1975) independently.

Theorem 24.13. *If a voting rule for three or more candidates is* onto *(that is, every candidate can be elected) and strategy-proof, then it is a dictatorship. That is, there is some agent i such that always agent i's most preferred candidate wins.*

The assumption that the voting rule is onto mirrors unanimity in Arrow's theorem. If a ranking rule fulfills unanimity, then, by definition, it can output every possible ranking, namely if all agents agree on this one. For a voting rule to be onto, a similar property is sufficient: If all agents agree that one candidate is the best, then this candidate gets elected.

We will not prove this theorem because the proof is not very enlightening on the technical level. The general argument is to use Arrow's theorem and to show that a voting rule fulfilling the assumptions of the theorem would be a contradiction to Arrow's theorem.

In more details, such a voting rule could be turned into a ranking rule as follows. The fact whether $A > B$ or not is determined by whether A would win a vote against B if they were both moved to the top of the preference profiles. From the assumptions, we get that this ranking rule fulfills unanimity and IIA. If the voting rule that we started from is not a dictatorship, then the created ranking rule isn't either: For every agent, there are preference profiles, such that some part of the social ranking does not correspond to the individual preference. This is a contradiction, therefore the voting rule cannot exist.