Mechanism Design without Money: Voting Systems

ETH Zürich Peter Widmayer, Paul Dütting

Voting systems are among the most well known mechanisms without money. It generally serves the purpose of aggregating preference lists into a group decision. In the simplest case this is a single winner. More generally, we could ask for a ranking of the candidates. Is the well known majority rule a good mechanism? Are there better ones? What does it even mean to be a good mechanism in this case?

1 Basic Definitions

We will study a setting with n players, called *voters* $1, \ldots, n$, each of whom has a complete individual *preference list* over all possible *alternatives* A. The total order of voter i over all alternatives is denoted by $\geq_i$, where $X \geq_i Y$ means that voter i prefers alternative X over alternative Y. Since we require a total order, the preference relation of each voter is requested to be transitive and antisymmetric. The objective of a voting system is to aggregate the preference lists of all voters into a single total order that reflects the voters' preferences. A voting system is often called a *social welfare function*, formally denoted as function $F : L^n \to L$, where L is the set of all linear orders over A (i.e., L is isomorphic to the set of all permutations of A). The input to F is called a *profile*, with $F(P)$ denoting the *aggregated votes* of profile P, or sometimes simply denoted as $\geq$. In situations where only a single winner needs to be elected, the aggregation of all votes into a linear order achieves more than needed. In this case, we talk about a *social choice function* $f : L^n \to A$, with $f(P)$ denoting the *winner* for profile P.

2 Warm-Up

Let us gain some intuition for the voting problem by approaching it "bottom up". That is let us start with the simplest settings that one can imagine: settings with only two or three alternatives.

2.1 Voting with Two Alternatives

For two alternatives, a social choice function tells just as much about the aggregated ranking as a social welfare function: With one alternative being the winner, the other is the loser. In this case, majority voting is the obvious solution to the rank aggregation problem. It trivially is incentive compatible. While one might think that there is not much to say about majority voting, May proved that it is the only reasonable voting system for two alternatives.

2.2 Voting with Many Alternatives

For three or more alternatives, voting becomes more tricky. As an example, let us consider three voters and a set of three alternatives $\{X, Y, Z\}$. The totally ordered preference list for the voters are X, Y, Z for voter 1, with the highest ranked alternative first in the list, Y, Z, X for voter 2, and Z, X, Y for voter 3. Interestingly, the same setting describes a situation in which a single individual aims to make a decision based on multiple criteria. A buyer's choice between three cars, for instance, based on the three criteria price, fuel consumption, and annual insurance cost might give the preference lists above, where the alternatives are the cars and the voters are the criteria.

	voter 1	voter 2	voter 3
1st choice	X	Y	Z
2nd choice	Y	Z	X
3rd choice	Z	X	Y

Figure 1: Example with three voters

2.2.1 Tournament Voting

As in a sports competition, one might opt for a tournament in which pairwise majority kicks out a loser (of a pair of alternatives) in each round, repeating until only one winner remains. In our example, we could first vote among alternatives X and Y, in which case X wins with two votes against one. Then, the vote among X and Z lets Z wins, again with two votes over one, and hence, Z is the overall winner. While this may sound reasonable, it provides room for strategizing: If we change the sequence of competitors and let first X and Z compete, Z wins the first round and then loses against Y in the second round. This dependence of the result on the sequence of "games" is clearly undesired.

2.2.2 Voting by Pairwise Majority

Instead of pairwise majority for just those pairs that the tournament dictates, one might opt for ranking all alternatives by taking a pairwise majority vote for each and every pair. In our example, this will lead to three majority votes for the three pairs: $X \geq Y$, $Y \geq Z$, and $Z \geq X$ with two votes against one in each case. The result relation, unfortunately, is useless, since it is not transitive (but cyclic), so it does not define a voting system. This phenomenon has been observed already around the times of the French revolution (appropriately) by Condorcet. It is known as the "Condorcet paradox".

2.2.3 Positional Voting

A well known voting system for parliament elections makes voters choose a few candidates out of many and aggregates this choice. More generally, positional voting assigns a weight to each position of the total order, and simply sums up the weights that an alternative gets over all voters. The social welfare function then merely arranges alternatives according to their sum of weights. Note that ties are possible, but are not a problem—one can think of resolving them in simple ways, like according to a predefined preference. One specific positional voting system, known as the *Borda count*, assigns weight i to the alternative in position $i + 1$ from below. That is, the lowest ranked alternative of a voter gets weight

0 for this voter, the next higher alternative gets weight 1, and so on. Let us look at an example. For two alternatives X and Y and five voters, where three voters rank X higher than Y and two voters prefer Y over X, we get Borda counts of 3 for X and 2 for Y, so X wins, as it should. Now, let an alternative Z come in that nobody likes at all, so everybody ranks last. With the new preferences, still X wins with a Borda count of 8 against Y with count 7. Unfortunately, this system has room for cheating: If the two voters who prefer Y to X place Z second on their lists (instead of third), the count for X drops to 6, and Y wins. This example tells that the Borda count is not incentive compatible: An irrelevant alternative, namely Z, can be used to influence the result.

3 Impossibility Results

The attempts above of identifying a satisfactory voting system for more than two alternatives failing, we should clarify what a satisfactory voting system should and should not do. We will now describe three properties that a voting system should have, none of which appear exaggerated—quite the opposite. We then explain that no voting system can satisfy all three properties in general. This result, due to Arrow, should not be viewed as the complete impossibility of democracy; it should rather be used as a guide for the search for good voting systems.

3.1 Desirable Properties

The first desirable property says that an irrelevant alternative should not matter, in the following sense:

Definition 8.1. *A social welfare function F satisfies* independence of irrelevant alternatives *(abbreviated* IIR*) if the following holds: For every pair of alternatives X and Y and every two profiles P and Q, we have: If every voter i has the same preference between X and Y in P as in Q, then also F must have the same preference between X and Y in $F(P)$ as in $F(Q)$.*

While IIA rules out the effect of irrelevant alternatives, it does not rule out that F goes against all voters: It could prefer X to Y even though every voter prefers Y to X. This would obviously go totally against the voters' preferences, violating even minimum democratic requirements. We therefore aim to rule it out:

Definition 8.2. *A social welfare function F satisfies* unanimity *(also called the* Pareto principle*) if the following holds: If for some pair of alternatives X and Y every voter prefers X to Y, then also F must prefer X to Y.*

Just like going against all voters is undesirable, so is a voting system that always simply goes with a single, specific voter (after the French revolution, one may think of the (former) king being one of the voters, and the voting system being arranged so that it always follows exactly the king's vote):

Definition 8.3. *For a social welfare function F, voter i is a* dictator *in F, if for any profile P, $F(P)$ equals voter i's preference order. A social welfare function F is a dictatorship if there exists a voter who is a dictator in F.*

3.2 Arrow's Theorem

We are now in the position to prove that these three plausible requirements are incompatible.

Theorem 8.4 (Arrow 1950). *For more than two alternatives, any voting system that satisfies IIA and unanimity is a dictatorship.*

Proof. It helps to express IIA in a slightly different way, by focussing on interesting pairs of alternatives and not talking about all others. We say that a ranking is *restricted to* X *and* Y if we erase all other alternatives from the ranking. That is, we are left with only X and Y, in the order given in the (full, unrestricted) ranking. If we now restrict every individual ranking to X and Y, we get a *profile restricted to* X *and* Y. We can now rephrase IIA as follows: For any two profiles P and Q that are identical when restricted to X and Y, the social welfare function F must produce rankings $F(P)$ and $F(Q)$ that are identical when restricted to X and Y.

Let now F be a voting system that satisfies IIA and unanimity. We will prove that F is a dictatorship in three steps. First, we show that F must place an alternative first or last in its ranking if every voter does—a polarizing alternative. Second, the reasoning of the first step points at a powerful voter, a potential dictator. Third, we show that this powerful voter indeed is a dictator.

Step 1: A polarizing alternative. Let us call an alternative X *a polarizing alternative* if X is ranked either first or last by every voter (not necessarily the same by every voter, though). We will show that F must place a polarizing alternative first or last, as follows. Let P be a profile in which X is polarizing. Assume for the sake of contradiction that X is neither first nor last in $F(P)$. That is, some alternative Y ranks higher than X and some other alternative Z ranks lower in $F(P)$. Now, we create a new profile by modifying P, and we study how F must behave. Within P, for each voter who ranks Y higher than Z, slide Z directly in front of Y. This manipulation creates a new profile Q, where for each affected voter the ordering of all other alternatives (except Z) remains the same, but the voter ranks Z now higher than Y (and higher than some other alternatives below it was before, but this is not our interest). So, X remains polarizing in Q, and it maintains its order w.r.t. Y and Z in each voter's list of preferences. By IIA, $F(Q)$ must therefore still (as in $F(P)$) rank Y above X above Z. But since every voter ranks Z above Y in profile Q, by unanimity $F(Q)$ must rank Z higher than Y, a contradiction.

Step 2: A potential dictator. Let us identify a powerful voter by inspecting how $F(P)$ changes as we make an alternative X jump from last to first position for one voter after another, starting with X last for each voter and ending with X first for each voter. Towards this end, let us choose an alternative X, and a profile P_0 that ranks X last for every voter. Define profile P_1 by modifying P_0 by moving X from last to first position for voter 1. Generally, define profile P_j by modifying profile P_{j-1} by moving X from last to first position for voter j. Note that P_j has X first for voters $1, \dots, j$ and last for voters $j+1, \dots, n$. The sequence of profiles ends with P_n that has X first for every voter. All other alternatives remain unchanged relative to each other throughout.

Now we start to observe how F behaves. By unanimity, X must be last in $F(P_0)$ and first in $F(P_n)$. Hence, there must be a voter i whose modification of X ranking last to first results in X going from last in $F(P_{i-1})$ to first in $F(P_i)$. Note that at this point, we cannot claim that i is unique, but this is of no concern to us, so we will simply think of the smallest such i. Hence, i is a very powerful voter whom we call a potential dictator.

Step 3: Indeed a dictator. We will now prove that voter i from Step 2 is indeed a dictator. We prove this by arguing that for any profile Q and for any alternatives Y and Z, the ordering of Y and Z in $F(Q)$ is the same as the ordering of Y and Z by voter i. First, we show this to be true whenever both Y and Z differ from X. Again, our proof works by modifying profile Q and observing how F must behave. We create a new profile R from Q by the following modifications: (1) move X to the first position for voters $1, \ldots, i$ (for voters for which it is already there, leave it there); (2) move X to the last position for voters $i+1, \ldots, n$ (for voters for which it is already there, leave it there); (3) for voter i, move Y to the first position (i.e., right before X that becomes second). Let us now study how F must act for R. We do this by making five simple observations.

Observation 1. When restricted to X and Z, R and P_i are the same. X is first in $F(P_i)$. By IIA, $F(R)$ must prefer X over Z.

Observation 2. When restricted to X and Y, R and P_{i-1} are the same. X is last in $F(P_{i-1})$. By IIA, $F(R)$ must prefer Y over X.

Observation 3. By transitivity, $F(R)$ must prefer Y over Z.

Observation 4. When restricted to Y and Z, R and Q are the same. By IIA, $F(Q)$ must prefer Y over Z.

Observation 5. (Conclusion from Observations 1 to 4.) The group ranking of F must be identical with the ranking of voter i for any profile Q and any pair of alternatives Y and Z that are different from X.

We still need to prove that also when alternative X is involved, i's ranking determines the group ranking. We do so by first noticing that the reasoning above shows that there also is a voter j whose ranking determines the group ranking when some alternative $W \neq X$ is considered instead of X. We will show that $i = j$, so we have a single voter who determines the entire group ranking, which then terminates the proof. We prove $i = j$ by contradiction. Assume therefore now that $i \neq j$. Note that X and W are fixed. Let V be an alternative different from X and W. Profiles P_{i-1}) and P_i differ only in voter i's ranking, but the ordering of V and X differs in $F(P_{i-1})$) and $F(P_i)$. Thus, in one of $F(P_{i-1})$) and $F(P_i)$, the order of V and X must differ from the order of V and X in j's ranking, and therefore voter j does not determine the order of V and X, a contradiction to $i \neq j$. Therefore, $i = j$ is the sole dictator. □

3.3 Gibbard-Satterthwaite Theorem

A different observation can be made for social choice functions, and can be proven in much the same way as Arrow's theorem. The notion of a dictator needs a slight adaptation: We call voter i a dictator if her most preferred alternative is chosen, regardless of all other voters' preferences. A social choice function can be *strategically manipulated* if some voter i who prefers X to Y can misreport her preferences so that socially, X is chosen rather than Y. Social choice funktion f is called *incentive compatible* if it cannot be strategically manipulated. Then one can prove:

Theorem 8.5 (Gibbard 1973, Satterthwaite 1975)**.** *Let f be an incentive compatible social choice function onto the set A of more than two alternatives. Then f is a dictatorship.*

Just like Arrow's theorem, the theorem by Gibbard and Satterthwaite should not be seen to destroy all hopes for acceptable election schemes, but instead guide the search for good schemes—which is indeed an active area of research.

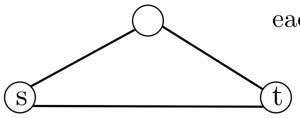
Recommended Literature

- David Easley and Jon Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Chapter 23: Voting Systems, Cambridge University Press, 2010. (General reading)
- Kenneth O. May, *A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decisions*, Econometrica, Vol. 20, Issue 4, pp. 680–684, 1952. (May's Theorem)
- Jean-Antoine-Nicolas de Caritat Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Imprimerie Royale, 1785. (Condorcet Paradox)
- Kenneth J. Arrow, *A Difficulty in the Concept of Social Welfare*, Journal of Political Economy, 58 (4): 328–346, 1950. (Arrow's Theorem)
- Allan Gibbard, *Manipulation of voting schemes: a general result*, Econometrica, 41 (4): 587–601, 1973. (Gibbard-Satterthwaite Theorem)
- Mark A. Satterthwaite, *Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions*, Journal of Economic Theory, 10: 187–217, 1975. (Gibbard-Satterthwaite Theorem)

Exercises
(during next exercise class - 12.11.2019)

We shall discuss and solve together these two exercises.

Exercise 1. *Consider the following scenario:*



each edge = player

and we want to sent T ***units of traffic*** *from s to t. Moreover:*

1. *Each player i has some* ***private*** *cost t_i and we know*

$$t_i \in \{L, H\}$$

   where t_i is the cost per unit of traffic, $L = Low$ and $H = High$.

2. *We give a* ***fixed compensation*** *per unit of traffic:*

$$F \cdot w_i$$

   is the payment to player i when he/she gets w_i units of traffic.

Question 1: *Model this game as a* ***single-peaked preferences*** *when $L < F < H$.*
Question 2: *Which outcomes are selected by the* ***median voter****?*