VOTING SYSTEMS AND ARROW'S THEOREM

AKHIL MATHEW

ABSTRACT. The following is a brief discussion of Arrow's theorem in economics. I wrote it for an economics class in high school.

1. Background

Arrow's theorem is a central result in modern economics and social science, which states that any voting system satisfying "reasonable" assumptions has inherent limitations, and in particular that the problem of spoilers cannot be resolved within this framework.

First, recall that voting is used to make decisions in democratic societies, but that many decisions can be made through the market system. The decision about how much corn will be produced in a given year is made not by central planning but by the cumulative choices of independent farmers and farming companies, each guided by market forces. However, many decisions must be made by the government: for instance, the provision of public goods such as military defense, or more simply the choice of a candidate. Some such decisions are between two alternatives (e.g. whether or not to support a given bill in a system of direct democracy), and there is a clear way to decide, namely, to make everyone vote and choose the side with the majority (choosing randomly if both options collect the same number of votes).

In practice, however, one often has at least three alternatives, and it is quite possible no option will have a majority. Of course, one can select the winning alternative as the one that receives a *plurality* of votes, as is usually done. The problem with this approach is the existence of *spoilers*, as demonstrated for instance in the 2000 presidential election. Suppose there are three candidates A, B, C. A is a conservative, B a moderate liberal, and C a staunch liberal. The country, leaning left, is split: 49% for A, 48% for B, and 3% for C. The plurality approach will give A the victory, though the liberal 51% of the country would have preferred B to A.[1] In 2000, Ralph Nader was candidate C.

Moreover, it is entirely possible that, according to pairwise match-ups, society will paradoxically prefer $A > B, B > C$, and $C > A$. In other words, even if each individual is rational—i.e., each individual imposes a *total* ordering on the candidates—society may be irrational. The *Concordet criterion*, named after the founder of voting theory, of choosing the candidate that defeats any other in a head-to-head match may not be applicable.

Date: November 25, 2010.

[1] In the American political system, the situation is complicated by the electoral college system.

There are various methods at ameliorating the situation. For instance, one could perform a run-off election between the top two candidates, though this incurs additional expense, and in fact need not fully resolve the spoiler problem—what if the split were $34-32-33$ between candidates A, B, C as above, and some of the moderately liberal B voters swung to A in the second (A versus C) round?

An improvement to this model is *instant-runoff voting* (IRV),[2] which works as follows. Each voter is asked to rank the candidates from favorite to least favorite. Then, the first choices of all voters are tabulated; a candidate with a majority here wins. If there is no candidate with a majority, the candidate with the *fewest* number of first-choice votes is eliminated. Those voters who voted for the eliminated candidate now have their second-choice votes counted, and the candidate left in the round with the fewest votes is eliminated. The process repeats. In general, at each step, one scans through the list of all ballots and marks a vote for the highest-ranked candidate on each ballot that has not been eliminated. Then one eliminates the candidate with the fewest votes and resets all vote counts to zero. One repeats until a candidate has a majority. Nevertheless, Arrow's theorem shows that this method is also imperfect.

Arrow's theorem itself is a purely mathematical statement, but it comes from the modelling of voting phenomena. Suppose we have a **society**, i.e. a finite set S (corresponding to the individuals of a population), and a finite set $\mathcal{C}$ of **candidates**. Each individual has a preference between candidates in a rational manner, i.e. if $A, B, C \in \mathcal{C}$ and an individual prefers A to B and B to C, then she prefers A to C. This means that each individual induces a **total ordering** on the set S by her preferences.

A **voting system** associates to each system of preferences by each member of society a *societal preference*, or equivalently a total ordering on $\mathcal{C}$. In other words, it is a function $V \colon \mathbf{L}(\mathcal{C})^S \to \mathbf{L}(\mathcal{C})$, where for a set T, we let $\mathbf{L}(T)$ denote the set of total orders on it.

We shall consider certain seemingly natural assumptions on V.

1. (Unanimity) Suppose $A, B \in \mathcal{C}$. If all voters prefer A to B, then society prefers A to B. I.e., if $A > B$ in each ordering in $\Delta \in \mathbf{L}(\mathcal{C})^S$, then $A > B$ in $V(\Delta)$.
2. (Independence of irrelevant alternatives) The societal preference of A relative to B depends only on the preferences (between A, B) of each individual in the society. In other words, if $\Delta, \Delta' \in \mathbf{L}(\mathcal{C})^S$ are societal preferences and $A > B$ for an ordering in Δ if and only if $A > B$ in the corresponding Δ'-ordering, then $A > B$ in $V(\Delta)$ iff $A > B$ in $V(\Delta')$.

The meaning of the first is clear. For the second, it means that whether society prefers A to B depends only on the individual preferences of A versus B (and not of any "irrelevant" alternatives $C \neq A, B$).

A simple way of constructing a voting system satisfying both conditions above is as follows. Pick $d \in S$ and define V by sending $\Delta \to \Delta(d)$. Equivalently, the

[2]See [3].

societal ranking is simply that of d. The individual d for which this occurs is evidently unique, and is called a **dictator**.

Theorem 1 (Arrow [5])**.** *Any voting system satisfying unanimity and the independence of irrelevant alternatives has a dictator, when $\mathcal{C}$ has at least three elements.*

The theorem stunned mathematicians and economists, because it demonstrates that (within this framework), no "perfect" voting system can exist! Various proofs, all somewhat technical, are presented in [6].

2. Proof of Arrow's theorem

Henceforth, we assume V is a voting system satisfying unanimity and the independence of irrelevant alternatives. We present a proof due to Geanakopolous that a dictator exists.

Step 1: An extremal lemma.

Lemma. *Suppose that each individual ranks $A \in \mathcal{C}$ either at the top or at the bottom. Then society ranks A either at the top or the bottom.*

Of course, this is an immediate consequence of unanimity when each individual chooses the same way (i.e. all top or bottom). For convenience, call a **profile** an element of $\mathbf{L}(\mathcal{C})^S$, i.e. a system of preferences.

Suppose, to the contrary, that $B > A > C$ in society, i.e. A is not in any extreme position. Since A is always at the extreme of each voter's list, we can tweak each voter's list so as to leave A fixed but shift the positions of B, C suitably such that $C > B$ for each voter. In this modified profile P', clearly $B > A, A > C$ societally because the relative preferences with respect to A have remained constant—A is at the extreme—and we have the independence of irrelevant alternatives. So $B > C$ in $V(P')$. But also $B < C$ for P' by unanimity, contradiction.

Step 2: Existence of a limited dictator. We now construct an element $d \in S$ that will become the dictator.

Lemma. *Fix $A \in \mathcal{C}$. There exists an individual $d \in S$ and a profile P such that:*

(1) *d ranks A at the bottom in P and $V(P)$ ranks A at the bottom.*
(2) *Consider the modified profile P' obtained by moving d's preference of A to the top. Then $V(P')$ ranks A at the top.*
(3) *In the profile P, each voter ranks A either at the top of the bottom*

In other words, there is a profile P and an individual d such that by *only* d's changing its preference of A from the bottom to the top, society changes its preference of A from the bottom to the top.

To see this, consider a profile P_0 where all individuals rank A at the bottom. Consider the profiles $P_1, \dots, P_s$ (where $s = \operatorname{card} S$) where to obtain P_s, we consider the profile P_{s-1} and, for one new voter, switch the preference of A from the bottom to the top. Then in P_0, A is at the bottom of all voters' preferences; in P_s, A is at the top of all voters' preferences. Since A is always at an extreme position in any voter in any profile P_i, A is at either the top or the bottom of each $V(P_i)$. When $i = 0$ it is at the bottom and when $i = s$ at the top, by unanimity. There must be some P_j such that A is at the bottom of $V(P_j)$ and at the top of $V(P_{j+1})$. Since P_{j+1} is obtained from P_j by shifting the preference of A from the top to the bottom at some voter (call it d), the lemma is now clear.

Step 3: Denoument. I claim that the individual $d \in S$ from the previous lemma is a dictator.

Step 3a. First, we show that:

Lemma. *If $B, C \in \mathcal{C}$ are distinct from A, then society (in any profile) ranks $B > C$ iff d does.*

We can express this by saying that d is a *dictator with respect to B, C*.

Consider the profiles P, P' in the previous lemma; recall that each voter ranked A at the top or the bottom.

Consider another profile Q in which, say, d ranks $B > C$. We will show that $V(Q)$ ranks $B > C$ as well. Now, modify P' in the following way to get a new profile R. At the voter d, B is at the top, A right below, and C further down. At other voters, A remains at the extreme position (either top or bottom), and B, C are ordered in the same way as they are in Q. Note then that a voter ranks $B > C$ in R iff she does so in Q. In particular, $B > C$ in $V(R)$ iff $B > C$ in $V(Q)$, by the independence of irrelevant alternatives.

I now claim that $B > C$ in $V(R)$. Indeed, I will show that $B > A$ and $A > C$. First, R and P have the same relative B, A preferences (e.g. at d, $B > A$), and A is at the bottom in $V(P)$. So $B > A$ in $V(P)$ and hence in $V(R)$. Second, R and P' have the same relative A, C preferences (e.g. at d, $A > C$), and in $V(P')$, A is at the top. So by necessity $B > A$ and $A > C$ in $V(R)$, implying $B > C$ in $V(R)$.

Since $V(R)$ and $V(Q)$ have the same relative B, C preference, we are done.

Step 3b. Now, I claim that if $B \in \mathcal{C}$, then society ranks $A > B$ if and only if d does. This handles the other case, and completes the proof that d is a (full) dictator.

First, there is an individual $d' \in S$ such that society (for any profile) ranks $A > B$ if and only if d' does; this can be seen by choosing $C \neq A, B$ (recall that $\mathcal{C}$ has at least three elements!) and using the construction of the previous lemma, interchanging C and A. So we get an individual $d' \in S$ which is a dictator respect to any pair of elements not containing C.

However, in the profiles P, P' constructed for A, the societal preference of A, B shifts, even though only at d is there a difference between P, P'. It follows that $d = d'$, proving the result. Arrow's theorem is proved.

3. Approval Voting and Near-counterexamples

It turns out, however, that there do exist instances of what one might colloquially call "voting systems" that do satisfy the conditions of unanimity, independence of irrelevant alternatives, and non-dictatorship. Consider the following procedure, called *approval voting*. Each citizen is given a ballot with a list of all the candidates and marks 1 or 0 for each one according as she approves or disapproves of the candidate. The candidate with the largest number of 1's wins. It is clear that this method of voting, in theory, satisfies the relevant conditions. It does not contradict Arrow's theorem because each individual does not choose a total ordering on the candidates.[3]

Approval voting is used in certain organizations such as the National Academy of Sciences and the Mathematical Association of America, but it lacks real-world testing. Moreover, a potential flaw in the model is that it may fail to elect a candidate who is the clear favorite of a majority of voters due to a new version of the "spoiler" problem. For instance, suppose there are three candidates A, B, C, and 51% of voters have the preference $A > B$ but approve of both A and B. Suppose 49% of voters have the preference $B > A$ but approve of only B. Then, B will win the election via approval voting. This is pointed out, for instance, in [1]. However, the problem diminishes in importance when $\{0, 1\}$ is replaced with a larger set, allowing more expressivity; sometimes this variant of approval voting is called *range voting*. Range voting, unlike strict approval voting, allows voters to register a preference.

Finally, another potential benefit of range voting is the boost given to third parties. Under the present system, it is unlikely that someone will cast a vote for a third-party candidate, but a range-wote allows her to express tepid support for a potentially too extreme but not out-of-the-question candidate. As a result, range-voting might increase the activity and strength of minor parties; given the disenchantment of many Americans with both the Republicans and the Democrats, this may be a serious consideration.

References

[1] Approval voting vs. instant runoff voting. `http://archive.fairvote.org/irv/approval.htm`.
[2] Citizens for approval voting. `http://www.approvalvoting.org`.
[3] Instant runoff voting. `http://instantrunoff.com/`.
[4] Rangevoting.org – Center for range voting. `http://rangevoting.org`.
[5] Kenneth J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):326–346, 1950.
[6] John Geanakoplos. Three brief proofs of Arrow's impossibility theorem. *Economic Theory*, 26:211–215, 2005. Available at `http://ideas.repec.org/p/cwl/cwldpp/1123r3.html`.
[7] Dana Mackenzie. Making sense out of consensus. 2000. `http://www.siam.org/news/news.php?id=674`.
[8] Ben Webster. Why Arrow's theorem is a scam, 2008. Blog post available at `sbseminar.wordpress.com`.

[3] See the discussion at [8].