The Mathematics of Elections

Kim Klinger-Logan

It’s not the voting that’s democracy; it’s the counting. – Tom Stoppard

This document is meant to give an explanation and proof of Arrow’s Impossibility Theorem.

Fairness Criteria and Arrow’s Impossibility Theorem

In the late 1940’s, American economist Kenneth Arrow attempted to answer the question:

What would it take for a voting method to at least be a fair voting method?

To answer this question, he set a minimum set of requirements that a method should have, these are called Arrow’s fairness criteria:

- **Non-dictatorship**: There is no one voter who determines the outcome of every election.
- **Universality**: Each election has a unique outcome.
- **Non-imposition**: Every election outcome is possible.

Majority Criterion: A majority candidate should always be the winner.

Monotonicity Criterion: If candidate X is the winner, then X would still be the winner had a voter ranked X higher in his preference ballot.

- **Unanimity**: If every voter prefers one candidate over another, the outcome will reflect that.
- **Independence-of-irrelevant-alternatives (IIA)** Criterion: If candidate X is the winner then X would still be the winner had one or more of the losing candidates not been in the race.

Condorcet Criterion: A Condorcet candidate (i.e. a candidate that beats each of the other candidate in a pairwise comparison) should always be the winner.

It turns out that every voting method is flawed. Arrow demonstrated that, in an election involving three or more candidates, it is mathematically impossible to satisfy all of these fairness criteria.

**ARROW’S IMPOSSIBILITY THEOREM**

For elections involving three or more candidates, a method for determining election results that is always fair is mathematically impossible.

Note though that this does not mean that every election is unfair or that every voting method is equally bad, nor does it mean that we should stop trying to improve the quality of our voting experience.

In order to prove such a theorem we need to introduce some notation and turn these intuitive criteria into something more formal. However, the rough idea of the proof is that there is no way guarantee that a transformation of an arbitrary preferences schedule to a ranking of candidates will satisfy all of the bulleted criteria.
Proof of Arrow’s Impossibility Theorem

We will begin by introducing some notation to formalize the concepts previously presented:
Let $V = \{v_1, \ldots, v_N\}$ be our set of voters. Let $C = \{c_1, \ldots, c_m\}$ $(m \geq 3)$ be our set of candidates. We take $L(C)$ to be the set of linear orderings of $C$. A profile on $C$ is a function $p : V \to L(C)$ (it can also be realized as an element of $L(C)^N$). We take $P(C)$ to be the set of profiles.

A social welfare function is a function $F : P(C) \to L(C)$. 

<table>
<thead>
<tr>
<th>Number of Voters:</th>
<th>14</th>
<th>10</th>
<th>8</th>
<th>4</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st:</td>
<td>A</td>
<td>C</td>
<td>D</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>2nd:</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>3rd:</td>
<td>C</td>
<td>D</td>
<td>B</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>4th:</td>
<td>D</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Note that a social welfare function is any way of sending a preference schedule to an election outcome. It will include any method of counting votes not just those we use regularly (i.e. not just the plurality method, plurality-with-elimination method, Borda count method, ...).

Arrow’s Fairness Criteria can be translated to properties of the social welfare function in the following ways:

- If $F$ is a social welfare function, then an $F$-dictator is a $v \in V$ so that for all $p \in P(C)$, $F(p) = p(v)$.
- If $p$ is a profile, and $a$ and $b$ are candidates, we say that $p$ unanimously ranks $a$ above $b$ when $a$ is above $b$ in $p(v)$ for all $v \in V$. $F$ satisfies the unanimity principle when $p$ unanimously ranks $a$ above $b$ then $F(p)$ ranks $a$ above $b$.
- Let $a$ and $b$ be candidates and $p$ be a profile. Then $p$’s $a, b$-subprofile is the profile $p_{a,b}$ on $\{a, b\}$ defined by saying for all $i$ $p_{a,b}(v_i)$ puts $a$ above $b$ if $p(v_i)$ puts $a$ above $b$. (This can be thought of as the two candidate subprofile involving only $a$ and $b$.) We say that $F$ satisfies the independence of irrelevant alternatives (IIA) condition when for all $a$ and $b$, if $p$ and $q$ have the same $a, b$-subprofile, then $a$ is above $b$ in $F(p)$ iff $a$ is above $b$ in $F(q)$.

We are also assuming that the social welfare function $F$ is a well-defined, surjective function (which codifies the universality and non-imposition criteria above).

Whew – That was a lot of definitions! Here are the intuitions to have: a profile is like a complete set of ballots for an election. A social welfare function is a way of turning complete sets of ballots into a ranking of candidates. Given a social welfare function $F$, an $F$-dictator is a voter whose decisions always decide the outcome of the election. We say $F$ satisfies the unanimity principle when $F$ respects whatever the voters unanimously agree on. Finally, $F$ satisfies the IIA condition just when if $a$’s performance among voters relative to $b$ in election 1 is no worse than $a$’s performance among voters relative to $b$ in election 2, $a$’s outcome in election 1 relative to $b$ is no worse than $a$’s outcome in election 2 relative to $b$.

Ok, back to our formal hijinks. The theorem we are going to prove is this:

\[ \text{Note that this version of IIA is actually somewhat stronger than that which was stated informally on page 1. The version on page 1 is more intuitive to understand; however, this version makes the proof simpler. There many versions of IIA that have been explored and the appropriate phrasing of such a criterion is up for debate.} \]
If the number of candidates, \( m \) is at least three, and \( F \) satisfies the unanimity principle and the IIA condition, then there is an \( F \)-dictator.

We’ll spend the rest of the document proving this theorem. We’ll do so in the special case in which \( m = 3 \). The modifications required to account for \( m > 3 \) are minor. We will also suppose that the three candidates are named \( a \), \( b \), and \( c \), rather than \( c_1 \), \( c_2 \), and \( c_3 \).

**Proof:**

Suppose \( F \) is a social welfare function that satisfies the unanimity principle and the IIA condition.

In this part we are identifying a voter that exists. Her name is \( k \):

To begin, let \( p_0 \) be any profile in which every voter ranks \( a \) above \( b \). For \( 0 < i \leq N \), let \( p_i(v_j) = p_{i-1}(v_j) \) for \( i \neq j \) and let \( p_i(v_i) \) be some ranking in which \( b \) is above \( a \). Thus, in \( p_0 \) nobody ranks \( b \) above \( a \); in \( p_1 \), only voter \( v_1 \) ranks \( b \) above \( a \); in \( p_2 \) only voters \( v_1 \) and \( v_2 \) rank \( b \) above \( a \); \ldots; and in \( p_n \) **everybody** ranks \( b \) above \( a \). Notice that since \( F \) satisfies the unanimity principle, \( a \) is ranked above \( b \) in \( F(p_0) \) and \( b \) is ranked above \( a \) in \( F(p_n) \). So the set of numbers \( i \) for which \( b \) is ranked above \( a \) in \( F(p_i) \) is nonempty. We let \( k \) be the least member of this set.

Now we are going to see how the voter \( k \)'s decisions about \( b \) and \( c \) impact the outcome of one particular election:

For the next part, we need to describe a particular profile that we will call \( q \). In \( q \), for \( 1 \leq i < k \), \( q(v_i) \) puts \( b \) above \( c \) and \( c \) above \( a \). To make life easier we will describe this by saying that \( q(v_i) \) has the form \( b \ c \ a \). We will also suppose for \( k \leq j \leq N \) that \( q(v_j) \) has the form \( a \ b \ c \). Notice that in \( q \), the first \( k - 1 \) voters rank \( b \) above \( a \), but none of the remaining voters do. Thus, the \( a, b \)-subprofile of \( q \) is the same as the \( a, b \)-subprofile of \( p_{k-1} \). So by IIA, since \( a \) is ranked above \( b \) in \( F(p_{k-1}) \), \( a \) is also ranked above \( b \) in \( F(q) \). Also notice that since every voter ranks \( b \) above \( c \), by unanimity \( F(q) \) must rank \( b \) above \( c \). So \( F(q) \) has the form \( a \ b \ c \).

Notice that this is exactly what \( k \) voted for. So we could say that \( k \) determined the outcome in this election.

Now we will consider a certain type of election \( r \) and will see how \( k \)'s choices about \( b \) and \( c \) impact the outcome.

Now say that a profile of type \( r \) is a profile with the following features:

- For \( 1 \leq i < k \), \( r(v_i) \) has either the form \( b \ c \ a \) or the form \( c \ b \ a \).
- \( r(v_k) \) has the form \( b \ a \ c \).
- For \( k < j \leq N \), \( r(v_j) \) has either the form \( a \ b \ c \) or the form \( a \ c \ b \).

Note that in any profile \( r^* \) with type \( r \), the first \( k \) voters all rank \( b \) above \( a \) and the remaining voters rank \( a \) above \( b \). Thus the \( a, b \)-subprofile of \( r^* \) is the same as the \( a, b \)-subprofile of \( p_k \). So by
IIA, since $b$ is ranked above $a$ in $\mathcal{F}(p_k)$, $b$ must be ranked above $a$ in $\mathcal{F}(r^*)$ as well. Finally, notice that the $a,c$-subprofile of $r^*$ is the same as the $a,c$-subprofile of $q$. So again by IIA, since $\mathcal{F}(q)$ ranks $a$ above $c$, $\mathcal{F}(r^*)$ must also rank $a$ above $c$. So $\mathcal{F}(r^*)$ has the form $\frac{b}{c}$.

Notice again that this is exactly what voter $k$ voted for!

We will now look at an arbitrary profile where votes $k$ ranks $b$ above $c$. We will see that the outcome ranks $b$ above $c$.

Now choose any profile $s$ in which $s(v_k)$ ranks $b$ above $c$. Then the $b,c$-subprofile of $s$ will be the same as the $b,c$-subprofile of some profile of type $r$. So by IIA, $\mathcal{F}(s)$ will rank $b$ above $c$. Stated otherwise, we have determined that for any profile $p$, if $b$ is above $c$ in $p(v_k)$ then $b$ is above $c$ in $\mathcal{F}(p)$.

What we have done so far is we have identified a particular voter $k$ if then rank $b$ over $c$ then the outcome will rank $b$ above $c$. Of course we can find some such voter (possibly different) for any pair of candidates, not just $b$ and $c$, however, there is no reason at this point that to suspect that these voter will all be voter $k$. In what remains we will show that in fact these voters are all voter $k$.

We will show that by permuting the roles of $a$, $b$ and $c$ above, for any pair of candidates, we can find a voter that always decides the ranking of those two candidates (just like $k$ did for $b$ and $c$).

Nothing in the argument just given relied on anything special about $b$ and $c$. So the number $k$ identified above will now be called $k_{b/c}$. Permuting $a$, $b$, and $c$ in the above argument we can see that for any pair $\langle x,y \rangle$ of distinct candidates, there is a number $k_{x/y}$ such that if $p(v_{k_{x/y}})$ ranks $x$ higher than $y$, then $\mathcal{F}(p)$ ranks $x$ higher than $y$ as well. First note that $k_{x/y}$ and $k_{y/x}$ are the same: But if $p$ were a profile in which $v_{k_{x/y}}$ ranked $x$ higher than $y$ and a profile in which $v_{k_{y/x}}$ ranked $y$ higher than $x$, then $\mathcal{F}(p)$ would have to both rank $x$ above $y$ and $y$ above $x$, which is impossible. So $k_{x/y}$ and $k_{y/x}$ must be the same number. So we can simplify and just write $k_{x,y}$. This leaves us with three partial dictators: $k_{a,b}$, $k_{a,c}$, and $k_{b,c}$.

But the three partial dictators cannot be distinct. To see this, suppose $p$ is a profile in which $p(v_{k_{a,b}})$ ranks $a$ above $b$, $p(v_{k_{b,c}})$ ranks $b$ above $c$, and $p(v_{k_{a,c}})$ ranks $c$ above $a$. Then $\mathcal{F}(p)$ must rank $a$ above $b$, rank $b$ above $c$, yet (impossibly) also rank $c$ above $a$. So either $k_{a,b} = k_{a,c}$ or $k_{a,b} = k_{b,c}$ or $k_{a,c} = k_{b,c}$.

Examining the first of these options, suppose $p$ is a profile in which $p(v_{k_{a,b}})$ has the form $\frac{c}{b}$ but $p(v_{k_{b,c}})$ ranks $b$ above $c$. Then $\mathcal{F}(p)$ must $c$ above $a$ (since $k_{a,b} = k_{a,c}$) and must rank $a$ above $b$, yet (impossibly) also rank $b$ above $c$. So in this case all three partial dictators are the same. Permuting the role of $a$, $b$, and $c$ gives the same result in all other cases as well, completing the proof.

References & Links:
Me: math.umn.edu/~kling202
CNN 2016 Election Results: https://www.cnn.com/election/2016/results
Radio Lab “Tweak the Vote”: https://www.wnycstudios.org/shows/radiolab