A reminder of the notation:

<table>
<thead>
<tr>
<th>English</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of candidates</td>
<td>$N$</td>
</tr>
<tr>
<td>Set of all candidates</td>
<td>$K = {A, B, C, \ldots}$</td>
</tr>
<tr>
<td>Ballot</td>
<td>Permutation (i.e., total ordering) $\sigma$ of $K$</td>
</tr>
<tr>
<td>where $\sigma(i)$</td>
<td>$i$th favorite choice</td>
</tr>
<tr>
<td>Set of all ballots</td>
<td>$S_K$ (cardinality $N!$)</td>
</tr>
<tr>
<td>Preference schedule (or profile)</td>
<td>a function $P : S_K \to \mathbb{N}$</td>
</tr>
<tr>
<td>where $P(\sigma)$</td>
<td>number of voters who cast ballot $\sigma$</td>
</tr>
<tr>
<td>Voting system</td>
<td>a function $V : {\text{preference schedules}} \to S_K$</td>
</tr>
</tbody>
</table>

Just to have it all in one place, here are the fairness criteria we need for a voting system $V$. Here $A, B, C$ are any candidates (I’m going to stop calling them $c_i$ and $c_j$ in order to use fewer symbols). As a reminder, $V_P(A) < V_P(B)$ means that with respect to a profile $P$, the voting system $V$ prefers $A$ to $B$, not vice versa—because $V_P(A) = i$ means that $A$ is the $i$th-place finisher.

- **Transitivity Criterion (TC):** If $V_P(A) < V_P(B)$ and $V_P(B) < V_P(C)$, then $V_P(A) < V_P(C)$. (Essentially, this just says that the output from a voting system should be a total order of the candidates.)

- **Unanimity Criterion (UC)** (a.k.a. weak Pareto criterion): If, in profile $P$, every ballot $\sigma$ has the property $\sigma(A) < \sigma(B)$—that is, every voter prefers $A$ over $B$—then $V_P(A) < V_P(B)$.

- **Independence of Irrelevant Alternatives (IIA):** Suppose that $V_P(A) < V_P(B)$. Construct a new profile $\hat{P}$ by disqualifying $C$ and keeping the relative orders of all other candidates unchanged. Then $V(\hat{P}, A) = V(\hat{P}, B)$.

And now, once again, in its full glory and terror, **Arrow’s Theorem:**

**Arrow’s Impossibility Theorem:**

Suppose $N \geq 3$. Then any voting system $V$ which always satisfies TC, UC and IIA is a dictatorship.

Proof. There are several steps. Essentially, the idea is to make the criteria fight with each other until we achieve a ridiculous conclusion. The profiles involved are going to sound somewhat strange, and arguably would not often be achieved in a real election—we’ll come back to this issue later.

Step 1: Suppose that the voting system $V$ satisfies IIA and UC, and that in profile $P$, the candidate $B$ is ranked either first or last on each individual ballot. That is, for every $\sigma$,

either $\sigma(1) = B$ or $\sigma(N) = B$ or $P(\sigma) = 0$.

Claim 1: $V_P(B) \in \{1, N\}$—that is, $B$ either wins or finishes last.

Suppose not—that is, there are candidates $A, C$ such that $V(A) < V(B) < V(C)$. Construct a new profile $P'$ as follows: every voter who preferred $A$ over $C$ now changes her ballot as follows:

$B, \ldots, A, x, y, z, p, d, q, C, \ldots \quad \Rightarrow \quad B, \ldots, x, y, z, p, d, q, C, A, \ldots$

or

$\ldots, A, x, y, z, p, d, q, C, \ldots, B \quad \Rightarrow \quad \ldots, x, y, z, p, d, q, C, A, \ldots, B$
By UC, we must now have

(1) \[ V_P(A) > V_P(C). \]

On the other hand, no voter has changed his preference between A and B, or between B and C. Therefore, IIA says that

(2) \[ V_P(A) < V_P(B) < V_P(C) \]

which together with (1) gives a contradiction. So that takes care of Claim 1.

Claim 2: There is some voter who is “extremely pivotal”, in the sense that she has the power to change B from bottom to top by changing her vote.

To see this, assume WLOG that \( V_P(B) = N \), and that

\[ \sigma_j(B) = \begin{cases} N & \text{for } 1 \leq j \leq n \\ 1 & \text{for } n + 1 \leq j \leq N \end{cases} \]

That is, B is ranked last on the first \( n \) ballots, and first on the other \( N - n \) ballots.

Let’s see what happens when we change the first \( n \) ballots, one by one. That is, let \( P_1 \) be the profile we obtain by changing \( \sigma_1 \) from

\[ (\ldots, \text{blah,blah,blah}, B) \]

to

\[ (B, \ldots, \text{blah,blah,blah}). \]

Let \( P_2 \) be the profile obtained by changing both \( \sigma_1 \) and \( \sigma_2 \) this way... and so forth: \( P_3, \ldots, P_n \). It is certainly true that \( V(P_n, B) = 1 \) by unanimity, so at some point, \( B \) must all of a sudden flip from least favored to most favored—that is, there must be some \( \ell \) such that

\[ V_{P_{\ell-1}}(B) = N \quad \text{and} \quad V_{P_\ell}(B) = 1 \]

so \( \ell \) is the voter we’re looking for. So Claim 2 is taken care of. For short (and as in class), let’s put \( Q = P_{\ell-1} \) and \( R = P_\ell \).

Claim 3: Suppose that \( A, C \) are candidates other than B. Then \( \ell \) is a dictator over the pair \( AC \)—that is, for every profile \( P \),

(3) \[ V_P(A) < V_P(C) \iff \sigma_\ell(A) < \sigma_\ell(C). \]

This seems strange, but here’s why. Right now,

\[ \sigma_\ell = (B, \ldots, \text{yada,yada}, A, \text{mumbo,jumbo,} \ldots) \]

Construct a new profile \( T \) by changing \( \sigma_\ell \) to

\[ (A, B, \text{yada,yada}, \text{mumbo,jumbo,} \ldots) \]

and allowing everyone not named \( \ell \) to rearrange their relative rankings of \( A \) and \( C \), without changing their (either top or bottom) ranking of \( B \)—that is, we allow the changes

\[ (B, \text{yak,yak,} A, \text{homina,homina,} C, \ldots) \quad \text{or} \quad (\text{yak,yak,} A, \text{homina,homina,} C, \ldots, B) \]

for each \( \sigma_k \) where \( k \neq \ell \).

Now everyone’s relative ranking of \( A \) and \( B \) in \( T \) is the same as that in \( Q \). Since

\[ V_Q(A) < N = V_Q(B), \]

we have by IIA

(4) \[ V_T(A) < V_T(B). \]
By the same reasoning, everyone’s relative ranking of $B$ and $C$ in $T$ is the same as that in $R$, and $V_R(C) > 1 = V_R(B)$, so

$$V_T(C) > V_T(B).$$

By transitivity, (4) and (5) imply together that $V_T(A) < V_T(C)$. But this doesn’t depend on how anyone except $\ell$ voted! By the exact same reasoning, if $\ell$ had chosen to put $C$ at the top instead of $A$, then we would have $V_T(A) > V_T(C)$. This establishes (3).

This argument shows that for every pair of candidates $X,Y$, there is some voter $\delta(X,Y)$ who is a dictator over that pair. If $X \neq B$, who can possibly be $\delta(X,Y)$? Well, we know for sure that

$$V_Q(B) > V_Q(X) \quad \text{and} \quad V_R(B) < V_R(X),$$

and the only voter who changed her mind between profiles $Q$ and $R$ was $\ell$ herself. Since $\ell$ can affect the relative ranking of $X$ and $B$, she must be the dictator over that pair! Therefore $\ell$ is the dictator over every pair—that is, she is the dictator, period. $\square$

Admittedly, we had to jump through quite a few hoops to prove Arrow’s Theorem. This is good news—it does say that the situations in which one of the fairness criteria is violated are, we hope, few and far between. OTOH, you can kinda get a sense from the proof that IIA is somewhat of a problematic fairness criterion; it sounds great in principle, but it can lead to some very strange situations, and in fact can be quite hard to satisfy in practice.

A related result, which we won’t prove (but, I am told, can be deduced from Arrow’s Theorem) is:

**The Gibbard-Satterthwaite Theorem:** If $N \geq 3$, then every voting system $V$ is either a dictatorship or is susceptible to manipulation.

That is, given $V$, there exists a profile $P$ such that some voter can achieve a more favorable outcome to himself by changing his ballot so as not to reflect his true preferences.

An entertaining side note: When someone pointed out that the Borda count election method was manipulable (two hundred years before Gibbard and Satterthwaite independently proved their result), Borda angrily responded, “My scheme is only intended for honest men!”