Arrow's Impossibility Theorem

By Joe Leventis

***Statement:** Any Preference Aggregation Rule which respects transitivity, unanimity, and independence of irrelevant alternatives is a dictatorship.*

Arrow's Impossibility Theorem, stated above, is a mathematical theorem dealing with voting systems. It essentially means that in a situation where voters must choose one out of three or more alternatives, no system where the voters rank the options by preference will produce a result that satisfies a series of conditions unless the final result depends on the preferences of a single voter (this is defined as a dictatorship). The theorem forms part of Social Choice Theory, a discipline which involves analysing how a society can reach collective decisions in a mathematical or logical theoretical framework.

In order to understand this theorem, it is necessary to explain several of the terms used in the statement:

Preference Aggregation Rule - This is essentially a fancy name for a voting system. It is so called because it takes in each individual's preferences and aggregates them into a single set of preferences (called the "social preferences", since they attempt to represent the aggregated preferences of society). For a single-winner election, the most preferred option in these social preferences is considered to have won the vote. So a Preference Aggregation Rule should function in this way:
Given a set of distinct options, e.g. $\{A,\ B,\ C\}$, it will aggregate the set of all individual preferences (individual preferences, such as individual i prefers option A to option B, are denoted $>_i$), e.g. $\{>_1, >_2, >_3, >_4, >_5\}$, and output the social preferences (often denoted >*), e.g. $B > A > C$.

Transitivity - The preferences must not "loop". This means that given a set of distinct options, e.g. $\{A,\ B,\ C\}$, it must not be possible for $B\ >^*\ A$ and $A\ >^*\ C$, but $C\ >^*\ B$.

Unanimity - If all individuals prefer one option to another, then the social preferences should prefer that option to the other.

Independence of Irrelevant Alternatives (IIA) - Whether A is preferred to B or not in the social preferences must depend only on individuals' preferences between A and B, and not any other factor. This means that if, from the set of options $\{A,\ B\}$, $A\ >^*\ B$, then the introduction of another option, Y, to the set, must not be able make it so that $B >^* A$.

Dictatorship - This is defined as where the social preferences do not take all voters into account and instead mimic the preferences of one individual.

While Arrow's Impossibility Theorem may seem of purely academic interest on the surface, it is of great relevance to society and especially to our current time. It has the effect of demonstrating that all voting systems which rely purely on ranking of candidates must suffer from one or several severe flaws. Analysis of voting systems through Social Choice Theory using the theorem must be a part of making any informed decision about the use of a single-winner voting system. This is incredibly important in the political climate we are experiencing, with many amongst the general public losing faith in the electoral process. The dire need for change can be seen especially clearly in the United States, where two individuals in the last 25 years have won the presidency while losing the popular vote, and the latest election is believed by many to have been illegitimate, inflaming tensions to the point where some were willing to storm the legislature in response. Some states in the US have started to consider reform, with Maine using Ranked-Choice Voting in all its elections since 2018. In the United Kingdom our current First-Past-The-Post system has given the government a large majority in Parliament despite not having won anywhere near a majority of the vote (though they were relatively far ahead of their closest rivals). With this context of malfunctions and mistrust, the careful analysis of voting systems through Arrow's Impossibility Theorem (and by other mathematical means) seems to be essential to the proper function and maintenance of stable democracies.

Now that the theorem and its importance can be somewhat understood, it may be useful and informative to delve into what it actually means for voting systems. We have seen that a voting system which does not respect transitivity has the potential to deliver nonsensical results in which no candidate wins outright. This aspect is relevant to systems designed to find the Condorcet winner of an election. An election has a Condorcet winner if one option would defeat every other option in a one-on-one contest. This can be determined by each voter ranking their choices in order of preference on their ballot. Isn't true, however that if each voter's preferences are transitive that the resulting social preferences will also be transitive. For example:

For the set of distinct options $\{A, B, C\}$, three voters set out their preferences thusly:

1. $B >_1 A >_1 C$
2. $C >_1 B >_1 A$
3. $A >_3 C >_3 B$

In two out of the three votes B is preferred to A, in two of the three C is preferred to B, and in two of the three A is preferred to C. So if we go by pairwise majorities our Preference Aggregation Rule will output $C >^* B >^* A$, but also $A >^* C$, which makes the result unusable.

This shows that any Condorcet method of voting will need a backup plan for where there is no option which would defeat every other in a one-on-one contest, or risk being unable to deliver a result.

Other ramifications of Arrow's Impossibility Theorem can best be delved into by looking at parts of an informal proof of the theorem. We first assume that the social preferences $\geq^*$ are always transitive, and always satisfy unanimity and IIA. Then we again say that there is a set of distinct options $\{A, B, C\}$. It is first necessary to prove that if all voters rank an option strictly first or last the social preferences must also do so. This can be done via contradiction.

> Assume that this is not true. Therefore, even though every voter has ranked option B either first or last, there is a set of individual preferences which would deliver the social preferences $A >^* B >^* C$ (or $C >^* B >^* A$, the order of the other preferences is irrelevant for our purposes).

> Then, the preferences are changed so that for every individual who ranked B first, C is moved to second, so that it is above A (if this was not already true). The preferences of those who ranked option B last are also changed so that C is moved to be ranked first, above A (again, if this wasn't already true).
> Since, despite these changes, no individual's preference between A and B has been changed (B is still ranked either first, above A, or last, below A), by IIA A must still be preferred to B in the social preferences. In the same way, no individual's preferences between B and C are changed, so the social preferences still state $B >^* C$. And since $A >^* B$ and $B >^* C$, then by transitivity $A >^* C$.
> However, we have seen that all individuals now rank C higher than A, so by unanimity $C >^* A$. This is a contradiction, so it must be true that if all voters rank an option either first or last, then the social preferences must also do so.

It must then be established that any Preference Aggregation Rule which respects transitivity, unanimity and IIA is a dictatorship. I will go through that proof in part, and then discuss the implications of this section of the theorem.

> As proved above, if every individual ranks option B from the set $\{A, B, C\}$ first or last, the social preferences will also rank Beither first or last. If we start of with the former case, but change each individual's preferences one by one to having B ranked last, then at some stage we will reach a tipping point where a change in the preferences of one individual makes the social preferences change from B being ranked first to being ranked last. This is true because in this situation, every voter has ranked B either first or last in their preferences.

Using this, it is possible to further prove that the social preferences must exactly mimic the preferences of this pivotal voter, making them a dictator according to the standard defined above. The proof itself is rather more lengthy, and not directly relevant to what will be discussed, so I will assume it to be true.

Is it possible to get around the limitations set out in Arrow's Impossibility Theorem, to find voting systems which come as close as possible to satisfying the criteria? In choosing which requirements to relax, one must make meaningful compromises as to what is considered acceptable in a functional democracy. The theorem only holds for more than two options. If voters are only choosing between two alternatives simple majority voting will satisfy the criteria set out, but the method by which all the alternatives are narrowed down to a final two to be voted on will obviously suffer from similar problems to any voting system, or be undemocratic and vulnerable to abuse. If one relaxes IIA the voting system is vulnerable to tactical voting and manipulation of the system by candidates running or dropping out to advance an agenda. Since the placement of irrelevant alternatives can now have an impact, voters more familiar with the voting system can use their preferences to give an extra advantage to their preferred candidate. This may be beneficial to those of us inclined to examine voting theory, but it can't be considered fair. Similarly, the presence of irrelevant additional candidates in a race, either purposely or unintentionally, can damage the chances of some of those candidates with a shot at winning. These disadvantages, though, could be considered acceptable seeing as is possible to deliver a system in this way which functions well otherwise.

One interesting path is to relax the prohibition on dictatorship. What is a dictatorship under Arrow? As stated earlier, a dictatorship is where the social preferences don't take all voters into account and instead mimic the preferences of a single individual. There is one obvious way this could happen, a traditional authoritarian dictatorship where one person holds all the decision-making power. An example of one such system is the fictional city of Ankh-Morpork in Terry Pratchett's *Discworld* books, which is rules by the calculating tyrant Lord Vetinari, the Patrician of the city: "Ankh-Morpork had dallied with many forms of government and had ended up with that form of democracy known as One Man, One Vote. The Patrician was the Man; he had the Vote". This system, whilst clear and simple, is obviously anti-democratic and can be dismissed for a modern functional nation-state. However the standard for dictatorship defined under Arrow's Impossibility Theorem is much broader, possibly too much so. As we saw above in the proof, in a system which respects transitivity, unanimity, and IIA, there will be a pivotal voter whose choices exactly reflect the social preferences. They are called the dictator. However, for most voting systems the dictator will not be self-selected, and will instead essentially be picked by the electoral system itself, like an elected representative. Their preferences may well reflect the views of a large section of society. It has been proven that there can exist democratic dictators, whose preferences reflect those of the majority, and so it seems that it is unnecessary to strictly prohibit these, as the theorem does. What will need to be prevented in an effective voting system is dictators whose preferences reflect only a minority of society.

Arrow's Impossibility Theorem is therefore, as well as being a rather neat and satisfying piece of mathematics, incredibly useful for understanding and ameliorating the flaws in our democracies so that they may survive and flourish. It also shows that no electoral system can be perfect, and it is necessary to consider what we consider fair in a system to make an informed choice.

Sources:
Arrow's Impossibility Theorem, Wikipedia:
https://en.wikipedia.org/wiki/Arrow%27s_impossibility_theorem

Original Proof:
Social Choice and Individual Values - Kenneth Arrow (1955)

Where I read the Proof:
https://www.ssc.wisc.edu/~dquint/econ698/lecture%202.pdf

Where they got the Proof from:
"Three Brief Proofs of Arrow's Impossibility Theorem" - John Geanakoplos (2005)
http://dido.econ.yale.edu/~gean/art/p1116.pdf