The Mathematics of Voting – Arrow's Impossibility Theorem

One of the greatest struggles of having a friend group nowadays is quite frankly knowing what to do together. You and your friends Bob and Milly decide that you guys deserve a day out together, but with your busy schedules, there's only enough time to do one proper activity. You've always been avid for football in the local park, saying it's much better than going to an arcade and even more preferable to sitting still in a cinema for 3 hours. But Bob has been waiting for Dune 3 all year, and he prefers going to the cinema to playing football to gaming in the arcade. Meanwhile, Milly has been practicing her bowling skills and will prefer arcade to cinema to football.

Voter	1st choice	2nd choice	3rd choice
You	Football	Arcade	Cinema
Bob	Cinema	Football	Arcade
Milly	Arcade	Cinema	Football

The friend group then decides to initiate a head-to-head voting, before realising that every option will lose to another option in this voting system by one vote. In the end, the group shrugs it off and decides to stay at home for the 5th weekend in a row. The results are cyclical – there's no crowning one option as the overall winner because there's always an option that could beat it, similar to the dynamic of rock-paper-scissors.

Voter	Football vs Cinema	Cinema vs Arcade	Arcade vs Football
You	Football	Arcade	Football
Bob	Cinema	Cinema	Football
Milly	Cinema	Arcade	Arcade

While at home, you think to yourself that it was simply 'bad luck' that nothing could get decided, and maybe the solution would be a better voting system in relation to deciding by majority votes (or even getting better friends…). But was the issue ever with the choice of voting system, or would a different voting system have failed regardless in our scenario? More importantly, does there exist a voting system where there is a decisive winner every time, regardless of the order of voter preferences?

Translating the real world into mathematics

It should come without surprise that, to approach this dilemma, we use our favourite logical powerhouse tool – mathematics. To begin, we should define a few key terms and notations: the finite **set of voters:** $\boldsymbol{V = \{1, 2, \ldots, v\}}$ and the finite **set of choices:** $\boldsymbol{C = \{a, b, c, \ldots\}}$ **where** $\boldsymbol{|C| > 2}$, since when there are only 2 choices, majority vote will always result in a defined winner.

Each voter also holds a **preference ordering** over the choices that forms a complete, transitive ranking. This assumption allows voters to have a preference for every pair of choices, and if they prefer a to b and b to c, they will inherently prefer a to c. The assumptions of completeness and transitivity helps us build the mathematical model for a rational agent.

We will also utilise a **function** $\boldsymbol{F}$ that acts as the societal voting rankings itself, which takes every voter's induvial preference ordering as input and outputs single collective preference ordering that represents the overall choice of society. Different voting systems are represented through different choices of F. As an example, plurality voting will utilise a function F that ranks choices by how many 1st place votes they receive. This numerical function F that represents the voting system is what we are fundamentally interrogating, detached from any specific implementation.

Notice how we have constructed a perfect, idealised model of a rational voting system, where voter preferences are fixed and consistent and F always yields a clean output. Inherently, we have created the most favourable conditions for a fair voting system to exist that is far from the reality. If we fail to construct a fair system here, we certainly would fail in our reality of behavioural biases and fluctuating preferences.

The 4 conditions of a voting system

In 1951, economist Kenneth Arrow posed a related question in his doctoral thesis: what properties should any reasonable voting system satisfy? He identified 4 key conditions, each one trivially reasonable on their own.

Unanimity – If every single voter prefers a to b, then the societal ordering must rank *a* above *b*. Formally, if $a \succ_v b$ for all voters v, then $a \succ_F b$ within the function F. This serves as the floor of democracy – if every person agrees, society should agree too.

Independence of irrelevant alternatives – The societal ranking of a versus b must depend only on how individuals rank *a* versus *b*, not where any third alternative *c* sits in anyone's ordering. Whether the friend group collectively prefers going to the arcade or the cinema should not change simply someone suggests the choice of football instead. It's this condition which happens to be violated by the majority of voting systems today – the phenomenon of tactical voting exists because our collective ranking of 2 candidates shift when a third candidate enters the race.

Non-dictatorship – There is no single voter *v* such that the societal ordering always matches *v*'s ordering regardless of others. Not merely a voter who gets their way, but someone whose preferences determine the outcome even if all other voters disagree unanimously.

Unrestricted domain – The function *F* most produce a valid collective ordering for every possible combination of individual preference orderings, regardless of how inconvenient preference orderings are. A voting system that only works when voters happen to have ideal preferences is simply an evasion of the fundamental problem of finding a voting system that fairly combines induvial preferences into a collective decision. The 4 conditions above are what Arrow defined to be the foundations of a 'fair' voting system.

Arrow's Impossibility Theorem – No function F simultaneously satisfies unanimity, independence of irrelevant alternatives, non-dictatorship, and unrestricted domain.

The proof – why no function can satisfy the 4 conditions of fairness

We begin by assuming that *F* satisfies unanimity, independence of irrelevant alternatives (IIA) and unrestricted domain. We're setting up for a proof of contradiction where we show that this function will have a dictator by nature, violating the condition of non-dictatorship.

The Extremes Lemma

We then fix a choice *b* by supposing every voter places *b* either at the top or bottom of their preferences P, and no one places it within the middle of the preferences. Through the condition of unanimity, we claim that the social ordering must do the same, placing *b* at either the first or last preferred choice.

If we suppose the opposite that *b* lies in the middle of the social ordering, then there exists choices a and c such that $a \succ_F b \succ_F c$. From this point, we will prove through contradiction that *b* must be at the top or bottom in the social welfare function if every voter places *b* at the extremes.

The Mathematics of Voting – Arrow's Impossibility Theorem

Now we construct a modified preference profile P′, where every voter moves c above a in their individual rankings, while leaving the ranking of b relative to everything else completely unchanged. Through unrestricted domain, voters can hold any preference ordering regardless of how ideal.

Position of b	Original preference P	Modified preference P′
With b at top	$b \succ a \succ c$	$b \succ c \succ a$
With b at bottom	$a \succ c \succ b$	$c \succ a \succ b$

By IIA, the societal ranking of a versus b and b versus c is dependent solely on the positions of a and c relative to b that remain unchanged between the original and modified preferences. Hence, $a \succ_F b \succ_F c$ in P′. Through unanimity of every voter ranking c above a in P′, $c \succ_F a$. However, the rule of transitivity tells us that $a \succ_F c$, since we established $a \succ_F b \succ_F c$ still holds in P′. This forms a contradiction, suggesting that if b lies at the top or bottom of every P then it cannot sit in the middle of the societal ordering. Unanimity tells us that if every voter had b at the top or bottom, then the societal rankings would reflect this. But unanimity does not cover the cases between these extremes where there's a split between voters of having b at the top or bottom. From the Extremes Lemma, we now know that at every intermediate step of voters moving b from the bottom to the top - such that voters are now split - , the social ranking must still place b at the top or bottom and not in the middle, suggesting that the transition from b at the bottom to top must happen in a single step. It's this single step that forms the basis of the next stage of the proof – the pivotal voter.

The pivotal voter

From the Extremes Lemma we know there exists a voter k whose move from placing b at the bottom to the top of their individual preference causes b to jump from the bottom to the top of the societal rankings. We call k the pivotal voter for b. Assume all voters in a set V now have b at the bottom of their preference, and each voter one by one moves b right to the top.

Voters 1, 2, …, k-1	Voter k	Voters k+1, k+2,…, n
b at bottom of F		b at top of F
Typical voter: $a \succ c \succ b / c \succ a \succ b$	$a \succ_k b \succ_k c$	Typical voter: $b \succ a \succ c / b \succ c \succ a$

For all voters but k, we assume a and c's relation to each other to be arbitrary. By constructing the pivotal voter k, the societal rankings ranks $a \succ_F b$ before k's move. After voter k moves, b also jump to the top, suggesting $b \succ_F c$. Transitivity hence yields us $a \succ_F c$.

Now we invoke IIA. The societal ranking of a versus c should solely depend on induvial rankings of a versus c only. However, even when a and b being ranked arbitrarily by all voters but k, voter k's rankings of a versus c will always match the societal rankings of a and c. This stems from us deriving $a \succ_F b$ and $b \succ_F c$ from voter k's position relative to b, regardless to where other voters placed the 2 alternatives. This shows that the only input that would consistently determine the output was voter k and their preferences of a versus c. Hence, voter k is a dictator. And thus, every fair voting system that satisfies Arrow's definition of fair cannot exist.

What we learn from this proof is that mathematically a perfectly fair voting system cannot exist. We often see modern voting systems sacrificing one of the 4 conditions of fairness, hence it's only expected to see phenomena such as tactical voting to harm a rival candidate, or abstention of voting entirely. Perchance if we used similar mathematical logic to derive a quote-on-quote 'optimal' voting system, we wouldn't have as much controversy over key political elections, and maybe next weekend you and your friend group would be able to decide on an activity more easily.