Chapter 25 Arrow's Theorem

25.1 Ranking Functions

The purpose of this chapter is to set out Arrow's Theorem, and its implications for the construction of group preferences from individual preferences. We'll also say a little about the implications of the theorem for the design of voting systems, though we'll leave most of that to the next chapter.

The theorem is a mathematical result, and needs careful setup. We'll assume that each agent has a **complete** and **transitive** preference ordering over the options. If we say $A >_V B$ means that V prefers A to B, that $A =_V B$ means that V is indifferent between A and B, and that $A \geq_V B$ means that $A >_V B \vee A =_V B$, then these constraints can be expressed as follows.

Completeness For any voter V and options A, B, either $A \geq_V B$ or $B \geq_V A$

Transitivity For any voter V and options A, B, the following three conditions hold:

- If $A >_V B$ and $B >_V C$ then $A >_V C$
- If $A =_V B$ and $B =_V C$ then $A =_V C$
- If $A \geq_V B$ and $B \geq_V C$ then $A \geq_V C$

More generally, we assume the **substitutivity of indifferent options**. That is, if $A =_V B$, then whatever is true of the agent's attitude towards A is also true of the agent's attitude towards B. In particular, whatever comparison holds in the agent's mind between A and C holds between B and C. (The last two bullet points under transitivity follow from this principle about indifference and the earlier bullet point.)

The effect of these assumptions is that we can represent the agent's preferences by lining up the options from best to worst, with the possibility that we'll have to put two options in one 'spot' to represent the fact that the agent values each of them equally.

A **ranking function** is a function from the preference orderings of the agent to a new preference ordering, which we'll call the preference ordering of the group. We'll use the subscript $_G$ to note that it is the group's ordering we are designing. We'll also assume that the group's preference ordering is complete and transitive.

There are any number ranking functions that don't look at all like the *group's* preferences in any way. For instance, if the function is meant to work out the results of an election, we could consider the function that takes any input whatsoever, and returns a ranking that simply lists by age, with the oldest first, the second oldest second, etc. This doesn't seem like it is the group's preferences in any way. Whatever any member of the group thinks, the oldest candidate wins. What Arrow called the citizen sovereignty condition is that for any possible ranking, it should be possible to have the group end up with that ranking.

The citizen sovereignty follows from another constraint we might put on ranking functions. If everyone in the group prefers A to B, then $A >_G B$, i.e. the group prefers A to B. We'll call this the **Pareto** constraint. It is sometimes called the **unanimity** constraint, but we'll call it the Pareto condition.

One way to satisfy the Pareto constraint is to pick a particular person, and make them dictator. That is, the function 'selects' a person V, and says that $A >_G B$ if and only if $A >_V B$. If everyone prefers A to B, then V will, so this is consistent with the Pareto constraint. But it also doesn't seem like a way of constructing the group's preferences. So let's say that we'd like a non-dictatorial ranking function.

The last constraint is one we discussed in the previous chapter: the **independence of irrelevant alternatives**. Formally, this means that whether $A >_G B$ is true depends only on how the voters rank A and B. So changing how the voters rank, say B and C, doesn't change what the group says about the A, B comparison.

It's sometimes thought that it would be a very good thing if the voting system respected this constraint. Let's say that you believe that if Ralph Nader had not been a candidate in the 2000 U.S. Presidential election, then Al Gore, not George Bush, would have won the election. Then you might think it is a little odd that whether Gore or Bush wins depends on who else is in the election, and not on the voters' preferences between Gore and Bush. This is a special case of the independence of irrelevant alternatives - you think that the voting system should end up with the result that it would have come up with had there been just those two candidates. If we generalise this motivation a lot, we get the conclusion that third possibilities should be irrelevant.

Unfortunately, we've now got ourselves into an impossible situation. Arrow's theorem says that any ranking function that satisfies the Pareto and independence of irrelevant alternatives constraints, has a dictator in any case where the number of alternatives is greater than 2. When there are only 2 choices, majority rule satisfies all the constraints. But nothing, other than dictatorship, works in the general case.

25.2 Cyclic Preferences

We can see why three option cases are a problem by considering one very simple example. Say there are three voters, V_1, V_2, V_3 and three choices A, B, C. The agent's rankings are given in the table below. (The column under each voter lists the choices from their first preference, on top, to their least favourite option, on the bottom.)

V_1	V_2	V_3
A	B	C
B	C	A
C	A	B

If we just look at the A/B comparison, A looks pretty good. After all, 2 out of 3 voters prefer A to B. But if we look at the B/C comparison, B looks pretty good. After all, 2 out of 3 voters prefer B to C. So perhaps we should say A is best, B second best and C worst. But wait! If we just look at the C/A comparison, C looks pretty good. After all, 2 out of 3 voters prefer C to A.

It might seem like one natural response here is to say that the three options should be tied. The group preference ranking should just be that $A =_G B =_G= C$. But note what happens if we say that and accept independence of irrelevant alternatives. If we eliminate option C, then we shouldn't change the group's ranking of A and B. That's what independence of irrelevant alternatives says. So now we'll be left with the following rankings.

V_1	V_2	V_3
A	B	A
B	A	B

By independence of irrelevant alternatives, we should still have $A =_G B$. But 2 out of 3 voters wanted A over B. The one voter who preferred B to A is making it that the group ranks them equally. That's a long way from making them a dictator, but it's our first sign that our constraints give excessive power to one voter. One other thing the case shows is that we can't have the following three conditions on our ranking function.

- If there are just two choices, then the majority choice is preferred by the group.
- If there are three choices, and they are symmetrically arranged, as in the table above, then all choices are equally preferred.
- The ranking function satisfies independence of irrelevant alternatives.

I noted after the example that V_2 has quite a lot of power. Their preference makes it that the group doesn't prefer A to B. We might try to generalise this power. Maybe we could try for a ranking function that worked strictly by consensus. The idea would be that if everyone prefers A to B, then $A >_G B$, but if there is no consensus, then $A =_G B$. Since how the group ranks A and B only depends on how individuals rank A and B, this method easily satisfies independence of irrelevant alternatives. And there are no dictators, and the method satisfies the Pareto condition. So what's the problem?

Unfortunately, the consensus method described here violates transitivity, so doesn't even produce a group preference ordering in the formal sense we're interested in. Consider the following distribution of preferences.

V_1	V_2	V_3
A	A	B
B	C	A
C	B	C

Everyone prefers A to C, so by unanimity, $A >_G C$. But there is no consensus over the A/B comparison. Two people prefer A to B, but one person prefers B to A. And there is no consensus over the B/C comparison. Two people prefer B to C, but one person prefers C to B. So if we're saying the group is indifferent between any two options over which there is no consensus, then we have to say that $A =_G B$, and $B =_G C$. By transitivity, it follows that $A =_G C$, contradicting our earlier conclusion that $A >_G C$.

This isn't going to be a formal argument, but we might already be able to see a difficulty here. Just thinking about our first case, where the preferences form a cycle suggests that the only way to have a fair ranking consistent with independence of irrelevant alternatives is to say that the group only prefers options when there is a consensus in favour of that option. But the second case shows that consensus based methods do not in general produce *rankings* of the options. So we have a problem. Arrow's Theorem shows how deep that problem goes.

25.3 Proofs of Arrow's Theorem

The proofs of Arrow's Theorem, though not particularly long, are a little tricky to follow. So we won't go through them in any detail at all. But I'll sketch one proof due to John Geanakopolos of the Cowles Foundation at Yale.[1] Geanakopolos assumes that we have a ranking function that satisfies Pareto and independence of irrelevant alternatives, and aims to show that in this function there must be a dictator.

The first thing he proves is a rather nice lemma. Assume that every voter puts some option B on either the top or the bottom of their preference ranking. Don't assume they all agree: some people hold that B is the very best option, and the rest hold that it is the worst. Geanakopolos shows that in this case the ranking function must put B either at the very top or the very bottom.

To see this, assume that it isn't true. So there are some options A and C such that $A \geq_G B$ and $B \geq_G C$. Now imagine changing each voter's preferences so that C is moved above A while B stays where it is - either on the top or the bottom of that particular voter's preferences. By Pareto, we'll now have $C >_G A$, since everyone prefers C to A. But we haven't changed how any person thinks about any comparison involving B. So by independence of irrelevant alternatives, $A \geq_G B$ and $B \geq_G C$ must still be true. By transitivity, it follows that $A \geq_G C$, contradicting our conclusion that $C >_G A$.

This is a rather odd conclusion I think. Imagine that we have four voters with the following preferences.

V_1	V_2	V_3	V_4
B	B	A	C
A	C	C	A
C	A	B	B

By what we've proven so far, B has to come out either best or worst in the group's rankings. But which should it be? Since half the people love B, and half hate it, it seems it should get a middling ranking. One lesson of this is that independence of irrelevant alternatives is a very strong condition, one that we might want to question.

The next stage of Geanakopolos's proof is to consider a situation where at the start everyone thinks B is the very worst option out of some long list of options. One by one the voters change their mind, with each voter in turn coming to think that B is the best option. By the result we proved above, at every stage of the process, B must be either the worst option according to the group, or the best option. B starts off as the worst option, and by Pareto B must end up as the best option. So at one point, when one voter changes their mind, B must go from being the worst option on the group's ranking to being the best option, simply in virtue of that person changing their mind.

We won't go through the rest, but the proof continues by showing that that person has to be a dictator. Informally, the idea is to prove two things about that person, both of which are derived by repeated applications of independence of irrelevant alternatives. First, this person has to retain their power to move B from worst to first whatever the other people think of A and C. Second, since they can make B jump all options by changing their mind about B, if they move B 'halfway', say they come to have the view $A >_V B >_V C$, then B will jump (in the group's ranking) over all options that it jumps over in this voter's rankings. But that's possible (it turns out) only if the group's ranking of A and C is dependent entirely on this voter's rankings of A and C. So the voter is a dictator with respect to this pair. A further argument shows that the voter is a dictator with respect to every pair, which shows there must be a dictator.

[1] The proof is available at http://ideas.repec.org/p/cwl/cwldpp/1123r3.html.