Game Theory, Welfare, and Applications
Supervision 4

Samuel Lee

Any reasonable social choice rule must satisfy the axioms used in Arrow's impossibility theorem, so all social choices are necessarily arbitrary. Discuss.

The common statement of Arrow's impossibility theorem has it that any social welfare function (SWF) which maps individual rational preferences over more than 2 alternatives onto a societal preference relation will violate at least one of 5 conditions. These conditions are:

1. Universality
   The domain of the SWF is the set of all possible rational individual preference relations. This basically means that the SWF has something to say about every possible combination of individual preferences that may be realized.

2. Rationality
   The societal preference relation that the SWF spits out is rational. That is, the resultant societal preference relation must be complete and transitive over the alternatives.

3. Pareto efficiency
   If $x \succ_i y \ \forall i$, then $x \succ y$. This means that among two alternatives, the SWF will rank an option which everyone prefers above the other option.

4. Independence of irrelevant alternatives
   For two sets of preference relations $\{\succeq_i\}_{i\in N}$ and $\{\succeq_i'\}_{i\in N}$, if $x \succeq_i y$ implies $x \succeq_i' y$ and vice versa for all i, then $x \succeq y$ implies $x \succeq' y$ and vice versa. In words, this means the societal preference between two alternatives should only depend on individual preferences between those two alternatives. Two sets of individual preference orderings which preserve the order of x and y will lead to the same ordering between those two alternatives in the societal preference relation.

5. Non-dictatorship
   There is no individual i such that $x \succeq y$ if and only if $x \succeq_i y$.

It is by now often said that the theorem shows there can be no "satisfactory" voting system. Some examples from the popular coverage shortly after Arrow's death include, from the Financial Times:

> Scholars had long known that voting systems could produce perverse results but Arrow went further, showing that the very idea of "what society prefers" was incoherent.

and The Economist:

> A colleague studying America's strategic contest with the Soviet Union had asked him whether it was safe to treat an entire country as an individual "player", with coherent preferences. What was required, Mr Arrow knew, was a robust, reasonable rule to translate the preferences of Americans, say, into the preferences of America. But to his surprise, he discovered that such a rule was "impossible" to find.

Rarely is it appreciated that Arrow actually called the theorem a "Possibility Theorem" in his papers (at Tjalling Koopmans's urging, according to Arrow himself[1]). This is a small bit of trivia which nevertheless suggests that the result might be better framed as "a preference aggregation rule can be found by relaxing one of the axioms".

A social choice rule does not necessarily have to come from aggregating individual preferences. For example, the Rawlsian SWF gives the societal welfare as the utility of the least well-off individual. However, the fixation on individual preferences probably arises because of an attractive property: rational choice theory is self-contained (some would say circular). From a revealed-preference perspective one need only observe that some alternative x is chosen to know that it is weakly preferred to all other alternatives in the choice set. In this way, utility representations of individual preferences are positive rather than normative by construction. If individual preferences could naturally be extended to preferences of society as a whole, this property would be preserved.

Instead, if as above our perspective is that "a preference aggregation rule can be found by relaxing one of the axioms", there is now the uncomfortable task of choosing one of the axioms to violate or moving away from the sole dependence on individual preference relations altogether. There are many ways of escaping the "impossibility" result, but these would depend on additional assumptions which are not immediately obvious (some are more obvious than others; for instance, it is doubtful that violating non-dictatorship is of practical use in this problem). This then becomes "arbitrary" in a broad sense.

This may not be the most preferred outcome, but "arbitrary" does not necessarily have to mean "meaningless". It seems inescapable that a choice of some sort is needed when deciding on a social welfare function, but as mentioned above some choices are more obvious than others. A possible way out of the problem is to allow for interpersonal comparisons of utility, which does not always have to be controversial (as in Sen's famous Emperor Nero example). And in any case, Arrow's theorem only states that all aggregation rules can fail one of the criteria, not that they necessarily will fail when put into practice. If, for example, all individuals happen to have single-peaked preferences, then the universality condition becomes irrelevant and a satisfactory preference aggregation rule can exist. The focus can instead be on which axioms are least costly to violate given what we know about the underlying distribution of preferences and alternatives, which turns this into more of an empirical question. When this is all said and done, the final choice may turn out to only be "arbitrary" in the broadest technical sense.

[1] JK: "Why did you call it a "Possibility" Theorem?"
KA: "That was Tjalling's idea. Originally I called it an impossibility theorem, but he thought that was too pessimistic! He was my boss and a very sweet man, so I changed it for him." (Kelly 1987)