ARROW'S IMPOSSIBILITY THEOREM OF SOCIAL CHOICE

HONI SANDERS

Abstract. A proof of Arrow's Impossibility Theorem based on the five conditions he imposed on the social-welfare function in his 1950 paper, "A Difficulty in the Concept of Social Welfare."

In order to make policy decisions, some algorithm must be used to choose one alternative out of all the available possibilities of action based on the preferences of the constituency. One might call this algorithm an election. The plurality election where the option with the most people preferring that option to any other is the one on which our mayorial, congressional, and most other elections are decided. One problem is that it could be that 30% of the population prefers a to b and b to c, 30% of the population prefers b to a and a to c, and 40% of the population prefers c to b and b to a. In this case, c would win, even though the majority of the population would rather anything but c. Many other election methods have been proposed as alternatives, but each either has outcomes that are unacceptable in certain cases or does not always give a winner. In his 1950 paper, "A Difficulty in the Concept of Social Welfare," Kenneth Arrow sets out guidelines for what he believes is a fair election method and proves that no such method exists. The social decision under consideration can be represented as a finite or infinite set S, and the options are represented as the elements $x_k \in S$. We can describe preference as a relation on the set. For an individual i, we will define relations R_i and P_i.

Definition 1. If individual i prefers x_k to x_j or is indifferent between x_k and x_j, then $x_k R_i x_j$.

Using this definition we can define P_i in terms of R_i:

Definition 2. $\sim x_k R_i x_j \equiv x_j P_i x_k$

The relations R_i and P_i on S are comparable to the relations $\geq$ and $>$, respectively, on $\mathbb{Z}$, so Definition 2 is comparable to defining $>$ in terms of $\geq$: $x \ngeq y \equiv y > x$.

We assume that all individuals are rational and preference is based on relative utility, so preference and indifference are transitive. From the transitivity property of R_i and P_i, and following directly from the definition, we have

(i) $x_k R_i x_j, x_j R_i x_m \rightarrow x_k R_i x_m$
(ii) $x_k P_i x_j, x_j R_i x_m \rightarrow x_k P_i x_m$
(iii) $x_k P_i x_j \rightarrow x_k R_i x_j$

Our goal is to find a social-welfare function, that is, an algorithm to determine relations R and P, representing the societal preference, on the set S given the relations $\{R_i\}$. Kenneth Arrow listed five conditions that he thought one should expect the social-welfare function and the relations R and P that it produces to fulfill.

Condition 1. The social-welfare function must have an output R and P for any set of relations $\{R_i\}$.

Condition 2. If $\{R_i\}$ changes s.t. all $x_k R_i x_j$ hold except for relations involving x_m, and $\forall R_i, \{x_k | x_m R_i x_k \text{ before}\} \subset \{x_k | x_m R_i x_k \text{ after}\}$ and $\{x_k | x_m P_i x_k \text{ before}\} \subset \{x_k | x_m P_i x_k \text{ after}\}$, then $\{x_k | x_m P x_k \text{ before}\} \subset \{x_k | x_m P x_k \text{ after}\}$. (That is, if people think more highly of x_m and nothing else changes, then x_m should not be thought of lower by society as a whole.)

Condition 3 (Independence of Irrelevant Alternatives)**.** $\forall C \subset S$ with relation R' on C ($\{R_i\}$ are relations on C by transitivity), for $x_k, x_j \in C, x_k R x_j \leftrightarrow x_k R' x_j$. (The ordering of given options should not change with the presence of other options.)

Condition 4 (Citizen Sovereignty)**.** $\nexists x_k, x_j$ s.t. $x_k R x_j \forall \{R_i\}$. (Any societal preference is allowed.)

Condition 5 (Nondictatorship)**.** $\nexists$ individual i s.t. $x_k P_i x_j \rightarrow x_k P x_j$.

Arrow then proved that these five conditions imply a contradiction. He used a society with two individuals and three elements in S. By Condition 3, disproving the existence of a social-welfare function on a set with three elements disproves it for all numbers of elements. He uses a mathematical wave of the hand on the generalizability of proofs involving societies with two individuals: "The restriction to two individuals may be more serious; it is conceivable that there may be suitable social welfare functions which cab be defined for three individuals but not for two, for example. In fact, this is not so, and the results stated in this paper hold for any number of individuals."

Lemma 1 (Pareto Efficiency)**.** $x_k P_i x_j \forall i \rightarrow x_k P x_j$.

Proof. Assume that $x_k P_i x_j \forall i \nrightarrow x_k P x_j$. Then, $\exists x_k, x_j$ s.t. $x_k P_i x_j \forall i$ but $x_j R x_k$. Condition 4 is equivalent to the statement $\forall x_k, x_j \exists \{R_i\}$ s.t. $x_k P x_j$. Take that $\{R_i\}$, which is not $x_k P_i x_j \forall i$ by assumption, and change only relations dealing with x_k so that $\forall i$ and $\forall x_m, x_k P_i x_m$. We now have a $\{R_i\}$ where $x_k P_i x_j \forall i$. But by construction following Condition 2, because $x_k P x_j$ before the change, it must be that $x_k P x_j$ after the change, which contradicts the original assumption. □

It is at this point that Arrow's proof begins to deal with only societies with two people.

Lemma 2. *If* $x_k P_1 x_j, x_j P_2 x_k, x_k P x_j$*, then* $x_k P_1 x_j \rightarrow x_k P x_j$.

Proof. Take R_1 where $x_k P_1 x_j$ and any R_2. Change R_2 only with respect to x_k to R_2' so that $\forall x_m, x_m P_2' x_k$, so $x_j P_2' x_k$. With R_1 and R_2', $x_k P' x_j$ by assumption. Change R_2' back to R_2 only by changing relations dealing with x_k, which is a change following Condition 2, meaning that $x_k P x_j$. □

Lemma 3. $x_k P_1 x_j, x_j P_2 x_k \rightarrow x_k R x_j, x_j R x_k$

Proof. Assume $x_k P_1 x_j, x_j P_2 x_k \nrightarrow x_k R x_j, x_j R x_k$. That is, $\exists R_1, R_2$ where $x_k P_1 x_j, x_j P_2 x_k$ and either $x_k P x_j$ or $x_j P x_k$.
With $S = \{a, b, c\}$, set $x_k = a, x_j = b$ without loss of generality. We will prove

a contradiction assuming aPb, and the case of bPa can be proved by the same process. By Lemma 2, we have

$$aP_1b \to aPb. \tag{1}$$

Take R_1 where aP_1b, bP_1c and R_2 where bP_2c, cP_2a, bP_2a. aPb by (1), and bPc by Lemma 1, so aPc by transitivity. By Lemma 2,

$$aP_1c \to aPc. \tag{2}$$

Take R_1 where bP_1a, aP_1c and R_2 where cP_2b, bP_2a. Then, bPa by Lemma 1 and aPc by (2); so by Lemma 2,

$$bP_1c \to bPc. \tag{3}$$

Take R_1 where bP_1c, cP_1a and R_2 where cP_2a, aP_2b, cPa by Lemma 1 and bPc by (3); so by Lemma 2,

$$bP_1a \to bPa. \tag{4}$$

Take R_1 where cP_1b, bP_1a and R_2 where aP_2c, cP_2b, cPb by Lemma 1 and bPa by (4); so by Lemma 2,

$$cP_1a \to cPa. \tag{5}$$

Take R_1 where cP_1a, aP_1b and R_2 where aP_2b, bP_2c, aPb by Lemma 1 and cPa by (5); so by Lemma 2,

$$cP_1b \to cPb. \tag{6}$$

Equations (1)-(6) can be summarized as $\forall x_m, x_n \in S, x_mP_1x_n \to x_mPx_n$ which establishes individual 1 as a dictator, and contradicts Condition 5. □

Now we can prove our main theorem:

Theorem 1 (Arrow's Impossibility Theorem)**.** *There is no social-welfare function which fulfills Conditions 1-5 and produces a rational societal preference ordering.*

Proof. Take R_1 where aP_1b, bP_1c and R_2 where cP_2a, aP_2b. By Lemma 1, aPb. $bP_1c, cP_2b \to bRc, cRb$ by Lemma 2, so aPc by (ii), but $aP_1c, cP_2a \to aRc, cRa$. □

References

[1] Kenneth J. Arrow. A Difficulty in the Concept of Social Welfare. The Journal of Political Economy, Vol. 58, No. 4. August, 1950 pp. 328-346.