Arrow's Impossibility Theorem on Social Choice Systems

Ashvin A. Swaminathan

January 11, 2013

Abstract

Social choice theory is a field that concerns methods of aggregating individual interests to determine net social preferences. Arrow's Impossibility Theorem states that no social choice system can satisfy all requirements in a collection of "reasonable" criteria. We make the notion of "reasonable" voting systems precise by stating the four properties that voting systems must have in order to be "reasonable." We then present two different proofs of Arrow's Theorem, both of which combine three of the "reasonable" properties to obtain a contradiction of the fourth. Finally, we discuss the theorem's implications for social choice theory and show how the theorem can be applied to real-world examples of voting systems.

1. Social Choice Theory

Kenneth Arrow's Impossibility Theorem, which he first introduced in his seminal 1951 book *Social Choice and Individual Values*, is considered to be the mathematical foundation of modern social choice theory, a field that concerns methods of aggregating individual interests to determine net social preferences [1]. The theorem essentially states that no social choice system can determine net social preferences without violating at least one condition in a specific set of "reasonable" criteria. These criteria have their roots in the democratic basis of social choice theory, the belief that social decisions should, in some "reasonable" way, depend on the preferences of individuals in the society [4] and on nothing else. As we discuss in Section 2, we can make the notion of "reasonable" criteria precise by considering a list of properties that we would like our social choice systems to satisfy.

The process of social decision-making has numerous real-world applications. The market system itself (assuming *laissez-faire*, of course) is a prime example of the importance of social choice: in this system, the price and quantity of a certain product is determined solely by the net demand (the choice to buy) of the product's buyers and the net supply (the choice to sell) of the product's sellers. Moreover, on the buyers' side itself, individuals makes choices among various products to maximize their own total utility. Likewise, on the sellers' side itself, individual firms make choices among various resources to maximize their own total profits. In particular, a firm is owned by its shareholders, so if the shareholders disagree about how profit can be maximized, the wishes of each individual shareholder need to be aggregated in a social choice system to yield the final production strategy (which cannot be done "reasonably," according to Arrow's Theorem) [3]. Also, some decisions are made in the government, in which case representatives make decisions on nation-wide issues through some kind of "voting system" (this is what goes on in governmental bodies like the U.S. Congress all the time). In sum, whether school administrators are seeking to maximize their students' total utility by selecting appropriate numbers of basketballs and footballs to purchase or citizens are choosing among various candidates for their country's presidency, people are constantly forced to make choices.

2. Voting Systems

For ease of discussion, we will consider only voting systems throughout the rest of the paper, but it is important to recognize that our subsequent arguments for voting systems can be easily applied to social choice systems in general.

Before we define voting systems precisely, we provide the necessary background definitions and assumptions. Let the finite set S be the set of all voters in society and the finite set C be the set of all candidates (candidates are often more generally referred to as alternatives or options). For any two candidates $a, b \in C$, each $s \in S$ has either $a >_s b$, meaning that s prefers a to b, or the opposite. As is true of economic theory in general, we make the assumption that the voters are "rational" individuals. Firstly, each rational voter must rank all candidates in a transitive manner, so that for any $s \in S$, if $a >_s b$ and $b >_s c$, then $a >_s c$ must hold. Because each voter's preferences are transitive, we can conclude that in making preferences, rational voters define total orderings upon elements of C. Secondly, rational voters must order their candidates without coercion, so that each voter composes his preferences independently of all other voters. Finally, let O_C denote the set of all total orderings of the elements of C.

Define a *voting system* to be any map $V : (O_C)^{|S|} \to O_C$ that takes some $t_1, t_2, \ldots, t_{|S|} \in O_C$ to some $t \in O_C$ (here, $(O_C)^{|S|} = O_C \times \cdots \times O_C$ is the set of all possible combinations, or *voter profiles*, of individual voters' orderings). In other words, a voting system V is a function that is given as input the preferences (the t_i's) of each voter and that returns as output some ordering, or *social ranking*, (t) of the candidates. Whether t is "reasonable" or not is independent of the definition of a voting system; we now consider properties that any "reasonable" voting system should satisfy along with abbreviations that we will use to refer to them throughout the rest of the paper:

1. *Unanimity* (**U**): If for some $a, b \in C$ we have $a >_s b$ for every $s \in S$, then $t = V(t_1, t_2, \ldots, t_{|S|})$ is an ordering that satisfies $a > b$. In other words, the output t of the voting system V, when applied to a society that unanimously prefers a to b, must rank a above b. So, if a society is in consensus about the ranking of a pair of candidates, then the society must choose to rank those two candidates in accordance with their common preference.

2. *Rationality* (**R**): The ranking $t = V(t_1, t_2, \ldots, t_{|S|})$ is a total ordering, so that transitivity is obeyed in the social ordering t of the candidates.

3. *Non-dictatorship* (**ND**): If for some $i \in 1, 2, \ldots, |S|$ and any $a, b \in C$ we have that whenever t_i satisfies $a > b$ and t_j satisfies $b > a$ for $j \neq i$, $t = V(t_1, t_2, \ldots, t_{|S|})$ satisfies $a > b$, then the voter whose ranking is t_i is said to be a *dictator*. A "reasonable" voting system should not have any dictators. Thus, the output t of the voting system should not satisfy the wishes of a

single voter by overriding the wishes of all the other voters. For property ND to make sense, we require *monotonicity*: if $a, b \in C$ such that t has $a < b$, then lowering the relative position of a (with respect to b) in any of the t_i should not alter the relative position of a in t.

4. *Independence of Irrelevant Alternatives* (**IIA**): For $a, b, c \in C$, the ranking between a, b in t should be identical to the corresponding ranking when the position of c in t is changed. That is, if a is ranked higher than b socially, then the position c has in t should be of no relevance whatsoever to the fact that $a > b$ in t.

Note that if there are only two candidates, then we get property IIA for free, because we do not have any third alternatives to be concerned about. In fact, as we discuss in Section 4, Arrow's Theorem does not even apply when $|C| = 2$; in this case, it is actually possible to have a voting system that satisfies all of the above properties.

3. Arrow's Theorem and its Proofs

In this section, we state Arrow's Theorem formally and provide two short proofs, one due to Terence Tao [5] and the other due to John Geanakoplos [2].

Theorem (Arrow's Theorem)**.** *If $|C| \geq 3$, then the properties U, R, ND, and IIA are inconsistent.*

We note that both proofs often utilize $a, b, c \in C$, thereby employing the assumption in the theorem statement that $|C| \geq 3$.

3.1. Tao's Proof

In Tao's original proof of the theorem, it is assumed that there is impartiality among the candidates; i.e., the output of the voting system does not change when the candidates are permuted. Nonetheless, as we show in the following proof, a similar approach can be used to obtain the theorem without the assumption of impartiality among the candidates. Furthermore, Tao's statement of property ND is different from the one presented in Section 2. Specifically, Tao says that a voter who can force his preference for some $a, b \in C$ is a dictator, while we would only consider that voter to be a dictator over a, b. The following proof thus possesses further modifications needed to make it work with our version of property ND. Tao's proof relies on the notion of a "quorum," a set of voters who can force their mutual preference on the social ranking.

For some particular pair $a, b \in C$, define a *quorum* to be a set Q of voters such that if the ranking of every $q \in Q$ satisfies $a >_q b$ and if the ranking of every $s \in S \setminus Q$ satisfies $b >_s a$, then the output ranking of the voting system satisfies $a > b$. This means that the members of a quorum for candidates a, b can always force $a > b$ in the social ranking by unanimously ranking $a > b$ in their individual rankings. We denote the set of all quorums of a, b by $R(a, b)$ (a pair of candidates might have multiple possible quorums). Note that by property U, S is a quorum for any two candidates.

Because of property IIA, whether or not $Q \in R(a,b)$ can force $a > b$ depends only on the individual rankings of a, b for each $q \in Q$. Thus, the notion of a quorum is well-defined; in other words, a quorum cannot force $a > b$ sometimes but fail to force $a > b$ at other times if all its members vote $a > b$. Notice that each pair of candidates has a distinct notion of quorum, for Q need not be a quorum for some other pair of candidates c, d (i.e., $Q \in R(a,b)$ does not imply $Q \in R(c,d)$) if impartiality exists among the candidates.

We are now ready to prove Arrow's Theorem. We first prove that if $Q_1 \in R(a,b)$ and $Q_2 \in R(b,c)$, then $Q_1 \cap Q_2 \in R(a,c)$ for any distinct $a, b, c \in C$; that is, the intersection of a quorum for candidates a, b and a quorum for candidates b, c is a quorum for candidates a, c. Suppose that the following are true:

1. All members of Q_1 vote $a > b$;
2. All members of Q_2 vote $b > c$; and
3. All members of $S \setminus (Q_1 \cap Q_2)$ vote $c > a$.

Clearly, the social ranking must have $a > b$ and $b > c$ (because Q_1 and Q_2 are quorums), and so, using property R, we obtain $a > c$. Thus, $Q_1 \cap Q_2$ is a quorum for a, c, and sticking to our notation, we can write $Q_1 \cap Q_2 \in R(a,c)$.

We next prove that for any $s \in S$, we have $S \setminus \{s\}$ is a quorum for any two candidates. Suppose that this is not the case. Then s is a dictator for at least one pair $a, b \in C$, meaning that s can force his preference between a, b in the social ranking. We claim that $S \setminus \{s\}$ must then be a quorum for a, c, where $c \in C$ satisfies $c \neq b$. This claim is true because s would otherwise be a dictator for a, c, implying that s must be a dictator for all pairs involving a, b, c. Then, by repeating this argument where a is replaced with each of the other elements of C, we see that s must be a dictator for all pairs of candidates, which violates property ND. Since $s \in R(a,b)$ and $S \setminus \{s\} \in R(a,c)$, we have $\{s\} \cap (S \setminus \{s\}) = \varnothing \in R(b,c)$, a result that clearly violates property U, as the empty set cannot possibly be a quorum. Therefore, for any $s \in S$, we have $S \setminus \{s\}$ is a quorum for any two candidates.

If $s_1, s_2, \ldots, s_{|S|}$ are all the elements of S and a, b, c are three elements of C (this is where the $|C| \geq 3$ condition comes in), then $S \setminus \{s_1\} \in R(a,b)$, $S \setminus \{s_i\} \in R(b,c)$ for all $i \neq 1$. So, $\bigcap_{i=1}^{|S|}(S \setminus \{s_i\}) = \varnothing$ is a quorum (it is either in $R(a,b)$ or $R(a,c)$), so the empty set is a quorum, which again violates property U. Thus, the properties U, R, ND, and IIA together are inconsistent, and we have the theorem. □

3.2. Geanakoplos' Proof

In [2], Geanakoplos gave two other short proofs of Arrow's Theorem, but in our view, the one presented below is the simplest of the three.

While Tao's proof uses the properties R, ND, and IIA to contradict property U, Geanakoplos' proof uses the properties U, R, and IIA to contradict property ND. For this proof, we first require the statement and proof of the Extremal Lemma. The main body of theorem's proof relies not only on the Extremal Lemma, but also on the notion of an "extremely pivotal" voter, one who can move a candidate from the very bottom to the very top of a social ranking. Notice that because the set of candidates is finite and because all rankings are total orderings, we can say that the least candidate in a ranking is at the "very bottom" and that the greatest candidate in a ranking is at the "very top."

Extremal Lemma. *For any $b \in C$, if every voter places b at the very top or bottom of his ranking, then the social ranking also has b at the very top or bottom.*

Proof. Suppose the contrary. Then for some $a, c \in C$, we have $a > b$ and $b > c$ in the social ranking, and by property R, $a > c$ must hold. Because b occupies an extremal position in each voter's ranking, we can change the individual rankings of a, c without affecting the position of b in the social ranking. So, we can make each individual ranking satisfy $c > a$. Now by property U, $c > a$ must hold, which is a contradiction. Thus we have the lemma. □

We are now ready to prove Arrow's Theorem. Let a voter $s \in S$ be called "extremely pivotal" for $b \in C$ at some voter profile $\mathcal{P}$ if s induces the social ranking of a, b to change by altering his own ranking of a, b as long as the rest of the rankings are fixed at their positions in $\mathcal{P}$.

We first prove that at some voter profile $\mathcal{P}$ and for some $b \in C$, we can find $s \in S$ such that s is extremely pivotal. Let us construct $\mathcal{P}$ in such a way that all voters rank b at the very bottom of their individual rankings. By property U, b is clearly at the bottom of the social ranking in this case. Now suppose the voters successively move b from the very bottom to the very top of their rankings, and let s be the first voter who, by moving b from the very bottom to the very top of his ranking, causes the position of b in the social ranking to increase. By the Extremal Lemma, s causes b to move to the very top of the social ranking, meaning that for any candidate $a \neq b$, s can alter the social ranking of a, b by altering his own ranking of a, b, keeping all other rankings fixed. So, s is extremely pivotal.

We next prove that s is a dictator for all pairs $a, c \in C$ that satisfy $a \neq b$ and $c \neq b$. Call the voter profile that results just before s moves b from the very bottom to the very top of his own ranking $\mathcal{P}_1$. Just after s moves b from the very bottom to the very top of his own ranking, suppose that no more voters execute this operation, and call the resulting voter profile $\mathcal{P}_2$. Further suppose that s then moves a above b in his own ranking, so that $a >_s b >_s c$, and let the rankings of a, c for all voters in $S \setminus \{s\}$ be arbitrary, keeping b in its extremal position. Call this new voter profile $\mathcal{P}_3$. By property IIA, we can say that (1) $\mathcal{P}_3$ is identical to $\mathcal{P}_1$ with respect to candidates a, b; and (2) $\mathcal{P}_3$ is identical to $\mathcal{P}_2$ with respect to candidates b, c. Therefore, $V(\mathcal{P}_3)$ satisfies $a > b$ and $b > c$, so by R, $V(\mathcal{P}_3)$ satisfies $a > c$ whenever $a >_s c$, showing that s is a dictator for a, c.

We conclude by proving that s is a dictator for all pairs of candidates. To do this, we need only show that s is a dictator for all pairs of candidates involving b, because we have already shown that s is a dictator for all other pairs. Some $d \in S$ is a dictator for all pairs involving b, because we could have repeated the arguments of the previous two paragraphs using some other candidate a instead of b. But because s can change the social ranking of a, b for any $a \in C$ in moving from profile $\mathcal{P}_1$ to profile $\mathcal{P}_2$, we must have that $s = d$, so s is a dictator for a, b, thereby violating property ND. Thus we have the theorem. □

4. Consequences of Arrow's Theorem

Arrow's Theorem unfortunately tells us that voting systems cannot satisfy all "reasonable" criteria at once. Therefore, all of the standard voting systems used to make decisions today must at times violate at least one of the properties U, R, ND, and IIA. We now consider the pitfalls of two common voting systems, *majority rule* and *plurality rule.*

Example 4.1. The Marquis de Condorcet, a French mathematician, was the first to study the fairness of majority rule in depth [3]. In the majority rule voting system, the winning candidate must be preferred by a majority to each other candidate. For example, if $a, b, c \in C$, suppose $a > b > c$ has 44%, $b > c > a$ has 34%, and $c > b > a$ has 22% of the vote. Then b is the winner of the election since he is ranked higher than a by 56% of the voters and since he is ranked higher than c by 78% of the voters, although a has the most first place votes at 44%. To show that majority rule violates one of the properties U, R, ND, and IIA, let us consider *Condorcet's paradox*, described as follows. Suppose for $a, b, c \in C$ the social ranking that 1/3 of the voters rank $a > b > c$, another 1/3 rank $b > c > a$, and the last third rank $c > a > b$. Then a majority prefer a to b, a majority prefer b to c, and a majority prefer a to c, a result that contradicts property R. Note that we do not obtain such paradoxes if $|C| = 2$; i.e., there are two candidates. If there is a majority, then the winner is obvious, and if not, then a random selection of the winner will suffice.

Example 4.2. Plurality rule is the voting system that is most widely used in the United States [3] today; it is used for electing members of congress, and the election of presidents is similar to a collection of plurality rule systems, one for each state. In the plurality rule voting system, the candidate who is ranked first by most voters is the winner. For example, if $a, b, c \in C$, suppose $a > b > c$ has 44%, $b > c > a$ has 34%, and $c > b > a$ has 22% of the vote. Then a is the winner of the election since he has the plurality of 44%. Note that this result is different from the same situation when majority rule was used, as plurality rule gave a as the winner but majority rule gave b as the winner. To show that plurality rule violates one of the properties U, R, ND, and IIA, let us consider the 2000 presidential election in the United States. There were three contestants in the election, Republican George Bush, Democrat Al Gore, and Green Party candidate Ralph Nader. In the case of Florida, which uses plurality rule and was the deciding state, Gore lost to Bush by less than 600 votes [3]. Had Nader not run in the election, it is quite likely that many Green Party supporters would have ended up voting for Gore on account of his more liberal policies. Thus, Gore would have probably taken Florida by storm had Nader, who is known in political terms as a *spoiler*, not run, a result that violates property IIA because the removal of a supposedly irrelevant candidate would have changed the outcome of the election.

Gloomy though the theorem may seem, "Impossibility Theorem" may actually be a misnomer for Arrow's result. While the theorem does state that no voting system satisfies the properties U, R, ND, and IIA, many voting systems actually come quite close to doing so. As we saw in Section 2, majority rule violates property R *if* the voter profile happens to result in a Condorcet paradox. However, if the situations inducing such paradoxes are rare, then we can exclude those situations, and voting systems like majority rule can actually end up satisfying all four properties [3]. For example, let $a, b, c \in C$. If unlikely rankings like, say, $a > b > c$ and $b > a > c$ are excluded, then the only possible rankings left are $a > c > b$, $c > b > a$, $c > a > b$, and $b > c > a$, no three of which form Condorcet paradoxes. Therefore, majority rule, when taken with an understanding of which outcomes are likely and which are not, can actually be "reasonable"; i.e., property R will not be violated, so a winner can be determined. This may very well be why Arrow originally gave his result the optimistic title "General Possibility Theorem" [4].

Acknowledgements

I thank Dr. Daniel Kane of the Stanford University Mathematics Department for enlightening discussions on Arrow's Theorem and for providing valuable feedback on the version of Tao's proof presented in this paper. I also thank Dr. Simon Rubinstein-Salzedo of the Dartmouth College Mathematics Department for his insightful comments on this paper.

References

[1] Kenneth Arrow. *Social Choice and Individual Values.* John Wiley and Sons, Inc., 1951.

[2] John Geanakoplos. Three brief proofs of arrow's impossibility theorem. Cowles Foundation for Research in Economics, Yale University, Paper No. 1116, 2005.

[3] Eric Maskin. The arrow impossibility theorem: Where do we go from here? Arrow Lecture, Columbia University, 2009.

[4] Amartya Sen. Arrow and the impossibility theorem. Arrow Lecture, Columbia University, 2009.

[5] Terence Tao. Arrow's theorem. http://www.math.ucla.edu/ tao/preprints/misc.html.