The Arrow Impossibility Theorem: Where Do We Go From Here?

Eric Maskin*

Institute for Advanced Study, Princeton

Arrow Lecture

Columbia University

December 11, 2009

* I thank Amartya Sen and Joseph Stiglitz for helpful comments on the oral presentation of this lecture. The NSF provided research support.

Giving a lecture in honor of Kenneth Arrow would be a high point for any economist, but there are two additional reasons why this occasion is a special pleasure for me.

First, Ken Arrow was my teacher and Ph.D. advisor, and most likely I would not have become an economist at all, had it not been for him. I was a math major in college and intended to continue in that direction until I happened to take a course of Ken's - - not on social choice theory, but on "information economics." The course was a hodgepodge - - essentially, anything that Ken felt like talking about. And it often seemed as though he decided on what to talk about on his way to the classroom (if then); the lectures had an improvised quality. But they were mesmerizing, and, mainly because of that course, I switched to economics.

Second, lecturing here with Amartya Sen brings back many happy memories for me, because, a number of times at Harvard, he and I taught today's subject—social choice theory—together in a graduate course. It's great to be renewing our pedagogical partnership.

Like Amartya, I will talk about the Arrow impossibility theorem, but I will concentrate on its implications for voting and elections; I will leave aside its broader implications for social welfare.

Now, by its very name, the impossibility theorem engenders a certain degree of pessimism: if something is "impossible," it's pretty hard to accomplish. As applied to voting, the theorem appears to say there is no good election method. Well, I will make the case that this is too strong a conclusion to draw; it's overly negative. But whether or not I persuade you of this, I want to argue that the theorem inspires a natural follow-up question, which oddly was not addressed until quite recently. And I will discuss that question and its answer at the end of this talk.

Let me begin by reviewing the impossibility theorem from the standpoint of elections. If there is a political office to fill, then a *voting rule* is a method of choosing the winner from a set of candidates (this set is called the *ballot*) on the basis of voters' rankings of those candidates.

Many different voting rules have been considered in theory and practice. Probably the most widely used method here in the United States is *plurality rule*, according to which the winner is the candidate who is more voters' *favorite candidate* (i.e., the candidate more voters rank first) than any other. Thus, if there are three candidates *X*, *Y* and *Z*, and 40% of the electorate like *X* best (i.e., 40% rank him first), 35% like *Y* best, and 25% like *Z* best (see table 1), then *X* wins because 40% is bigger than 35% and 25% - - even though it is short of an over-all majority. Plurality rule is the method used to elect Senators and Representatives in the U.S. and Members of Parliment in Britain (where it's called "first-past-the-post.")

40%	35%	25%
X	*Y*	*Z*

Table 1 - - *X* is the plurality winner

Another well-known method is *majority rule*, which the eighteenth-century French mathematician and philosopher Condorcet was the first to analyze in detail. The winner under majority rule is the candidate who is preferred by a majority to each other candidate. For instance, suppose there are again three candidates, *X*, *Y*, and *Z*. 40% of voters rank *X* first, then *Y*, and then *Z* ; 35% rank *Y* first, then *Z*, and then *X* ; and 25% rank *Z* first, then *Y*, and then *X* (see Table 2). Based on these rankings, the majority winner is candidate *Y*, because a majority of voters (35% + 25% = 60%) prefer *Y* to *X*, and a majority (40% + 35% = 75%) prefer *Y* to *Z*.

40%	35%	25%
X	*Y*	*Z*
Y	*Z*	*Y*
Z	*X*	*X*

Table 2 - - *Y* is the majority winner

Notice that plurality rule and majority rule lead to different outcomes: For the voter rankings of Table 2, plurality rule elects candidate *X*, whereas majority rule chooses *Y*. This difference prompts an obvious question: which outcome is "right." Or, put another way, which voting rule is better to use? Indeed, there is no reason to stop with plurality or majority rule: we can ask which among all *possible* voting rules is best.

Arrow provided a framework for answering these questions. He proposed that we should first try to articulate what we *want* out of a voting rule, that is, what properties—or *axioms*—we want it to satisfy. The best voting rule will then be the one(s) that fulfill all those axioms.

Here are the axioms that Arrow considered. As we will see, each is highly desirable on its own, but collectively they lead to impossibility. Because I am particularly concerned with elections, I will give versions that are particularly suited to such contests.

The first is the requirement that an election be *decisive*, i.e., that there always be a winner and that there shouldn't be more than one winner. The second is what an economist would call the *Pareto principle* and what a political theorist might call the *consensus* principle: the idea that if all voters rank candidate *X* above candidate *Y* and *X* is on the ballot (so that *X* is actually available), then we oughtn't elect *Y*. The third axiom is the requirement of *nondictatorship* - - no voter should have the power to always get his way. That is, it should not be the case that if he likes candidate *X* best, then *X* is elected, if he likes candidate *Y* best, then *Y* is elected, and so on. Otherwise, that voter would be a dictator.

The final Arrow axiom is called *independence of irrelevant alternatives*, which in our election context could be renamed "independence of irrelevant candidates." Suppose that, given the voting rule and voters' rankings, candidate *X* ends up the winner of an election. Now look at another situation that is exactly the same except that some other candidate *Y*—who didn't win—is no longer on the ballot. Well, candidate *Y* is, in a sense, "irrelevant;" he didn't win the election in the first place, and so leaving him off the ballot shouldn't make any difference. And so, the independence axiom requires that *X* should still win in this other situation.

I think that, put like this, independence seems pretty reasonable but its most vivid justification, probably comes from actual political history. So, for example, let's recall the U.S. presidential election of 2000. You may remember that in that election everything came down to Florida: if George W. Bush carried the state, he would become president, and the same for Al Gore. Now, Florida—like most other states—uses plurality rule to determine the winner. In the event, Bush got somewhat fewer than six hundred more votes than Gore. Although this was an extraordinary slim margin in view of the nearly six million votes cast, it gave Bush a plurality (and thus the presidency). And, leaving aside the accuracy of the totals themselves (hanging chads and the like), we might reasonably ask whether there was anything wrong with this outcome. But a problem was created by a third candidate in Florida, Ralph Nader. Nearly one hundred thousand Floridians voted for Nader, and it is likely that, had he not been on the ballot, a large majority of these voters would have voted for Gore (of course, some of them might not have voted at all). That means that Gore would probably not only have won, but won quite handily, if Nader had not run.

In political argot, Nader was a *spoiler*. Although he got less than two percent of the vote in Florida—he was clearly "irrelevant" in the sense of having no chance to win himself - - he ended up determining the outcome of the election. That seems highly undemocratic.

The independence axiom serves to rule out spoilers. Thus, because plurality rule was quite spectacularly vulnerable to spoilers, we can immediately conclude that it violates independence. Majority rule, by contrast, is easily seen to satisfy independence: if candidate *X* beats each other candidate by a majority, it continues to do so if one of those other candidates is dropped from the ballot.

Unfortunately, majority rule violates our first axiom, decisiveness - - it doesn't always produce a clear-cut winner (this is a problem that Condorcet himself discussed). To see what can go wrong, consider an election with three candidates *X*, *Y*, and *Z*, and an electorate in which 35% of the population rank *X* first, *Y* second, and *Z* third; 33% rank *Y* first, *Z* second, and *X* third; and 32% rank *Z* first, *X* second, and *Y* third (see table 3).

35%	33%	32%
X	*Y*	*Z*
Y	*Z*	*X*
Z	*X*	*Y*

Table 3 - - Indecisiveness of majority rule

Observe that *Y* beats *Z* by a majority (68% to 32%), and *X* beats *Y* by a majority (67% to 33%). But *Z* beats *X* by a majority (65% to 35%)- - and so there is no candidate who beats each of the other two. This phenomenon is called the Condorcet paradox.

Interestingly, Kenneth Arrow wasn't aware of the Condorcet paradox when he started work on social choice theory. He came across it while studying how firms make choices. In economic textbooks, firms choose production plans to maximize their profit. But in reality, of course, a firm is not typically a unitary decision-maker; it's owned by a group of shareholders. And even if every shareholder wants to maximize profit, different shareholders might have different beliefs about which production plans will accomplish that. So, there has to be a choice method—a voting rule—for selecting the actual production plan.

Ken's first thought was to look at majority rule as the method, but soon discovered—or, rather rediscovered—the Condorcet paradox. Now, he knew that majority rule had been around for a long time, and so assumed that his discovery couldn't possibly be novel. Indeed, when he wrote up the work, he referred to it as the "well-known" paradox of voting. It was only after publication that readers directed him to Condorcet.

Although majority rule violates decisiveness and plurality rule violates independence, Ken felt that surely there must be other voting rules that satisfy all four axioms: decisiveness, consensus, nondictatorship, and independence. But after trying out rule after rule, he eventually came to suspect that these axioms are collectively contradictory. And that's how the impossibility theorem was born; Ken showed that there is *no* voting rule that satisfying all four axioms.

Now, in fact, the nondictatorship axiom is very undemanding. For instance, if instead of one voter, *two* voters out of the entire electorate have all the power in determining the winner, we probably still won't be terribly happy with the election method, but nondictaorship will then be satisfied. The (stronger) condition that we normally want in democratic societies is *equal treatment of voters*, the requirement that all voters count the same. Equal treatment of voters is called *anonymity* in voting theory, reflecting the idea that voters' *names* shouldn't matter; only their *votes* should. Indeed, just as we require that voters be treated equally, we ordinarily do the same for candidates too: we demand *equal treatment of candidates* (called neutrality in the voting theory literature). But because Arrow showed that impossibility results from requiring decisiveness, consensus, independence, and nondictatorship, we get impossibility *a fortiori* from imposing the more demanding set of axioms: decisiveness, consensus, independence, equal treatment of voters, and equal treatment of candidates.

The impossibility theorem has been the source of much gloom because, individually, each of these five axioms seems so compelling. But, as I suggested in my opening remarks, there is a sense in which the theorem overstates the negative case. Specifically, it insists that a voting rule satisfy the five axioms *whatever* voters' rankings turn out to be. Yet, in practice, some rankings may not be terribly likely to occur. And if that's the case, then perhaps we shouldn't worry too much if the voting rule fails to satisfy all the axioms for those improbable rankings.

For an example, let's go back to the U.S. presidential election of 2000. The three candidates of note were Bush, Gore, and Nader. Now, many people ranked Bush first. But the available evidence suggests that few of these voters ranked Nader second. Similarly, a small but significant fraction of voters placed Nader first. But Nader aficionados were very unlikely to rank Bush second.

Indeed, there is a good reason why the rankings

Bush		Nader
Nader	or	Bush
Gore		Gore

appeared to be so rare. In ideological terms, Nader was the left-wing candidate, Bush was the right-wing candidate, and Gore was somewhere in between. So, if you liked Bush's proposed policies, you were likely to revile Nader's, and vice versa.

Yet, if we can rule out the two rankings above (or, at least, assign them low enough probability), then it turns out that majority rule *is* decisive - - it always results in a clear-cut winner. That is, majority rule satisfies all five axioms—decisiveness, consensus, no spoilers, and the two equal treatment properties—when rankings are *restricted* to rule out the rankings Bush/Nader/Gore and Nader/Bush/Gore.

That's the sense in which the impossibility theorem is too gloomy: if rankings are restricted in an arguably plausible way, then the five axioms are no longer collectively inconsistent. But regardless of whether you accept the plausibility of this particular restriction, the impossibility theorem prompts a natural follow-up question: Given that no voting rule satisfies the five axioms all the time, which rule satisfies them *most often*? In other words, if we can't achieve the ideal, which voting rule gets us closest to that ideal and maximizes the chance that the properties we want are satisfied?

Perhaps, surprisingly, this question wasn't posed in the literature until many years after the publication of *Social Choice and Individual Values.*[1] In an effort to provide an answer, let me say that a voting rule *works well* if, for a particular restricted class of rankings, it satisfies the five axioms whenever voters' rankings adhere to the restriction. So, for example, majority rule works well in the U.S. presidential election example if rankings are restricted to exclude the two rankings Bush/Nader/Gore and

[1] See E. Maskin, "Majority Rule, Social Welfare Functions, and Games Forms," in K. Basu, P. Pattanaik, and K. Suzumura (eds.), *Choice, Welfare, and Development* (essays in honor of Amartya Sen), Oxford University Press, 1995, pp.100-109 , and P. Dasgupta and E. Maskin, "On the Robustness of Majority Rule," *Journal of the European Economic Association*, 2008, pp. 949-973.

Nader/Bush/Gore. The goal then becomes to find the voting rule that works well for as many different restricted classes of rankings as possible.

It turns out that there is a sharp answer to this problem, provided by a "domination theorem."[2] The theorem can be expressed as follows. Take any voting rule that differs from majority rule, and suppose that it works well for a particular class of rankings. Then, majority rule must also work well for that class. Furthermore, there must be some other class of rankings for which majority rule works well and the voting method we started with does not. In other words, majority rule dominates every other voter rule: whenever another voting rule works well, majority rule must work well too, and there will be cases where majority rule works well and the other voting rule does not.

I noted before that Kenneth Arrow himself began with majority rule when he set off on his examination of social choice theory. He was soon led to consider many other possible voting rules too. But it turns out that, using the criteria he laid out, there's a sense in which we can't do better than majority rule after all.

[2] See the Dasgupta-Maskin article.