# THE ARROW IMPOSSIBILITY THEOREM: WHERE DO WE GO FROM HERE?[1]

ERIC MASKIN

Giving a lecture in honor of Kenneth Arrow would be a high point for any economist, but, in my case, there were two additional reasons why the occasion of the second annual Arrow lecture was such a special pleasure.

First, Ken Arrow was my teacher and PhD 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 with Amartya Sen brought back many happy memories for me, because he and I taught this book's subject—social choice theory—together several times in a graduate course at Harvard. It's great to renew our pedagogical partnership.

Like Amartya, my essay is on 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 do. As applied to voting, the theorem appears to say that 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.[2] Thus, if there are three candidates *X*, *Y*, and *Z*, and 40 percent of the electorate like *X* best (i.e., 40 percent rank him first), 35 percent like *Y* best, and 25 percent like *Z* best (see table 1), then *X* wins because 40 percent is bigger than 35 percent and 25 percent—even though it is short of an overall majority. Plurality rule is the method used to





Table 1 *X* is the plurality winner

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

elect senators and representatives in the United States and members of parliament in Britain (where it's called "first-past-the-post.")

Another well-known method is *simple majority rule*,[3] which the eighteenth-century French mathematician and philosopher Condorcet was the first to analyze in detail.[4] 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*. Forty percent of voters rank *X* first, then *Y*, and then *Z*; 35 percent rank *Y* first, then *Z*, and then *X*; and 25 percent 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 percent + 25 percent = 60 percent prefer *Y* to *X*, and a majority (40 percent + 35 percent = 75 percent) prefer *Y* to *Z*.

Table 2 *Y* is the plurality winner

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

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 suggest reformulations of the axioms that are particularly suited to such contests.

The first is the requirement that an election be *decisive*, i.e., that, whatever voters' rankings turn out to be, there should always be a winner and there shouldn't be more than one winner.[5] 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 ought not to 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 necessarily elected, regardless of how others feel about *X*. 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.[6]

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[7]—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 extraordinarily slim margin in view of the nearly 6 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 the answer to this question is made more complicated by the presence of 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 2 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.[8]

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 percent of the population rank *X* first, *Y* second, and *Z* third; 33 percent rank *Y* first, *Z* second, and *X* third; and 32 percent rank *Z* first, *X* second, and *Y* third (see table 3).

Table 3 *Y* is the plurality winner

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





Observe that *Y* beats *Z* by a majority (68 percent to 32 percent), and *X* beats *Y* by a majority (67 percent to 33 percent). But *Z* beats *X* by a majority (65 percent to 35 percent)—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 Condorcet's paradox when he started work on social choice theory. He rediscovered it while studying how firms might make choices. In economics 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 it up, 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 satisfies all four axioms.[9]

Now, 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, even though *nondictatorship* will then formally be satisfied. Democratic societies usually insist on the stronger condition *equal treatment of voters*, the requirement that all voters have the same weight. 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 (i.e., for an unrestricted domain of rankings). 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/Gore (Bush ranked above Nader and Nader ranked above Gore) or Nader/Bush/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 the Condorcet paradox cannot occur, and majority rule is decisive after all—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 the six logically possible rankings of Gore, Bush, and Nader 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 doesn't seem to have been formally posed in the literature until many years after the publication of *Social Choice and Individual Values*.[10] (Note that a paper I did with Partha Dasgupta that addresses this question is reprinted in part 2 of this book.) In an effort to provide an answer, let me define 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 Nader/Bush/Gore.[11] The question then becomes: What is 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."[12] 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.[13]





As mentioned in footnote 2, I have been assuming that voters report their rankings *sincerely*—that there is no difference between their expressed rankings and their true rankings. But for many voting systems, strategic voting (ranking A above B even if you prefer B to A) may at times be advantageous. Nevertheless, the domination theorem I stated goes through unchanged if we add an additional axiom, *strategy-proofness* (the requirement that voters should find it in their interest to vote sincerely), to the list. Indeed, if majority rule satisfies decisiveness for a class of rankings, it also satisfies strategy-proofness for that same class.

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 is a sense in which we can't do better than majority rule after all.

## NOTES

1. I thank Amartya Sen and Joseph Stiglitz for helpful comments on the oral presentation of this lecture. The NSF provided research support.
2. Here I make no distinction between a voter's *expressed* ranking of the candidates and his *actual* ranking. In other words, I suppose voters vote "sincerely": If a voter says that candidate X is his favorite, then X really is his favorite. Still, the possibility that

voters might vote differently from their actual rankings—i.e., vote strategically—is an interesting and realistic possibility that I will return to at the end of the lecture.

3. There are many variants of majority rule, but here I will consider only Condorcet's version. I shall, therefore, omit the word *simple* from now on.
4. See M. J. Condorcet, *Essai sur l'application de l'analyse à la pluralité des voix* (Imprimerie Royale, 1785).
5. In Arrow's own framework, decisiveness is formulated as *transitivity*, the requirement that if, given voters' rankings, candidate *X* is elected over candidate *Y* and *Y* is elected over *Z*, then *X* should be elected over *Z*. (Note that if transitivity is not satisfied—so *Z* is elected over *X*—and *X*, *Y*, and *Z* are all on the ballot, then *none* of them is elected, a failure of decisiveness.)
6. There are two closely related conditions that go under the name "independence of irrelevant alternatives": Arrow's axiom and the condition formulated by J. Nash ("The Bargaining Problem," *Econometrica* 18(2), 1950: 155–162). Here I am using the Nash formulation, because it is somewhat more convenient for my purposes.
7. All other states, in fact, except, currently, for Maine and Nebraska.
8. I have singled out the election of 2000. But the same problem has recurred many times in U.S. presidential election history. For example, Ross Perot may well have spoiled the 1992 election for George H. W. Bush, enabling Bill Clinton to win.
9. In fact, Arrow also imposed a fifth axiom, *unrestricted domain of preferences*: the requirement that a voting role be defined regardless of what voters' preferences turn out to be. This axiom is implicit in the way I have formulated the other axioms.
10. See E. Maskin, "Majority Rule, Social Welfare Functions, and Games Forms," in *Choice, Welfare, and Development* (Essays in Honor of Amartya Sen), eds. K. Basu, P. Pattanaik, and K. Suzumura (Oxford: Oxford University Press, 1995), 100–109; and P. Dasgupta and E. Maskin, "On the Robustness of Majority Rule," *Journal of the European Economic Association*, 2008: 6: 949–973.





11. But that is not the only case in which majority rule works well—excluding Gore/Nader/Bush and Bush/Nader/Gore, for example, would also do.
12. See the Dasgupta/Maskin article.
13. Of course, if we are to use majority rule in practice, we must have a tie-breaking rule in place in case the rankings of the Condorcet paradox arise—even if they are unlikely to do so. One possibility is to use plurality rule to break the tie. So, to return to the example of table 3—where candidates *X*, *Y*, and *Z* are essentially tied according to majority rule—we would then choose *X*, who enjoys the highest plurality, as the winner.