ARROW AND THE IMPOSSIBILITY THEOREM[1]

AMARTYA SEN

I

It was wonderful for me to have the opportunity to pay tribute to Kenneth Arrow, who is not only one of the greatest economists of our time but also one of the finest thinkers of our era. That itself made the occasion of the second annual Arrow lecture very special for me, but on top of that, it was marvelous to have the company of Eric Maskin, with whom I used to teach a most enjoyable joint course on social choice theory at Harvard, until he deserted us for the Institute for Advanced Study at Princeton.[2] And it was very pleasing for me to have Joe Stiglitz as the *participating* chair of the meeting (having known Joe for many years, I can assure you that there was no danger of Joe being an aloof chair) and to know that Akeel Bilgrami's intellectual vision was behind the planning of this event. I was in admirable company at the lecture and want to begin this piece by expressing my appreciation of that, but most especially by thanking Ken Arrow himself, for making us all think in new lines, and personally for me, for being such a major influence on my own intellectual life.

I shall be particularly concerned in this essay with Arrow's pathbreaking "impossibility theorem," for which Arrow managed to find, in line with his sunny temperament, a rather cheerful name: "General Possibility Theorem."[3] This result, and with it the formulation of the demands of mathematical social choice theory, were real watersheds in the history of welfare economics as well as of voting theory and collective choice.

The informational foundation of modern social choice theory relates to the basic democratic conviction that social judgments and public decisions must depend, in some transparent way, on individual preferences, broadly understood. (I have investigated the implications of this perspective in social choice theory in my paper "The Informational Basis of Social Choice," which is reprinted in part 2 of this book.)[4] The emergence of this democratic instinct relates closely to the ideas and events that surrounded the European Enlightenment. Even though the pursuit of democratic social arrangements drew also on various earlier sources and inspirations, it received a definitive delineation and massive public acknowledgment only during the Enlightenment, particularly—but not exclusively—during the second half of the eighteenth century, which also saw the French Revolution and American independence.

What can be called "preferences" of persons can, of course, be variously interpreted in different democratic exercises, and the differences are well illustrated by the contrasts between (1) focusing on votes or ballots (explored in the classic works of Borda and Condorcet), (2) concentrating on the interests of individuals (explored, in different ways, in the pioneering writings of David Hume, Jeremy Bentham, and John Stuart Mill), and (3) drawing on the diverse judgments and moral sentiments of individuals about societies and collectivities (explored by Adam Smith and Immanuel Kant, among many others over the centuries). These contrasts, between alternative interpretations of preferences, can be very important for some purposes. I shall visit that territory before long. However, for the moment I shall use the generic term *preference* to cover all these different interpretations of individual concerns that could be invoked, in one way or another, to serve as the informational bases of public decisions and of social judgments.[5]

In contemporary social choice theory, pioneered by Kenneth Arrow, this democratic value is absolutely central, and the discipline has continued to be loyal to this basic informational presumption. For example, when an axiomatic structure yields the existence of a dictator as a joint implication of chosen axioms that seemed plausible enough (on this more presently), this is immediately understood as something of a major embarrassment for that set of axioms, rather than being taken to be just fine on the ground that it is a logical corollary of axioms that have been already accepted and endorsed. We cannot begin to understand the intellectual challenge involved in Arrow's impossibility theorem without coming to grips with the focus on informational inclusiveness that goes with a democratic commitment, which is deeply offended by a dictatorial procedure. This is so, even when the dictatorial result is entailed by axiomatic requirements that seem reasonable, taking each axiom on its own.

So let me begin by discussing what Arrow's impossibility theorem asserts and how it is established. The theorem has the reputation of being "formidable," which is a good description of its deeply surprising nature as well as of its vast reach, but the air of distanced respect is not particularly helpful in encouraging people to try to understand how the result emerges. It is, however, important for people interested in political science, in welfare economics, or in public policy to understand the analytical foundations of Arrow's far-reaching result, and there is no reason it should be seen as a very difficult result to comprehend and appreciate. A closer understanding is also relevant for seeing what its implications are and what alleged implications, often attributed to it, may be misleading.

The proof of the Arrow theorem I shall present follows Arrow's own line of reasoning, but through some emendations that make it agreeably short and rather easy to follow. However, it is a completely elementary proof, using nothing other than basic logic, like Arrow's own.[6] The important issue here is not just the shortness of getting to the Arrow theorem but the ease with which it can be followed by anyone without any technical reasoning or any particular knowledge of mathematics or advanced mathematical logic. So I have spelled out fully the reasoning behind each step (perhaps too elaborately for some who are very familiar with this type of logical reasoning).

II

The basic engagement of social choice with which Arrow was concerned involved evaluating and choosing from the set of available social states ($x, y, \ldots$), with each x, y, etc., describing what is happening to the individuals and the society in the respective states of affairs. Arrow was concerned with arriving at an aggregate "social ranking" R defined over the set of potentially available social states x, y, etc. With his democratic commitment, the basis of the social ranking R is taken to be the collection of individual rankings $\{R_i\}$, with any R_i standing for person *i*'s preference ranking over the alternative social states open for social choice. It is this functional relation that Kenneth Arrow calls the "social welfare function." Given any set of individual preferences, the social welfare function determines a particular aggregate social ranking R.

That there could be problems of consistency in voting rules was demonstrated by the Marquis de Condorcet in the eighteenth century. It is useful to recollect how the problem comes about, for example, for the method of majority of decision.

Take three persons 1, 2, and 3 with the following preferences over three alternatives x, y, and z.

1	2	3
x	y	z
y	z	x
z	x	y

In majority decisions, x defeats y, which defeats z, which in turn defeats x. The R generated by majority rule violates transitivity and even weaker conditions of consistency than that (such as acyclicity). And since each alternative is defeated by another available alternative in the available set, there is no majority winner—no "choice set"—for the available set $\{x, y, z\}$.

Majority rule is of course a very special rule, though highly appealing. Arrow's theorem, among other things, generalizes the problem for any voting rule, and indeed it does much more than that (as I shall presently discuss).

Consider now the following set of axioms, which are motivated by Arrow's original axioms but are in fact somewhat simpler—and also somewhat less demanding—which, taken together, are nevertheless adequate for the impossibility theorem.

- *U* (*unrestricted domain*): For any logically possible set of individual preferences, there is a social ordering *R*.
- *I* (*Independence of irrelevant alternatives*): The social ranking of any pair $\{x, y\}$ will depend only on the individual rankings of *x* and *y*.
- *P* (*Pareto Principle*): If everyone prefers any *x* to any *y*, then *x* is socially preferred to *y*.
- *D* (*Nondictatorship*): There is no person *i* such that whenever this person prefers any *x* to any *y*, then socially *x* is preferred to *y*, no matter what others prefer.

The General Possibility Theorem: If there are at least three distinct social states and a finite number of individuals, then no social welfare function can satisfy *U*, *I*, *D*, and *P*.

One common way of putting this result is that a social welfare function that satisfies unrestricted domain, independence, and Pareto Principle has to be dictatorial. This is a repugnant conclusion—antithetical to the democratic commitment—emanating from a collection of reasonable-looking axioms.

III

In proving this theorem we can go by two intermediate results ("lemmas," if you want to sound "high-tech"). Crucial to this line of reasoning is the idea of a set of individuals, *G*, being "decisive." This could be "local" in the sense of applying over some particular pair of alternatives {*x, y*}, when *x* must be invariably socially preferred to *y*, whenever all individuals in *G* prefer *x* to *y (no matter how the individuals not in G rank x* and *y*). Much more demandingly, *G* is globally "decisive," if it is locally decisive over every pair. The first intermediate result is about what we can understand as the spread of decisiveness from one pair of alternatives to all others. Central to the strategy of demonstration used here is to postulate only a specific partial ordering for individuals not in *G*, over particular pairs (allowed by Unrestricted Domain), and the reasoning works for every complete ordering of individuals compatible with that partial ordering.

SPREAD OF DECISIVENESS

If G is decisive over any pair {x, y}, then G is [globally] decisive.

Proof: Take any pair {*a, b*} different from {*x, y*}, and assume that everyone, *without exception,* prefers *a* to *x*, and *y* to *b.*

For those who do not belong to G, we do not impose any other condition on the rest of their preferences (in particular they can rank a and b in any possible way). But we assume that all members of group G also prefer x over y; that is, they subscribe to the descending order: a, x, y, b. By the Pareto Principle, a is socially preferred to x, and y is socially preferred to b. By the decisiveness of G over $\{x, y\}$, x is socially preferred to y. Putting them together (that is, a preferred to x, that to y, and that to b), we have, by transitivity of strict preference, the result that a is socially preferred to b. By the Independence of Irrelevant Alternatives (condition I), this must relate only to individual preferences over $\{a, b\}$. But only the preferences of individuals in G have been specified (they rank a above b); all others can rank a and b in any way they would like. So G is also decisive over the pair $\{a, b\}$, and not just over $\{x, y\}$. And this applies to all pairs $\{a, b\}$ distinct from $\{x, y\}$. So G is indeed [globally] decisive.[7]

CONTRACTION OF DECISIVE SETS

If a set G of individuals is decisive (and if it has more than one individual), then some reduced part (a "proper subset") of G is decisive as well.

Proof: Partition G into two subsets G_1 and G_2. Let everyone in G_1 prefer x to y, and x to z, with the ranking of y and z unspecified, and let everyone in G_2 prefer x to y, and z to y. Others not in G can have any set of preferences. By the decisiveness of G, we have x socially preferred to y. If now z is taken to be socially at least as good as x for some configuration of individual preferences over $\{z, x\}$, then we must have z socially preferred to y (since x is socially preferred to y) for that configuration of preferences over $\{z, x\}$. Since no one's preference over $\{z, y\}$ other than those in G_2 has been specified,[8] and those in G_2 prefer z to y, G_2 is decisive over $\{z, y\}$, and thus, by the Spread of Decisiveness, G_2 must be [globally] decisive. Since that shows that some reduced part of G is indeed decisive, then we have got what we want to show, for that particular case. To avoid this possibility, we must assume that our initial presupposition that z is at least as good as x must be eschewed. But then x must be preferred to z. However, since no one's preference over $\{x, z\}$ has been specified for this, other than those in G_1 who prefer x to z, clearly G_1 is decisive over $\{x, z\}$. Thus by the Spread of Decisiveness, G_1 is [globally] decisive. So either G_1 or G_2 must be [globally] decisive, which establishes the Contraction of Decisive Sets.

Now, Arrow's impossibility result.

PROOF OF THE GENERAL POSSIBILITY THEOREM

By the Pareto Principle the set of all individuals is decisive. By the Contraction of Decisive Sets, some proper subset of all individuals must also be decisive. Take that smaller decisive set, but some proper subset of that smaller set must also be decisive. And so on. Since the set of individuals is finite, we shall arrive, sooner or later, at one individual who is decisive. But that violates the nondictatorship condition—hence the impossibility.

IV

I end with a few observations on the nature of Arrow's result. First, the combination of unrestricted domain, independence, and the Pareto Principle—each of which is individually somewhat innocuous—seems to produce both the spread of decisiveness and the contractibility of decisive sets. One lesson to draw is that it is hard to judge the plausibility of axioms unless we also consider with what other axioms they are to be harnessed together.

Second, even though the basic axioms for the Arrow theorem concentrate on individual preferences over the set of alternative social states, it is not directly presumed by any of the axioms that we are not allowed to take note of the nature of the alternatives and their relations to the persons involved in arriving at the social ranking R. That ruling out of information about the nature of the alternatives results from the combination of the three axioms—U, I, and P—that together entail Spread of Decisiveness, ruling out the relevance of the nature of the alternatives involved in the choice, and of the comparative predicaments of different individuals in different states of affairs. This is what causes the permitted social welfare functions to be confined to the class of voting rules. So, in this sense, it is wrong to think of the Arrow result as merely extending the Condorcet paradox to all voting rules. It first establishes that the permitted social welfare functions must be voting rules (that is the big intermediate result), and *then* generalizes the Condorcet paradox.

Third, the result can be easily extended to social choice reasoning that avoids talking about any social relation *R* and concentrates instead on what can be chosen from particular sets of alternatives. This is called, in the literature, "social choice functions." The ultimate choice-functional extension is to impose no internal consistency conditions on social choice functions at all, except those that come from the relation between individual preferences and the social choice function. To give an example, the choice-functional version of the Pareto Principle would be: If everyone prefers *x* to *y*, then *y* must not be chosen if *x* is available to be chosen. The Arrow theorem can indeed be extended to choice functions as well, without any internal consistency condition, as I have shown elsewhere (but, I fear, that proof is too complicated to be aired here!).[9]

Fourth, the individual preferences are just the orderings of individuals considered separately, without any interpersonal comparisons. Once interpersonal comparisons are allowed, various positive possibilities open up. This is important for welfare economics and for judgments of equity and aggregate welfare with which welfare economists are rightly concerned. Consider a nasty proposal to take some of the income of the poorest person and divide it over several others. In a society of selfish persons, this unprepossessing proposal will be a majority improvement. So the problem here is not the lack of consistency of majority rule or any other voting rule. It is that we are in wrong territory by concentrating only individual preference orderings, and then—with the help of combining *U*, *I*, and

P—getting to the Spread of Decisiveness. It is the wrong informational base for many welfare economic concerns, and it is perhaps all to the good that the majority rule is also—in addition to being obtuse—inconsistent.

Fifth, the way to tackle the Arrow theorem in the context of welfare economics certainly includes making use of interpersonal comparisons in our judgments. Indeed, all public policy tends to bring in interpersonal comparisons in one way or another. But that route is not easily available when we are dealing with a voting process, such as elections of candidates for, say, political positions. For the political exercise and voting theory, we have to think in different lines, as indeed Eric Maskin does, with his usual skill and elegance, in his own lecture here.[10]

Finally, even for political processes, one problem that remains does not involve interpersonal comparisons but for which the Arrovian axioms are inadequate. This is the problem of liberty and rights. If we follow John Stuart Mill in standing up for the rights of minorities and of individuals in their personal domain, then we must not be too impressed by how many people oppose the minorities being able to choose their own life styles or how many try to eliminate the exercise of personal liberties by individuals. The importance of liberty may demand going against the objections of numerous "busybodies." Not to be able to accommodate liberty is a political limitation of majority rule. It is not accommodated in Arrow's axioms either.

All these problems, and many more, open up as we follow—and follow up on—Arrow's pioneering result, including his demonstration that serious problems can arise with axiomatic combinations even when each individual axiom, seen on its own, seems just fine. That was indeed a game-changing contribution, which affected a large variety of disciplines. It is difficult to think that Arrow was just a graduate student when he transformed the intellectual world of social thought.

NOTES

1. I am grateful for the comments on my presentation at the meeting at Columbia, including those from Eric Maskin and Joseph Stiglitz.
2. Happily, Maskin rejoined the faculty at Harvard in 2012.
3. Kenneth J. Arrow, *Social Choice and Individual Values* (New York: Wiley, 1951; republished in an extended form, 1963).
4. Amartya Sen, "The Informational Basis of Social Choice," in *The Handbook of Social Choice and Welfare,* vol 2, eds. Kenneth Arrow, Amartya Sen, and Kotaro Suzumura (Amsterdam: Elsevier, 2010).
5. These distinctions and their far-reaching implications were investigated in my essay "Social Choice Theory: A Re-examination," *Econometrica* 45 (1977); reprinted in *Choice, Welfare and Measurement* (Oxford: Blackwell, and Cambridge, Mass.: MIT Press, 1982; republished, Cambridge, Mass.: Harvard University Press, 1997), and further in my Presidential Address to the American Economic Association: "Rationality and Social Choice," *American Economic Review* 85 (1995); reprinted in *Rationality and Freedom* (Cambridge, Mass.: Harvard University Press, 2002).
6. Providing short proofs of Arrow's theorem is something of a recurrent exercise in social choice theory, and one must not make a cult of it, since all the proofs draw in one way or another on Arrow's trail-blazing insight. One has to be careful, however, not to be too convinced by the alleged "shortness" of some proofs when they draw on other mathematical results (as some of the

proofs do) that are invoked and used but not established in the course of the proof. Arrow's proof was entirely elementary and did not presume anything other than the basic rules of logic, which is the route followed here. I presented an earlier version of this proof in footnotes 9 and 10, in "Rationality and Social Choice" (1995); reprinted in *Rationality and Freedom* (2002), p. 267.

7. We are cutting a corner here by assuming that a, b, x, and y are all distinct social states. The reasoning is exactly similar when two of them are the same alternative.
8. Note that this lack of specification of individual preferences over $\{z, y\}$ is consistent with what has been assumed about individual preferences over $\{z, x\}$.
9. See my Presidential Address to the Econometric Society, "Internal Consistency of Choice," *Econometrica* 61 (1993); reprinted in *Rationality and Freedom* (2002).
10. Eric Maskin, "The Arrow Impossibility Theorem: Where Do We Go from Here?" in this volume.