Arrow’s impossibility theorem

Bibliography

Arrow’s impossibility theorem
Kenneth J. Arrow (b. 1921, Nobel Prize in Economics 1972) is the author of this celebrated result which first appeared in Chapter V of Social Choice and Individual Values (1951). Paradoxically, Arrow called it initially the ‘general possibility theorem’, but it is always referred to as an impossibility theorem, given its essentially negative character. The theorem establishes the incompatibility among several axioms that might be satisfied (or not) by methods to aggregate individual preferences into social preferences. I will express it in formal terms, and will then comment on its interpretations and on its impact in the development of economics and other disciplines.

In fact, the best known and most reproduced version of the theorem is not the one in the original version, but the one that Arrow formulated in Chapter VIII of the 1963 second edition of Social Choice and Individual Values. This chapter, entitled ‘Notes on the theory of social choice’, was added to the original text and constitutes the only change between the two editions. The reformulation of the theorem was partly justified by the simplicity of the new version, and also because Julian Blau (1957) had pointed out that there was a difficulty with the expression of the original result.

Both formulations start from the same formal framework. Consider a society of agents, which has to express preferences regarding the alternatives in a set A. The preferences of agents are given by complete, reflexive, transitive binary relations on A. Each list of $n$ such relations can be interpreted as the expression of a state of opinion within society. Rules that assign a complete, reflexive, transitive binary relation (a social preference) to each admissible state of opinion are called ‘social welfare functions’.

Specifically, Arrow proposes a list of properties, in the form of axioms, and discusses whether or not they may be satisfied by a social welfare function. In his 1963 edition, he puts forward the following axioms:

- Universal domain ($U$): the domain of the function must include all possible combinations of individual preferences;
- Pareto ($P$): whenever all agents agree that an alternative $x$ is better than another alternative $y$, at a given state of opinion, then the corresponding social preference must rank $x$ as better than $y$;
- Independence of irrelevant alternatives ($I$): the social ordering of any two alternatives, for any state of opinion, must only depend on the ordering of these two alternatives by individuals;
- Non-dictatorship ($D$): no single agent must be able to determine the strict social preference at all states of opinion.

Arrow’s impossibility theorem (1963) tells that, when society faces three or more alternatives, no social welfare function can simultaneously meet $U$, $P$, $I$ and $D$.

By Arrow’s own account, the need to formulate a result in this vein arose when trying to answer a candid question, posed by a researcher at RAND Corporation: does it make sense to speak about social preferences?
A first quick answer would be to say that the preferences of society are those of the majority of its members. But this is not good enough, since the majority relation generated by a society of \( n \) voters may be cyclical, as soon as there are more than two alternatives, and thus different from individual preferences, which are usually assumed to be transitive. The majority rule (which otherwise satisfies all of Arrow's requirements), is not a social welfare function, when society faces more than two alternatives. Arrow's theorem generalizes this remark to any other rule: no social welfare function can meet his requirements, and no aggregation method meeting them can be a social welfare function.

Indeed, some of the essential assumptions underlying the theorem are not explicitly stated as axioms. For example, the required transitivity of the social preference, which rules out the majority method, is included in the very definition of a social welfare function. Testing the robustness of Arrow's theorem to alternative versions of its implicit and explicit conditions has been a major activity of social choice theory for more than half a century. Kelly's updated bibliography contains thousands of references inspired by Arrow's impossibility theorem.

The impact of the theorem is due to the richness and variety of its possible interpretations, and the consequences it has on each of its possible readings.

A first interpretation of Arrow's formal framework is as a representation of voting methods. Though he was not fully aware of it in 1951, Arrow's analysis of voting systems falls within a centuries-old tradition of authors who discussed the properties of voting systems, including Pliny the Young, Ramón Lull, Borda, Condorcet, Laplace and Dodgson, among others. Arrow added historical notes on some of these authors in his 1963 edition, and the interested reader can find more details on this tradition in McLean and Urken (1995). Each of these authors studied and proposed different methods of voting, but none of them fully acknowledged the pervasive barriers that are so well expressed by Arrow's theorem: that no method at all can be perfect, because any possible one must violate some of the reasonable requirements imposed by the impossibility theorem. This changes the perspective in voting theory; if a voting method must be selected over others, it must be on the merits and its defects, taken together, none can be presented as an ideal.

Another important reading of Arrow's theorem is the object of Chapter IV in his monograph. Arrow's framework allows us to put into perspective the debate among economists of the first part of the twentieth century, regarding the possibility of a theory of economic welfare that would be devoid of interpersonal comparisons of utility and of any interpretation of utility as a cardinal magnitude, Kaldor, Hicks, Seitzovsky, Bergson and Samuelson, among other great economists of the period, were involved in a discussion regarding this possibility, while using conventional tools of economic analysis. Arrow provided a general framework within which he could identify the shared values of these economists as partial requirements on the characteristics of a method to aggregate individual preferences into social orderings. By showing the impossibility of meeting all these requirements simultaneously, Arrow's theorem provided a new focus to the controversies: no one was closer to success than anyone else. Everyone was looking for the impossible. No perfect aggregation method was worth looking for, as it did not exist. Trade-offs between the properties of possible methods had to be the main concern.

Arrow's theorem received immediate attention, both as a methodological criticism of the 'new welfare economics' and because of its voting theory interpretation. But not everyone accepted that it was relevant. In
particular, the condition of independence of irrelevant alternatives was not easily accepted as expressing the desiderata of the new welfare economics. Even now, it is a debated axiom. Yet Arrow's theorem has shown a remarkable robustness over more than 50 years, and has been a paradigm for many other results regarding the general difficulties in aggregating preferences, and the importance of concentrating on trade-offs, rather than setting absolute standards.

Arrow left some interesting topics out of his monograph, including issues of aggregation and mechanism design. He mentioned, but did not elaborate on, the possibility that voters might strategically misrepresent their preferences. He did not discuss the reasons why some alternatives are on the table, and others are not, at the time a social decision must be taken. He did not provide a general framework where the possibility of using cardinal information and of performing interpersonal comparisons of utility could be explicitly discussed. These were routes that later authors were to take. But his impossibility theorem, in all its specificity, provided a new way to analyze normative issues and established a research program for generations.

Salvador Barbera

Bibliography


See also: Bergson's social indifference curve, Bordley's rule, Chipman-Moore-Samuelson compensation criterion, Condorcet's criterion, Hicks compensation criterion, Kaldor compensation criterion, Schotter's compensation criterion.

Arrow's learning by doing

This is the key concept in the model developed by Kenneth J. Arrow (b. 1921, Nobel Prize 1972) in 1962 with the purpose of explaining the changes in technological knowledge which underlie intertemporal and international shifts in production functions. In this respect, Arrow suggests that, according to many psychologists, the acquisition of knowledge, what is usually termed 'learning', is the product of experience ('doing'). More specifically, he advances the hypothesis that technical change depends upon experience in the activity of production, which he approaches by cumulative gross investment, assuming that new capital goods are better than old ones; that is to say, if we compare a unit of capital goods produced in the time $t_1$ with one produced at time $t_2$, the first requires the cooperation of at least as much labour as the second, and produces no more product. Capital equipment comes in units of equal (infinitesimal) size, and the productivity achievable using any unit of equipment depends on how much investment had already occurred when this particular unit was produced.

Arrow's view is, therefore, that at least part of technological progress does not depend on the passage of time as such, but grows out of 'experience' caught by cumulative gross investment, that is, a vehicle for improvements in skill and technical knowledge. His model may be considered as a precursor to the further new or endogenous growth theory. Thus the last paragraph of Arrow's paper reads as follows: 'It has been assumed that learning takes place only as a by-product of ordinary production. In fact, society has created institutions, education and research, whose purpose is to enable learning to take place more rapidly. A fuller model would take account of these as additional variables.' Indeed, this is precisely what more recent growth literature has been doing.

Carmela Martín