The Making of Arrow's Impossibility Theorem

Adam Hamdan

December, 2021

Abstract

This paper contextualize Arrow's Impossibility Theorem within the history of social choice and welfare economics while briefly discussing its impact on the literature.

Keywords: arrow's impossibility theorem, social choice, welfare economics, voting systems.

Introduction

The field of social choice refers to the dedicated study of collective decision making and the aggregation of societal preferences. Since its formal inception in the 18th century, its theoretical richness and philosophical relevance have garnered the attention of economists, philosophers and political scientists alike (List, 2013). Within the vast body of literature making up this field, one can find various axiomatic approaches and attempts to answer some of the following problems: How can we model the relationship between individual choices and collective outcomes? Which voting systems and ranking systems produce the highest level of social welfare (List, 2013)? What is the role of individual rights in the optimal allocation of public resources (Sen, 1998)?

Of the discipline's various findings, we focus our attention on one particularly influential discovery: Kenneth J. Arrow's famous "Impossibility Theorem." In an attempt to adequately capture the impact of this theorem on social choice, this paper presents a chronological account of the field's development leading up to Arrow's finding, with a particular emphasis on its relationship with welfare economics. In doing so, we also provide a simple explanation of Arrow's impossibility theorem, along with a brief contextual analysis of his 1951 monograph *Social Choice and Individual Values*, within the history of economic thought (Arrow, 1951).

Fathers of Social Choice

While the natural problem of reconciling individual freedoms with a society's general wellbeing has long been at the center of philosophical investigations, the formal discipline of social choice is said to be 'initiated' by the 18th century French mathematicians Jean-Charles de Borda and Nicolas de Condorcet (Sen, 1998), whose mathematical treatments of voting systems distinguished themselves from previous discussions on voting theory, which include the writings of Aristotle, Catalan theologian Ramon Lull, and German mathematician Nikolaus of Cusanus (McLean, 2014). Taking place during the onset of a French Revolution, Borda and Condorcet's research was partly fuelled by a desire to reduce "instability and arbitrariness" in the crafting of political systems (Sen, 1998). Both of their works offered different ways to improve the standard method of "plurality-voting", according to which the candidate with the most 1st place votes would win an election. In 1781, Borda devised a simple scoring method, known today as *Borda Count*, which attributes a certain number of points to each ranking and subsequently elects the candidate with the highest score (McLean, 2014). Condorcet, on the other hand, defended a method of pair-wise majority voting, which functions as follows: for any two candidates x and y, x is preferred to y if a majority of voters prefer x over y; the candidate with the most pair-wise wins is named the *Condorcet winner* (Condorcet, 1785).

Despite their ingenuity and the improvements they made over the standard

method of plurality voting, both Borda and Condorcet's voting schemes were not entirely devoid of problems. As Condorcet himself would point out in his *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, the Condorcet system presents a *Condorcet Paradox*, in which his majority rule can, in some cases, violate the basic requirement of Transitivity (Condorcet, 1785). In such cases, the aggregation of individually "rational" preferences can yield "irrational" outcomes (List, 2013). In a future publication, Condorcet would also explicitly criticize the Borda method, which successfully avoids the *Condorcet Paradox*, but does so at the expense of another crucial principle called the Independence of Irrelevant Alternatives (IIA) (McLean, 2014). Interestingly enough, this trade-off between the Condorcet and Borda method would effectively foreshadow future debates surrounding the significance of Arrow's theorem (List, 2013). Moreover, these two results would only mark the beginning of a long wave of theoretical pessimism – arguably a lasting characteristic of 20th century social choice research.

Utilitarianism & The Birth of Welfare Economics

In the 19th century, several thinkers would re-explore some of the concepts that were introduced by Borda and Condorcet. These scholars most notably include the famous fiction author, mathematician, and clergyman C.L. Dodgson (better known as Lewis Carroll), as well as the English mathematician E.J. Nanson, both of whom explored hybrid voting systems that combined Borda's rule with a majority vote. In his research, Dodgson was also the first to use the term 'cycles' when referring to a *Condorcet Paradox* (McLean, 2014). Today, one popular method of breaking these Condorcet cycles is called the *Schulze Method*, introduced by Markus Schulze in 1997 (Schulze, 2018). This procedure consists of iteratively breaking the weakest link of a Condorcet cycle, by ignoring the pair-wise majority preference supported by the smallest majority, until the Condorcet cycle is definitively broken.

It was also around the turn of the 19th century that philosophical utilitarianism began bleeding deeply into the works of neoclassical economists. Indeed, prominent thinkers such as Jeremy Bentham, John Hicks and Henry Sidgwick had a great deal of influence on what would be known as the traditional school of welfare economics (Fleurbaey & Salles, 2021). In particular, Bentham's use of "utilitarian calculus" to study the general welfare of a society became a common framework for the likes of Edgeworth and later Marshall and Pigou (Sen, 1998). At that time, welfare economists were primarily concerned with the possibility of evaluating a society's aggregated utility, and primarily made use of ordinal preferences to do so. In fact, several neoclassical economists, including Arrow himself, held the view that "[...] interpersonal comparison of utilities has no meaning [...]" (List, 2013). These beliefs would likely play a major role in shaping Arrow's 1951 theorem (List, 2013). According to Sen, this "almost unquestionable adherence to utilitarian calculus" also fuelled the scrutiny that

welfare economics would face in the 20th century (Sen, 1998).

Changing Tides

In the decades leading up to Arrow's famous theorem, a series of important writings would completely redefine the status of welfare economics, and subsequently, its relationship with social choice research. One important figure in this theoretical shift is Lionel Robbins, who believed that the comparison of utility across individuals was neither possible, nor relevant to the *verifiable* science of economics (Sen, 1998). In the final chapter of his *Essay on the Nature and Significance of Economic Science*, Robbins firmly argues: "There is no means of testing the magnitude of A's satisfaction as compared with B's" (Robbins, 1932:124). A few years later, Robbins published a short text labelled "Interpersonal Comparisons of Utility: A Comment" in which he declares that despite his admiration for Benthamite utilitarianism, he could no longer believe in a "continuity between politics and economic analysis" (Robbins, 1938). As a result, a "new welfare economics" began solely relying on Pareto comparisons (named after Vilfredo Pareto) as a means of assessing social welfare, and consequently neglected the distributive implications of its methods (Sen, 1998).

In parallel to Robbins' publications however, economists like Hicks, Kaldor, Bergson, and Samuelson kept investigating various ways to assess a community's total sum of welfare (Fleurbaey & Salles, 2021). This effort amounted in several discoveries, including the introduction of "social welfare functions" in Bergson's 1938 QEJ paper, "A Reformulation of Certain Aspects of Welfare Economics" (Bergson, 1938). Nearly a decade later, Paul Samuelson would build upon Bergson's work and refine the concept of a social welfare function, as a function of individual utility functions (Samuelson, 1947). As it turns out, the Bergson-Samuelson model would be extremely influential on Arrow's formulation of the basic problem of social choice in his 1950 paper, "A Difficulty of the Concept of Social Welfare" (Arrow, 1950).

Modern Social Choice Theory

In the period spanning 1948 to 1951, a handful of papers by Black and Arrow mark the "birth" of a modern social choice theory (Fleurbaey & Salles, 2021). The first of these papers is Duncan Black's often overlooked 1948 publication in the Journal of Political Economy, "On the Rationale of Group Decision Making" (Black, 1948), in which Black uses a geometric representation of individual preferences to study the property of single-peakedness and develop a method of preference-aggregation. Black later refines his theory in his paper, "The Decision of a Committee, Using a Special Majority" (Black, 1948). Broadly speaking, both of these papers represent Black's attempt at a general theory of preference ordering whose main objective was to serve as "[...] a basis for the

development of a pure science of politics" (Black, 1948). Shortly after Black's publication, Arrow tackles the same challenge of preference aggregation using a very different approach.

In his 1948 RAND discussion "The Possibility of a Universal Social Welfare Function," Arrow introduced an axiomatic method of analyzing preference relations which filled a gap that in the welfare economics literature of the time (Arrow, 1948). Arrow's framework differed from previous investigations in at least two crucial ways: it borrowed several concepts from logic (List, 2013), and made it possible to study the limitations of *any* "social welfare function," in Arrow's sense of the term (Fleurbaey & Salles, 2021). In brief, his axiomatic method consisted in identifying "[...] a set of plausible necessary and sufficient conditions that uniquely characterize a particular solution (or a class of solutions) to a given type of collective preference problem" (List, 2013), and allowed for powerful claims to be made about any preference-aggregation method. The most surprising of these claims, is Arrow's "Impossibility Theorem", which is first alluded to in his 1948 RAND paper (Arrow, 1948). The theorem would later be mentioned in his 1950 paper "A Difficulty in the Concept of Social Welfare", before being taken up in greater length in his 1951 book, *Social Choice and Individual Values* (Arrow, 1951).

Modern Social Choice Theory

We can define a social welfare function as a "preference-aggregation rule," which takes any combination of votes (preference orderings) as input and gives a social decision (social state) as output (List, 2013). Arrow's "General Possibility Theorem," later known as Arrow's impossibility theorem revealed "astonishing" results concerning the properties of such social welfare functions (Sen, 1998). Accordingly, Arrow's theorem is most easily understood as a discrepancy between a social welfare function's possible set of properties, and its *desired* set of properties. In his 1951 book, Arrow describes five subjectively essential properties, formally known as *Arrovian conditions*, which every social welfare function ought to satisfy:

1) *Universal Domain*: "The domain of F is the set of all logically possible profiles of complete and transitive individual preference orderings" (List, 2013)

2) *Ordering*: "For any profile [R1, R2, ..., Rn] in the domain of F, the social preference relation R is complete and transitive" (List, 2013);

3) *Weak Pareto Principle*: "in case of unanimous preferences of a social state a over a social state b, then the social state a must be socially preferred to social state b" (Fleurbaey & Salles, 2021);

4) *Independence of Irrelevant Alternatives (IIA)*: "[...] the choice made by society from any given set of alternatives should be independent of the very existence of alternatives outside the given set" (Arrow, 1950);

5) *Non-dictatorship*: "The social welfare function is not to be dictatorial [...]" (Arrow, 1951:30).

Given these definitions, Arrow manages to prove that if there are at least two individuals voting over three or more alternatives (social states), then there exists no social welfare function which simultaneously satisfies all five conditions (List, 2013). An equivalent, slightly more pessimistic, interpretation of the theorem reads as follows: "If we exclude the possibility of interpersonal comparisons of utility, then the only methods of passing from individual tastes to social preferences which will be satisfactory and which will be defined for a wide range of sets of individual orderings are either imposed or dictatorial." (Arrow, 1951:59)

On The Relevance & Repercussions of Arrow's Theorem

Within the field of social choice, few findings have had as monumental of an impact as Arrow's theorem whose "breathtaking elegance and power" spurred a drastic expansion in social choice research (Sen, 1998). In fact, some social choice scholars made use of Arrow's methods to construct new possibility and impossibility theorems of their own (May, 1952). Much of the literature that followed Arrow's book, however, questioned the necessity of arrow's conditions, and explored ways to "avoid this impossibility result" altogether (Sen, 1998). In 1967, Tullock is amongst the first to attack Arrow's theorem on the grounds of practical irrelevance, stating that cyclical paradoxes rarely occur in practical settings therefore rendering Arrow's impossibility "insubstantial" (Tullock, 1967). In a similar fashion, Black later showed that, given a sufficiently large committee, one could construct a voting procedure "[...] which gives rise to a fraction of intransitivities so trivial that for every practical purpose it can be disregarded" (Black, 1969). Others, such as Harsanyi and Sen, constructed ways to advance the general theory of social choice beyond its Arrovian limitations. In 1955, Harsanyi borrowed a game theoretic method to devise a more forgiving weighted social welfare function (Harsanyi, 1955) whose arbitrariness in the selection of weights was called a "probabilistic veil of ignorance" (Harsanyi, 1953). In 1970, Sen went as far as proving the "Impossibility of a Paretian Liberal," which highlighted the inherent conflict between "liberal values" and "pareto efficiency" (Sen, 1970), and gave rise to a "non-welfaristic approach to social choice" (Fleurbaey & Salles, 2021).

Somewhere between the blind acclaim and brash criticism of Arrow's work, there exists a third, more nuanced view, which considers the power of Arrow's theorem without upholding its pessimistic implications. In his 1972 Nobel-winning lecture, Arrow himself concluded by saying that the paradox of social choice should be treated "[...] as a challenge rather than a discouraging barrier" (Arrow, 1972). Similarly, Sen has interpreted Arrow's work as more of a "general approach to thinking about social decisions," and by considering the numerous inputs one can give a social welfare function, sees the field of social choice as being "[...] as much a world of possibility as of conditional impossibilities" (Sen,

1999:253). By contextualizing Arrow's work within the evolution of social choice and welfare economics, this paper hopes to contribute to this alternative view on the topic, and encourages readers to appreciate the impossibility theorem as, above all, an unforgettable moment in the history of thought.

References

Arrow, K. (1948, September 26). The Possibility of a Universal Social Welfare Function. Retrieved from https://www.rand.org/pubs/papers/P41.html

Arrow, K. J. (1950). A Difficulty in the Concept of Social Welfare. Journal of Political Economy, 58(4), 328-346. doi:10.1086/256963

Arrow, K. J. (1951). Social choice and individual values. New York: Wiley & Sons.

Arrow, K. (1972). General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice - 1972 Nobel Prize Lecture.

Bergson, A. (1938). A Reformulation of Certain Aspects of Welfare Economics. The Quarterly Journal of Economics, 52(2), 310. doi:10.2307/1881737

Black, D. (1948). On the Rationale of Group Decision-making. Journal of Political Economy, 56(1), 23-34. doi:10.1086/256633

Black, D. (1948). The Decisions of a Committee Using a Special Majority. Econometrica, 16(3), 245. doi:10.2307/1907278

Black, D. (1969). On Arrows Impossibility Theorem. The Journal of Law and Economics, 12(2), 227-248. doi:10.1086/466667

Condorcet, N. D. (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.

Fleurbaey, M., & Salles, M. (2021). A Brief History of Social Choice and Welfare Theory. Studies in Choice and Welfare Conversations on Social Choice and Welfare Theory - Vol. 1, 1-16. doi:10.1007/978-3-030-62769-0-1

Harsanyi, J. C. (1953). Cardinal Utility in Welfare Economics and in the Theory of Risk-taking. Journal of Political Economy, 61(5), 434-435. doi:10.1086/257416

Harsanyi, J. C. (1955). Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. Journal of Political Economy, 63(4), 309-321. doi:10.1086/257678

List, C. (2013, December 18). Social Choice Theory. Retrieved from https://plato.stanford.edu/entries/social-choice/

May, K. O. (1952). A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision. Econometrica, 20(4), 680. doi:10.2307/1907651

Mclean, I. (2014). The strange history of social choice, and the contribution of the Public Choice Society to its fifth revival. Public Choice, 163(1-2), 153-165. doi:10.1007/s11127-014-0161-7

Robbins, L. C. (1932). An essay on the nature significance of economic science. London: Macmillan.

Robbins, L. (1938). Interpersonal Comparisons of Utility: A Comment. The Economic Journal, 48(192), 635. doi:10.2307/2225051

Samuelson, P. A. (1947). Foundations of economic analysis. London: Oxford University Press.
Schulze, M. (2021, October 10). The Schulze Method of Voting. Retrieved from https://arxiv.org/abs/1804.02973
Sen, A. (1970). The Impossibility of a Paretian Liberal. Journal of Political Economy, 78(1), 152-157. doi:10.1086/259614
Sen, A. (1998). The Possibility of Social Choice - 1998 Nobel Prize Lecture.
Sen, A. (1999). Development as freedom. Oxford: Oxford University Press.
Tullock, G. (1967). The General Irrelevance of the General Impossibility Theorem. The Quarterly Journal of Economics, 81(2), 256. doi:10.2307/1879585