Kenneth Arrow's Contributions to Social Choice Theory

Eric Maskin

Harvard University and Higher School of Economics

Kenneth Arrow created the modern field of social choice theory, the study of how society should make collection decisions on the basis of individuals' preferences. There had been scattered contributions to this field before Arrow, going back (at least) to Jean-Charles Borda (1781) and the Marquis de Condorcet (1785). But earlier writers all focused on elections and voting, more specifically on the properties of *particular* voting rules (I am ignoring here the large literature on utilitarianism – following Jeremy Bentham 1789 – which I touch on below). Arrow's approach, by contrast, encompassed not only *all possible* voting rules (with some qualifications, discussed below) but also the issue of aggregating individuals' preferences or welfares, more generally.

Arrow's first social choice paper was "A Difficulty in the Concept of Social Welfare" (Arrow 1950), which he then expanded into the celebrated monograph *Social Choice and Individual Values* (Arrow 1951). In his formulation, there is a society consisting of n individuals, indexed $i = 1,...,n$, and a set of social alternatives A (the different possible options from which society must choose). The interpretation of this set-up depends on the context. For example, imagine a town that is considering whether or not to build a bridge across the local river. Here, "society" comprises the citizens of the town, and A consists of two options: "build the bridge" or "don't build it." In the case of pure distribution, where there is, say, a jug of milk and a plate of cookies to be divided among a group of children, the children are the society and A

includes the different ways the milk and cookies could be allocated to them. If a committee must elect a chairman, society is the committee and *A* comprises the various candidates for chair. Thus, the formulation covers a huge number of possible applications.

Arrow's definition of a *social welfare function* (SWF) is likewise very general: it is a rule that determines society's preferences over *A* on the basis of individuals' preferences or welfares. Formally, if $\mathfrak{R}_i$ is a set of possible orderings of *A* for individual *i* (an ordering is a complete, reflexive, and transitive binary relation on *A*), then a SWF *F* is a mapping

$$F : \mathfrak{R}_1 \times \ldots \times \mathfrak{R}_n \to \mathfrak{R},$$

where $\mathfrak{R}$ is also a set of orderings.

Although highly permissive, this formulation still excludes some important possibilities. First, it rules out making use of *intensities* of individuals 'preferences (or other cardinal information). For example, it disallows a procedure in which each individual assigns a numerical "utility" (or "grade") to every alternative (say, on a scale from 1 to 5) and alternatives are then ordered according to the sum (or median) of utilities (see Balinski and Lariki 2010 for a recent cardinal approach along these lines). Arrow's rationale for excluding cardinality – following Lionel Robbins (1932) – is that such information cannot be reliably obtained empirically unless individuals trade off alternatives in *A* against some other good like money (in which case *A* does not fully describe the alternatives). Second, the formulation doesn't allow for *interpersonal comparisons*: it takes no account of the possibility that, say, individual 1 may gain more in going from alternative *a* to *b* than individual 2 loses (thus, Arrow excluded classical utilitarianism à la Bentham, according to which *a* is socially preferred to *b* if the sum of individuals utilities for *a* is greater than that for *b*). Arrow ruled out making such comparisons because, again, he argued

they lack an empirical basis. Finally, the requirement that social preferences constitute an *ordering* (in particular, that preferences are *transitive*) seems to attributes a degree of rationality to society that is questionable (see, for example, James Buchanan 1954). Arrow's reason for positing social transitivity, however, was purely pragmatic: it allows us to say what society ought to choose even when the *feasibility* of various alternatives in *A* is unknown in advance. Specifically, with transitive preferences, society should choose the alternative at the top of the preference ordering – call it *a* – if *a* is feasible, the next best alternative *b*, if *a* is infeasible, and so on. But if *a* is socially preferred to *b*, *b* is preferred to *c*, and *c* to *a* (i.e., an intransitivity arises), then it is not clear what society should choose.

Indeed, social intransitivity is a potential problem for the best known way of determining social preferences, *majority rule*, in which alternative *a* is socially preferred to *b* if a majority of individuals prefer *a* to *b*. Imagine, for example, that 35% of society prefer *a* to *b* and *b* to *c*; 33% prefer *b* to *c* to *a*; and 32% prefer *c* to *a* to *b*. Then, under majority rule, *a* is socially preferred to *b* (67% prefer *a*); *b* is preferred to *c* (68% prefer *b*), and *c* is socially preferred to a (65% prefer *c*). Majority rule doesn't work in this case.

As far as we know, the possibility of majority rule's intransitivity was first noted by Condorcet (himself a strong proponent of majority rule); social preferences over *a, b,* and *c* in the example form a "Condorcet cycle." Condorcet cycles were Arrow's starting place. Interestingly, he was, at that time, unaware of Condorcet's work but rediscovered the intransitivity for himself. It made him wonder whether there is some other reasonable way of determining social preferences that *does* always work.

By "reasonable," Arrow meant, first, that social preferences should be *transitive*, as I have just discussed. Thus, majority rule does not technically constitute a SWF if the preference domains $\Re_1, \ldots, \Re_n$ admit a Condorcet cycle. Arrow also required:

Unrestricted Domain (U): The SWF must determine social preferences for all possible preferences that individuals might have. Formally, for all $i = 1, \ldots, n$, $\Re_i$ must consist of *all* orderings of A.

Pareto Property (P): If all individuals prefer a to b, then a must be socially preferred to b. Formally, for all $(R_1, \ldots, R_n) \in \Re_1 \times \ldots \times \Re_n$ and all $a, b \in A$, if for all i, aP_ib (which means aR_ib but not bR_ia), then aPb, where $R = F(R_1, \ldots, R_n)$.

Independence of Irrelevant Alternatives (IIA): Social preferences between a and b should depend only on individuals' preferences between a and b, and not on their preferences concerning some third ("irrelevant") alternative c. Formally, for all $(R_1, \ldots, R_n), (R'_1, \ldots, R'_n) \in \Re_1 \times \ldots \times \Re_n$ and all $a, b \in A$, if, for all i, $(aR_ib$ if and only if $aR'_ib)$ and $(aP_ib$ if and only if $aP'_ib)$, then $(aRb$ if and only if $aR'b)$ and $(aPb$ if and only if $aP'b)$, where $R = F(R_1, \ldots, R_n)$ and $R' = F(R'_1, \ldots, R'_n)$.

Nondictatorship (ND): There exists no individual who always gets his way in the sense that social preferences must coincide with his own regardless of others' preferences. Formally, there does *not* exist i such that, for all $(R_1, \ldots, R_n) \in \Re_1 \times \ldots \times \Re_n$ and all $a, b \in A$, if aP_ib, then aPb, where $R = F(R_1, \ldots, R_n)$.

On the face of it, these conditions may seem natural and undemanding. But, stunningly, Arrow showed that they cannot be satisfied simultaneously: his Impossibility Theorem establishes that there is *no* SWF satisfying all of U, P, IIA, and ND, if *A* includes 3 or more alternatives (note that majority rule satisfies all the conditions when there are only two alternatives).

A crude measure of a book's influence is the number of times it is cited. By that standard, *Social Choice and Individual Values* has been immensely influential, with over 17,000 Google Scholar citations as of this writing. But perhaps a better criterion of influence is how long a book continues to inspire new work and thought, i.e., its intellectual longevity. Kenneth Arrow's great monograph measures up extremely well there too: sixty-six years on, it is still alive and kicking.

Note: This essay draws substantially from my Foreword to the third (2012) edition of *Social Choice and Individual Values.*

References

Arrow, K.J. (1950), "A Difficulty in the Concept of Social Welfare," *Journal of Political Economy* 58(4), 328-46.

Arrow, K.J. (1951), *Social Choice and Individual Values,* New York: Wiley. Second edition (1963), Yale University Press. Third edition (2012), Yale University Press.

Balinski, M. and R. Laraki (2010), *Majority Judgment*, Cambridge, MA: MIT Press.

Bentham, J. (1789), *An Introduction to the Principles of Morals and Legislation*, London.

Borda, J-C. (1781), "Mémoire sur les Elections au Scrutin," *Memories de L'Academie Royale des Sciences.*

Buchanan, J.M. (1954), "Social Choice, Democracy, and Free Markets," *Journal of Political Economy* 62, no 2, 114-23.

Condorcet, Marquis de (1785), *Essai sur L'application de L'analyse à la Probabilité des Décisions Rendus à la Pluralité des Voix,* Paris: L'Imprimerie Royale.

Robbins, L. (1932), *An Essay on the Nature and Significance of Economic Science,* London: Macmillan.