CHAPTER

# 1 INTRODUCTION

Social aggregation theory is concerned with investigating methods of clustering values that individuals in a society attach to different social or economic states into values for the society as a whole. Loosely speaking, a social state, a state of affairs, represents a sketch of the amount of commodities possessed by different individuals, quantities of productive resources invested in different productive activities and different types of collective activities (Arrow 1950). The values that individuals attach to different social states are reflections of respective preferences. Consequently, the problem of social aggregation is to combine individual preferences into a social preference in an unambiguous way. In this monograph, we will use the terms "social aggregation" and "social choice" interchangeably.

Modern social aggregation theory started with the publication of Kenneth J. Arrow's pioneering contribution *Social Choice and Individual Values*, his PhD dissertation, in 1951. It can be regarded as the foundation to laying the groundwork of social aggregation theory in view of its innovative facture and revolutionary influence. The idea of aggregating individual preferences into a collective choice rule predates Arrow (1950) by more than 150 years. In 1785, the French mathematician and philosopher Marie-Jean de Condorcet considered the problem of collective decision-making with regard to majority voting. According to majority voting, in a choice between two alternatives $x$ and $y$, $x$ is declared as the winner if it gets more votes than $y$. He established that the method of pair-wise majority voting may give rise to cyclicality in social preference. This paradoxical result, popularly known as the *Condorcet voting paradox*, appears to draw inspiration, to a certain extent, from an earlier contribution by the French mathematician Jean-Charles de Borda (de Borda 1781). In this alternative voting system, known as the Borda count method, voters rank candidates in order of preference.[1]

One of the major goals of this monograph is the analysis of the Arrovian approach to the theory of collective aggregation and later developments on it. We, therefore, focus now on Arrow's impossibility theorem, which is generally acknowledged as the formative basis of modern social aggregation rules. Arrow's seminal work examines the possibility of the aggregation of individual preferences into a social preference in order to obtain a social ranking of alternative states of affairs. Each individual preference relation (or ordering) designed with the objective of ranking alternative social states is assumed to satisfy *completeness*, *reflexivity*, and *transitivity*. Each of these assumptions may be regarded as a value judgment, a subjective statement that cannot be verified by factual evidence. Completeness means that the individual can compare and rank any two social states. That is, between any two state of affairs $x$ and $y$, the person regards either $x$ as at least as good as, $y$, $y$ as at least as good as $x$ or both. Reflexivity demands that each social state should be always as good as itself. According to the transitivity of three social states $x$, $y$, and $z$, if a person regards $x$ as at least as good as $y$, $y$ as at least as good as $z$, then he must regard $x$ as at least as good as $z$. If a continuity assumption is made, the preference relation can be represented by a utility function. Continuity means that of the three social states $x$, $y$, and $z$, if $x$ is treated as better than $y$ and $y$ is treated as better than $z$, then any curve connecting $x$ and $z$ must cross the indifference curve

[1] Contributions along this line also came from Pierre-Simon Laplace (1812), Charles Lutwidge Dodgson (better known as Lewis Carroll) Dodgson (1873, 1874, 1876), Isaac Todhunter (1865), Edward J. Nanson (1882), and Francis Galton (1907). Two important references for detailed discussions in this context, which are beyond the scope of this monograph, are Black (1958) and Suzumura (2002).

containing $y$. For a continuous preference ordering, there can be no sudden jump from being better than an alternative to being worse than the alternative. Continuity ensures that if the person prefers all states that are close to $x'$ to $y'$, then $x'$ should be preferred to $y'$. Thus, utility expresses the intensity of individual preferences.

Arrow's fundamental result on preference aggregation shows that there is no way to aggregate person-by-person preferences for arriving at a social preference relation or social welfare ordering (social ordering, for short), satisfying five highly plausible value judgments such that all social states can be ranked unambiguously by the ordering. The five value judgments that are required to be fulfilled by an Arrovian social ordering are: $(a)$ completeness, reflexivity, and transitivity, $(b)$ universality or unrestricted domain, $(c)$ weak Pareto principle, $(d)$ non-dictatorship, and $(e)$ independence of irrelevant alternatives. According to universality, an Arrovian social preference ordering should work irrespective of what individual preferences happen to be. In other words, there is a lack of restriction on the domain. The weak Pareto principle demands that of two alternatives $x$ and $y$, if everybody strictly prefers $x$ to $y$ then $x$ must be regarded as socially better than $y$. That is, a gain by each individual must be acknowledged as a social enrichment. A social ordering is called a dictatorship if there is someone whose strict preferences depict social preferences. Non-dictatorship requires that a social ordering must make nobody a dictator. Independence of irrelevant alternatives demands that a social ranking between any two states of affairs should be independent of the individual orderings over other states of affairs. Thus, Arrow's theorem looks for a complete, reflexive, and transitive social ordering that can be represented as a social welfare function under continuity. While individual and social orderings are presented analytically in Chapter 2, Chapters 4 and 6 analyze Arrow's theorem with preferences and utilities respectively.

Arrow's theorem casts doubts on all concepts that implicitly or explicitly incorporate a societal preference. Examples that can be included within this purview are "a social contract," "a social benefit," "a public good," and so on. Evidently, any notion that casts so much perplexity will invite a lot of feedback. Consequently the investigations advanced along the lines of looking for aggregations of individual preferences so that alternative states of affairs can be evaluated in a satisfactory way. Some prior value judgment is required to be made for pooling of individual preferences into an aggregated preference. One such value judgment in the current context is the assumption about measurability and comparability of individual preferences.

For expositional ease, the remaining discussions on these lines, will be in terms of individual and social utility functions.[2] In the Arrovian framework when individual preferences are portrayed in terms of utility functions, it is said that utility functions are ordinally measurable and interpersonally non-comparable. The measurability of a utility function refers to the meaningfulness of the real numbers attached to a given person's utility levels. It is formalized by considering the type of transformations that can be applied to the utility function such that the information conveyed by the original utility function is retained by its transformed counterpart. Since such a transformation maintains the original information, it is referred to as an information invariance assumption. Sen (1970a) and many others have relaxed the non-comparability assumption and demonstrated the possibilities of existence of social aggregation rules.

To understand the measurability and comparability notions more explicitly, let the real valued function $U_i$ denote person $i$'s utility function defined on the set of alternatives. Thus, between two states of affairs $x$ and $y$, if the person regards $x$ as at least as good as $y$, then in terms of utility we express this as $U_i(x) \geq U_i(y)$. Consider the least restrictive measurability assumption that $U_i$ is measurable on an ordinal scale. Under the ordinal measurability assumption, we can say that for any increasing transformation $f_i$ defined on the set of real numbers, $U_i(x) \geq U_i(y)$ if and only if $f_i(U_i(x)) \geq f_i(U_i(y))$. In words, ordinal scale measurability allows the utility function of an individual to be rescaled using any ordinal or increasing transformation of it. For instance, $f_i(t)$ can be $t^r$, where $r > 0$ is any positive real number. (The real number $t$ is any element in the range of $U_i$.) A second example can be $\log(t)$ given that $t > 0$ and so on. A stricter type of

2 A social utility function, expressed as a function of individual utilities, is referred to as a social welfare or evaluation functional.

measurability assumption is cardinal scale measurability. Under this notion of measurability, the only allowable transformation is an affine transformation, that is, $f_i(t) = a_i + b_i t$, where $b_i > 0$ and $a_i$ are arbitrary real numbers. The relationship between centigrade and Fahrenheit temperature scales is a standard example of cardinal equivalence. Thus, if $C$ and $F$ stand for temperatures on the centigrade and Fahrenheit scales respectively, then the relationship between the two scales is given by $C = -160/9 + 5F/9$. Another strict notion of measurability is ratio scale measurability, which claims that the only admissible transformation under which the utility and its transformed counterpart convey to us the same information is $f_i(t) = b_i t$, where $b_i > 0$ is any arbitrary real number. An example can be the measurement of weight of a person. It does not matter whether we measure the weight of a person in grams or in kilograms, since we can convert the latter into the former by multiplying with 1,000.

The problem of interpersonal comparisons of utility is simply the problem of comparing different person's utilities. It means how the real numbers attached to different individuals' utility levels can be compared in a meaningful way. A simple example can be as follows. Suppose a person C is given the option of being person A or person B in a situation. He claims that he prefers to be person B rather than be person A in the situation. Implicit under this comparability by person C is an interpersonal comparison.

However, the idea of interpersonal comparison has been criticized on the grounds that exact numerical scales of utility cannot be contrived or there may be difficulties involved in the process of such comparisons.[3] Nevertheless, if we cannot compare different individuals' utilities, it may be difficult to evaluate situations where a change in the social state increases the utility of one or more individuals at the cost of reduction of the utility of at least one other individual. This means that we may not be able

[3] See, for example, Robbins (1932, 139–142), and Arrow (1963, 9). Harsanyi (1955, 317n20) offered the logical basis of such comparisons. The possibility of interpersonal comparisons was also discussed by Little (1957). Waldner (1972) analyzed the problem of interpersonal comparisons using the notion of empirical meaningfulness. For generalization and justifications of Waldner's approach, see List (2003).

to settle distributive justice. Nonetheless, it has been noted in the literature that in many situations social evaluation involves only limited comparability.

Alternative notions of interpersonal comparisons can be formalized by making assumptions about the information invariance transformations applied to individual utility functions. For instance, under ordinal scale measurability, if the increasing transformations $f_i$s are not necessarily identical across persons, then we have ordinally measurable, non-comparable utilities. If the increasing transformations $f_i$s are assumed to be same for every person, then the situation is the one of ordinal scale measurability combined with full comparability of utilities. While under full comparability the utility levels are comparable across persons, under non-comparability this is not so. More precisely, under full comparability, for any two individuals $i$ and $j$, it is possible to make claims such as $U_i(x) \geq U_j(x)$ if and only if $f(U_i(x)) \geq f(U_j(x))$, where the transformation $f$ is increasing. Likewise, with cardinally measurable utilities, if the scalars $a_i$s and $b_i$s are non-identical across persons, then we have full non-comparability under cardinal scale measurement. In this case, intrapersonal, but not interpersonal, comparability of utility differences is allowed. That is, for any person $i$ one can compare utility gains or losses of the form $U_i(x) - U_i(y)$ for $i$ only, not across persons. If, however, the multiplicative scalar $b_i > 0$ is the same for all persons but $a_i$s are distinct, then the setting comes to be the one of cardinally measurable unit-comparable utilities. In this case of interpersonal comparison, comparability of utility differences across persons is allowed, but comparison of utility levels is not permitted. More precisely, for any two individuals $i$ and $j$, differences of the type $U_i(x) - U_i(y)$ and $U_j(x) - U_j(y)$ can be compared but not utility levels such as $U_i(x)$ and $U_j(x)$. The utilitarian social evaluation functional, the sum of individual utilities, has been characterized using this notion of information invariance (d'Aspremont and Gevers 1977). Following Bentham's (1789) usage, the classical economists John Stuart Mill, Alfred Marshall, Francis Y. Edgeworth, Henry Sidgwick, and Arthur Pigou employed the sum of cardinal utilities to evaluate public policies from the viewpoint of increase or decrease in the sum of satisfaction.

A scheme of classification of different measurability and comparability assumptions was provided by Sen (1974, 1977, 1986).[4] Chapter 6 of this monograph deals with a detailed taxonomy of alternative notions of measurability and comparability. Aggregation theorems, along with the Arrow theorem, under alternative notions of measurability and comparability are presented in Chapter 6.

In the Arrovian framework, the objective was to find whether a profile of individual preference relations can be aggregated into a social preference relation. This goal turned out as an impossibility. An alternative natural line of inquiry can be investigating whether a profile of relations can be converted into a single winner or a single best alternative. The single-winner problem arises in many practical situations. For instance, in a single-winner election the election process has to declare exactly one of the contestants as the winner. A rule that transforms preference profiles into a single winner is called a social choice function. The question now boils down to this: Is there a trustworthy social choice function which can claim that this alternative is on top?

To understand this in greater detail, we need to figure out what we mean by trustworthy. For illustrative purposes, let us consider the problem of (private) provision of a public good using the Lindahl tax scheme. A vector of tax shares and an output level $g$ for the public good is said to constitute a Lindahl equilibrium if $g$ maximizes individual utility functions subject to respective budget constraints. At a Lindahl equilibrium, the marginal utility of a person, evaluated at $g$, is equated with his tax share, establishing that what the person receives is the same as what he pays. The public good output quantity $g$ is the optimal level of the public good as well and the tax shares are optimal. The determination of Lindahl equilibrium, thus, needs information on the utility maximizing tax shares. Now, if a person truthfully reveals his preference about public good production quantity, then he will have to pay what the public good is worth to him. Consequently, if he is asked by the public good authority to communicate his preference for the public good, he may behave strategically and underreport his preference with the anticipation that others will pay for it and he can take

[4] See also Blackorby, Donaldson, and Weymark (1984); and d'Aspremont and Gevers (2002).

a free ride. Therefore, the Lindahl tax scheme is not trustworthy. Equivalently, we can say that it is not foolproof or cheat proof.[5]

If a social choice function does not come up with any inducement for strategic behaviour in the sense of falsification of preference by anybody irrespective of what others are doing, then it can be regarded as strategyproof or non-manipulable. Therefore, in our public good example, strategyproofness or non-manipulability requires everybody to reveal their preferences truthfully irrespective of what others are doing.

Apart from non-manipulability, the two other value judgments we impose on a social choice function are non-degeneracy and universality. According to non-degeneracy, any state of affairs must be included in the range of the social choice function. Universality is the same as in the Arrovian case. A social choice function is called dictatorial if the social choice is always the favourite alternative of a person. A social choice function is non-dictatorial if it is not dictatorial.

Gibbard (1973) and Satterthwaite (1975) demonstrated that there is no non-dictatorial social choice function that satisfies universality, non-degeneracy, and non-manipulability. Equivalently, the Gibbard–Satterthwaite theorem says that any social choice function satisfying universality, non-degeneracy, and non-manipulability must be dictatorial. The theorem definitely does not claim that there can be no useful social choice functions. The theorem does not even claim that no social decision can be taken in a given situation. There can be some choice functions that are superior to dictatorship. We provide extensive discussion along this line in Chapter 10 of the monograph. Moreover, in Chapter 10, we also impose a domain restriction called single-peakedness and provide two non-dictatorship results due to Moulin (1980). Another kind of domain restriction is achieved by allowing for side-payments and assuming that the agents have quasi-linear preferences. In Chapter 11, we assume quasi-linear preferences and discuss the Vickrey–Clarke–Groves mechanisms (see Vickrey 1961; Clarke 1971; Groves 1973). In Chaper 11, we also discuss Roberts'

[5] See Feldman and Serrano (2006).

mechanisms that generalize the Vickrey–Clarke–Groves mechanisms (see Roberts 1980).

A possibility result on the existence of a social choice function in the two-alternative case was demonstrated by May (1952). May's theorem says that a two-candidate group decision function is the simple majority rule if and only if it satisfies *decisiveness*, *anonymity*, *neutrality*, and *strong monotonicity*. Decisiveness requires a group decision function to be well defined for all profiles of individual voters' preferences and single valued. Anonymity means that the individuals should be treated symmetrically; a reordering of voters should not change the outcome. Neutrality demands symmetric treatment of alternatives. A monotonicity condition stipulates that increased support for a candidate may help it to win. According to May's theorem, in a preference between two alternatives, if the number of individuals liking the former over the latter is more than the number of individuals liking the latter over the former, then the group recommends the choice of the former. Thus, May's theorem identifies a winner on a majority rule basis. A rigorous discussion on May's theorem is presented in Chapter 3 of this text.

Harsanyi (1955, 1977) investigated the social aggregation problem from a different standpoint. In the Harsanyi framework, social evaluation judgments are made behind a veil of ignorance. Both individual ex-ante and social utilities, defined over a set of lotteries, are assumed to fulfill the von Neumann–Morgenstern expected utility axioms. Harsanyi assumed ex-ante Pareto indifference for a social preference over lotteries, which claims that if two lotteries are judged as equally valuable by all individuals then it should be treated as equally valuable by society as well. It then turns out that the social evaluation functional can be expressed as a weighted sum of individual utilities. The literature refers to this as *Harsanyi's social aggregation theorem*. According to Harsanyi, this social evaluation functional is of the utilitarian type. As expected, Harsanyi's contributions have invited a lot of responses from different angles. We provide extensive discussions on Harsanyi's aggregation theorem and related issues in Chapter 7 of this monograph.

The social aggregation theorems do not tell us anything about the welfare implications of inequality resulting from unequal distribution implicit in the states of affairs. Often from the policy point of view it becomes worthwhile to look at distributional implications of states of affairs in terms of equity. This issue is addressed from single and multidimensional perspectives in Chapters 8 and 9, respectively. In this context, by a state of affairs we mean a description of individuals' achievements in a single dimension, say income, or in multiple dimensions, depending on whether our concern is a unidimensional or a multidimensional analysis. More precisely, while in the single dimensional structure a state of affairs describes a distribution of income, in the multidimensional structure it represents the achievements of different individuals in different dimensions of well-being, say income, wealth, health, literacy, and so on. A state of affairs here gives information on individual achievements but not on tastes and preferences.

The two different approaches to welfare evaluations adopted here are direct and inclusive measures of well-being. In the former, welfare is defined directly on dimensional achievements while the latter parallels an idea implicit under a social evaluation functional rule. Thus, the latter methodology assigns each person a well-being number, as indicated by his utility, by aggregating all the welfare-relevant dimensions in his life while taking account of his achievement in each dimension. For the sake of convenience, we use the common term social evaluation function for welfare metric in the current context. Three value judgments that are assumed for a social evaluation function are efficiency (size), equity (distribution), and anonymity. Efficiency is taken care of by the strong Pareto principle, which requires welfare to increase if the achievement of a person, under *ceteris paribus* assumption, increases. Well-defined notions of equity that ensure welfare increase under equitable redistributions are assumed. Anonymity means that any feature other than achievements is irrelevant to welfare assessment. Intrinsic to the notion of multidimensional welfare evaluation is inter-dimensional correlation of achievements that enables us to distinguish among dimensions in terms of substitutability, complementarity, and independence.

While welfare evaluation is concerned with both size and distribution, the only concern of inequality is distribution. Size independence property of inequality analysis can be taken care of by considering inequality metrics of both relative