Social Choice Theory

Jacob M. Nebel and John A. Weymark

1. Arrovian Social Choice

In Kenneth Arrow's *Social Choice and Individual Values* (1951), collective decision-making is viewed as a problem in preference aggregation: the objective is to determine a collective ranking of a set of alternatives based on individual preferences. The procedure for aggregating preferences, a *social welfare function*, is designed before the preferences are known so that it will be applicable whatever these preferences turn out to be.

Prior to Arrow's book, the formal analysis of collective decision-making studied particular voting procedures, such as majority rule. Much of this analysis consisted of identifying problems with existing voting rules and of designing new rules to overcome them. For example, pairwise majority rule is subject to *Condorcet's Paradox*. There are three candidates for an election, x, y, and z, and three voters. Voter 1 prefers x to y to z, voter 2 prefers y to z to x, and voter 3 prefers z to x to y. Using majority voting a pair at a time results in a cycle with x beating y, y beating z, and z beating x.

Arrow changed the focus from specific collective decision-making procedures to the identification of properties that any preference aggregation procedure should exhibit. Once these properties are identified, the task is then to determine what aggregation rules satisfy them. Unfortunately, the properties that Arrow thought reasonable are incompatible, as shown in his famous impossibility theorem.

Arrow's analysis is formal, abstract, and axiomatic. Individuals are characterized by their preferences, the set of alternatives is some set whose members are unspecified, and the properties of aggregation rules are formalized as axioms. Alternatives can be anything: quantities of a public good, economic allocations, candidates for an election, etc.

A *preference* is a binary relation R on the set of alternatives X, with xRy meaning that alternative x is weakly preferred to alternative y. The corresponding strict preference relation is P. R is *complete* if for any two alternatives at least one of them is weakly preferred to the other and *transitive* if whenever one alternative is weakly preferred to a second and the second is weakly preferred to a third, then the first is weakly preferred to the third. A preference is an *ordering* if it satisfies these two "rationality" properties. It is *acyclic* if there are no strict preference cycles. The set of all orderings on X is $\mathcal{R}$.

There are a fixed finite number of individuals, n, with $n \geq 2$. Each individual i has a preference ordering R_i on X. A *preference profile* $\mathbf{R} = (R_1, \ldots, R_n)$ is a list of preferences, one for each individual. The set of all possible preference profiles is $\mathcal{R}^n$, the n-fold Cartesian product of $\mathcal{R}$. A *preference domain* $\mathcal{D}$ is some subset of $\mathcal{R}^n$. A *social welfare function* f assigns a social preference on X to each preference profile in $\mathcal{D}$.

Arrow considered five social welfare function axioms.

> *Social Ordering*. For each preference profile $\mathbf{R}$ in $\mathcal{D}$, the corresponding social preference R is an ordering.

Social Ordering is Arrow's *collective rationality* axiom. One rationale for it is that social preferences should exhibit the same degree of rationality as individual preferences. By requiring social preferences to be orderings, cycles like in the Condorcet Paradox are precluded.

If there are no restrictions on the individual preferences, then f should produce a social ordering for every conceivable preference profile.

> *Unrestricted Domain*. The preference domain $\mathcal{D}$ is $\mathcal{R}^n$.

In order for the social preference to be responsive to individual preferences, Arrow required that if everybody strictly prefers one alternative to a second, then the former should be socially preferred to the latter, a principle due to Vilfredo Pareto.

> *Weak Pareto*. For any alternatives x and y and any preference profile $\mathbf{R}$ in $\mathcal{D}$, if xP_iy for every individual i, then xPy.

Arrow contends that when socially ranking a pair of alternatives, all that should matter is what they are and how individuals rank them. Everything else, including the rest of the individual preference orderings, is irrelevant.

> *Independence of Irrelevant Alternatives*. If two preference profiles $\mathbf{R}$ and $\mathbf{R}'$ in $\mathcal{D}$ coincide on a pair of alternatives, then the corresponding social preference R and R' also coincide on this pair.

Arrow's final axiom ensures that nobody has undue influence in determining social preferences. Someone is a *dictator* if the social preference agrees with this person's preference on any pair of alternatives for which he expresses a strict preference.

> *Nondictatorship*. There is no dictator.

Arrow's Theorem shows that these axioms are incompatible if there are at least three alternatives. With only two alternatives, majority rule satisfies all four axioms. In this case, the requirement that social preferences be transitive is vacuous.

The Arrow axioms are logically independent and any four of them are consistent. There is an extensive literature that considers the implications of dropping or weakening some of the Arrow axioms. For the most part, this has not been a promising way to escape from the nihilism of Arrow's Theorem.

Arrow's axioms are not meant to be an exhaustive list of the desirable properties for a social choice function. For example, it may be desirable to treat individuals symmetrically and alternatives neutrally.

> *Anonymity*. For any preference profiles $\mathbf{R}$ and $\mathbf{R}'$ in $\mathcal{D}$ for which $\mathbf{R}'$ is obtained from $\mathbf{R}$ by permuting the individual preferences, the corresponding social preferences R and R' coincide.

> *Neutrality*. For any preference profiles $\mathbf{R}$ and $\mathbf{R}'$ in $\mathcal{D}$ and any alternatives w, x, y, and z, if wR_ix if and only if yR'_iz for every individual i, then wRx if and only if $yR'z$.

Pairwise majority rule satisfies both of these axioms. With Anonymity, social preferences are invariant to permutations of the preferences among the individuals. With Neutrality, descriptions of alternatives

are irrelevant—only the pattern of preferences (person by person) matters. Neutrality implies Independence of Irrelevant Alternatives. Anonymity is inappropriate if someone should have the right to decide between two particular alternatives (e.g., which clothes he wears). Similarly, Neutrality is inappropriate if some alternative, such as the *status quo*, should be given special treatment.

Arrow's Theorem can be reformulated in choice-theoretic terms. For each profile in the preference domain, a *social choice correspondence* chooses a subset of each *agenda* (a subset of X) in a domain of agendas. If a choice must be made from all two- and three-alternative agendas, the choice correspondence analogues of the Arrovian social welfare function axioms are incompatible.

2. Social Welfare Functionals

Amartya Sen (1977) has argued that the information used by a social welfare function is inadequate for ranking alternatives in terms of their contribution to social welfare. Specifically, the definition of a social welfare function prevents the social ranking from considering non-preference welfare information. Many social decision rules used in social and political philosophy such as utilitarianism or leximin are thereby ruled out from the outset. These rules can be formalized using Sen's concept of a social welfare functional.

In Sen's framework, the well-being of individual i is measured by a *utility function* U_i that assigns a number to each alternative in X. A *social welfare functional* F determines a social preference ordering of X for each profile of utility functions $\mathbf{U} = (U_1, \ldots, U_n)$ in some domain $\mathcal{D}$ of *profiles of utility functions*. With the *utilitarian* rule, x is socially strictly preferred to y if and only if the sum of individual utilities with x is larger than the corresponding sum with y. With the *leximin* rule, this social

ranking obtains if and only if the smallest utility that is not the same for both alternatives is larger for x than with y. Leximin underlies John Rawls' (1971) difference principle, but with utility substituting for an index of primary goods.

Welfarism is the principle that requires only utilities to matter when socially ranking alternatives. Given a profile **U**, the *utility vector* of alternative x is the ordered list of n numbers $U(x) = (U_1(x), \ldots, U_n(x))$ indicating each person's utility in x. The set of all such vectors is the n-dimensional Euclidean space $\Re^n$. A *social welfare ordering* is an ordering R^* of $\Re^n$. A social welfare functional F is *welfarist* if there exists a social welfare ordering R^* for which for every profile **U** in $\mathcal{D}$ and every pair of alternatives x and y in X, the social ranking of x and y is given by the ranking of the utility vectors $U(x)$ and $U(y)$ by R^*. Hence, in socially ranking two alternatives, the only information used is the utilities obtained with them, and not their descriptions or the utility functions used to generate these utilities. Utilitarianism and leximin, and many other social decision rules, are welfarist.

Welfarism is characterized by three axioms for F.

> *Unrestricted Domain*. $\mathcal{D}$ is the set of all possible profiles of utility functions.
>
> *Pareto Indifference*. For any alternatives x and y and any profile of utility functions **U** in the domain, x is socially indifferent to y if $U(x) = U(y)$.
>
> *Independence of Irrelevant Alternatives*. For any alternatives x and y and any profiles of utility functions **U** and **U**′ in the domain, if $U(x) = U'(x)$ and $U(y) = U'(y)$, then the social ranking of x and y coincides for these profiles.

Given Unrestricted Domain, a social welfare functional is welfarist if and only if it satisfies Pareto Indifference and Independence of Irrelevant Alternatives. With Pareto Indifference, the descriptions of alternatives are irrelevant, whereas with Independence of Irrelevant Alternatives, the utility functions used to generate the utility consequences are irrelevant.

The ability to make intrapersonal and interpersonal utility comparisons may be limited, which places constraints on what social welfare functionals can be employed. Informational restrictions are usually modeled by specifying what transformations to the utility functions preserve the meaningful utility information. A utility function U_i is *ordinal* if the only meaningful statements are those that are preserved by applying an increasing transform g_i to it. That is, the utility $U_i(x)$ is replaced by $V_i(x) = g_i(U_i(x))$. With an ordinal utility function, it is meaningful to make intrapersonal comparisons of utility levels. If the same increasing transform must be applied to each person's utility function, then utility levels are also interpersonally comparable. U_i is *cardinal* if the only meaningful statements are those that are preserved by applying an affine transform $g_i(U_i) = a_i + b_i\, U_i$, where a_i and b_i are constants with $b_i > 0$. With such transforms, both utility levels and differences in utility are intrapersonally comparable. If the b_i are the same for everyone, then utility differences are interpersonally comparable and if the a_i are also the same, then so are utility levels.

U_i implicitly defines a preference R_i by setting xR_iy if and only if $U_i(x) \geq U_i(y)$. Thus, if utility is ordinal and interpersonally noncomparable, Sen's framework reduces to that of Arrow because only individual preference information is meaningful.

Social welfare functionals have been used to axiomatize various social decision rules, including utilitarianism and leximin. Utilitarianism requires utility differences to be interpersonally comparable. With this informational assumption, Claude d'Aspemont and Louis Gevers (1977) axiomatize

utilitarianism by supplementing the welfarism axioms with the social welfare functional analogues of Weak Pareto and Anonymity. Leximin requires utility levels to be interpersonally comparable. Peter Hammond (1976) axiomatizes leximin by combining the welfarism axioms with Anonymity, Strong Pareto (any change that makes someone better off with hurting anyone else is a social improvement), and an equity axiom that regards any change that reduces the inequality in the utilities of any two individuals as being a social improvement.

Leximin gives absolute priority to the worst off, whereas utilitarianism gives no priority to anyone. Various social welfare functionals lie in between these two extremes. *Prioritarian* social welfare functionals compare alternatives according to the sum of transformed utilities $g(u_i)$, where g is an increasing, strictly concave transformation common to all individuals. This reflects Derek Parfit's (2000) *priority view* that we have stronger reason to benefit people the worse off they are, regardless of how well off others are. Matthew Adler (2019) has defended a particular class of prioritarian rules. *Sufficientarian* social welfare functionals give absolute priority to individuals below some threshold value.

Sen's framework has been criticized on the grounds that the utilities being considered contain no information about the scales on which they are measured. Sen (1977: 1542) himself observed that his informational invariance conditions are "unable to distinguish between (i) everyone having more welfare … and (ii) a reduction in the unit of measurement of personal welfares." Michael Morreau and John Weymark (2016) propose a generalization of a social welfare functional that takes account of the unit in which utility is measured so as to distinguish between these two kinds of changes. Jacob Nebel (2021) argues that once utilities are thought of as being dimensioned quantities, the standard arguments for the invariance conditions fail.

In order to address problems in population ethics, Sen's framework has been extended to allow for different population sizes by Charles Blackorby and David Donaldson (1984). Alternatives are now complete histories of the world, with possibly different people alive in different histories. For each history, only the individuals who ever live in it have utilities. Utilities are lifetime utilities, with a utility of zero representing a life that is just worth living. Welfarism now requires that the social ranking of alternatives be determined by a social welfare ordering of all possible vectors of individual utilities for all possible population sizes.

Familiar variable-population social welfare orderings are *total* and *average utilitarianism*, which compare alternatives according to their utility sums and their average utilities, respectively. Total utilitarianism is subject to Parfit's (1984) *repugnant conclusion*: for any alternative in which everybody has the same arbitrarily high utility, there is a socially preferred alternative with a larger population in which everybody has lives barely worth living. Average utilitarianism has the drawback that it recommends that an individual should not be brought into existence if by doing so average utility decreases, even if this person is extremely well off and nobody else is affected.

To overcome these problems, Blackorby and Donaldson propose a *critical-level generalized utilitarian* rule, which ranks alternatives by first subtracting $g(c)$ from individual i's transformed utility $g(u_i)$ before summing, where g is a strictly increasing, concave transform with $g(0) = 0$ and c is a non-negative constant—the *critical level*. Adding someone with the critical level of utility to any society is a matter of social indifference. Total utilitarianism is the special case in which g is the identity transform and $c = 0$. For fixed population comparisons, prioritarian rules are the special case in which g is strictly concave. There is now an extensive literature that explores alternative formal approaches to population ethics.

3. Harsanyi on Utilitarianism

John Harsanyi (1955) uses expected utility theory to provide decision-theoretic foundations for utilitarianism. For the set of alternatives $X = \{x_1, \ldots, x_m\}$, a *lottery* is a probability distribution $p = (p_1, \ldots, p_m)$ that yields outcome x_j with probability p_j. Receiving x_j for certain is equivalent to the lottery in which $p_j = 1$. A preference R on the set of lotteries L on X is *represented* by the utility function U on L if pRq if and only $U(p) \geq U(q)$ for all lotteries p and q. R is an *expected utility preference* if it can be represented by a utility function for which the utility of a lottery is the sum of the utilities for each outcome in X weighted by the probabilities of their occurrence. Expected utility preferences are characterized by axioms that Harsanyi regarded as requirements of rationality.

For Harsanyi, welfare judgments are impersonal preferences expressed by a moral agent who ranks lotteries on social alternatives based on a sympathetic and impartial concern for each member of society. He modeled such preferences using a *veil of ignorance* argument. The moral agent engages in a thought experiment in which he imagines having an equal chance of being each person in society, complete with that person's preferences and objective circumstances. In this way, the problem of ranking social alternatives is transformed into a problem of individual decision-making under uncertainty in which the outcome once the uncertainty has been resolved is an *identity-outcome pair* consisting of being person i with social alternative x. Assuming that the preferences of each individual and the moral agent satisfy the expected utility axioms, Harsanyi argued that the moral agent must rank alternatives using the average utilitarian rule. In developing his principles of justice, Rawls (1971) employed a thicker veil in which it is not known what the preferences are outside the veil or who in fact exists.

Harsanyi's utilitarian conclusions have been challenged by Sen (1976) using an informal argument that was later formalized by John Weymark (1991). Only increasing affine transforms preserve the expected utility form of a utility function. If these are the only permissible transforms of the utility functions, which is what Harsanyi mistakenly thought to be the case, then the moral agent's preferences over identity-outcome lotteries would supply the interpersonal comparisons of utility differences required for utilitarianism to be meaningful. However, an expected utility preference need not be represented by a utility function with the expected utility form; any increasing transform of it represents the same preferences. This ordinality makes comparisons of utility differences meaningless, thereby undermining Harsanyi's utilitarian conclusions.

Harsanyi offered an alternative justification for a weighted form of utilitarianism based on aggregating a single profile of preferences over lotteries. It is also subject to the Sen–Weymark critique.

Not everybody regards the Sen–Weymark critique as fundamentally undermining Harsanyi's defense of utilitarianism. Hilary Greaves (2017) attempts to vindicate Harsanyi by arguing that terms like "well-being" and "utilitarianism" are semantically indeterminate. Nebel (2022) argues that the Sen–Weymark critique implicitly assumes that utilities are dimensionless.

4. Non-welfarist Criteria

Welfarist approaches to social choice focus only on the consequences of collective decision-making. However, non-welfarist criteria such as rights, fairness, and inequality are also important.

As Sen's (1977) *liberal paradox* demonstrates, some kinds of rights are inconsistent with the Pareto principle and a weak collective rationality requirement. Sen illustrated this conflict with an example of

two individuals deciding who should read an erotic novel. In alternative x, the prudish person 1 reads it; in y, it is read by lascivious person 2; and in z, nobody reads it. The prude prefers z to x to y because he thinks that nobody should read this smut, but if someone must, better him than the impressionable person 2. Person 2 prefers x to y to z because he thinks that it would be best if the prude expanded his horizons by reading the book, but if this is not possible, he would like to read it. If an individual has the right to determine the social ranking between alternatives that differ only in whether he reads the book, then socially y is preferred to z and z is preferred to x. If unanimous preferences are also respected, then x is preferred to y. Hence, the social preferences are cyclic.

Sen's version of the liberal requirement that everybody has some rights is given by the following social welfare function axiom.

> *Liberalism*. For each individual, there exist distinct alternatives for which the social preference agrees with this person's preference whenever it is strict.

Sen's Theorem shows that no social welfare function can satisfy Unrestricted Domain, Weak Pareto, and Liberalism if social preferences are required to be acyclic. Sen's response to this result is to abandon the Pareto principle. In terms of his example, he does not think that unanimity should be respected because the preferences are held for inappropriate reasons; they are meddlesome.

Allan Gibbard (1974) shows that Weak Pareto can be dispensed with in Sen's Theorem if Liberalism is strengthened by letting each individual determine the social ranking on *any* pair of alternatives that differ only in some features that lie within his private sphere (e.g., in what books he reads). Thus, the fundamental conflict is between liberal rights and having an unrestricted preference domain, not with Paretianism. Gibbard's solution to this conflict is to either only give an individual these kinds of rights

when his preferences for the alternatives in his private sphere do not depend on what anybody else gets or to allow individuals to waive some of their rights.

When reformulated as a problem of choosing one of the available social alternatives, the kinds of rights that Gibbard and Sen have considered give individuals the right to exclude some alternatives from being chosen. Robert Nozick (1974) does not believe that this is the correct way to model rights. He regards an individual as having the right to choose the features of a social alternative that lie within his private sphere, but not to make any decision about the complete social alternative. Unfortunately, as shown by Wulf Gaertner, Prasanta Pattanaik, and Kotaro Suzumura (1992), Nozick's approach does not circumvent the problems that Gibbard and Sen have identified.

Issues of fairness and inequality have been explored in the literature on axiomatic models of resource allocation. This literature considers a number of axioms that can only be defined using the special structure provided by economic models. By dispensing with the Arrovian independence axiom in either its social welfare function or social choice correspondence formulation, many allocation mechanisms with nice properties have been characterized. This framework has been used to examine the foundations of egalitarian theories and to investigate the implications of different fairness criteria. For example, Ronald Dworkin's (1981) idea that individuals should be compensated for inequalities due to factors beyond their control, but not for choices for which they are responsible, has been formalized and analyzed using this framework.

5. Judgment Aggregation

Judgment aggregation is concerned with making consistent collective judgments on a set of logically connected propositions based on the relevant group members' judgments on them. The formal analysis

of judgment aggregation originated with Lewis Kornhauser and Lawrence Sager's (1993) *doctrinal paradox*. A three-judge court considers three propositions:

p: the defendant contracted not to do x,

q: the defendant did x,

r: the defendant is in breach of contract.

All judges agree that r is true if and only if p and q are both true. Suppose that only judges 1 and 2 think that p is true, whereas only judges 1 and 3 think that q is. Thus, only judge 1 thinks that the defendant is in breach of contract. A majority agree on each proposition in $\{p, q, r \leftrightarrow (p \wedge q), \neg r\}$, but these are jointly inconsistent.

Christian List and Philip Pettit (2002) have generalized this example by identifying a set of mutually inconsistent properties for a procedure that aggregates judgments on logically connected propositions into a collective judgment. Their result has spawned a vast literature on judgment aggregation.

Judgment aggregation raises epistemic considerations, but it is not necessarily supposed that there is a correct judgment. When a jury engages in a collective decision, its objective is to make a correct decision about the guilt of the defendant. This raises the problem of how well different decision procedures track the truth. The *Condorcet jury theorem* shows that larger juries are more likely to make the correct decision using majority rule when there are only two possible choices if jurors make their decisions independently and have a better than even chance of making the correct decision.

6. Strategic Considerations

Arrow assumed that individuals' preferences are truthfully reported. However, if preferences are private information, it may be advantageous for an individual to misrepresent his preference so as to

obtain a preferred outcome. To address this issue, the social choice problem is modelled using a social choice function f which chooses one alternative from the set of alternatives X for each preference profile in the domain $\mathcal{D}$.

Individuals must be provided with incentives to truthfully report their preferences if the desired outcome $f(\mathbf{R})$ is to be chosen when the true preference profile is $\mathbf{R}$. A social choice function is *strategy-proof* if nobody can ever obtain a preferred outcome by reporting a false preference. A social choice function that takes account of the fine structure of individual preferences is typically manipulable.

Point voting procedures such as the Borda rule are manipulable. The *Borda rule* for m alternatives has each person assign m points to his most-preferred alternative, $m-1$ points to his second-best alternative, and so on, with the chosen alternative being the one with the most points (ties are broken according to some fixed rule). Suppose that the preferences of two individuals for four alternatives are given in descending order by x, z, y, w for person 1 and y, x, w, z for person 2. For this profile, x is chosen. Person 2 can manipulate the outcome so that his most-preferred alternative y is chosen by falsely reporting that x is his worst alternative.

An alternative is *Pareto optimal* if there is no other alternative that is unanimously preferred to it. *Pareto Optimality* is the requirement that only Pareto optimal alternatives are chosen. Allan Gibbard (1973) and Mark Satterthwaite (1975) have independently shown that if a strategy-proof social choice function satisfies Unrestricted Domain and Pareto Optimality, then it must be dictatorial when there are three or more alternatives ever chosen. There is now a vast literature that considers the implications of strategy-proofness when the domain of preferences is not unrestricted.

Agenda manipulation is another way in which strategic considerations may affect what is chosen. For example, sequential voting applied to the alternatives and preference profile used in Condorcet's Paradox is subject to agenda manipulation. Suppose that majority rule is first used to determine a winner from one pair of alternatives and then used to choose between the first-round winner and the third alternative. Depending on the order in which alternatives are compared, any outcome may result. For example, x is ultimately chosen if the first vote is between y and z. Thus, whoever determines the voting order can ensure that his favored outcome is chosen. This scenario arises in legislative and committee settings in which the standard voting rules result in three alternatives being considered: the *status quo* x, a motion y, and an amendment z. Had z been the motion and y the amendment, a different outcome may result. To avoid this kind of manipulation, it is sufficient for social preferences to be acyclic.

7. Majority Rule

The nihilistic conclusions of the Arrow and Gibbard–Satterthwaite Theorems can be avoided if the set of alternatives is one-dimensional and preferences are single-peaked. When this is the case, majority rule has many attractive properties.

The set of alternatives X is *one-dimensional* if the alternatives can be identified with points on the real line $\Re$. For example, the alternatives might be nonnegative numbers, each of which specifies a possible expenditure on some public good; or they might be numbers assigned to candidates for an election, with higher numbers assigned to more right-wing candidates. In such applications, it is natural to suppose that individual preferences are single-peaked. A preference R is *single-peaked* if there is a unique best alternative in X, the *peak*, and alternatives on the same side of the peak are worse the farther they are from it.

With single-peaked preferences, Duncan Black (1948) has shown that pairwise majority rule results in a social preference that is an ordering if the number of individuals is odd. With an even number, the social strict preference is transitive, but the social indifference relation need not be. Black has used these observations to establish his *median voter theorem*, which shows that for any single-peaked preference profile, a median preference peak defeats every other alternative in a pairwise majority vote and that choosing a median peak is strategy-proof. This theorem has gained widespread use in formal models of political decision-making.

Majority rule loses these attractive features when the set of alternatives is multidimensional. For example, suppose that there are $m \geq 2$ issues each of which can take on any value in $\Re$ and that everybody has a *Euclidean spatial preference*. For such a preference, there is a best alternative x, with y strictly preferred to z if and only if it is closer to x. A Euclidean spatial preference is a multidimensional generalization of a single-peaked preference that is often used in political models. Charles Plott (1967) has shown that when individuals have Euclidean spatial preferences on a multidimensional issue space, typically there is no alternative than can beat every other alternative in a pairwise majority vote. For such an alternative x to exist, it must be the case that individuals who do not have x as their ideal points can be paired up in such a way that the individuals in each pair most prefer to move away from x in exactly opposite directions.

A *top cycle* T is a set of alternatives for which (i) for any finite subset of T, there is a strict preference cycle among them and (ii) no alternative not in T is strictly preferred to any alternative in T. Plott's result would not be particularly troubling if majority rule reliably generated a top cycle T confined to a small region of the issue space. For then it would not make much difference which alternative in T is chosen because all of them are majority preferred to any alternative outside T. Unfortunately, Richard

McKelvey's (1976) *Chaos Theorem* shows that except in the rare circumstances in which Plott's conditions hold, the top cycle set is the whole set of alternatives. Thus, with sequential pairwise majority vote, an agenda setter can devise a sequence of votes that ensures that his most-preferred alternative is chosen.

8. Democratic Theory

James Buchanan's (1954) review of Arrow's book examines the philosophical assumptions of Arrow's framework. He argues that Arrow's Social Ordering condition implicitly treats society as an organic entity with its own values, which runs counter to Buchanan's methodological individualism. For a methodological individualist, Social Ordering is irrelevant because there is no need for a collective entity with its own ends. Majority rule intransitivities arise because different individuals comprise a majority on different pairwise choices. For Buchanan, this has the virtue of preventing a tyranny of the majority. Accordingly, Buchanan regards transitivity as being an undesirable consistency condition on collective choices. Unlike Arrow, Buchanan does not take individual values as given but instead regards them as being shaped by democratic deliberation. In his view, deliberation may lead to consensus on constitutional principles but should not allow for majorities to coalesce on non-constitutional issues. Buchanan further argues that Arrow's analysis is inapplicable to market mechanisms, because "the market does not call upon individuals to make a collective decision at all" (p. 122). Nozick's entitlement theory of justice shares this individualistic perspective.

William Riker (1982) contrasts two views of democracy: a liberal view in which voting serves to constrain official tyranny and a populist view in which voting is used to articulate a Rousseauian general will. He argues that the Arrow and Gibbard–Satterthwaite Theorems show there to be no

such thing as the general will. According to Riker, this is because (i) alternative fair and reasonable procedures for aggregating individual values produce different outcomes, none of which has a better claim to representing a general will than the others, and (ii) even if there is an agreed-upon aggregation procedure, the possibility of cycles and of mispresenting preferences or manipulating the agenda makes it dubious that what is chosen reflects the true will of the populace. Riker further argues that liberalism (as he interprets it) is compatible with social preference cycles and manipulation, and so is immune to the social choice impossibility theorems. Riker contends that a liberal polity with veto powers is difficult to maintain because it is vulnerable to populist threats. In his view, overcoming these threats requires the kind of political institutions established in the American Constitution with its checks and balances and division of responsibilities.

Riker's critics have raised several concerns about his analysis. The most thoroughgoing critique has been offered by Gerald Mackie (2003). He argues that majority rule does not produce cycles in practice and so can be thought of as a procedure that identifies a general will. He supports this conclusion by disputing the claims Riker makes about the ubiquity of cycles and manipulation in the history of American politics.

Deliberative democrats regard the legitimacy of collective decisions as depending on public deliberation supported by publicly justifiable reasons. John Dryzek and Christian List (2003) have provided a systematic analysis of how deliberation may attenuate social choice theory's negative findings. For example, they argue that deliberative democracy may encourage cooperative behavior and sanction untruthfulness, thereby circumventing the Gibbard–Satterthwaite Theorem. Further, the need to publicly justify one's views may facilitate agreement on a salient dimension on which preferences are single-peaked, thereby circumventing Arrow's

Theorem. The social welfare functional literature abounds with positive results, but it is concerned with the aggregation of welfare-representing utilities, which are subject to disagreement and difficult to measure. Dryzek and List suggest reinterpreting this formal framework using a different evaluative variable to represent individual interests (e.g., an index of primary goods), one that is susceptible to public verification. Public deliberation could serve to identify this variable and to determine how the individual values of this variable should be aggregated.

The axiomatic method of social choice theory might be thought, when successful, to justify the outcomes of certain collective decision-making mechanisms based only on intrinsic properties of those procedures, as opposed to external, substantive standards of good political decisions or truths of political morality. David Estlund (2008) questions this view. The authority and legitimacy of the jury system relies on its propensity to make correct decisions. Similarly, Estlund argues that democratic institutions are authoritative and legitimate because they are more conducive to making correct decisions than any other generally acceptable procedure.

Related Topics

Discourse Theory; Intergenerational Distributive Justice; Rational Choice Theory; The Difference Principle; Utilitarianism and Consequentialism

References

Adler, M. D. (2019) *Measuring Social Welfare: An Introduction.* New York: Oxford University Press.

Arrow, K. J. (1951) *Social Choice and Individual Values*. New York: Wiley. Second Edition, 1963.

Black, D. (1948) "On the Rationale of Group Decision-Making," *Journal of Political Economy* 56: 23–34.

Blackorby, C. and Donaldson, D. (1984). "Social Criteria for Evaluating Population Change," *Journal of Public Economics* 25: 13–33.

Buchanan, J. M. (1954) "Social Choice, Democracy, and Free Markets," *Journal of Political Economy* 62: 114–123.

d'Aspremont, C. and Gevers, L. (1977) "Equity and the Informational Basis of Collective Choice," *Review of Economic Studies* 44:199–209.

Dryzek, J. S. and List, C. (2003) "Social Choice Theory and Deliberative Democracy: A Reconciliation," *British Journal of Political Science* 33: 1–28.

Dworkin, R. (1981) "What is Equality? Part 2: Equality of Resources," *Philosophy and Public Affairs* 10: 283–345.

Estlund, D. M. (2008) *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press.

Gaertner, W., Pattanaik, P. K., and Suzumura, K. (1992) "Individual Rights Revisited," *Economica* 59: 161–167.

Gibbard, A. (1973) "Manipulation of Voting Schemes: A General Result, *Econometrica* 41: 587–601.

Gibbard, A. (1974) "A Pareto Consistent Libertarian Claim," *Journal of Economic Theory* 7: 388–410.

Greaves, H. (2017) "A Reconsideration of the Harsanyi–Sen–Weymark Debate on Utilitarianism," *Utilitas* 29: 175–213.

Hammond, P. J. (1976) "Equity, Arrow's Conditions, and Rawls' Difference Principle," *Econometrica* 44: 793–804.

Harsanyi, J. C. (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63: 309–321.

Kornhauser, L. A. and Sager, L. G. (1993) "The One and the Many: Adjudication in Collegial Courts," *California Law Review* 81: 1–59.

List, C. and Pettit, P. (2002) "Aggregating Sets of Judgments: An Impossibility Result," *Economics and Philosophy* 18: 89–110.

Mackie, G. (2003) *Democracy Defended*. New York: Cambridge University Press.

McKelvey, R. D. (1976) "Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control," *Journal of Economic Theory* 12: 472–482.

Morreau, M. and Weymark, J. A. (2016) "Measurement Scales and Welfarist Social Choice," *Journal of Mathematical Psychology* 75: 127–136.

Nebel, J. M (2021) "Utils and Shmutils," *Ethics* 131: 571–599.

Nebel, J. M. (2022) "Aggregation Without Interpersonal Comparisons of Well-Being," *Philosophy and Phenomenological Research* 105: 18–41.

Nozick, R. (1974) *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, D. (1984) *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. (2000) "Equality or Priority?" in M. Clayton and A. Williams (eds.), *The Ideal of Equality*, Basingstoke: Palgrave Macmillan, 81–125.

Plott, C. R. (1967) "A Notion of Equilibrium and Its Possibility Under Majority Rule," *American Economic Review* 57: 787–806.

Rawls, John (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Riker, W. H. (1982) *Liberalism Against Populism*. San Francisco: W. H. Freeman.

Satterthwaite, M. A. (1975) "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorem for Voting Procedures and Social Choice Functions," *Journal of Economic Theory* 10: 187–217.

Sen, A. K. (1970) "The Impossibility of a Paretian Liberal," *Journal of Political Economy* 78: 152–157.

Sen, A. K. (1976) "Welfare Inequalities and Rawlsian Axiomatics," *Theory and Decision* 7: 243–262.

Sen, A. K. (1977) "On Weights and Measures: Informational Constraints in Social Welfare Analysis," *Econometrica* 45: 1539–1572.

Weymark, J. A. (1991) "A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism," in J. Elster and J. E. Roemer (eds.), *Interpersonal Comparisons of Well-Being*, Cambridge: Cambridge University Press, pp. 255–320.

Further Reading

Textbooks include D. Austen-Smith and J. S. Banks, *Positive Political Theory I: Collective Preference* (Ann Arbor: University of Michigan Press, 1999); W. Gaertner, *A Primer in Social Choice Theory*, Second Edition (Oxford: Oxford University Press, 2009); J. E. Roemer, *Theories of Distributive Justice* (Cambridge, MA: Harvard University Press, 1996); and A. K. Sen, *Collective Choice and Social Welfare, Expanded Edition* (Cambridge, MA: Harvard University Press, 2017). Topics in social choice theory are surveyed in K. J. Arrow, A. K. Sen, and K. Suzumura (eds.), *Handbook of Social Choice and Welfare* (Amsterdam: North-Holland; Vol. 1, 2002; Vol. 2, 2011). For welfarist approaches to social evaluation, see Adler (2019) and W. Bossert and J. A. Weymark, "Utility in Social Choice," in S. Barberà, P. J. Hammond, and C. Seidl (eds.), *Handbook of Utility Theory: Vol. 2, Extensions* (Boston, Kluwer, 2004), 1099–1177. For population ethics, see C. Blackorby, W. Bossert, and D. Donaldson, *Population Issues in Social Choice Theory, Welfare Economics, and Ethics* (Cambridge: Cambridge University Press, 2005. For judgment aggregation, see C. List (2012) "The Theory of Judgment Aggregation: An Introductory Review," *Synthese* 187: 197–207. For applications of social choice theory to democratic theory, see Dryzek and List (2003) and J. W. Patty and E. M. Penn, *Social Choice and Legitimacy* (Cambridge: Cambridge University Press, 2014).