Social Choice

Josh Cohen

1. The Voting Paradox

The puzzles in the area of social choice begin with the paradox of voting. So suppose we as a group are trying to decide when the course should meet: 9, 10, or 11. Suppose that choosing rationally (whether for a group or person) is choosing the best alternative from a consistently ordered set of feasible alternatives, and that the group ranking is fixed by majority support: A is ranked higher than B by the group iff A is ranked higher than B by a majority of group members. Assume finally that consistency is a matter of transitivity of “better than”: that if A is better than B and B is better than C, then A is better than C (we will see later that there are ways to weaken this notion).

Assume that we divide into three groups of equal size:

	Best	Second	Worst
I	9	10	11
II	11	9	10
III	10	11	9

Lets ask how the group ranks the three possibilities, using majority rule as the procedure for group ranking: 9 beats 10, 10 beats 11, and 11 beats 9. Group rationality (i.e. transitivity) requires that 9 beat 11. But 11 is a majority rule winner over 9. In fact matters are worse: putting aside the failure of transitivity, notice that each alternative loses to another alternative. Thus we have not just a failure of transitivity, but further a failure of acyclicity, which is an even more minimal condition of rational decision:

$$\textit{Acyclicity}\text{: If } x_1Px_2Px_3...x_{n-1}Px_n\text{, then not } x_nPx_1$$

A violation of acyclicity means that the feasible set contains no best element. If rationality means choosing the best from the feasible, then there is no basis here for rational choice.

One might wonder whether there is something very special about this case. The answer appears to be: No. Assume we have at least three people and a set of alternatives that vary continuously on two dimensions (e.g., guns and butter; spending on environmental regulation and economic growth). Two things turn out to be true: Using majority rule as the basis of the collective ordering, there is virtually never a best element—that is, everything loses to something.[1] And, what is worse, the majority rule cycle is global. That is, beginning from any alternative, we can arrive at any other point by finitely many applications of majority choice: including points that are worse for everyone that the original (outside the pareto set). So if I tell you the preferences of the members of a group and that the group will make its decisions by majority rule, this information, taken on its own, will not enable you to narrow the class of outcomes at all. The product of majority-rule procedures will simply a result that reflects features of circumstances (say, the order of decision-making, permissible amendment procedures, institutions of decision-making), but will lack any deeper rationale in

---

[1] Special symmetry conditions are required to prevent this.

a rational group decision based (via majority rule) on individual judgments, values, and tastes.

2. Social Choice Theory

Faced with these majority rule troubles, one might wonder if there are special troubles with rational group decisions using majority rule as a method of aggregation. Perhaps there is some alternative to majority rule that enables us to base a group or collective ranking of alternatives on rational individual rankings of those alternatives, where the collective ranking is itself rational. Thus Arrow says (SCIV, 2): “The methods of voting and the market…are methods of amalgamating the tastes of many individuals in the making of social choices. The methods of dictatorship and convention are, or can be, rational in the sense that any individual can be rational in his choices. *Can such consistency [rationality] be attributed to collective modes of choice, where the wills of many people are involved*?”

Social choice theory explores this possibility: of basing rational collective choices on rational individual choices. The field of social choice theory provides a way to formalize these two fundamental ideas: (1) the idea that a group or collective ranking is constructed from rational individual rankings, with the members treated as equals; and (2) the idea that the group acts rationally. We can then study whether the difficulties in the case of majoritarian conception of group preference can be circumvented if we opt for some other account of group judgment, while preserving the idea that this judgment must bear some systematic connection to the rankings of alternatives by individual members.

Framework

I want now to sketch out the main elements of the social choice theory framework. In each case, I will make some comments on interpretation that will figure in later discussion.

1. *Alternatives*. Individuals face a set of alternatives (policies or outcomes) that they personally rank and over which we are looking to characterize a group ranking. Let the set of alternatives be X={x, y, z,...}. In Arrow, the alternatives ordered by individuals are vectors that represent total social states: complete specifications of all the goods of each person, the resources devoted to each economic activity, the level of government activity, etc.[2]

2. *Individuals*. There is a set of individuals I={1,2,...,n}, each of whom has an ordering R_i (complete, reflexive, transitive) over X: thus, we read xR_iy as: for individual i, x is at least as good as y; xP_iy as x is preferred to y; and xI_iy as individual i is indifferent between x and y. We define R in terms of P and I: xRy $=_{df}$ xPy or xIy.

Completeness: For all x,y in X, either xR_iy or yP_ix.

Reflexivity: for all x in X, xR_ix

Full Transitivity: for all x,y,z in X, if xPy and yPz, then xPz; if xIy and yIz, then xIz; (and similarly for PI/P and IP/P)

---

[2] Arrow, 17.

An array of such orderings, one for each individual, is called a profile: $[R_i]$. Because X is finite, the ordering can be represented by an ordinal utility function. For every person, there is a function u_i from the set X of alternatives to the real numbers, such that xP_iy iff $u_i(x)>u_i(y)$. And of course if there is one such function, there an infinitely many others.

3. *Social Preference Relation*. Let R be the social preference relation: so 'xPy' is to be read as "x is socially better than/preferred to y;" 'xRy' that "x is socially at least as good as y." What "social preference" comes to is not so clear, and two interpretations are available. One is that the social ranking represents a ranking of alternative social states to be used by officials—or by citizens when making public political judgments—in assessing policies that will be made and enforced by the state (107).

Alternatively, the social ranking can be thought of as the product of some institution for making decisions (market, elections, committee) in which the inputs to the institution are individual preferences or evaluations of alternatives. Here "x is socially better than y" comes out as: institutions are so arranged that y is not chosen when x and y are both available (analogous to a revealed preference idea for individual preferences).

4. *Aggregation Function*. A collective choice rule is a function (many-one) that assigns a social preference ranking to profiles of individual rankings (may be undefined for some rankings, unless it satisfies unrestricted domain):

Collective Choice Rule: $R=f([R_i])$

If the function’s range is restricted to full transitive social preference relations, then it is called a "social welfare function." If it is confined to acyclic social preference relations, it is called a "social decision function."

To see the difference between social welfare functions and social decision functions, consider the following social preference relation that is not transitive: xPy, yPz, xIz. While it fails to be transitive (or even quasi-transitive[3]), it is acyclic since z is not preferred to x; only one element is at least as good as every other (last man standing). Intuitively, this seems sufficient for arriving at a rational decision: for choosing the best from the feasible.

Corresponding to the two interpretations of the social preference relation, there are two interpretations of a collective choice rule. On the one hand, we can think of it as a way of determining what the social choice ought to be, given any array of individual preferences: that is, as a political principle or basis of political advice. So, for example, take Majority Rule as the SWF: then, the collective choice ought to be the alternative that is majority-preferred over all alternatives. This might be implemented in a variety of ways: for example, a fully informed benevolent dictator might pick the majority preferred alternative. On the other hand, we might think of majority rule as an institution which picks the majority-supported alternative as the winner.

In an interesting paper from the early 1950s, James Buchanan criticizes the idea of a social welfare (or social decision) function, on both interpretations. The idea of the social welfare function as an aggregation rule is objectionable

---

[3] Quasi-transitive: strict preference is transitive, but indifference is not.

because the very idea of a social preference or collective judgment requires that we think of society as an entity with "an existence apart from that of its individual components" (116). You might alternatively (see Arrow's discussion of ideas of consensus) think that the idea of a collective judgment makes sense, but reject the idea of constructing it from individual orderings, rather than from expressions of opinions of what is socially best. On the institutional interpretation: Buchanan argues that an apparently elementary condition of "rationality"—consistency or coherence in decisions—is not a desideratum for a social decision process. He claims that a consistency requirement fosters rule by consolidated majorities rather than shifting coalitions of minorities, and that it stands in the way of social experimentation (whereby "competing alternatives may be experimentally and provisionally adopted, tested, and replaced by new compromise alternatives").

Some Conditions on Social Choice Rules

The aim of social choice theory is to formalize requirements on collective choice rules—on the aggregation of individual preferences into a collective will—and to consider whether or not it is possible for a collective choice rule to jointly satisfy those conditions. The requirements are normative (as Arrow puts it, they express value judgments), and the concern is with the consistency of those judgments). In the cases of particular interest to us, they reflect the idea that the collective choice is to combine the interests of individuals, understood as equals. What we want to know is whether there is any aggregation that both treats individuals as equals—where the conditions provide precise expression to that intuitive idea—and that results in a rational collective decision.

1. *Unrestricted Domain*: This condition says that the domain of the collective choice rule includes all logically possible individual profiles, for i=1,2,...,n over X. Recall that a profile is collection (an n-tuple) of orderings. So unrestricted domain does not require that the collective choice rule yield a social ranking over all possible rankings, but over all possible rankings that meet the minimal rationality conditions.

U is one possible condition on a collective choice rule, but not a condition mandated by the social choice framework. So for example, an alternative condition would be to say that the domain of a collective choice rule is to be limited to collections of rankings that human beings might plausibly hold, or to substantively reasonable individual rankings.

It is sometimes said (by way of objection) that social choice theory takes preferences as given. And in a way that is true. It treats preferences and values as the domain of the collective choice rule, and does not model the ways that processes of collective decision may shape preferences and values: "individual values are taken as data and are not capable of being altered by the nature of the decision process itself" (7). But U takes preferences and values as given in a stronger sense that is not part of the social choice framework itself. It requires that a collective choice rule give us rational decisions, whatever those inputs are.

The next four conditions are "responsiveness" requirements. These conditions require that collective preferences depend positively on the preferences (orderings) of individuals. The intent of these conditions is to capture the idea that individuals are to be treated as equals in the aggregation of preferences, but—as we will see—the conditions in effect render the idea that individuals are to be treated as equals by requiring that the aggregation give equal weight to the individual orderings.

2. *Pareto*: if one alternative if higher than another in each individual ordering, then it is higher in the social ranking (for all i, xP_iy then xPy). This is an apparently weak condition of positive responsiveness; I don't want now to explore objections to it, only to say that it is weaker than the condition of responsiveness that is commonly known as monotonicity.

3. *Monotonicity*: if an alternative improves in at least one individual ranking, then it should not decline in the social ranking: if it is already the winner, it should remain so; if it in a tie, it should not lose. This condition is stronger than Pareto because it constrains the collective choice even in cases in which there is no unanimous judgment among individuals.

You can see the power of monotonicity by noting that systems with elections that occur in stages violate it: for example, systems with a runoff between the top two candidates in a first round election. Thus suppose we have an election with three candidates: Bush, Kerry, and Nader. In case 1, the vote is 8-7-6 (respectively) in the first round; in the second, the Nader voters—who are evenly divided on their second choice—split their votes and Bush wins 11-10. In case 2, the preferences are the same except that two of Kerry's supporters have shifted to Bush: so Bush wins 10-5-6. But in round 2, we have a run-off between

Bush and Nader, and all five Kerry supporters support Nader as their second choice. So Nader wins 11-10: Bush does worse by doing better, and this violates the monotonicity condition, which rules out electoral schemes that permit such negative responsiveness.

3a. *Strong monotonicity*: this condition treats everyone as a tie-breaker. So if the division of support is such as to generate a tie or victory for Jones, then if one vote shifts to Jones, then Jones wins.

4. *Liberty*: for each person, there exists at least one pair of alternatives (x,y) such that the social ranking (which is socially better) is fixed by the individual's own ranking. if xP_iy then x is socially better (xPy) and if yP_ix then y is socially better (yPx).

Note that Liberty does not capture the idea of an individual's having *control* over a decision: indeed, Liberty would be satisfied by an omniscient, benevolent, non-paternalistic dictator (in the familiar non-social choice theoretic sense of dictator). Thus if the dictator knew what subjects preferred, was concerned with their welfare, and non-paternalistically accepted their own preferences as dispositive as to their welfare, then the dictator's judgments would satisfy Liberty. So it is another condition on responsiveness, this time to particular persons.

4a. *Minimal Liberty*: for at least two individuals (though perhaps not for all) there exists at least one pair of alternatives such that the social ranking is fixed by the individual's own ranking: two persons, and two pairs of alternatives, such that if xP_iy then xPy and if yP_ix then yPx.

5. *Non-Dictatorship*: it is not the case that there exists an individual whose ordering matches the social ranking in this sense: xP_iy if and only if xPy. Notice that this is a very special sense of "dictatorship," in a way that parallels the special sense of "right" in the Liberty condition. Thus the dictator is not someone who controls outcomes, but someone who rankings match the social ranking: put otherwise, a dictator is someone who has authority over all pairs, in just the way that the right-bearers in Liberty have authority over one pair.

The final three conditions are "invariance requirements." Formally speaking, an invariance requirement is a condition on a function, such that the output of the function remains fixed so long as certain input conditions remain fixed, not matter what happens to other features of the world: so that output is invariant under various sorts of transformation. In the case at issue, invariance requirements require that the aggregation function give the same result so long as certain aspects of the individual orderings are held fixed. So they are invariance conditions because they require that the function from individual rankings to a social ranking be invariant even as the characteristics of individuals or their circumstances change. The invariance conditions have the effect of limiting the range of information that the collective choice rule can look at by requiring that the output of the function remain fixed, even if certain conditions change. Less abstractly put, the effect of the invariance conditions is to force collective decisions to pay attention only to certain facts about individual preferences.

6. *Independence of Irrelevant Alternatives*: roughly, the social ranking of a pair of alternatives is to depend only on the rankings of that pair by individuals. So if we take two profiles in which all the individual rankings of the two alternatives are identical, then the social ranking from those profiles must be the same, no matter how much the other features of the situation vary. To see the effect of this condition, consider the following array:

Smith x,y,z

Jones x,y,z

Brown y,x,z

Suppose now that the collective choice works this way: assign 3 points to the best, then 2, then 1; add up, and determine social rankings by the scores. So with this array, x is socially better than y since x beats y 8 to 7. But now imagine an array the same but for Brown, who reverses the order of x and z. With this array, x and y now tie for the socially best. But the change in the social ranking of x and y (from preference to indifference) does not reflect any change in the individual orderings of x and y, which remain precisely as before. So this collective choice rule violates I. So does a utilitarian collective choice rule, since cardinal utilities are necessary (though not sufficient) for the summation of utilities, but constructing a cardinal utility measure requires information (roughly, about intensities) not contained in pairwise comparisons.

I is very strong, particularly if we think of a collective choice rule as an aggregation rule. It says, in effect, that the only information relevant to social choice—the only facts relevant as a foundation for political advice—are facts about pairwise individual orderings. So collective choice cannot consider preference intensities, or depend on interpersonal comparisons of utilities: this is a point that Arrow emphasizes, when he says that "the viewpoint will be taken here that interpersonal comparison of utilities has no meaning" (9). But I is much stronger, because it says that the collective ordering must remain invariant whatever changes there may be in objective conditions as well: position in the distribution of resources; their health, abilities, needs, etc..[4] The collective preference is to be fully fixed by individual pairwise orderings, in the following sense: if we impose the Independence condition, then facts other than individual preferences can freely vary without affecting the collective preference, which is fixed simply by the order in individual preferences: same orderings, same collective decision.

Independence is a natural condition to impose if we identify treating individuals as equals with giving equal weight to their preferences. But suppose we think that the idea that individuals are to be treated as equals supports giving additional weight to the well-being of the person whose circumstances are worst. Then we need to reject I, because it does not permit us to look at any factors beyond preferences.

7. *Anonymity*: This condition requires that the collective choice depend only on the preferences (individual rankings), not on who has the preferences. Thus, if we hold the array of preferences fixed, while permuting the individuals who have the preferences, the result must be the same. This condition appears

---

[4] See Amartya Sen's review of Arrow's collected papers [ref].

to be associated with a requirement of equality: for example, it excludes schemes of weighted voting of the kind favored by John Stuart Mill, who thought that educated voters should get multiple votes. But notice that it also excludes a scheme in which a winner is chosen on the basis of the number of districts the candidate wins. Suppose there are 15 voters equally divided into three districts. Then if the division of the vote is 3-2/3-2/3-2 in favor of Jones over Smith, Jones wins. But now suppose it is 5-0/2-3/2-3. Then Smith wins: even though the set of preferences is the same, the fact that the preferences are differently distributed across voters changes the result. Notice, too, that it is not obvious that this scheme—as distinct from a system of weighted voting—violates the requirement of treating people as equals. Indeed, the scheme of district voting assigns no one, ex ante, greater impact on the outcome than anyone else.

8. *Neutrality (Outcome Indifference)*: The intuitive idea is that the social choice function should not itself favor one or another alternative. The social choice should, so to speak, not depend on intrinsic features of the alternatives but only on how they are ranked by individuals. So if each person's preferences over a pair of alternatives reverses, the resulting social choice should also reverse. This requirement also seems intuitively plausible. But notice that it is violated if we require, as a way of protecting majorities—including the political rights of minorities—agreement from a large majority before passing a law restricting those rights (say, amending the constitution): if the vote is 60-40 against, the amendment loses; but if it is 60-40 in favor, the amendment still

loses. In such a case, it seems entirely appropriate that the procedure for making decisions violate neutrality.

Some Results

Having described the framework of collective choice theory, and some of the conditions that have been proposed as appropriate to impose on collective choices rules, lets now consider some of the central results, all of which are about the implications of imposing these conditions.

1. *May's Theorem*: May proved that, when there are two alternatives, simple majority rule is the only social choice function that satisfies Unrestricted Domain, Strong Monotonicity, Anonymity, and Neutrality: the alternative with one more than half the support wins. If there are more than two alternatives, then majority rule is no longer a social decision function (as the voting paradox indicates).[5]

2. *Sen's Liberalism Theorem*. The Paretian-Liberty Theorem states that there is no social decision function that satisfies conditions U, P, and L: that is, no social decision function that, whatever the individual rankings, produces a social ranking that both respects the rankings of individuals over the alternatives that lie within their personal domains, and at the same time is responsive to the rankings of members, taken collectively.

---

[5] For discussion of May's theorem, see Ackerman, *Social Justice in the Liberal State;* Dahl, *Democracy and Its Critics*; Beitz, *Political Equality*.

For reasons that will become clear, all of the examples that illustrate the problem have an artificial air. Suppose, then, that the legislature is to have a debate that will issue in a vote. The following conditions obtain:

1. There are two members (parties) each of whom will be given an opportunity to speak.

2. The rules of the body set aside 2 hours for debate.

3. There must be some debate before the vote, but all the time need not be used up.

4. Remarks may last for either 75 minutes or 45 minutes, but there cannot be two 45 minute speeches.

5. Each party has the right to respond: if A uses some time, then B has the right to use some time.

6. Although each party would prefer to keep the debate short each also wants to talk longer than the opposing party, and in fact thinks that as between talking shorter and not talking at all, better to not talk at all. Here then are the preferences of the parties:

Party A	Party B
(45, 0)	(0,45)
(75,45)	(45,75)
(0,45)	(45,0)
(45,75)	(75,45)

Now, the problem: Since A has the right to respond (0,45) is beaten by (75,45); since B has the right to respond, (45,0) is beaten by (45,75). But (75,45) is beaten by (45,0): if there, in any case, to be an imbalance of time, reduce the time spent. And (45,75) loses to (0,45): same theory. So everything loses: two alternatives by the exercise of rights, two by the application of Pareto. Put otherwise, if A faces (0,45) then A will exercise the right to reply, giving (75,45). But both prefer (45,0). But if they face (45,0), then B will exercise the right of reply, and talk for 75 minutes. But then they would both prefer that B be briefer and A not bother at all.

That's an example. I will not go through the proof, which is pretty straightforward.

3. *Arrow's Theorem*. We come now to Arrow's theorem. What Arrow showed is that there is no *social welfare function* that satisfies Unrestricted Domain, Independence, Pareto, and Non-Dictatorship. Put otherwise, if a social welfare function satisfies Pareto, Independence, and Unrestricted Domain, then there must be a Dictator in the special sense noted earlier: not someone who actually controls the decisions (we may not even be aware who the dictator is), but someone whose ranking, whatever it is, corresponds to the social ranking.

The conditions on a social welfare function are meant to capture the intuitive idea that a coherent collective ranking, suited to guiding collective decisions, should depend on individual rankings, where individuals are treated as equals. So the conclusion is that the very idea of rational collective choice may be incoherent, where such choice is understood to require picking the best element from a consistent social ranking that reflects individual rankings. Thus understood the result may appear to face a problem: if we think that the condition of rationality is captured by a range condition of acyclicity rather than transitivity (find a social decision function), then Arrow's conditions can be satisfied. But this will not do as a reply because there are related theorems—due to Gibbard, and Brown—that show that a social decision function that satisfies Arrow's conditions will guarantee a quasi-transitive or acyclic social order only if there exists a group in the population that can veto the choice of the rest of the group.

The large issue about the theorem, then, turns on the reasonableness of the four conditions on the social welfare function. Arrow's theorem causes trouble if we accept the idea that is enforced through I—that the standards to be used for evaluating collective decisions and arrangements for making such decisions are to be constructed from individual preferences (binary orders), taken as having equal weight. But there are good reasons for doubting this condition: in the special setting of binding collective choice, preferences as such may have little if any weight; whatever their role in explaining conduct, they do not, as such, provide a basis for claims on others that a collective decision needs to take into account.

An alternative understanding of the idea of a collective decision, suggested by Rousseau's notion of a general will based on shared or "generalizable" interests. Suppose that we aim to construct the collective will by confining attention to certain fundamental interests of individuals—for example, in Rousseau's case, interests in protection of person and goods, and in personal independence, or in the case of Rawls's theory of justice, interests in the primary goods of basic liberty, opportunity, income and wealth, and self-respect. The notion of fundamental interests is a normative notion: these are the interests in the name of which persons may reasonably make demands on others, acknowledging those others as equals who may make similar demands. The key point is that the individual preferences, interests, and values that are to play a role in justifying political principles that apply to arrangements of binding collective choice must satisfy certain conditions suited to the special problem of justifying such arrangements, and—more particularly—suited to justifying arrangements that capture the idea that people are to be treated as equals in the making of binding collective decisions. Since we are concerned with justifying norms, not explaining the operation of institutions of collective choice, such constraints are appropriate. Once we confine ourselves to inputs that pass through such a filter, then, depending on how these notions are developed—what precisely the interests are, and the extent to which they are shared—we may well be able to generate a set of principles that define a coherent collective ranking.