ARROW'S THEOREM AND DEMOCRATIC POLITICAL PRACTICE

Thomas F. Pyne
California State University, Sacramento

Suppose that we set out to devise an institution that will produce social policies genuinely derived from the preferences of the individuals constituting the society, and not just mimic or replicate the derivation.[1]   Such an institution will have certain features.   Kenneth Arrow's General Possibility Theorem shows that, under certain assumptions about preference relations, any institution with a particular set of quite reasonable "democracy-introducing" features will have another, unwelcome feature: it will yield inconsistent social policies. It will produce "voting paradoxes".[2]

The usual response by those who accept the validity of Arrow's derivation is to relax one or another either of the assumptions about preference or the democracy-introducing features until the paradox goes away. The argument of this paper is that a certain understanding of human psychology, in particular what it is for an individual mental state to have a certain content, will help to show what we should relax.

Democracy-Introducing Features

Democracies are institutions that take individual preferences expressed as votes and enact social policies derived from those votes. Such an institution can be represented as a function taking individual votes as arguments and yielding a social preference (policy) as value. So the social decision question reduces, in effect, to determining what the democracy-introducing features of such a function are.[3]

Let us represent individual and social preferences thus[4]:

$x\mathrm{P}_n y$ represents "person n prefers x to y"

$x\mathrm{P}y$ represents "the society prefers x to y"

Some democracy-introducing features are obvious, some less so. Taking the obvious ones first, the following seven properties would seem to be necessary for any function with a claim to represent democratic policymaking:

1. The social decision, the value of the function $x\mathrm{P}y$, is a function of the preferences of all the individuals.
2. The social decision is a function only of the preferences of the individuals. The concept of a purely democratic decision seems intuitively to involve this: that the social policy enacted should derive from the preferences of the individuals regarding the issue at hand *and on nothing else*: not on traditional policies or solutions, nor on individual preferences regarding other issues.
3. If all individuals prefer x to y, then the social decision is a preference for x over y.
4. Equally, if all individuals prefer y to x, then the social decision prefers y to x.
5. If any individual changes her preferences, it is possible for that change to affect the outcome.
6. There is no fixed preference which the function yields whatever the preferences and wishes of the citizens. "Traditional" social preferences may be overturned by the wishes of the citizens.

7. For any issue there is no individual whose preference always settles that issue. *A fortiori* there is no individual whose preference settles every issue.

Some properties are less obvious:
8. Every individual has only one chance to affect the social decision.
In principle, (8) would seem to be a requirement of any institution that we would call "democratic". An individual with two votes would in effect be a partial dictator. For some properties it is less obvious that they are necessary for democracy.
9. The social decision function is capable of yielding all possibilities. No possibility that individuals might prefer is ruled out.
Feature (9) is in fact too strong. Consider an alternative that individuals would prefer if they could understand it, but which is in fact too complex for us to grasp. Democracy should not require greater than human cognitive powers. Furthermore, some possibilities which democracy might yield violate well-entrenched moral principles. Society's moral principles will constrain the preferences of individuals in such a way that outcomes incompatible with them would never be preferred.[5]

Some features are controversial:
10. Every individual has an equal chance to affect the social decision.
Shouldn't the intensity with which individuals desire a particular social policy, or the degree to which the decision affects those individuals, be taken into account? Is simple majority rule the only possible democratic form? After all, many democratic institutions use two-thirds ballots in certain circumstances. Weighting the preferences of certain individuals over others or requiring super-majorities makes certain individuals partial dictators. This may be reasonable in some circumstances. But that is precisely the point: diverging from equality requires a *reason* in a way that the equality embodied in simple majority rule does not. Consequently, a necessary condition for democracy in *general* will be simple majority rule.
11. If all individuals are indifferent about the choice between x and y, then the social decision is one of indifference.
Intuitions about democracy do not seem force feature (11) on us.[6] But again there would have to be a *reason* for not allowing a democratic outcome, even if that outcome is indifference.

Arrow's Assumptions

Arrow's social decision function incorporates the democracy-introducing features.
Arrow makes two assumptions regarding preference. Call them the "rationality assumptions". The first is that every individual has a ranking for every possible pair of outcomes. That is, for every pair of alternatives x and y in the set of alternatives, each individual either prefers x to y, prefers y to x, or is indifferent between x and y. No alternatives are "incommensurable."

The second embodies the most obvious way for a social decision function to represent simple democracy, by having the value of the social decision depend on the *sum* of individual decisions in the following fashion: If individual m prefers x to y (xP_my) then the value of the individual preference function = 1; if m prefers y to x, then yP_mx, and the value of the function = -1. If m is indifferent ($\sim xP_my$ & $\sim yP_mx$), the value = 0. Summing all the individuals' preferences will yield one of three values, 1, -1, or 0. In a society of three people, Moe, Larry, and Curley, if all prefer x to y, then the sum of individual preferences equals 3, and the social decision function takes the value 1. If they unanimously prefer y to x, then the sum of their preferences equal -3, and the social decision function takes the value -1. And so on for all the combinations.

This function obviously incorporates features (3) and (4). The following five conditions which Arrow places on the function represent judgments about the other properties among (1) - (11). Arrow intends these conditions to be ones even the weakest democracy should satisfy.

Arrow's Conditions

I. Collective Rationality

> Assuming that an issue includes a triple of alternatives (x, y, z) on which every possible combination of individual preferences is possible, every individual preference structure is symmetric and transitive; every social decision is symmetric and transitive.

This is a reasonable weakening of requirement (9), that all possible alternatives be realizable. Arrow assumes only that there will be at least three alternatives the individuals are free with respect to. Collective Rationality stipulates that our preference structure regarding those three will be individually and collectively rational. It also constitutes a denial of (11): indifference is a permissible social decision in Arrow's function.

II. Positive Association

> If every individual's decision concerning an alternative x remains unchanged or changes in favor of x and the decision concerning any other alternatives is unchanged, then a social decision concerning x also remains unchanged or changes in favor of x.

This is a more precise condition expressing the intuition embedded in requirement (5). It seems to be necessary if democracy is really to derive from the preferences of individuals.

III. Independence of Irrelevant Alternatives

> Take any pair of alternatives in the triple (x, y, z) -- say, x and y. A social decision on (x, y) depends on, and only on, the individual decisions concerning that pair.

Arrow requires that a social decision be based on a pair-wise comparison. This is a means of embodying feature (2) – that democracy should settle issues based upon individual preferences regarding that issue and on nothing else.[7]

IV. Citizen Sovereignty

> For any pair of alternatives x and y, there exists a combination of individual decisions which induces xPy and there exists a combination of individual decisions which induces yPx.

Condition IV does not require that there actually be such combinations of individuals. It requires only that such combinations be possible and, if one such combination exists, that it yield the corresponding social policy. So this condition constitutes a version of property (6), that nothing else besides combinations of individual preferences induces a social decision.

V. Non-Dictatorship

> There is no individual whose preference is always adopted by the society for any pair of alternatives.[8] Nor is there for any pair of alternatives some individual whose preference concerning that issue is always adopted.[9]

Arrow's Conclusion

Call the two rationality assumptions and the five conditions collectively the 'Assumptions'. Arrow's theorem amounts to the claim that a social decision function satisfying the Assumptions will produce inconsistent social decisions. To put it another way, if the existence of a social decision function depends upon its yielding consistent values for all permissible distributions of preferences over individuals, the class of functions satisfying Arrow's Assumptions is empty.

> Consider the following pattern of preferences among the three individuals comprising a rudimentary society:

Moe prefers x to y and y to z	$(xP_{Moe}y)$ and $(yP_{Moe}z)$
Larry prefers y to z and z to x	$(yP_{Larry}z)$ and $(zP_{Larry}x)$
Curley prefers z to x and x to y	$(zP_{Curley}x)$ and $(xP_{Curley}y)$

On the issue (x, y), Moe and Curley form the majority: the social decision function has xPy as its value. On the issue (y, z), Moe and Larry form the majority, so the social decision function has yPz as its value. But then on the issue (x, z), Larry and Curley win; so the social decision function has zPx as its value.

However, the Collective Rationality condition (Condition I) requires that social decisions, not just individual preferences, be transitive: if x is socially preferred to y and y to z, then x must be socially preferred to z. By symmetry (xPz) if and only if $\sim(zPx)$. So by yielding both zPx and $\sim(zPx)$ the social decision function produces a "voting paradox." The paradox arises even though the example satisfies all of Arrow's Assumptions:

- Every individual has a ranking for every possible pair of outcomes.
- Distribution of individual preferences is "rational" in the sense of Condition I.
- If Curley were to change his mind and prefer x to z and z to y, then the social decision on the pair x and z would also change; so Condition II is satisfied.
- The social decision is made for each pair entirely on the individual preferences concerning that pair, satisfying Condition III.
- Condition IV is satisfied, since there is a combination of individual decisions that produces every possible result for each pair.
- Finally, no one is a dictator.

The voting paradox can be eliminated only by denying or relaxing one or more of the Assumptions. But the Assumptions were designed to represent a minimal set of conditions for democracy among rational choosers. If the only way to preserve some version of democratic decision making is to relax one of them, which one?

Relaxing Conditions IV and V is out of the question. (*Of course* problems specific to democracy can be solved by eliminating citizen sovereignty or instituting dictatorship...) Likewise it is difficult to imagine a genuine democracy that did not allow for Condition II.

The most arbitrary and artificial-seeming assumption is Condition III. Why shouldn't our decision regarding x and y take z into account?

Relaxing Condition III does dissolve the paradox. But it does so at the cost of introducing another serious problem for democratic theory. If democracy is intended to be the means by which actual individual preferences are embodied in social policy, it now becomes possible to change the result not by expressing one's preferences, but by disguising or misrepresenting them. Consider two cases:

Case 1 involves pair wise voting in which individuals vote twice The first ballot determines the social preference regarding one pair; the second ballot is a runoff between the winner of the first ballot and the remaining alternative.[10] Consider this non-paradoxical preference ordering:

Moe prefers x to y and y to z	$(xP_{Moe}y)$ and $(yP_{Moe}z)$
Larry prefers y to z and z to x	$(yP_{Larry}z)$ and $(zP_{Larry}x)$
Curley prefers z to x and x to y	$(zP_{Curley}x)$ and $(xP_{Curley}y)$

Suppose the first ballot is on the pair (x, y): alternative x wins. In the second ballot, on (x, z), alternative z wins. So z is the social policy.

But with the relaxation of Condition III, individuals can take their preferences regarding the third alternative into account. Alternative z is Moe's least favorite, so his strategy should be

to disguise his preference and vote for *y* over *x*, even though he really prefers *x*. Then *y* wins the first ballot, and the runoff between y and *z* produces *y* as the social policy. (Not Moe's favorite, but it's better than the hated *z*.)

Case 2 involves dropping the pair wise restriction and voting on all three alternatives together. The most obvious way to do this is the "Borda count": Each individual orders his preferences among *m* alternatives, assigning m points for every first ranking, m-1 points for each second ranking, and so on.[11] Consider this preference distribution:

> Moe prefers *x* to *y* to *z* $(x\ P_{Moe}\ y\ P_{Moe}\ z)$
> Larry prefers *x* to *y* to *z* $(x\ P_{Larry}\ y\ P_{Larry}\ z)$
> Curley prefers *y* to *z* to *x* $(yP_{Curley}\ z\ P_{Curley}\ x)$

The result is $x = 7$, $y = 7$, $z = 4$. The alternatives *x* and *y* are stalemated, the social policy indifferent between them, even though a clear majority favors *x*.

However, Moe or Larry can break the stalemate by strategic voting, disguising their preferences and voting *x* over *z* over *y*. If only one changes, the result is: $x = 7$, $y = 6$, $z = 5$. So the social policy becomes *x*.[12]

It seems itself paradoxical to suggest that democracy, which is intended to enact as social policy the genuine preferences of individuals, can be made non-paradoxical only by disguising those preferences.

The only other apparent alternative is to relax the rationality assumptions somehow. For instance, if we consider four alternatives we may wish to conclude that it is possible for an individual to prefer *x* to *y*, *y* to *z*, and *z* to *w*, but neither prefer or not prefer *x* to *w*. Maybe preference operates like resemblance.[13] Preference, like resemblance, might allow for non-transitivities and "incommensurabilities". An individual may consider *x* and *w* just too different from each other for the question of preferring one over the other to have sense.

This response is self-defeating for democratic theory. It would limit the scope of democratic institutions to those decisions which individuals find commensurable. But democratic theory, like economic theory, should have a wider scope than that. In economics incommensurable-sounding comparisons like,

Which do you prefer, a good night's sleep or a new Honda Accord?

make sense and have answers: the price of a Honda can be compared with the price of a prescription for sleeping tablets. Likewise, social questions like

Which do you prefer, full jury trials in civil cases or music programs in the high schools?

may also have -- even demand -- an answer.

The Rationality Assumptions Reconsidered

Finding a way to retain the rationality assumptions while relaxing Condition III is the key to understanding the implications of Arrow's Theorem for substantive political practice. But Condition III introduces the possibility that individuals can influence social policy by strategic voting, raising considerations irrelevant to the issue at hand, misrepresenting preferences, misleading others – all the things we criticize in democratic politicians.

Is there an understanding of the rationality assumptions that can help solve this troubling puzzle? I argue that there is. The rationality assumptions can be misunderstood: it may be true that every individual has a preference ranking for all alternatives; it may also be true that an individual's behavior may be explained and predicted by appeal to that preference ranking. However, the assumptions do not involve any particular psychological commitments about how that preference structure is embodied within the individual. In particular the assumptions do not require explicit awareness on the individual's part. Indeed it is so implausible to imagine that we

are aware of our complete preferences ranking that any theorem requiring such awareness could be dismissed out of hand.

Traditional epistemological assumptions leave us only with an unpalatable choice between making the implausible assumption of conscious awareness or assuming that the preference ranking is somehow psychologically present, complete in detail, but not present to consciousness awareness. On the first choice the process of disguising or misrepresenting preferences involves deliberate dissimulation, so that the ordinary processes of democratic decision-making can appear incompatible with the openness and fraternity that democratic decision-making also requires. On the second choice, the preferences driving our conduct lie beneath the level of our conscious, rational control, making democratic decision-making appear incompatible with the rational conduct of our private and public affairs that democracy is supposed to foster. Are we forced to conclude that democratic decision-making must be characterized either by deception or by self-deception?

A recent development in epistemology and philosophy of mind known as "externalism" reveals a way out of the difficulty. Externalism undercuts the assumption prevailing since the early modern period that the nature and identity of an individual's psychological state is determined entirely and exclusively by properties intrinsic to that individual's mind. Thus, if an individual can truly be said to, say, prefer some alternative x over another alternative y, it must be because there is in her mind some representation of the two alternatives sufficient to identify them and distinguish them from other relevant alternatives. Even if she cannot bring the representations to conscious awareness, such representations are nonetheless present as intrinsic contents of her state.

Externalism responds that psychological states are *intentional*: that is, they refer to -- are *about* -- external objects.[14] It is relation to those objects, not any intrinsic feature of a mental state, that makes the objects part of the content of that state. David Houghton illustrates the point by a vivid example:

> Shoppers deliberate about what to buy, then make a record of their prospective purchases...As they go about their business, there need be nothing inside their heads, no trace of previous deliberations, which makes it true that they intend one alternative rather than another, nothing inside them which determines exactly what their agenda are. Were God to peer into these shoppers' brains, he would find nothing there to tell him that it was basket of goods *A* rather than basket of goods *B* that they intend to purchase. The all-seeing eye would be better focused on points around their persons -- in pockets, shopping bags or glove compartments.[15]

That is, a shopper may be in the mental state of *intending to buy Kleenex* there may be nothing in her psychological state itself that represents Kleenex. Likewise, an individual may be in the psychological state of *preferring x to y*, there may be no representation of those alternatives "in the head". The facts which make it true that this is her preference are her relations, actual or potential, to external states of affairs. In fact, she may be in the state of *intending to buy Kleenex* because, in her pocket, she has a list which she will not consult until she arrives at the market.

On this conception of psychological states it is unsurprising then that the individual may no conscious access to her preference structure. But it doesn't then follow that it is unconscious. Her preference structure is revealed only in her relations with her environment, including her[16] relations with other individuals. In particular it may be revealed only in political dealings with others.

That is, individuals may be revealing their preference structures to others – *as well as to themselves* – by deliberating in a democratic setting. This suggests a new model of democratic political practice stemming from considerations about Arrow's Theorem. Ruth Grant suggests that the two prevailing models of democratic political practice might be called "Legislative Deliberation" and "Pluralist Bargaining." On the Legislative Deliberation model, democratic

politics involves answering the question, what do our moral principles require of us in distributing the goods of which government disposes? The question is answered by *dialogue* among fellow citizens who are striving for the same goal – satisfying the demands of impersonal justice. On the Pluralist Bargaining model the question is, how do we resolve the competition among us for those goods? The question is settled by *negotiation*, each of us striving as hard as possible for as much as possible consistent with maintenance of the democratic institution. If everyone takes care of herself, justice will also take care of itself.

The second model accepts the need for disguising preferences, withholding information, even misleading others, in short the strategies acceptable in any negotiations. It does so, however, at the cost of making democratic politics look too much like ordinary bargaining. Surely strategies appropriate for negotiation even among enemies should not automatically be acceptable among fellow citizens.

The first model sees the dissimulation that arises when Condition III is relaxed as constituting a moral problem in democratic political practice even if it constitutes a way to avoid paradox or stalemate. Is there a conception of democratic practice on which such "strategic" behavior is both morally acceptable in itself and supportive of democracy as correctly understood. We could call the third model, deriving from an externalist conception of our psychological states, "Discovery".

On this third view of democratic political practice each individual has a preference structure, but cannot be assumed to know what it is. An individual may in fact have false beliefs about her preferences. Since in a democracy individuals are confronted with others who have different preferences, yet some social policy acceptable to all must be enacted, a process of refining preferences in order to achieve a result is inevitable. Call the process, not deliberation or negotiation, but "persuasion". Deliberation is an attempt to derive, from general principles what those principles call for in the current circumstances. Negotiation, on the other hand, presupposes self-knowledge about one's preferences in the attempt to get as much as one can. Persuasion, in contrast, is a species of *discernment*: the process of clarifying even to oneself what one's real preferences are by being confronted directly with the alternatives of others in a public forum. Since each individual is engaged in revealing to herself what her preference structure is, what looks like disguising preferences, or appealing to considerations irrelevant to the issue at hand, really aren't. You can't be said to be disguising preferences if you are in a process of learning yourself what they are. You can't be said to be introducing considerations irrelevant to the issue, because what *is* relevant to the issue may be precisely what is in question among the parties.

In conclusion, then, the "Discovery" conception of democratic political practice also throws new light on that most common accusation in democratic politics, hypocrisy.[17] If I try to persuade you to vote a particular way by appealing to considerations which have force with you but not with me, am I being dishonest? In giving those arguments I am leaving myself open to their force in a situation where my grasp of my own preferences is still up in the air. I could say to myself as I rehearse my "hypocritical" argument, "Yes, that is a good reason. And I am obliged to acknowledge that anyone deciding for that reason would be deciding well. But that does not forbid me from deciding differently. What Legislative Deliberators and Pluralist Bargainers fail to see is this: recognizing your reason as a good reason does not oblige me to take it as my reason. Furthermore, if I can persuade you to adopt my choice by appeal to a reason which you take to be good, but which is not mine, then I will. And both of us will be deciding well."

Bibliography

Arrow, Kenneth J. *Social Choice and Individual Values*. New Haven, CT: Yale University Press, 1963.

Houghton, David, "Mental Content and External Representations." *The Philosophical Quarterly*, Vol. 47, No. 187, (April 1997), 159-177.

Murakami, Y. *Logic and Social Choice*. London: Routledge & Kegan Paul Ltd, 1968.

Ordeshook, Peter C. *Game Theory and Political Theory*. Cambridge: Cambridge University Press, 1986.

Endnotes

[1] A "Populist Philosopher King" might use his superior knowledge to enact policies his subjects would vote for, while allowing them no say in the matter. This would merely replicate democracy.

[2] See Arrow, Chapter V, Section 3.

[3] And so we would establish the institutions that will produce the policies reflecting the values of this function. It may be that very different-seeming institutions are "functionally" equivalent in this sense.

[4] I am using Murakami's notation here, and his formulations throughout.

[5] And if they were still preferred, this would constitute evidence that they weren't really the society's moral principles, but rather that society was divided on the question.

[6] While a democracy in which a simple majority prefers to drive on the right will produce a social policy to drive on the right, a democracy where people have no preference for right or left (or where exactly half prefer each result) will still need to enact one policy or the other.

[7] The rationality assumptions ensure that each individual's ranking of all alternatives will be a 'weak' ordering. That is, the preference relation need not rank every alternative individually. After all, individuals might be indifferent regarding some of them. The preference order will instead place all alternatives into ranked disjoint *classes*. The indifference relation, an equivalence relation, will partition all alternatives into equivalence classes. The equivalence classes induced by the indifference relation are identical to the ranked disjoint classes induced by the preference relation. Alternatives toward which you are indifferent are equivalent.

[8] That is there is no "strong" dictator: one who decides every issue.

[9] That is, there is no "weak" dictator: one who decides a particular issue.

[10] Observe that, even without individuals manipulating preferences, the *order* of the voting is decisive. If they vote on (y, z) first, y wins. Then, in the runoff between (x, y) alternative x becomes the social decision. This is simply the voting paradox in another form.

[11] Ordeshook, 68-71. Indeed, either the Borda count or the inverse Borda count (assigning the first ranked alternative the number 1 and them proceeding down to the mth-ranked alternative...), are the only two non-arbitrary methods for taking all alternatives into account. Assigning numbers to ranks by, say, *weighting* the ranks, will be arbitrary in the sense that it will be possible to get any result, depending on the weighting used.

[12] If both do so, the score is: $x = 7$, $y = 5$, $z = 6$, again yielding their favored choice x but reversing the second and third choices (x P z P y).

[13] A resemblance relation need not be transitive. Color x may resemble color y and y resemble color z, but x may not resemble z. Objects x and w may be so dissimilar as to be incommensurable: x neither resembles nor doesn't resemble w. The question of resemblance might seem to make no sense.

[14] See Hilary Putnam, "The Meaning of Meaning," *Mind, Language, and Reality: Philosophical Papers*, Vol. II (Cambridge University Press, 1975); Tyler Burge, "Individualism and the Mental," *Midwest Studies in Philosophy*, 5 (1979), 73-122.

[15] David Houghton, "Mental Content and External Representations," *The Philosophical Quarterly*, Vol. 47, No. 187, (April 1997), 159-177. (162)

[16] Ruth W. Grant, *Hypocrisy and Integrity: Machiavelli, Rousseau, and the Ethics of Politics* (University of Chicago, 1997), 181.

[17] These considerations thus constitute further support for Grant's conclusion that certain forms of hypocrisy are "positive and necessary".