— What sphinx of cement and aluminum bashed open their skulls and ate up their brains and imagination? (Ginsberg 1956, 17)

Contents

List of figures
List of tables
Acknowledgments

1 A long, dark shadow over democratic politics
2 The doctrine of democratic irrationalism
3 Is democratic voting inaccurate?
4 The Arrow general possibility theorem
5 Is democracy meaningless? Arrow’s condition of unrestricted domain
6 Is democracy meaningless? Arrow’s condition of the independence of irrelevant alternatives
7 Strategic voting and agenda control
8 Multidimensional chaos
9 Assuming irrational actors: the Powell amendment
10 Assuming irrational actors: the Depew amendment
11 Unmanipulating the manipulation: the Wilmot Proviso
12 Unmanipulating the manipulation: the election of Lincoln
13 Antebellum politics concluded
14 More of Riker’s cycles debunked
15 Other cycles debunked
16 New dimensions
preference rankings fall outside the domain restriction. We would have to "restrict entry into the community to those having preference orderings that do make collective choices possible" or if it is already too late "they must somehow be isolated and excluded from the community, or an impossibility result can again emerge" (Mueller 1989, 392–393). In the next chapter, I shall argue that if the question is about the public interest, then individual preferences are naturally sufficiently similar to one another's to avoid cycling most of the time. And if the question is fixed-sum redistribution, then destructive self-seeking preferences should be excluded from public consideration (as taught in kindergarten).

5 Is democracy meaningless? Arrow's condition of unrestricted domain

Introduction

Given theoretically predicted instability, why the empirically observed stability? There are several types of answers: that stability is an illusion because we are unable to detect the manipulation that occurs (Riker 1982); that stability is due to institutional devices (e.g. Shepsle and Weingast 1984); that such institutional devices are themselves pervasively unstable (Riker's rejoinder 1980a); that stability is due to similarity in preference rankings among the population, and to preferences for fair distribution; or is due to some other defect in the models. The counterfactual outcome of Arrow's theorem puts us on notice that one or another of its conditions must be misconceived.

I begin with similarity among individuals' preference rankings, a challenge to the realism of Arrow's condition of unrestricted domain. The theorem's impossibility result is a logical possibility but not an empirical probability, I shall argue. One kind of similarity in preference rankings is disastrous though: if majority-rule voters divide up a fixed good, and if each is motivated solely by self-interest and not at all by fairness, then we are guaranteed instability. Contrary to theoretical prediction, however, democratic legislators are typically universalistic rather than factional on distributional questions. This may be due to uncertainty about the future, or a direct concern for fairness, or independently motivated reciprocity, or public deliberation, or due to some combination of these devices. Empirical work shows that citizens vote judgments of general welfare rather than personal welfare. An individual voter almost never affects the outcome of an election, hence she is free to express her disinterested sentiments rather than her interests, and this may explain the empirical finding. Further, there is empirical evidence that Americans overestimate the prevalence of self-interest. If there is not sufficient fairness in the population to tame Condorcet-voting instability, I argue, then there are other acceptable voting rules that avoid cycles altogether.
Table 5.1. Probability of Condorcet winner, impartial culture, strong preference order

<table>
<thead>
<tr>
<th>Voters' Alternatives</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>49</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.944</td>
<td>0.931</td>
<td>0.925</td>
<td>0.922</td>
<td>0.914</td>
<td>0.912</td>
</tr>
<tr>
<td>5</td>
<td>0.840</td>
<td>0.800</td>
<td>0.785</td>
<td>0.776</td>
<td>0.752</td>
<td>0.749</td>
</tr>
<tr>
<td>7</td>
<td>0.761</td>
<td>0.704</td>
<td>0.682</td>
<td>0.670</td>
<td>0.638</td>
<td>0.631</td>
</tr>
<tr>
<td>25</td>
<td>0.475</td>
<td>0.379</td>
<td>0.345</td>
<td>0.327</td>
<td>0.281</td>
<td>0.270</td>
</tr>
<tr>
<td>Limit</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: Table adapted from data in Gehrlein (1983, 174). It is important to distinguish two forms of the Condorcet Paradox: first, when there is no Condorcet winner; second, when there are more than three alternatives, and there is a Condorcet winner, but there is a majority cycle on alternatives below the Condorcet winner. The Table, and the discussion, unless otherwise indicated, refer to the first form of the paradox.

Simulations of homogeneity

In the last chapter, we saw that voting cycles are rare in measures of real-world preferences. Why might that be so? Simulations suggest that it is because of natural similarities among individuals’ preference rankings. Unrestricted domain is any combination of individual preference orders. For lack of any more salient and tractable assumption, calculations of the probability of cycling therefore tend to assume an impartial culture, just as defined in the discussion of Nurmi (1992) above: each of the AI! orders of alternatives is equally likely. Under the impartial-culture assumption, the probability of cycling increases slowly as the number of voters increase, and increases more quickly as the number of alternatives increase, as seen in Table 5.1.

Simulations assume strong preference orders (no indifference among alternatives) for computational convenience. If weak preference orders (voters may be indifferent among some candidates) and impartial culture are assumed, then the probability of a Condorcet winner changes. If we permit ties for top place, in the three-voter, three-alternative case the probability of Condorcet winners increases from 0.944 to 0.995 for weak preference orders (Van Deemen 1999). If we require a unique majority choice, then a Condorcet winner among weak preference orders is less likely than among strong preference orders with smaller numbers of voters, and a Condorcet winner among weak preference orders is more likely than among strong preference orders with larger numbers of voters (Jones et al. 1993). These gyrations are simply due to how the impartial-culture assumption behaves in the contrasting simulations; what is important, as we shall now see, is that the more similar are preference orders the less likely are cycles.

Regenwetter, Adams, and Grofman (2000) remark that the empirical literature finds evidence of virtually no cycles, and that the few cycles observed tend to be over low-ranked alternatives not sharply discriminated among by voters. Tsetlin, Regenwetter, and Grofman (2001) demonstrate that for three voters the impartial-culture assumption maximizes the probability of cycles in samples drawn from an infinite population of individuals with transitive weak preferences. In this setting, any departure from the impartial-culture assumption, realistic or not, reduces the probability of cycling. Recall that Tangian (2000) showed that as we depart from the impartial-culture assumption, and as the number of voters increases, then the outcomes selected by summed cardinal utilities, the Condorcet method, and the Borda method tend to converge. Since there are no cycles with summed utilities or with the Borda count, that implies that with a large number of voters cycles are extremely unlikely, according to Tangian. Similarly, List and Goodin (2001, Appendix 3) show that, given suitably systematic, however, slight, deviations from the impartial culture assumption, the probability of a cycle converges either to 0 (more typically), or to 1 (less typically) as the number of voters increases. The impartial culture assumption is often used in textbook demonstrations of the incidence of cycling (e.g., Shepsle and Bonchek 1997, 53, who on the assumption of impartial culture also claim that the likelihood of cycles goes up as the number of voters increases, contrary to Tangian), but it is the assumption most likely to yield cycles.

There are several ways of conceptualizing the similarity of preference orders. I will begin with the probabilistic. Kuga and Nagatani (1974) develop an antagonism index, a ratio of the number of antagonistic choice pairs (e.g.,  \( x > y \) and \( y > x \)) to the number of pairs of persons in the population, normalized by the number of persons in the population. Under the impartial-culture assumption the antagonism index would equal one. When that antagonism index is \( \frac{2}{3} \) or less, there is no voting paradox (by analysis), between \( \frac{2}{3} \) and 1 the probability of paradox increases as antagonism increases (by simulation). Williamson and Sargent’s (1967, 811) probabilistic method concludes, “where unimodal preferences exist… and a large vote can be expected, the probability of intransitivity seems to vanish.” Berg’s Polya-type urn model claims to subsume most remaining prior measures of homogeneity of preferences, and confirms prior studies: “greater preference similarity among voters, or stronger mutual influence among voters, leads to a lesser chance for the paradox of occurring” (Berg 1985, 379). The model contains a contagion or
Table 5.2. Probability of Condorcet winner, increasing homogeneity, three alternatives

<table>
<thead>
<tr>
<th>Homogeneity parameter # of voters</th>
<th>0</th>
<th>1/2</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.9444</td>
<td>0.9643</td>
<td>0.9750</td>
<td>0.9815</td>
<td>0.9925</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.9306</td>
<td>0.9524</td>
<td>0.9665</td>
<td>0.9602</td>
<td>0.9938</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.9250</td>
<td>0.9470</td>
<td>0.9626</td>
<td>0.9690</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>0.91226</td>
<td>0.91919</td>
<td>0.92578</td>
<td>0.93750</td>
<td>0.95493</td>
<td></td>
</tr>
</tbody>
</table>

homogeneity parameter, and as that parameter increases, the probability of a Condorcet winner increases. Suppose an urn with chips of six different kinds corresponding to the six strong preference orders resulting from three alternatives. When the homogeneity parameter is zero, voters are independent of one another, and chips are drawn from the urn with no replacement (known in the literature as the impartial-culture assumption). When the homogeneity parameter is one, we replace a chip drawn with one of the same kind (known elsewhere in this literature as the anonymous-culture assumption). When the homogeneity parameter is two, we replace a chip drawn with two of the same kind, and so on, reflecting the extent to which voters influence one another. Gehlren (1995) further developed estimations of Condorcet efficiency and social homogeneity under the urn model.

In Chapter 3 on the irrationalists’ claim that democracy is arbitrary, I argued that natural similarity among individual preference orders reduces divergence of outcomes among alternative reasonable voting rules. I argued that people have similar preference orders because they live in the same world and they have similar interests in that world; for example, most prefer prosperity to torture of kittens to suicidal nuclear war. If again we move from logical possibility to empirical probability, then the irrationalists’ claim that democracy is meaningless also seems to miss the target. Even if there is no similarity among individuals’ preference rankings, the probability of a cycle is very low. With a more natural similarity among preference rankings, such as what we found in our empirical examination in the last chapter, the probability of a cycle shrinks further. It is nature that provides the cycle-shrinking domain restriction, not human institutions. To relax the domain restriction does not require a crushing uniformity in the population, nor the horrifying picture of preference police hounding out those with topsy-turvy rankings. Should everyone have topsy-turvy rankings, then of course there is no mere aggregation function that can bring order out of disorder. In that case, however, the fault lies in the distribution of preference rankings, not in the voting rule.

Condition U was primarily intended as a matter of methodological convenience, according to Brennan and Hamlin (2000, 106–107). Whatever the intent, the condition is normative in effect, they continue. The requirement that all and only individual orderings be admitted for social choice is not neutral, but rather is a particular view on the question. The condition bites even if everyone were to agree to exclude some kinds of preferences — such as the unfair, the malevolent, or even the evil — from consideration. Such an agreement would violate the condition; the condition is imposed even if all citizens were to reject it. The imposition of a rule dictating the consideration of evil preferences is not normatively compelling. Condition U “imposes a substantive and questionable ethical view on members of society, while at the same time failing to reflect the patterns of common interest that might be expected to arise within genuine societies” (107).

Table 5.3. Egomaniacal redistributional instability

<table>
<thead>
<tr>
<th>Voter</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ABD</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>2. AB</td>
<td>300</td>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td>3. BC</td>
<td>0</td>
<td>301</td>
<td>299</td>
</tr>
<tr>
<td>4. AC</td>
<td>299</td>
<td>0</td>
<td>301</td>
</tr>
<tr>
<td>5. AB</td>
<td>300</td>
<td>300</td>
<td>0</td>
</tr>
</tbody>
</table>

Egomaniacal redistributional instability

I said that natural resemblance of preferences from individual to individual almost eliminates cycling, but in another way similarity of preferences might force Condorcet instability. Asked what the disagreement was between himself and King Francis I of France, Charles V replied: “My cousin Francis and I are in perfect accord — he wants Milan, and so do I.” If Condorcet majority-rule voters are engaged in division of a fixed good, let’s say pie, then instability is inevitable if each insists only on getting the most pie for himself (Sen 1995, 10). In Table 5.3 we begin in stage one with an equal distribution of 200 each among majority-rule voters A, B and C. But next in stage two the coalition AB can form, offering each of A and B more than they would receive under equal distribution. But
in stage three, the coalition \( BC \) could offer each of \( B \) and \( C \) more than they would receive in stage two. But in stage four, the coalition \( AC \) could offer each of \( A \) and \( C \) more than they would receive in stage three. But in stage five, the coalition \( AB \) could return and offer each of \( A \) and \( B \) more than they would receive in stage four. And so on.

Say that there are three piemen: Jaime, Lyman, and Simon. Any majority-rule coalition of two or three is unstable. Condorcet voting cannot transform the egomaniacal “more-for-me” preferences of these three into a fair outcome. Suppose that each of the three piemen has a preference in the first rank of all the pie for himself, and then in the second and third ranks has random preferences over all the pie going to one of the other two, such that there is a cyclical profile over the “more-for-me” alternatives, and in fourth rank has preferences for fair division of the pie. There is a cycle over the “more-for-me” alternatives, each of which is preferred over fair division. (In contrast to cycling with Condorcet voting, Borda voting would declare a tie among all-for-Jaime, all-for-Lyman, and all-for-Simon.) If, however, the first rank remains the same, but fair division is moved up to everyone’s second rank, then fair division becomes the Condorcet winner, and the more-for-me alternatives cycle in the remaining ranks of the social ordering (Borda would rank fair division first, and a tie among the all-for-me alternatives). Thus, even if everyone’s first preference is more-for-me, if their second preference is fair division, then fair division prevails as the social decision. Impartial alternatives drive out partial alternatives.

To illustrate more explicitly, assume there are five individuals, labeled 1 through 5. There are six alternatives: \( v \) gives the whole world to individual 1, \( w \) gives the world to individual 2, \( x \) gives the world to individual 3, \( y \) gives the world to individual 4, \( z \) gives the world to individual 5, and \( a \) divides the world equally among the five individuals.

The collective ranking is \( a > (x > w > v > z > y > x) \). Fair division prevails (and would prevail even if fairness were anywhere in the upper half of the rankings). If preferences for fair division were originally ranked lowest, and if there were costs to cycling, then perhaps it would develop that preferences for fair division would come to rank higher. Since there are many redistribution issues in politics, some commentators are confident that there must be much majority-rule instability, even if it is difficult to detect empirically. But I think they are making an empirical assumption, controversial and not widely warranted, that political participants do not at all value fair outcomes. As I have told the story, each voter could still in a childish way want all the pie for himself, but just so long as each ranks fair division better than random distributions of the pie then there is a tendency for the social choice to be fair division. They would also have to rank universalistic outcomes over majority-faction outcomes, but that would be likely if majority coalitions were unstable and themselves random in incidence.

In earlier work based on cooperative game theory, Riker (1963) formulated his size principle: that in the majority-rule division of a fixed good, only minimum-winning coalitions would form. If the voting rule is half the voters plus one, then all distributional measures would be won only by that margin. The idea is that adding more members to the coalition would dilute the amount of the good going to each member of the coalition. If there are 100 voters, dividing the pie among 51 of them gives each of the 51 more than she would get than dividing the pie among 52 or among all 100. A problem developed with the theory: actual legislatures infrequently passed measures by the minimum number of votes, rather they frequently passed measures by large, almost unanimous majorities, especially on distributional issues. Empirical studies concluded that “distributive policy making was characteristically associated with the development of an extremely large and at times unanimous coalition that included virtually everyone with a stake in the outcome” (Collie 1988, 430).
Congressional porkbarrel spending on projects at the Congressional-District level resembles such a constant-sum game. If there is cycling, then spending by district should equalize over time, as losers can always create new winning coalitions; if there is no cycling, then spending by district should remain stable from year to year. Stratmann (1997) proposes a measure for this, applies it to federal-funding data, and infers an absence of cycles. If his finding is correct, it may be due to any number of reasons, for example institutional constraints in Congress, or perhaps there are substantive reasons for some districts consistently to obtain more funding than others, and it is not necessarily due to a similarity among preference orders. Additionally, Poole and Rosenthal's (1997; 1999) data suggest that although the one dimension of importance in Congress is the struggle between the poor and the rich, still, redistributive struggles rarely take a multidimensional form. Brennan and Lomasky (1993, 44–45, 83) say that the instability story strains credulity: one never observes the degree of volatility in spending or taxation that the cycling model requires.

The empirical tendency was termed the norm of universalism. Weingast (1979) sought to explain universalism within the inherited cooperative game-theoretic framework. The set of minimum-winning coalitions dominates all coalitions not in the set. But, Weingast argued, benefit-seeking legislators would be uncertain, as the vote approached, of which of the many possible minimum-winning coalitions would form and win. Because of this uncertainty, the chance of ending up in a losing coalition, each legislator beforehand would prefer the universalistic coalition to any other coalition. The Weingast model assumed that the benefits of projects funded exceeded their costs, which, because of the view that many government-funded projects are economically unjustifiable, occasioned further models by him and others that would explain inefficient projects.³

Ferejohn, Fiorina, and McKeelvey (1987) modeled the same problems with noncooperative game theory. In the one-shot game, the cheapest projects would be funded, and by a minimum-winning coalition. If that stage game were indefinitely repeated then, as with social-dilemma models of indefinite repetition, any Pareto-optimal outcome is sustained in equilibrium, from minimum-winning to universalistic. Baron and Ferejohn (1989) modeled a sequential multilateral bargaining game. Now the pie rots as the legislators debate how to divide it up. Under a closed rule, a minimum–coalition wins. Under an open rule, the outcome can be universalistic, but only if the legislature is quite small in number of voters (or in blocks) and if the pie rots quickly. Glazer and McMillan (1982) propose a nifty noncooperative model that yields a universalistic equilibrium when legislators want to spend their limited time, in my metaphor, baking new pies rather than fighting over the division of old ones. Grose and Snyder (1996) propose another model showing that supermajority coalitions can be cheaper than minimal-winning coalitions.

There is an alternative hypothesis to explain the observed norm of universalism, however, which enjoys the advantage of consistency with a controlled experiment on the topic, and more importantly consistency with legislative discourse on redistributive taxing and spending tasks.⁴ The alternative hypothesis is that a sufficient number of voters are directly concerned to attain a fair distribution. Miller and Oppenheimer (1982) conducted experimental tests of Weingast's model of universalism. In controlled experiments, five-person committees were formed, and individuals' preferences were induced by cash reward. Committees operated by majority rule, any three of the five members determined the outcome, after 15 minutes available for discussion; and subjects were prohibited from redistributing rewards outside the terms of the experiment. Coalitions of three or more could choose among one alternative that provided an equal distribution to all five committee members and other alternatives that would provide various payoffs to the members. In the first experimental condition, the value to each member of equal distribution was $12.40, and, à la Weingast, the expected value of belonging to any one of the decisive three-member coalitions was $8.40. If a three-member coalition were actually to win on any alternative other than equal distribution, each of its members would make more than $12.40. Subjects rapidly agreed that equal distribution was the best choice. Typically some members pointed out that three members could be made better off, and typically some members responded that an alternative three-member coalition could be formed leaving two of the three in the original coalition worse off, and then members would settle for the safety of equal distribution. Subjects were aware of the fragility of equal distribution, and typically truncated discussion so as to discourage temptation. Four of five committees were unanimous for equal distribution, in the fifth, the vote for equal distribution was four against one economics student.

In the second condition, the value of equal distribution was made $8.50, just above the expected value of belonging to any one of the decisive three-member coalitions. In four groups out of five, members chose the equal distribution. In the fifth group, discussion secured unanimous consent to explore coalition formation, and a three-member coalition formed. Weingast's model seems to be supported. Further experiments were run, however, with the universalistic alternative worth less than the expected value of $8.40 for belonging to one of the decisive three-person coalitions. In the third condition, equal distribution was worth $6.72. Still, in four out of five committees the universalistic alternative was chosen.
In the fifth committee, three members wanted universalism, and two insisted on coalition formation. In response, the three members who originally proposed universalism voted for their own three-member coalition, thereby punishing the two members for advocating an unfair outcome. In a fourth condition, equal distribution was lowered to $4.20. Again, four out of five committees chose equal distribution. The fifth committee devised an alternative universalism: they decided to draw lots among all alternatives, with an expected value of $7.70 for each member. They then voted unanimously for the randomly drawn outcome. The third and fourth conditions do not support the Weingast model. In a fifth condition, equal distribution was worth only $2.10; finally, members agreed that the equal distribution was not worthwhile. I have gone on at length because I think it is important to remember the possibility that on distributional questions many people are motivated directly by fairness.

More than a decade of human-subject experiments, say Fehr and Fischbacher (2002), demonstrate robustly that many humans possess social preferences beyond material self-interest, and, they continue, some economic models that neglect this fact fail in their predictions. Although the models of economists permit complete heterogeneity of taste with respect to material consumption, they insist that individuals are motivated only by self-interest and dogmatically exclude heterogeneity in social preferences. The experiments show, however, that roughly 40 to 50 percent of subjects are motivated by reciprocal fairness. Moreover, the presence of a critical mass of those motivated by reciprocal fairness changes the incentives of the selfish types, resulting in constraints and interaction patterns quite different from those that would result if the population were wholly selfish. A reciprocal individual responds to kind action with kind action, and to hostile action with hostile action. Such reciprocity is carried out even if costly; it is not motivated by expectation of future material benefit; it differs from retaliatory behavior in repeated interactions. It also differs from altruism, which is kindness in response to either kindness or hostility. The primary experiment in this paradigm is the ultimatum game. A pair of subjects are assigned the task of dividing a given sum of money, in an unrepeated, one-shot game. The first makes an offer to the second of from 0 to 100 percent of the sum. The second either accepts or rejects the offer. If the second accepts, then the second gets the share offered, and the first gets the remainder. If the second rejects, then both get nothing. The rational-choice prediction is that the first will offer the smallest possible unit of money, say a penny, and the second will accept. The experiments show, however, that proposals offering less than 20 percent are rejected with 0.4 to 0.6 probability. Also, it appears that proposers anticipate rejection of lower offers, which motivates them to make higher offers. This is shown by comparison to dictator games, where the first mover simply decrees the division. The rational-choice prediction is that first movers would give nothing; in dictator experiments, however, some give, but on average less than in the ultimatum game. Henrich et al. (2001) conducted ultimatum, dictator, and social dilemma games across fifteen nonindustrial cultures, and found that humans consistently deviate from the model of textbook economics. Many appear to care about fairness and reciprocity and reward cooperators and punish noncooperators, even when such actions are costly to the individual. Mean offer per culture in ultimatum games ranges from 0.26 to 0.58.

The Los Angeles County Board of Supervisors, report Cox and Tutt (1984), formally adopted a rule that nonmandated federal funds be divided equally among the five supervisorial districts — again, the norm of universalism. Observe that any three supervisors superficially have an incentive to vote to depart from the norm, but that they do not. One of the things I did before going to graduate school in political science was to serve for three years as the directly hired policy aide (aka political hack) to one of three elected commissioners of a county of about 250,000 people — the retail politics of roads, sewers, zoning, garbage, and dogs. The norm of universalism was common in allocative decisions, with arguments around the edges. One day, however, there was a palace revolution. We policy aides went into a meeting to discuss allocation of millions of dollars of CETA make-work funds among the many applicants. The previous year the norm of universalism was informally followed. In the year under discussion, a representative advisory committee had ranked proposals. In the meeting, the aides of the other two commissioners junked the advisory committee proposal, and devised a list of grantees which excluded any applicant of interest to the third commissioner, my boss. Moreover, the gang of two continued this exclusive collaboration on all further issues. Consider that when the norm of universalism is a possibility, everyone is strongly motivated to maintain comity. A permanent loser, however, has no such motivation and, spared some of the burdens of governance, has lots of spare time on his hands. There is more to politics than votes on propositions. My boss's constituencies were extraordinarily mobilized because of the unfairness of the "CETA massacre," he gained unexpected allies motivated solely by considerations of fairness because of it, and he triumphed against a well-funded opponent in his next reelection campaign. It also turned out that one of the gang of two became subject to a recall election. He survived the recall, but was permanently damaged politically, and later did not run for reelection. The other of the gang of two was accused of defrauding the county on a land sale; the criminal prosecution did not secure a conviction, but the political career
of a gubernatorial aspirant was ended, and later a civil action against him to recover defrauded value did succeed. New commissioners after that operated more by the norm of universalism. Perhaps this is an example of reciprocity as a motivator of equal distribution.

In the famous prisoners’ dilemma or social dilemma used widely to model many collective action problems, it is in each individual’s interest not to cooperate, which makes everyone worse off than they would be if all did cooperate. In terms of material payoffs, the problem may appear to be a social dilemma, but if there is a critical mass of individuals independently motivated by reciprocity to punish noncooperators, then the dilemma is transformed into a coordination game. In a coordination game there is more than one possible equilibrium, and the question is whether everyone will coordinate on an equilibrium that makes all or most better off or on an equilibrium that makes all or most worse off.

In this model, cooperation could fail in one of two ways. First, given a critical mass of conditional cooperators, promoting cooperation involves management of people’s beliefs: “If people believe that others cooperate to a large extent, cooperation will be higher compared to a situation where they believe that others rarely cooperate” (Fehr and Fischbacher 2002, C16). Second, even if there is a critical mass of conditional cooperators, and people believe there is one, then it still may happen that people have inherited a welfare-inferior coordination equilibrium that can only be escaped by coordinated exit (see Mackie 1996). Notice that inheriting a welfare-inferior equilibrium, say a tradition of political corruption, might cause a false belief that there is not a sufficient mass of conditional cooperators in the population, when in fact there are.

A nice feature of a model like this is that it might explain both the presence and the absence of fair cooperation, for example, of democratic success in Australia, and of democratic failure in Argentina: identical institutions containing otherwise identical individuals might vary in fair cooperation just because people vary in their beliefs of what to expect from one another. This is one reason why some people deplore the rhetoric of the material-egoist version of rational choice theory: the more hegemonic is the theory, the more self-fulfilling it is, by causing people to believe that the proportion of unconditional and conditional cooperators in the population is less than it actually is, thereby making people worse off than they would have been in the absence of the theory. The correlation between self-interest (personal financial situation) and voting in US presidential elections is usually 0.08, but after four years of self-interest rhetoric in President Reagan’s first term of office, and his call to “Ask yourself, are you better off today than you were four years ago?”, the correlation jumped to 0.36 (Miller and Ratner 1996, 31).

Traditional social choice theory is solely concerned with voting, and excludes democratic discussion about the content of alternatives and reasons for favoring one alternative over another. Some strains of deliberative democratic theory seek to escape the social choice-problems of democratic voting by sole reliance on unanimous agreement attained by democratic discussion. In democratic institutions, I submit, voting and discussion are complements, not substitutes. A voting rule such as Condorcet’s is, or should be, adopted in the first place due to considerations of fairness. Majority rule would be a hideous institution if it were confined only to the casting of votes on alternatives without content, or if it were carried out by unfair people, or both. Sen (1982, 21) illustrates. In case A, person 1 is very rich, and persons 2 and 3 are very poor; 2 and 3 vote to take some from 1 and give to 2 and 3, reducing inequality. In case B, person 1 is very poor, and persons 2 and 3 are very rich; again 2 and 3 vote to take some from 1 and give to 2 and 3, increasing inequality. In the Arrovian noncomparabilist framework, Case A is identical to Case B. Some people might judge Case A to be morally wrong, but most people would judge Case B to be morally wrong.

Recall Rousseau (1668/1762, 72): “There is often a great difference between the will of all and the general will; the general will studies only the general interest while the will of all studies private interest, and is indeed no more than the sum of individual desires.” If we could exclude egomaniacal preferences from consideration, then cycling would be of little worry. Is there some way to do this? Robert Goodin and Jon Elster propose that public discussion helps:

The conceptual impossibility of expressing selfish arguments in debates about the public good, and the psychological difficulty of expressing other-regarding preferences without ultimately acquiring them, jointly bring it about that public discussion tends to promote the common good. (Elster 1986b)

The full argument is found in Goodin’s (1995/1986) landmark essay, “Laundering Preferences.” Devices such as vested rights filter the output of collective decisions. Another kind of device is to filter the input to collective decision, via the exclusion of some individual preferences, the laundering of them. Acceptable grounds for laundering preferences include: for one individual, some preferences are more fully considered than conflicting others; some individuals agree amongst themselves to reciprocal forbearances, such as conditional toleration of the others; individuals may explicitly have preferences over what kind of preferences they should have, and the second-order preference should be honored over the first-order preferences; individuals may implicitly have such second-order preferences; and finally, “our very choice of aggregating preferences as
a way of making social decisions carries consequences for the kind of preferences that we can count” (140).

Goodin suggests further that an individual might possess multiple preference orderings, selfish or fair, material or moral, according to context, and that, in the context of collective decision making, role rationality dictates the expression of publicly oriented, ethical preferences and the suppression of privately oriented, egoistic ones. It’s pragmatically impossible to argue publicly that an alternative which affects others should be adopted solely because it is good for the arguer. Further, that someone voting in the first place indicates an ethical context, because voting is instrumentally irrational according to rational choice theory (to be discussed in the next section). In addition to these pragmatic mechanisms, some or many individuals may possess simply moral motivations to support impartial outcomes. Elsewhere, Goodin (1992) cautions against expecting too much from this model of publicity and discursive defensibility.

If a body of people is such that, on redistributive questions, most individuals and factions politically insist at all costs on more than their fair share of the social offices and product, then there will be instability, probably trouble, maybe even violence. Formally, there would be Condorcet-voting instability, but that is not the main problem. The flaw is in the egomaniacal preferences of the individual citizens, not in the fact that those unfair preferences are aggregated by majority rule. For example, Didymus Mutasa, organizational secretary of the ruling party in Zimbabwe, ZANU-PF, in response to accusations that his government was withholding food aid from famine areas which did not vote for it in the election, is reported to have said that, “We would be better off with only six million people [out of a population of 12 million], with our own people who support the liberation struggle. We don’t want all these extra people.” No aggregation procedure can turn unfair individuals into fair ones. For someone such as Mutasa, public discussion probably wouldn’t do much good either.

Voting and self-interest
The empirical evidence indicates that democratic voters are not motivated by material self-interest, or only minimally in limited circumstances. Kinder and Kiewiet (1979, 1981) were the first to investigate “sociotropic” voting. Generally, Americans vote for incumbent candidates when the economy is good, and for opposition candidates when the economy is bad. Most earlier work had been done on an aggregate basis, and thus it was wrongly concluded that citizens vote their pocketbooks: I vote against the incumbent when I’m not doing well. Alternatively, citizens could be sociotropically motivated: I vote against the incumbent when the country is not doing well. By studying individual rather than aggregate responses, Kinder and Kiewiet (1979, 524) showed that:

With respect to economic issues, voters appear to choose between congressional candidates "sociotropically." Voters are not egocentric in any narrow sense -- they do not vote their own pocketbooks. Rather, their preferences follow a more collective reckoning.

Many similar empirical investigations were undertaken in reaction to the rise of the materially self-interested rational actor in the social sciences. Lewin (1991, 45) reviews the literature on self-interest or public interest in western politics, and finds that the "the results point overwhelmingly against the self-interest hypothesis." He suggests that it is odd that the self-interest hypothesis is so tenaciously adhered to, when the great bulk of evidence goes against it. An even more detailed and wide-ranging review by Sears and Funk (1991, 47; abridged in Mansbridge's excellent 1990 volume on self-interest), including survey data on two dozen diverse issues, finds "that personal self-interest generally has not been of major importance in explaining the general public's social and political attitudes." There are a few exceptions, and "self-interest is most potent when the issue provides large, clear, certain, and salient costs or benefits" (63). Citrin and Green (1990, 16) conduct a similar survey, and conclude that the literature reviewed appears to be "devastating for the claim that self-interest, defined narrowly as the pursuit of immediate material benefits, is the central motive underlying American public opinion." A test of whether individuals are concerned about whether or not their own group is doing well found another sociotropic effect, that citizens are concerned about whether economic gains and losses are distributed equitably among groups (Mutz and Mondak 1997).

There are a great many confusions and equivocations about self-interest in the rational-actor tradition. Current economic theory has it that individuals maximize their utility, but the only content that claim has is that an individual chooses the most highly ranked alternative feasible from her consistent ranking of alternatives. Those aims could be egoistic or altruistic, smart or stupid, or apparently random, just as long as their ordering is consistent. The traditional utilitarian idea that individuals somehow maximize an underlying satisfaction is wholly abandoned. Individual aims, if they are concerned about the well-being of others, are nevertheless contained within the individual’s preference ranking; thus, it is sometimes said that the theory assumes that individuals are self-interested. But they are self-interested merely by definition: any alternative that St. Francis
chose, no matter how selfless, was self-interested in this tautologous sense (Riker 1990b). It is more useful to say that current economic theory assumes that people are purposive rather than that they are self-interested. This is sometimes called the thin theory of rationality (Green and Shapiro 1994, 17–19).

A thick theory of rationality goes further, and assumes an underlying dimension to those choices, often of objective content, in contrast to the purely subjective value assumed by the thin theory. The paradigm is the profit-maximizing firm in the competitive market. Economic competition forces the successful firm to pursue solely profit and to do so with perfect information. The rationality of the firm is measured by the objective criterion of profit, and there is a background theory as to why on average firms should possess this thick rationality. Confusion arises when we move from firms to ordinary individuals, however. It could be assumed as a matter of modeling and measuring convenience that on average ordinary individuals engaged in economic activity are concerned to maximize material-egoistic wealth. This is a limited theoretical simplification, however, not a descriptive claim, and should never be permitted to mislead, especially outside the economic sphere. Casual observation suggests a tendency to self-wealth maximization. But only a tendency, and primarily in the economic sphere: many individuals are also motivated by many other aims, including relations with family and friends, morality, justice, honor, and self-expression. Unlike firms in perfect competition, there is no background theory as to why ordinary individuals should exclusively maximize material-egoistic wealth; individuals who do not, are not irrational. Those who do – the C. Montgomery Burns of the world – seem to be more pathological outliers than healthy specimens of humanity, or so the wisdom of the ages suggests. I hasten to add that both self-interest and the pursuit of material wealth, in proper proportion and place, are consistent with if not necessary to the moral life.

The reign of self-interest in popular economics is all the more puzzling in that the intellectual founders of the economic tradition themselves opposed Mandeville's simplistic reduction of all human motivation to self-interest. Holmes's (1990) marvelous essay on the history of self-interest teaches us that David Hume, Adam Smith, and their fellow travelers believed both that an ideal of commercial self-interest was better for the common good than the ideals it displaced of a passion for military glory and the religious dogma of original sin, and that calculated self-interest was one human motivation among many. Calculation is contrasted with destructive impulse and folly, and self-interest is contrasted with varieties of benevolence on the one side and varieties of malevolence on the other. Some disinterested motivations are good, and, Holmes teaches, plenty

of disinterested motivations are evil: these days, one thinks of the selfless cruelty of suicide bombers.

Olson's logic of collective action assigns a certain preference ranking to each individual, together these preferences amount to a social dilemma, and thus nonrivalous and nonexcludable public goods tragically are undersupplied, the story goes. Rational individuals, Olson claims, will not voluntarily contribute to public goods, instead they will free-ride on the efforts of others, a finding that irks many. Olson and his followers believe that, since the contribution of any one agent does not affect provision of the public good, each agent would first prefer to avoid her own costly contribution while all others contribute, second prefer mutual cooperation, third prefer mutual defection, and last prefer being the only person to contribute. The social outcome is mutual defection, even though mutual cooperation is better for everyone. Pellikaan and van der Veen (2002) show that Olson's logic wrongly assumes that the objective interest of firms in the competitive market should apply as well to ordinary individuals. The thick rationality of competitive firms, for whom that ranking is appropriate, is confounded with the thin rationality of ordinary individuals. They further show that on certain environmental questions Dutch citizens do not have the preference order leading to a social dilemma, in fact, many are shown to be unconditional cooperators, willing to contribute even if others do not, just as some moralities would recommend.

We have seen assumptions of self-interest and perfect information justified within a model of firms in a competitive market unjustifiably transplanted into the ordinary individual. Next, one variety of public choice theory goes one step worse: it transplants these same assumptions into the voter, for whom they are not justified at all. There is no background theory as to why the ordinary individual should be a maximizer of material self-interest, but there is or should be a background theory of why the ordinary voter should not be a maximizer of material self-interest. If there are more than a few voters in an election, then it's unlikely that my vote would make any difference to the outcome, and the more voters there are the more unlikely it is that I would be decisive. Since voting takes effort, no instrumentally rational person would bother, and according to rational choice theory almost no one would vote, which we shall term the paradox of participation. Goodin and Roberts (1975) may have been the first to point out that this unleashes the ethical voter. As the argument about public deliberation, suppose that the individual possesses multiple preference orderings, selfish or fair, material or moral. Since the individual's vote lacks decisiveness, egoistic preferences are made futile, and thereby expression of ethical preferences comes to the fore. This is fully developed by Brennan and Lomasky (1993). In the market, the agent is
decisive; if she chooses A over B, she gets A and forgoes B. When the voter chooses A over B, however, the outcome is not a function of how she votes, but of how everyone votes. For each voter, electoral outcome is detached from electoral choice. The market responds to interests, but democratic voting invites the expression of disinterested ethical and ideological principles. Thus, the empirical finding of sociotropic voting may have a theoretical foundation in the account of voting as expressive rather than instrumental action. As we have learned from Holmes, however, and as Goodin, Roberts, Brennan, and Lomasky confirm, the disinterested need not be the good: justice could be unleashed, but so could jingoism. All the more important then is sound public deliberation over the content of political alternatives.

Sen's (1995, 15) presidential address to the American Economic Association deplores a trend among some economists to reduce all socially motivated action to some kind of cunning attempt to maximize purely private gains, and wonders whether the trend is more prevalent in America than in Europe, citing Tocqueville (1969/1851, Book II, Chapter VIII):

The Americans... enjoy explaining almost every act of their lives on the principle of self-interest properly understood. It gives them pleasure to point out how enlightened self-love continually leads them to help one another and dispose of their time and wealth for the good of the state. I think that in this they often do themselves less justice, for sometimes in the United States, as elsewhere, one sees people carried away by the disinterested, spontaneous impulses natural to man. But the Americans are hardly prepared to admit that they do give way to emotions of this sort. They prefer to give credit to their philosophy rather than to themselves.

Miller and Ratner (1996; 1998) update Tocqueville's observation. They undertook a series of social-psychology experiments which measured both the actual extent that self-interest guided action, and people's estimates of the extent of self-interest motivation in the situation. For example, they asked subjects whether or not they would be willing to donate blood for free and whether or not they would be willing to donate blood for $15, and they also asked subjects to estimate what percentage of their fellow students would donate blood for free and for $15. Material payoffs only mildly affected willingness to give: 73 percent said they would donate for a price, 63 percent would donate for free. But the same subjects underestimated by half the number of students willing to donate blood for free: they thought that 61 percent of the population would donate blood for a price, and that only 32 percent would donate blood for free. Similar experiments studied effect of gender on support for abortion coverage, effect of racial status on concern for minority needs, and effect of payment on willingness to participate in future experiments, and found the same overestimation of self-interested motivation.

We are enthralled by a myth of self-interest, according to Miller and Ratner. Public discourse equating rationality with self-interest normalizes actions consistent with self-interest and pathologizes actions inconsistent with it. People experience discomfort when they take action inconsistent with self-interest: one study found that people's decision about whether or not to pursue an injury claim was based much more strongly on expectations of financial gain than on expectations of fair treatment, but that people's satisfactions with the claiming process were based far more on fair treatment than on degree of financial gain. Thus, people ignore their own sentiments in favor of the norm that material self-interest should predominate, and they are led "to willingly adopt decision strategies they know will not maximize their satisfaction" (1996, 35). People fear social isolation when they take actions inconsistent with self-interest. People justify their actions in terms of self-interest: they explain their charitable activity, for example, in instrumental terms -- "It gave me something to do," "I liked the other volunteers," "It got me out of the house" (38).

Miller and Ratner do not indicate awareness of Holmes's message that there are worse things than self-interest. Otherwise, their findings suggest that a dominant ideology of self-interest does obscure recognition of motivations of benevolence and fairness in the population. It does not take sainthood to overcome egomaniacal redistributional instability. Merely a secondary preference for fair distribution, should one be unable to grab everything for oneself, would do. Further, among a small number of people, independently motivated reciprocity would maintain equal distribution. Empirically, legislators are universalistic on redistributional questions. Among a larger number of people, the inefficacy of voting could liberate sentiments of fairness. Empirically, voters are sociotropic.

If there is a worry that there is not enough fairness in the population to avoid egoistic redistributional instability under the Condorcet voting rule, then there are good voting rules that avoid cycles.

**Cyclebusters**

We have seen that cycles are rare. This section will offer two further claims. First, most cycles are trivial. Second, if there are cycles that are not trivial, then there exist accurate and fair voting rules that eliminate such cycles.

1. The closer alternatives are to one another, the less structure there is between individuals' preference orders. Permit me to explain. Suppose that the three alternatives under consideration are education appropriations
of $a, $110,000, $10,000 or $g, $0. There are 99 people in our assembly and all 99 rank $a > $f > $g, $110,000 > $10,000 > $0; there is just general agreement that an education expenditure closer to $110,000 than to $10,000 is good for everyone in the community. Individual preference orderings are perfectly correlated with one another. As we introduce more fine-grained alternatives the less structure there is between individual preference orders, but also the less does the absence of structure matter. Suppose now that the three alternatives under consideration are $a, $110,000, $c, $100,000, and $e, $90,000. There is unanimity that educational expenditure should be within the range of $110,000 and $90,000, but there is not unanimity on a figure within that range: 33 voters rank $c > $e > $a, 33 rank $a > $c > $e and 33 rank $e > $c > $a. On superficial inspection the three different rankings don’t seem to resemble one another, but they are somewhat correlated and can be arrayed along one dimension; they enjoy the property of single-peakedness which will be elucidated below. The social ordering by Condorcet order and Borda count is $c > $e > $a, which discloses the resemblance between the individual orderings. Now suppose that the three alternatives are $b, $100,000, $c, $100,000, and $d, $99,900. Everyone thinks that $100,000 is about right, especially when compared to $10,000, but people’s preferences over these last three alternatives $100 apart from one another are not well-considered, there is no structure between individual preference orders, which happens to be perfectly uncorrelated for the sake of my illustration: 33 voters rank $b > $c > $d, 33 rank $d > $b > $c, and 33 rank $c > $d > $b. The collective ordering is a cycle $b > $c > $d > $b, which discloses the absence of structure between the individual preference orders.

The closer alternatives are to one another, the less likely is it that individual preference orders resemble one another, the more likely is it that social cycle results, but the more trivial is the resulting cycle. A cycle over closely adjacent alternatives is not as troubling as a cycle over remotely distant alternatives. I want to distinguish between benign cycles and malign cycles. This distinction would be meaningless in the framework of ordinal and noncomparable utility, but will be of interest to those whose common sense is intact. For the noncomparabilist, that a senator prefers highway expenditures of $15,000,000,000.00 to $15,000,000,000.01 to $14,999,999,999.99 carries just as much information as that a senator prefers a policy of preemptive nuclear war to isolationism to charitable internationalism – the only information the Arrovian permits is the order of preferences. For the Arrovian, a benign cycle among three alternatives a penny apart from each other on $15 billion in highway expenditures is indistinguishable from a malign cycle among world war, neutrality, and charity. I claim, however, that a cycle among three alternatives a penny apart is of no importance compared to a cycle among war, neutrality, and charity. If an arbitrary method of choice were applied to select one from the three highway expenditure alternatives, or even an unfair method such as one agent with unfair control over order of consideration of alternatives, it would be of no practical concern. If, however, a random method, or worse, an unfair method were used to select among the alternatives of war, neutrality, and charity, we would rightly be horrified. My hypothesis is that what empirical cycles there are will much more likely be of the benign rather than the malign variety.

Now we are in a position to propose a possible explanation for the sharp disjunctive identified by Gehlein: almost no cycles with eight or fewer candidates, more cycles with nine or more candidates. Formal theory does not predict this disjunct; the likelihood of a cycle should increase smoothly as we move from three to many candidates. Gehlein’s data were mostly from single-transferable vote elections where voters must rank all candidates, regardless of whether their preferences are fuzzy, indifferent, or incomplete. There is a rule of seven, plus or minus two, for identification of items by humans in short-term memory and in unidimensional absolute judgment tasks (e.g., sounds varying only in loudness, lights varying only in brightness, forms varying only in size; G. Miller 1944/1956; Baddeley 1994; Shiffrin and Nosofsky 1994). Beyond the capacity limit of about seven, humans commit errors, even on the simplest of tasks, and this is a most robust effect. I suggest that the voters in Gehlein’s study committed such errors when judging nine or more candidates, and that these individual errors, structureless ranking over a portion of the range of candidates, sometimes aggregated to cycles. Humans bypass the information-capacity limit by means of three methods. First, relative judgment rather than absolute judgment. Rather than judging brightness on an absolute ten-point scale, instead make relative judgments, $A$ brighter than $B$, $B$ brighter than $C$, and so on; then we can string together the pairwise comparisons. Second, increase the number of dimensions along which stimuli can differ. We can identify hundreds of faces, but faces are sorted over multiple dimensions. Third, make a sequence of several absolute judgments in a row. With respect to memory, for example, a ten-digit phone number is impractical to remember, but chunked into three portions, an area code of three digits, a prefix of three digits, and a suffix of four digits, it is practical to remember. A fourth method not mentioned by the psychologists cited is the use of external technologies.

This is uncanny. As a speculative hypothesis, it seems that the capacity limit and the methods of bypassing the limit are both reflected by
voting procedures, as are the intuitions that modelers of voting strive to formalize. Arrow is intuitively certain that compounding of pairwise comparisons should be the way to rank many alternatives – even when that drives him to dictatorship. Maybe the root of his intuition is a tacit understanding that absolute judgments cannot distinguish more than about seven items along a single dimension. Why did the spatial model of voting, based, so far as I can tell, on no evidence, arise and have such widespread appeal? Maybe the root of the intuition that underwrites the spatial model is again a tacit understanding that humans can better sort multiple alternatives by mapping them to several dimensions (even if those dimensions are highly correlated with one another). We also seem to like handling multiple alternatives by chunking them into sequential procedures: the amendment procedure in legislatures, plurality runoff in elections, federalism. The formal models of voting may be appealing because they appeal tacitly to the capacity-limit problem, but otherwise be misleading because they do not expressly recognize and incorporate the problem.

Here is a new concept. To return to the example of the education appropriations, if the alternatives available for consideration were those between $100,100 and $0, then the collective ordering would be $b > c > d > b > e > f > g$, a so-called top cycle simply because $g$ is at the top of the social ranking. The collective choice rejects $0$ and $10,000$, but cycles among the closely adjacent alternatives $100,000, 110,000, 120,000$. Does this matter? It violates Condition O, which requires an ordering over all alternatives. It would matter perhaps if it resulted in a cyclic deadlock and no action (effectively choosing the last of the three). Saari, a formidable advocate of the Borda count, reports that a class of fresh-minded fourth-graders patiently explained to him that the notorious three-voter cycle was nothing but a tie (Saari 1995a, 50–51). Now consider the four-voter two-alternative majority-rule case where the Condorcet paradox does not apply. Sometimes, perhaps even frequently for ranked $0$ as the assembly’s votes cycled among the top three alternatives, it would mean that no action could be remedied by an arrowian method, and in most instances of a nonbinary method, as we shall see. The arbitrary method would be to define cycles as ties. Then the collective ordering would be $b > c > d > e > f > g$. Further, it may be disconcerting for there to be a tie among two or three alternative single outcome rather than merely formal decisiveness were required. From then we could add the refinement of some kind of fair, perhaps random, tie-breaking institution, and of $b, c, d$ would be chosen. This is known as Schwartz’s method, or the method of transitive closure (Sen 1982, 162–163, 180–183). As a matter of logical possibility, it violates one of Arrow’s conditions. If the transitive closure is taken on the set of all possible alternatives $X$, then the relation satisfies contraction, consistency, and transitivity, also called Sen’s alpha (to be explained in Chapter 6), and applying the method of transitive closure to this unbalanced profile would be taken on the subset $S$ of considered alternatives, then Condition I holds $x \sim y \sim z$. There are 98 out of 99 voters, however, who favor $z$.
Table 5.5. An unbalanced cycle

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>50</td>
<td>50</td>
<td>1</td>
<td>(51)</td>
</tr>
<tr>
<td>y</td>
<td>49</td>
<td>50</td>
<td>1</td>
<td>(99)</td>
</tr>
<tr>
<td>z</td>
<td>98</td>
<td>49</td>
<td>1</td>
<td>(147)</td>
</tr>
</tbody>
</table>

over x, and can we really say that the social relation between z and x should be one of indifference? If the one individual who ranks x first were an evil agenda-controller among foolishly sincere voters, and she pit an alternative y against alternative z, then first y would beat z and next z would lose to x. Alternative x would be the winner even though 98 of the voters favor z over x. The Borda count and the Young-Kemeny rule, however, properly identify z as the winner. For someone who believes in the Arrow theorem, it would make no sense to say that z is better than x. For someone without such beliefs, it seems unfair for z to lose. The method of transitive closure would be unsatisfying because the cycle is unbalanced, z is better than y, and not as good as x.

There are several remedies for this unbalanced cycle. If there is pairwise majority voting, if there is an open agenda such that any alternative can be proposed in any order, and if the members of the assembly prefer any choice over eternal cycling, then z will be selected. How? Before proceeding, an aside. The normative argument against cycling appeals to the intuition that it is bad for a voting procedure to cycle eternally among alternatives. That judgment is external to the model, however. Inside the formal setup no one loses anything from cycling. The normative argument illicits appeal to our real-world experience that indecisiveness is wasteful, and could be doing something better with our time.

Cycling, and we also need a decision now. In order to bring the setup in line with our intuitions, we have to amend our description of preferences. Call the costly absence of a decision arising from an eternal cycle an alternative o. Now we have one individual who ranks x > y > z > o, 49 who rank z > x > y > o, and 49 who rank y > z > x > o. To continue, we will examine three agendas, one beginning with y against z, one beginning with x against y, and one beginning with x against z. I will show that in the final choice on each agenda.

In order to obtain her second-ranked z in the second stage and overall avoid her last-ranked o, then z would win and the cycle would be broken. The incentive of the 49 who rank x first and z > x > y > o overall to vote sincerely in order to obtain z. If in a third stage x is moved against z, then z beats x by 98 to 1. Thus, z is stable. Next, if the first-stage contest is between x and y, then x wins the first stage and loses in the second stage to z. If voters are sincere in the first and second stages, and the third-stage vote is between y and z, and z beats y and the cycle is broken just as before. Would the 49 who rank z > x > y > o vote strategically in the first stage for y over x? No, that would gain them y, their last-ranked alternative and lose them z, their first-ranked alternative. Would the one who ranked x > y > z > o vote strategically in the first stage for y over x? By doing so, this lone individual might gain y, her second-ranked alternative and avoid x, her last-ranked alternative. Therefore, she might contemplate voting strategically. If she did so, however, the next choice would again be between y and z, but we have already seen that z would win that contest among strategic voters for whom a cycle is costly, so she would refrain from voting strategically. Finally, if the first-stage contest is between x and z, then z beats x by 98 to 1, and again z goes on to beat y among our strategic voters for whom any choice among x, y, and z is more important than no choice o. The most strongly favored alternative, y, is picked from the cycle by strategic voting. The same analysis holds if this were a barely unbalanced cycle: one who ranks x > y > z > o, two who rank z > x > y > o, and two who rank y > z > x > o.

The balanced cycle is similar but not identical. Assume three voters each with orders x > y > z > o, z > x > y > o, and y > z > x > o. Say that the first-stage vote is between alternatives y and z. The voter with ranking z > x > y > o could end the cycle by voting for her third-ranked alternative y in order to avoid her fourth-ranked alternative o. Why should she do this? She would rather let y beat z and go up against x. Then the voter with ranking x > y > z > o can end the cycle by voting for her third-ranked alternative. Why should voter x > y > z > o be the one to make the sacrifice, however? And so on. The three face a situation resembling an assurance game. The three need to devise and play in advance a fair tie-breaking institution, not any more challenging, technically or normatively, than to devise and to justify such an institution conventionally.
than some of those other things, but that Sharon wants some of those other things more than she wants B. Ozzie could trade some of those other things to Sharon in exchange for A, thereby breaking the tie. Now (example adapted from Hovenkamp 1990a) suppose a balanced cycle, full knowledge of one another’s utilities, and preferences and associated cardinal but incomparable utilities as follows: Huey prefers $A(6) > B(5) > C(1)$, Dewey prefers $B(60) > C(40) > A(20)$, and Louie prefers $C(600) > A(400) > B(200)$. Suppose the three nephews have no tie-breaking procedures, and that the value to each of perpetual cycling is zero. Then, Huey will vote for B, B will win, and Huey will get 5 rather than zero; and no cycle. Suppose the nephews do have a fair tie-breaking procedure. Then the expected value to Huey of voting for $A$ is $\frac{1}{3} \times 6 + \frac{1}{3} \times 5 + \frac{1}{3} \times 1 = 4$, and the expected value to Huey of voting for $B$ is $1 \times 5 = 5$; hence, Huey votes for $B$, $B$ wins; and no cycle.

There are good voting rules that remedy unbalanced cycles. When the number of voters inhabiting each of the three cyclical rankings are not exactly equal to one another, when there is an unbalanced cycle, then the Borda count almost always declares unique winners, not ties. Riker (1982, 120) is not clear about this property of the Borda count. In discussing the frequency of cycles he says that the Borda count produces a tie among cycled alternatives, but he does not note that the Borda count produces a tie almost always in the rare instance of a perfectly balanced cycle, an equal number of voters across the cyclical rankings. To further illustrate, in Table 5.6 there are 301 voters and three alternatives distributed across three cyclical rankings, followed by the pairwise comparison matrix. By pairwise comparison we have a cycle: $A > B > C > A$. The Borda count, however, ranks the alternatives $A > B > C$, and the Borda count tells us as well, if we are not dogmatic ordinalists, the social judgment is that there is not much difference between the alternatives (if the number of votes had been in three equal groups of 100, the Borda count would have yielded a tie). The Young–Kemeny rule has us break the cycle at the weakest link: the pairwise majority for $C > A$ is 199 votes, less than either the 201 votes for $B > C$ or the 202 votes for $A > B$, and thus the Young–Kemeny ranking is $A > B > C$ just like the Borda count.

In Table 5.7 there is a more ragged distribution of preferences (borrowed from Young 1997), together with the associated pairwise-comparison matrix, involving five of the six possible rankings of three alternatives, that more strongly distinguishes Borda and Young–Kemeny from Condorcet order.

Pairwise comparison delivers a cycle: $A > B > C > A$. Should we stop there and declare a tie? Or should we use the ranking information we have to try and find a winner with a better method than pairwise comparison? The Borda order is $B > A > C$ and the Young–Kemeny order is $B > C > A$ (recall that on assumptions Borda picks the most likely winner and Young–Kemeny the most likely ranking).

If cycles are indeed a threat to the very intelligibility of democracy, then shouldn’t we eagerly seek for voting rules that resolve cycles? If that is our quest, then the accuracy and fairness properties of Borda count and perhaps of Young–Kemeny are distinguished, and thus we should adopt one of those rules and thereby save the republic from Arrovian dictatorship. If, in contrast, the prospect of cycles is rare due to mild homogeneity among voters’ rankings of alternatives, and mild preferences for fairness in distribution, then we may relax, knowing that we
could adopt a more accurate voting rule if required by circumstances, but otherwise confident that simpler voting rules may be adequate for many purposes. Why would anyone reject the cycle-busting voting rules? They violate Arrow’s independence condition, the subject of the next chapter.

Is democracy meaningless? Arrow’s condition of the independence of irrelevant alternatives

---

Introduction

The interpretation that Arrow’s Condition I, independence of irrelevant alternatives, prohibits the use of individuals’ intensities of preference in the construction of social choices is not precise. Rather, it is the social-welfare function (as defined in Chapter 4) which demands both individual and social orderings, and thereby prohibits cardinal utility inputs. Condition I, as Arrow wrote it, redundantly requires individual orderings, but goes further and demands that, even given the ordinal data from individual orderings, the social choice over any two alternatives not be influenced by individuals’ preferences involving any third alternatives. This is explicit in Arrow (1963/1951, 59, emphasis added):

It is required that the social ordering be formed from individual orderings and that the social decision between two alternatives be independent of the desires of individuals involving any alternatives other than the given two. These conditions taken together serve to exclude interpersonal comparison of social utility either by some form of direct measurement or by comparison with other alternative social states.

Arrow’s is a strong independence condition. Slight weakenings of it allow the Borda count or the Young–Kemeny rule as possible social welfare functions and further weakenings permit further voting procedures. Barry and Hardin (1982, 217–218) agree that Arrow’s IIA is a powerful condition. “Part of its power is that one cannot easily intuit what it means or why it matters. Perhaps because of its subtlety, condition I is apparently the condition that is most readily taken for granted in the proof of Arrow’s and related theorems.” Its content is frequently misunderstood. Justifications of the condition are typically thin and dogmatic, often no more than an assertion that its appeal is intuitively obvious. My search for justifications of the condition found thicker arguments mostly by Arrow (1963/1951; 1952; 1967; 1987; 1997), Sen (1970; 1982), some by Riker (1961; 1965; 1982), and otherwise mostly repetition of points made by Arrow without further justificatory development.