**Part Four: Collective Choice Theory**

    Collective Choice Theory is the theory of how one selects a rule to go from a set of individual preference orders over alternatives available to a society of those individuals to a collective or social preference order over those same alternatives. [Or, as they say in the trade, how to "map a set of individual preference orders onto a social preference order."] There is a long history of debates about how to make social or collective decisions, going back at least two and a half millennia in the West. The simplest answer is to identify one person in the society and stipulate that his or her preference order will **be** the social preference order. *L'etat, c'est moi,* as Louis XIV is reputed to have said. A variant of this solution is the ancient Athenian practice of rotating political positions. One can also choose a person by lot whose preferences will thereupon become the social preference. A quite different method is that used by the old Polish parliament, which consisted of all the aristocrats in the country [there were quite a few, the entry conditions for being considered an aristocrat being low]. Since each of them thought of himself as answerable only to God, they imposed a condition of unanimity on themselves. If as few as one Polish aristocrat objected to a statute, it did not become law.

    These rules for mapping individual preference orders onto a social preference order, unattractive as they may be on other grounds, all have one very attractive feature in common: They guarantee that if all of the individual preference orders are **ordinal orderings**, which is to say if each of them is complete, reflexive, and transitive [you see, I told you we would use that stuff], then the social preference order will also be an **ordinal ordering**, and that is something you really, really want. You want it to be complete, so that it will tell you in each case how to choose. And you want it to be transitive, so that you do not get into a situation where your **Collective Choice Rule** tells the society to choose a over b, b over c, and c over a.

    To sum it all up in a phrase, the aim of Collective Choice Theory is to find a way of mapping **minimally**  
**rational individual preferences** onto a **minimally rational social preference.**

    For the past several hundred years, everybody's favorite candidate for a Collective Choice Rule has been **majority rule**. This is a rule that says that the social preference between any two alternatives is to be decided by a vote of all those empowered to decide, with the alternative gaining a majority of the votes being preferred over the alternative gaining a minority of the votes. Should two alternatives, in a pairwise comparison, gain exactly the same number of votes, then the society is to be **indifferent** between the two.

         Enter the Marquis de Condorcet, who published an essay in 1785 called [in English] *Essay on the Application of Analysis to the Probability of Majority Decisions*. In this essay, Condorcet presented an example of a situation in which a group of voters, each of whom has perfectly rational preferences over a set of alternatives, will, by the application of majority rule, arrive at an **inconsistent** group or social preference. This is, to put it as mildly as I can, a tad embarrassing. Indeed, it calls into question the legitimacy of majority rule, which lies at the heart of every variant of democratic theory that had been put forward at that time, or indeed has been put forward since.

    Let us take a moment to set out the example and examine it. In its simplest form, it involves three voters, whom we shall call X, Y, and Z, and three alternatives, which we shall call a, b, and c. We may suppose that a, b, and c are three different tax plans, say. Let us now assume that the three voters have the following preferences over the set of alternatives S = (a, b, c).

    X prefers a to be and b to c. Since X is minimally rational, he also prefers a to c.

    Y prefers b to c and c to a. Since she is also minimally rational, she prefers b to a.

    Z prefers c to a and a to b. As rational as X and Y, she naturally prefers c to b.

    Now they take a series of pairwise votes to determine the collective or social preference order among the three alternatives. When they vote for a or b, X and Z vote for a, Y votes for b. Alternative a wins. When they vote for b or c, X and Y vote for b, Z votes for c, alternative b wins. Now, if the social ordering is to be **transitive**, then the society must prefer a to c. What happens when X, Y, and Z choose between a and c? X prefers a to c. But Y and Z both prefer c to a. So the society must, by majority rule, prefer c to a. Whoops. The society's preference order violates transitivity.

    And that is the whole story. The selection of a social or collective preference order by majority rule cannot guarantee the transitivity of the social preference order, and therefore does not even meet the most minimal test of rationality. There are, of course, lots and lots of sets of individual preference orders that generate a consistent social preference order when Majority Rule is applied to them. The problem is that here is at least one, and actually many more, that are turned by Majority Rule into an inconsistent preference order.

    If you have never encountered this paradox before [the so-called **paradox of majority rule**], you may be inclined to think that it is a trick or a scam or an illusion. Alas, not so. It is just as it appears. Majority Rule really is capable of generating an inconsistent social preference ordering.

    All of this was well known in the eighteenth century, and was, as we shall see later on, the subject of some imaginative elaboration by none other than the Reverend Dodgson, better known as Lewis Carroll. Enter now the young, brilliant economist Kenneth Arrow in the middle of the twentieth century. Coming out of a tradition of economic theorizing called Social Welfare Economics, to which a number of major figures, such as Abram Bergson, had contributed, Arrow conceived the idea of analyzing the underlying structure of the old Paradox of Majority Rule and generalizing it. The result, which he presented in his doctoral dissertation no less, was The General Possibility Theorem. Arrow published the theorem in 1951 in a monograph entitled *Social Choice and Individual Values*.

    Another great economist and fellow Nobel Prize winner, Amartya Sen, in 1970 published *Collective Choice and Social Welfare*, in which he generalized and extended Arrow's work in astonishing ways. Sen's book is difficult, but it is simply beautiful, and deeply satisfying. I strongly urge you, if you have a taste for this sort of thing, to tackle it. Sen has written widely and brilliantly on a host of extremely important social problems, including economic inequality, famine, and the demographic imbalance between men and women in the People's Republic of China. His little series of Radcliffe Lectures, published in 1973 as *On Economic Inequality*, is the finest use of formal methods to illuminate and analyze a social problem of which I am aware. It is a perfect example of the *proper* use of formal methods in social philosophy, and as such deserves your attention.

    In *Collective Choice and Social Welfare*, Sen gives a simpler and more elegant proof of Arrow's General Possibility Theorem. Nevertheless, I have chosen in this blog to expound Arrow's original proof. Let me explain why. It often happens that the first appearance of an important new theorem is somewhat clumsy, valid no doubt, but longer and more complicated than necessary. Later theorists refine it and simplify it until what took many pages can be demonstrated quickly in a few lines. Sometimes, this development is unambiguously better, but at other times, the original proof, clumsy though it may be, reveals the central idea more perspicuously than the later simplifications do. I find this to be true in the case of Arrow's theorem. Sen's simplification serves several purposes, not the least of which is to set things up formally for his extremely important extension and elaboration of Arrow's work. Therefore, I urge you to look at it, once you have worked with me through Arrow's original proof.

    Now let us begin. This is going to take a while, so settle down. Before we get into the weeds, let me try to explain in general terms what Arrow is doing. He asks, in effect, what are the underlying general assumptions of majoritarian decision making? What is it about voting with majority rule that appeals to us? He identifies five conditions or presuppositions [later reduced and simplified to four] that capture the logic of majority rule in a general way, and then shows that *no* way of making collective decisions that satisfies all four of them guarantees that the resulting social or collective choice will be consistent. This way of thinking about the problem accomplishes three things simultaneously. First, it unpacks majority rule voting into its component parts so that we can look at it and understand it better. Second, it generalizes the Paradox of Majority Rule so that we realize we cannot avoid it simply by tweaking Majority Rule a bit [for example by requiring a two-thirds majority.] And finally, it allows us to see just exactly what Majority Rule does *not* do -- in other words, it gives us insight into what would be totally different ways of making collective decisions.

    We start with a series of assumptions, definitions, and notational conventions, some of which are already familiar to you from the opening segments of this general tutorial. This is going to be tedious, but learning these up now will make it infinitely easier to follow the proof. Here they are:

(a) We start with a set of mutually exclusive alternatives, x, y, z, ..... These may be all of the possible candidates in an election [i,.e., every single person who is eligible to hold office under the rules governing the election], every possible tax scheme that might come before Congress, all of the various possible decisions a City Council might take concerning zoning regulations, and so forth. The point of the phrase "mutually exclusive" is to rule out, for example, "Obama" and "Obama or Clinton" as two of the available alternatives.

(b) On any give occasion when a decision is to be made, there is a subset, S, of the available alternatives, which will be called The Environment. This might be, for example, the relatively small number of people who have stated publicly that they would like to be elected to that office, or all the people who have formed campaign committees, or all the people who survive the primary season and are on the final ballot. Each of these is a subset of all the people eligible to hold the office [not necessarily a proper subset -- i.e., not necessarily smaller than the total set of alternatives. All that is required is that S be included in the set of all alternatives, not that it be smaller than that set].

(c) There is a set of individuals ["voters"], identified by numerical subscripts, 1, 2, 3, 4, ....

(d) Each individual is assumed to have a **complete, transitive ranking** of the entire set of alternatives, which we indicate using the notation introduced earlier -- the binary relations R, I, and P. Just to review, xRiy means that individual i considers alternative x to be as good as or better than alternative y. xPiy and xIiy are derived from R in the way indicated in the opening segments of this tutorial. What we are aiming for, of course, is a collective or social ranking, and that is indicated by the same letters, R, P, and I without the subscripts. So xPy means that the *society* prefers s to y. The whole point of this exercise is to start with complete, transitive **individual** rankings of the alternatives and then see whether there is any way of going from the individual rankings to a social ranking that satisfies certain conditions [see below] and results in a **social** ranking that is complete and transitive.

(e) Ri all by itself refers to individual i's ranking of the entire set of alternatives, x, y, z, .... Correspondingly, R all by itself refers to the society's ranking of the entire set of alternatives.

(f) We shall have occasion to refer to different possible rankings, by an individual i, of the set of alternatives. We will indicate these different rankings by superscripts. So, for example, Ri is one ranking by individual i of the entire set of alternatives. Ri' is a second ranking. Ri'' is a third ranking. And Ri\* is a fourth ranking. A ranking Ri can be thought of either as a list showing the way individual i ranks the alternatives, including ties [indifference], or as a set of all the ordered pairs (x,y) such that xRiy.

(g) A Social Welfare Function [ an SWF ] is a function that maps sets of individual rankings onto a social ranking. Such a mapping function qualifies as an SWF just in case **both** the individual rankings, the Ri, **and** the social or collective ranking, R, satisfy Axioms I and II below -- which is to say, just in case the rankings, both individual and social, are **complete** and **transitive**.

(h) A Social Welfare Function is said to be Dictatorial if there is some individual i such that, for all x and y, xPiy implies xPy regardless of the orderings of all of the individuals other than i. Thus, in particular, to say that an SWF is dictatorial is to say that there is some individual who can impose his or her will on the society with regard to the choice between any pair, x and y, even if everyone else in the society has the opposite preference as between those two alternatives.

     (i) Finally, we define something called a Social Choice Function [ symbolized as C(S).] C(S) is the set of all alternatives x in the Environment S such that for every y in S, xRy. In other words, C(S) is the set of top alternatives or best alternatives in S. Quite often, C(S) will contain only one alternative, the one that the society prefers over all the others. But it may include more than one if the society is indifferent as among several best alternatives.

    Those are the nine definitions and stipulations. The key new ones that we have not met before are S, the set of available alternatives, R, the social ranking, SWF, a Social Welfare Function, and C(S), the Choice Function. Now Arrow lays down two Axioms governing the social ordering, R. These are:

Axiom I: For all x and y, xRy or yRx [Completeness]

Axiom II: For all x, y, and z, if xRy and yRz then xRz. [Transitivity]

    O.K. So much for the preliminary throat clearing. I want you to go over these definitions and stipulations until you are comfortable with them. The proof is going to be a formal argument couched in terms of these symbols and appealing to these assumptions and axioms. You will find it impossible to follow if you do not have a solid grasp on these preliminary definitions and so forth. While you are doing that, I want to talk for a bit about several important points that are implicit in what we have just laid down, but may not be obvious.

From here on, I am going to break the exposition into short bits, because this is hard, and I do not want to lose anyone. My apologies to those of you who are having no trouble following it.

    First of all, notice that Arrow assumes only ordinal preference. This means that there is no way in the proof to take account of intensity of preference, only order of preference. Let me give an example to make this clearer. In 1992, George H. W. Bush, Bill Clinton, and H. Ross Perot ran for the Presidency. There were some devoted followers of Perot who were crazy about him, and almost indifferent between Bush and Clinton, whom they viewed as both beltway politicians. Let us suppose that one of these supporters ranked Perot first, way ahead of the other two, and gave the edge slightly to Bush over Clinton, perhaps because Bush was a Republican. A second Perot supporter might have been rather unhappy with the choices offered that year, but preferred Perot slightly over Bush, while hating Clinton passionately. From Arrow's perspective which is that of ordinal preference, these two voters had identical preferences, namely Perot > Bush > Clinton, and an Arrovian SWF would treat the two individual preference orders as interchangeable.

    Now, there are many ways in which citizens in America can give expression to the intensity of their preferences, as political scientists are fond of pointing out. One is simply by bothering to vote. Voter enthusiasm, in a nation half of whose eligible voters routinely fail to go to the polls, is a major determinant of the outcome of elections. A second way is by contributing to campaigns, volunteering for campaign work, and so forth. Yet another way is through a vast array of voluntary organizations dedicated to pursuing some issue agenda or advantaging some economic or regional group. *None of this can find expression in the sort of Social Welfare Function Arrow has defined.* This is a very important limitation on the method of collective decision that we call voting. Now, there are voting schemes that allow voters to give expression to the intensity of their preferences [such as giving each voter a number of votes, which he or she can spread around among many candidates or concentrate entirely on one candidate], but these too are ruled out by Arrow, who only allows the SWF to take account of individual ordinal preferences.

    The second thing to note is that the requirement of completeness placed upon the SWF rules out partial orderings, such as those established by Pareto-Preference. It is often the case that every individual in the society prefers some alternative x to some other alternative y, and if there are a number of such cases, a robust partial ordering might be established that, while not complete, nevertheless allows the society to rank a sizeable number of the available alternatives. This option too is ruled out by Arrow's two axioms. These observations have the virtue of helping us to understand just how restricting a collective decision-making apparatus like majority rule is.

    We are now ready to state the four conditions that Arrow defines as somehow capturing the spirit of majoritarian democracy. Arrow's theorem will simply be the proposition that there is no Social Welfare Function, defined as he has in the materials above, that is compatible with all four conditions. In the original form of the proof, the conditions were, as you might expect, called Conditions 1, 2, 3, 4, and 5. In the revised version, which I shall be setting forth here, they are called Conditions 1' [a revised version of Condition 1], Condition 3 [which also is sometimes called the Independence of Irrelevant Alternatives], Condition P [for Pareto], and Condition 5. Here they are. I will tell you now that Condition 3 is the kinky one.

Condition 1': All logically possible rankings of the alternative social states are permitted. This is a really interesting condition. What it says, formally speaking, is that each individual may order the alternatives, x, y, z, ... in any consistent way. What it rules out, not so obviously, is any religious or cultural or other constraint on preference. For example, if among the alternatives are various dietary rules, or rules governing abortions, or rules governing dress, nothing is ruled in or ruled out. The individuals are free to rank alternatives in any consistent manner.

Condition 3: Let R1, R2, ......, Rn and R1', R2', .... Rn' be two sets of individual orderings of the entire set of alternatives x, y, z, .... and let C(S) and C'(S) be the corresponding social choice functions. If, for all individuals i and all alternatives x and y in a given environment S, xRiy if and only if xRi'y, then C(S) and C'(S) are the same.

    OK, this is confusing, so let us go through it slowly step by step and figure out what it means. To get to the punch line first, this condition says that the society's eventual identification of best elements in an environment is going to be determined solely by the rankings by the individuals of the alternatives in that environment, and not by the rankings by the individuals of alternatives not in the environment. [Remember, the Environment, S, is a subset of all the possible alternatives.] Now, take the condition one phrase at a time. First of all, suppose we have two different sets of individual rankings of all the alternatives. The first set of rankings is the Ri [there are as many rankings in the set as there are individuals -- namely, the first individual's ranking, R1, the second individual's ranking, R2, and so forth.] The second set of rankings is the Ri', which may be different from the first set.

    Now, separate out some subset of alternatives, which we will call the Environment S, and focusing only on the alternatives in S, take a look at the way in which the individuals rank those alternatives, ignoring how they rank any of the alternatives left out of S. If the two sets of individual orderings, Ri and Ri', are exactly the same for the alternatives in S, then when the Social Welfare Function cranks out a social ranking, R, based on the individual orderings Ri and a social ranking, R', based on the individual orderings Ri', Condition 3 stipulates that the set of best elements [The Social Choice set] will be the same for R and for R'.

    Whew, that still isn't very clear, is it? So let us ask the obvious question: What would this Condition rule out? Here is the answer, in the form of an elaborate example. Just follow along.

    Suppose that in the 1992 presidential election, there are just three voters, whom we shall call 1, 2, and 3. Also, suppose there are a total of four eligible candidates: George H. W. Bush, Bill Clinton, H. Ross Perot, and me. Now suppose there are two alternative sets of the rankings of these four candidates by individuals 1, 2, and 3.

Ri:     Individual 1: Wolff > Clinton > Bush > Perot

    Individual 2: Bush > Perot > Wolff > Clinton

    Individual 3: Wolff > Clinton > Bush > Perot

Ri':    Individual 1: Clinton > Bush > Perot > Wolff

    Individual 2: Bush > Perot > Clinton > Wolff

    Individual 3: Clinton > Bush > Perot > Wolff

    The crucial thing to notice about these two alternative sets of rankings is that they are identical with regard to the environment S = (Bush, Clinton, Perot). The only difference between the two sets is that in the second set, Wolff has been moved to the bottom of everyone's list. [The voters find out I am an anarchist.]

    Now let us consider the following Social Welfare Function: For each individual ranking, assign 10 points to the first choice, 7 points to the second choice, 3 points to the third choice, and 2 points to the fourth choice. Then, for any Environment, S, selected from the totality of available alternatives, determine the social ranking by adding up all of the points awarded to each alternative by the individual rankings. Got it?

    Go ahead and carry out that exercise. If you do, you will find that for the first set of rankings, the Ri, and for the Environment S = (Bush, Clinton, Perot). the SWF gives 16 points to Clinton, 16 points to Bush, and 11 points to Perot. So, C(S), the society's decision as to which candidates are at the top, is (Clinton, Bush), because they each have the same number of points, namely 16. But if you now carry out the same process with regard to the second set of individual rankings, the Ri', and the same Environment S, you will discover that the SWF assigns 23 points to Clinton, 24 points to Bush, and 13 points to Perot, which means that C'(S) is (Bush). So the social choice in the Environment S has changed, despite the fact that the relative rankings of the elements in S have not changed, because of a change in the rankings of an element not in S, namely Wolff. And this is just what Condition 3 rules out. It says that the Social Welfare Function cannot be one that could produce a result like this.

    All of us are familiar with this sort of problem from sports meets or the Olympics. When we are trying to decide which team or country has done best, we have to find some way to add up Gold medals and Silver medal and Bronze medals, and maybe fourth and fifth places as well. And, as we all know, you get different results, depending on how many points you award for each type of medal. Arrow's Condition 3 rules out SWFs like that.

Condition P: If xPiy for all i, then xPy. This just says that if everyone strongly prefers x to y, so does the society. This is a very weak constraint on the SWF.

Condition 5: The Social Welfare Function is not dictatorial.

    Remember the definition of "dictatorial" above. This rules out "l'état c'est moi" as a Social Welfare Function.

    So, we have the definitions, etc., and we have the four Conditions that Arrow imposes on a Social Welfare Function. Remember that a Social Welfare Function is *defined* as a mapping that produces a social ranking that satisfies Axioms I and II. Now Arrow is ready to state his theorem. It is quite simple:

**There is no Social Welfare Function that satisfies the four Conditions**.

This is really a devastating theorem. Basically, it says that there is no voting mechanism that gets around the Paradox of Majority Rule. The proof proceeds as follows. First Arrow states a set of little results about the relations R, I, and P. You are already familiar with them. They are trivial, as we shall see. Then he proves a little Lemma about the choice function. Then he proves a big important Lemma that is really the guts of the theorem. Finally, he uses the Lemmas to prove what is essentially an extension of the Paradox of Majority Rule, and he is done. We are going to go through this slowly and carefully. Let us start with the two little lemmas. Lemma 1 and Lemmas 2.

Lemma 1:    (a)    For all x, xRx

        (b)    If xPy then xRy

        (c)    If xPy and yPz then xPz

        (d)    If xIy and yIz then xIz

        (e)    For all x and y, either xRy or yPx

        (f)    If xPy and yRz then xPz

    These all follow immediately from the definitions of R, I, and P, the assumptions of transitivity and completeness, and truth functional logic. Arrow includes them as an omnibus Lemma because at one point or another in his proof he will appeal to one or another of them. You should work through all the little proofs as an exercise. I will go through just one to show you what they look like.

        (e) xRy or yRx [completeness]

         So if not xRy, then yRx.

         But the definition of yPx is yRx and not xRy

         Therefore, either xRy or yPx

Lemma 2:    xPy if and only if x is the sole element of C([x,y])

    If you review the definition of the Choice set, you will see that this Lemma is intuitively obvious. It says that in the little environment, S, consisting of nothing but x and y, if xPy, then x is the only element in the Choice set, C(S). Since this is a bi-conditional [if and only if], we have to prove it in each direction.

    a. Assume xPy. Then xRy, by Lemma 1(b). [See, this is why he put those little things in Lemma 1]. Furthermore, xRx, by Lemma 1(a). So x is in C([x,y]), because it is at least as good [i.e., R] as each of the elements of S, namely x and y. But if xPy then not yRx. Therefore, y is not in C([x,y]). So x is the sole element of C([x,y]).

    b. Assume x is the sole element of C([x,y]). Since y is not in C([x,y]), not yRx. Therefore xPy.

Lemma 3:    If an individual, i, is decisive for some ordered pair (x,y) then i is a dictator.

    This is a rather surprising and very important Lemma. It is the key to the proof of Arrow's theorem, and shows us just how powerful the apparently innocuous Four Conditions really are. To understand the Lemma, you must first know what is meant by an ordered pair and then you must be given three definitions, including one for the notion of "decisive."

    Easy stuff first. An ordered pair is a pair in a specified order. An ordered pair is indicated by curved parentheses. Thus, the ordered pair (x,y) is the pair [x,y] in the order first x then y. As we shall see, to say that individual is decisive for some ordered pair (x,y) is to say that i can, speaking informally, make the society choose x over y regardless of what anyone else thinks. But a person might be decisive for x over y and not be decisive for y over x. We shall see in a moment how all this works out. Now let us turn to the three definitions that Arrow is going to make use of in the proof of Lemma 3.

    Definition 1: "A set of individuals V is decisive for (x,y)" =df "if xPiy for all i in V and yPjx for all j not in V, then xPy"

    In other words, to say that a set of individuals V is decisive for the ordered pair (x,y) is to say that if everyone in V strongly prefers x to y, and everyone not in V strongly prefers y to x, then the society will strongly prefer x to y. Under majority rule, for example, any set of individuals V that has at least one more than half of all the individuals in the society in it is decisive for every ordered pair of alternatives (x,y).

    Definition 2: "xḎy for i" or "i dictates over (x,y)" =df "If xPiy then xPy"

    In words, we say that individual i dictates over the ordered pair (x,y) if whenever individual i strongly prefers x to y, so does the society regardless of how everyone else ranks x and y. [Notice that the capital letter D has a little line underneath it.]

    Definition 3: "xDy for i" or "i is decisive for (x,y)" =df "If xPiy, and for all j not equal to i, yPjx, then xPy."

    In words, i is said to be decisive for the ordered pair (x,y) if when i strongly prefers x to y and everyone else strongly prefers y to x, the society prefers x to y. [Notice that in this definition, the capital letter D does not have a little line underneath it.]

    Ok. Now we are ready to state and prove the crucial Lemma 3.

Lemma 3: If xDy for i, the zḎw for i, for all z,w in S

    In words, what this says is that if any individual, i, is decisive for some ordered pair (x,y) then that individual i is a dictator [i.e., dictates over any ordered pair (z,w) chosen from S]. This is an astonishing result. It says that if the Social Welfare Function allows someone to compel the society to follow her ranking of some ordered pair, no matter what, against the opposition of everyone else, then the Social Welfare Function makes her an absolute dictator. [*L'ėtat c'est moi*]. Here is the proof. It is going to take a while, so settle down. In order to make this manageable, I must use the various symbols we have defined. Let me review them here, so that I do not need to keep repeating myself.

    An ordered pair is indicated by curved parentheses: (x,y), as opposed to a non-ordered pair, which is indicated by brackets: [x,y].

    xḎy for i, which is D with a line under it, means "i dictates over (x,y)" (an ordered pair)

    xDy for i, which is D with no line under it, means "i is decisive for (x,y)"

Proof of Lemma 3:        Assume xDy for i [i,e,, i is decisive for x against y]

    The proof now proceeds in two stages. First, for an environment [x,y,z], constructed by adding some randomly chosen third element z to x and y, we show that i is a dictator over [x,y, z].

    Then we show how to extend this result step by step to the conclusion that i is a dictator over the entire environment S of admissible alternatives.

First Stage: Proof that i is a dictator over the environment [x,y,z]

(step i)    Construct a set of individual orderings over [x,y,z] as follows.

    Ri: x > y > z [i.e., individual i's ordering of the three]

    All the other Rj: yPjx yPjz Rj[x,z] unspecified

    In other words, we will prove something that is true regardless of how everyone other than i ranks x against z.

(step ii)    xPiy by construction. But, by hypothesis xDy for i. Therefore xPy

    In words, i is assumed to strongly prefer x to y, and since by hypothesis i is decisive for x against y, the society also strongly prefers x to y.

(step iii)     For all i, yPiz, by construction Therefore, yPz, by Condition P, and xPz by Lemma 1(c). In words, since everyone strongly prefers y to z, so does the society. And since the society strongly prefers x to y and y to z, it strongly prefers x to z [since Axiom II, which is used to prove Lemma 1(c), stipulates that the SWF is transitive.]

(step iv) So xPz when xPiz, regardless of how anyone else ranks x and z. [check the construction of the individual orderings in step (i) ]

(step v)    Hence xḎz for i, which is to say that i dictates over the ordered pair (x,z)

(step vi)    Now consider (y,z) and assume the following set of individual orderings:

    Ri:    y > x > z

    All the other Rj: yPjx zPjx and Rj[y,z] unspecified.

(step vii)    yPix for all i. Therefore yPx by Condition P

(step viii)    xḎz for i, by (v). Hence xPz.

(step ix)    So yPz by Lemma 1(c). Thus yḎz for i.

    In words, we have now shown that i dictates over the ordered pair (y,z). Let us take a minute to review what is going on here. We are trying to prove that if i is decisive for a single ordered pair, (x,y), then i is a dictator over an environment consisting of x, y, and some randomly chosen z. If we can show that i is a dictator for every ordered pair in the environment [x,y,z] then we shall have shown that i is a dictator over that environment. There are six ordered pairs that can be selected from the environment, namely (x,y), (x,z), (y,x), (y,z), (z,x), and (z,y). So we must establish that i dictates over every single one of these ordered pairs. We have already established that i dictates over (x,z) in step (v) and over (y,z) in step (ix).

(step x) We can now extend this argument to the other four ordered pairs that can be selected from the environment [x,y,z]. In particular, let us do this for the ordered pair (y,x). Construct the following set of orderings:

    Ri:    y > z > x

    All the other Rj: zPjy zPjx Rj[x,y] unspecified.

(step xi)      zPix for all i. Hence zPx by Condition P

(step xii)    yḎz for i by (step ix). Hence yPz

(step xiii)    So yPx by Lemma 1(c). Thus yḎx for i.

    So we have proved [or can do so, by just iterating these steps a few more times] that i dictates over every ordered pair in [x,y,z], and therefore i is a dictator over the environment [x,y,z]. So much for Stage One of the proof of Lemma 3. Now, take a deep breath, review what has just happened to make sure you understand it, and we will continue to:

Stage Two: The extension of our result to the entire environment, S, of available alternatives. Keep in mind that S, however large it may be, is finite.

    Assume xDy for i [our initial assumption -- just repeating for clarity] and also assume the result of Stage One. Now consider any ordered pair of alternatives (z,w) selected from the environment S. There are just seven possibilities.

    1. x = z     w is a third alternative

    2. x = w     z is a third alternative

    3. y = z     w is a third alternative

    4. y = w     z is a third alternative

    5. x = z     y = w

    6. y = z    x = w

    7. Neither z nor w is either x or y

Case 1:        We have an environment consisting of three alternatives: [x=z, y, w]. Stage One shows that if xDy for i, then x=zḎw for i.

Case 2, 3, 4:    Similarly

Case 5:        Trivial

Case 6:        Add any other element v to form the environment [x=w, y=z, v]. From x=wDy=z for i, it follows that y=zḎx=w for i. [In words, just in case you are getting lost: In the case in which y is element z and x is element w, from the fact that i is decisive for x against y, which is to say for w against z, , it follows that i dictates over y and x, which is to say over z and w. This is just a recap of Stage One.

case 7: This is the only potentially problematic one case, and it needs a little explaining. We are starting from the assumption that i is decisive for x against y, and we want to show that i is a dictator over some totally different of alternatives z and w, so we are going to creep up on that conclusion, as it were. First we will add one of those two other alternatives, z, to the two alternatives x and y to form the environment [x,y,z]. From Stage One, if xDy for i then xḎz for i. But trivially, since xḎz for i, it follows that xDz for i. [The point is that if i dictates over x and z, then of course i is decisive for x against z].

    Now add w to x and z to form the environment [x,z,w]. Since xDz for i, it follows that zḎw for i, by Stage One. In words, if i is decisive for x against z, then in the environment [x,z,w], i dictates over z and w. This follows from Stage One. What this shows is just how powerful Lemma 3 really is.

    Thus we have demonstrated that xDy for i implies zḎw for i, for all z and w in S. In other words, if i is decisive for some ordered pair (x,y), then i is a dictator over S. But Condition 5 stipulates that no individual may be a dictator. Therefore:

    An acceptable Social Welfare Function does not permit any individual to be decisive for even a single ordered pair of alternatives in the environment S of available alternatives.

    Can we all say Ta-Da? This is the heavy lifting in Arrow's theorem. Using this Lemma, we can now fairly quickly prove that there is no SWF satisfying Axioms I and II and all four Conditions, 1', 3, P, and D.

Proof of Arrow's Theorem

Step 1.     By Condition P, there is at least one decisive set for each ordered pair, namely the set of all the individuals. From all the decisive sets, choose a smallest decisive set, V, and let it be decisive for some ordered pair (x,y). What I mean is this: Consider each set of individuals that is decisive for some ordered pair or other. Since there is a finite number of individuals, each of these sets must have some finite number of individuals in it. And the sets may have very different numbers of individuals in them. But one or more of them must be the smallest set. So arbitrarily choose one of the smallest, call it V, and label the pair of alternatives over which it is decisive (x,y).

Step 2:     By Condition P, V cannot be empty. [Go back and look at Condition P and make sure you see why this is so. It is not hard]. Furthermore, by Lemma 3, V cannot have only one member [because Lemma 3 proved that no single individual, i, can be decisive for any ordered pair (x,y) ]. Therefore, V must have at least two members.

Step 3: Partition the individuals 1, 2, ......, n in the following way:

The set of all individuals

---------------------------------------------------------------

| |

V V3

---------------------

V1 V2

    Where V1 = a set containing exactly one individual in V

        V2 = the set of all members of V except the one individual in V1

        V3 = the rest of the individuals, if there are any.

    Is this clear? V is a smallest decisive set. It must have at least two individuals in it. So it can be divided into V1 containing just one individual, and V2 containing the rest of V. V3 is then everyone else, if there is anyone else not in the smallest decisive set V.

Step 4: Now let the individuals in the society have the following rankings of three alternatives, x, y, and z. [And now you will see how this is an extension of the original Paradox of Majority Rule with which we began.]

    V1: x > y and y > z

    V2: z> x and x> y

    V3: y>z and z>x

[You see? This is one of those circular sets of preference orders: xyz, zxy, yzx]

    V1 is non-empty, by construction.

    V2 is non-empty, by the previous argument.

    V3 may be empty.

Step 5: a) By hypothesis, V is decisive for x against y. But V is the union of V1 and V2, and xPiy for all i in V1 and V2. Therefore, xPy. [i.e., the society prefers x to y.]

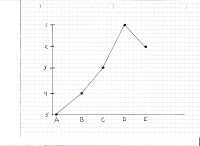
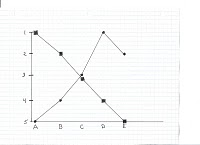
    b) For all i in the union of V1 and V3, yPiz. For all j in V2, zPjy. If zPy, then V2 is decisive for (x,y). But by construction, V2 is too small to be decisive for anything against anything, because V2 is one individual smaller than a smallest decisive set, V. Therefore not zPy. Hence, yRz [see the definitions of P and R].

    c) Therefore xPz by Lemma 1(f) [go back and look at it].

    d) But xP1z and zPix for all i not in V1, so it cannot be that xPz, because that would make V1 decisive for (x,z), which contradicts Lemma 3. Therefore, not xPz.

Step 6: The conclusion of Step 5d) contradicts Step 5c). Thus, we have derived a contradiction from the assumption that there is a Social Welfare Function that satisfies Conditions 1', 3, P, and 5. Therefore, there is no SWF that satisfies the four Conditions. Quod erat demonstrandum.

    OK. Everybody, take a deep breath. This is a lot to absorb. Arrow's Theorem is a major result, and it deserves to be studied carefully. Go back and re-read what I have written and make sure you understand every step. It is not obscure. It is just a little complicated. If you have questions, post them as a comment to this blog and I will answer them.

With this segment, I conclude my discussion of Arrow's General Possibility Theorem. And I think this will also conclude this tutorial on the use and abuse of formal methods in political philosophy. I will be happy to respond to questions, if there are any, but I think enough is enough. Thank you all for staying with me on this, for pointing out errors, and for asking questions. It has been fun for me, revisiting material I have not taught for twenty years or more, and I hope it has been informative and fun for you.  
  
An extremely interesting result concerning the consistency of majority rule was produced by the Australian political scientist Duncan Black. In a book called The Theory of Committees and Elections, published in 1958, Black proved an important theorem about circumstances under which majority rule is guaranteed to produce a transitive social preference ordering. In a moment, I am going to go through the proof in detail, but let me first explain intuitively what Black proved. Ever since the French Revolution, political commentators have adopted a convention derived from the seating arrangement in the National Assembly. In that body, Representatives belonging to each party were seated together, and the groups were arrayed in the meeting hall in such a manner that the most radical party, the Jacobins, sat on the extreme left of the hall, and the most reactionary party, the Monarchists, sat on the right, with the other groups seated between them from left to right according to the degree to which their policies deviated from one extreme or the other. Thus was born the left-right political spectrum with which we are all familiar. [Of course, in the U. S. Senate, there are no Communists and only one Socialist, but, as the reign of George W. Bush shows, there are still plenty of Monarchists.]  
  
The interesting fact, crucial for Black's proof, is that wherever a party locates itself on the spectrum, it tends to prefer the positions of the other parties, either to the left or to the right, less and less the farther away they are seated. So, if an individual identifies himself with a party in the middle, he will prefer that party's positions to those of a party a little bit to the left, and he will prefer the policies of the party a little bit to the left over those of a party farther to the left, and so on. The same is true looking to the right. Notice that since only ordinal preference is assumed, you cannot ask, "Is a party somewhat to the left of you farther from you than a party somewhat to the right of you?" [Make sure you understand why this is true. Ask me if it is not.]  
  
Consider contemporary American politics. If I am a moderate Republican [assuming there still is one], I will prefer my position to that of a conservative Republican, and I will prefer that position to a right wing nut. I will also prefer my position to that of a Blue Dog Democrat [looking to my left rather than to my right], and that position to the position of a Liberal Democrat, and that position in turn to the position of a Socialist [Bernie Sanders?].  
  
This can be summarized very nicely on a graph, along the X-axis of which you lay out the left-right political spectrum, while on the Y-axis you represent the order of your preference. Pretty obviously, the graph you draw will have a single peak -- namely, where your first choice is on the X-axis -- and will fall away on each side, going monotonically lower the farther you get on the X-axis from your location on it. In short, your preference, when graphed in this manner, will be single-peaked. Here is an example of a person's preference order graphed in this manner. For purposes of this example, there are five alternatives, (a, b, c, d, e), and the individual has the following preference order: d > e > c > b > a   
  
  
  
[](http://1.bp.blogspot.com/_lOTQWmy99zo/TE_5VEkTa7I/AAAAAAAAADQ/3eJKnNCDg7U/s1600/duncan+black.jpg)  
  
Let us suppose that there is a second person whose preference order is a > b > c > d > e. It is obvious that if we posted this person's preferences on the same graph, the two together would look like this:  
  
[](http://3.bp.blogspot.com/_lOTQWmy99zo/TE_6MDqFM3I/AAAAAAAAADY/PgVntFa7OGM/s1600/duncan+black+2.jpg)  
  
Notice that each of these lines has a single peak. The first individual's line peaks at alternative D; the second's at alternative A. If you do a little experimenting, you will find that if you change the order in which the alternatives are laid out on the X-axis, sometimes both lines are still single peaked, sometimes one remains single peaked and one no longer is. Sometimes neither is single peaked. Fr example, if you change the order slightly so that the alternatives are laid out on the X-axis in the order a B E D C, the first individual's line will still be single peaked, but the second individual's line will now be in the shape of a V with one peak at A and another peak at C. [Try it and see. It is too much trouble for me to draw it and scan it and size it and insert it.]  
  
Suppose now that we have an entire voting population, each with his or her own preference order, and that we plot all of those preference orders on a single graph, a separate line for each person. There might be some way of arranging the alternatives along the X-axis so that everyone's preference order, when plotted on that graph, is single peaked. Then again, there might not be. For example, if you have three people and three alternatives, and if those three people have the preferences that give rise to the Paradox of Majority Rule, then there is no way of arranging the three alternatives along the X-axis so that all three individuals' preferences orders can be plotted on that graph single-peakedly. [Try it and see. Remember that mirror images are equivalent for these purposes, so there are really three possible ways of arranging the alternatives along the X-axis, namely xyz, xzy, and yxz.]  
  
Duncan Black proved that if there is some way of arranging the available alternatives along the X-axis so that everyone's preference order, plotted on that graph, is single peaked, then majority rule is guaranteed to produce a consistent social preference order. Notice, in particular, that if everyone's preferences can be mapped onto the familiar left-right spectrum, with each individual preferring an alternative less and less the farther away it is in either direction from the most preferred alternative, then everyone will on that graph have a single peaked order [because it will peak at the most preferred alternative and fall away monotonically to the right and to the left.]  
  
The proof is fairly simple. It goes like this.  
  
Step 1: Assume that there are an odd number of individuals [the proof works for an even number of individuals, but in that case there can be ties, which produces social indifference, which then requires an extra couple of steps in the proof, so I am trying to make this as simple as possible.] Assume that their preferences can be plotted onto a graph so that all of the plots are single-peaked.  
  
Step 2: Starting at the left, count peaks [there may be many peaks at the same point, of course, showing that all of those people ranked that alternative as first] and keep counting until you reach one more than half of the total number of peaks, i.e. (n/2 + 1). Assume there are p peaks to the left of that point, q peaks at that point, and r peaks to the right, with (p+q+r) = n. Now, by construction, (p+q) > n/2 and pn/2, because if (q+r)n/2, which by construction it is not.  
  
Step 3: Let us call the alternative with the q peaks alternative x. Clearly, there is a majority of individuals who prefer x to every alternative to the right of x on the graph, because there are p+q individuals whose plots are downward sloping from x as you go to the right, which means they prefer x to everything to the right, and p+q is a majority. But there are q+r individuals who prefer x to everything to the left of x, because their plots are downward sloping as you go to the left, and q+r are a majority. So alternative x is preferred in a pairwise comparison by a majority to every other alternative.  
  
Step 4: Remove alternative x from the graph, remove alternative x from everyone's preference order, and then redraw all of the plots. They will all still be single-peaked. Why? Well, there are three possible cases: Either the dot representing the individual's ranking of x was the peak, or it was to the left of the peak, or it was to the right. In each case, when you reconnect the remaining dots, the graph remains single-peaked [try it and see. It is too hard to draw it and scan it and upload it. But it is intuitively obvious.]  
  
Step 5: You now have a new set of single-peaked plots on a single graph, so go through Steps 2 and 3 all over again. The winning alternative is preferred to every other remaining alternative, and is of course inferior to the first winner. If you now iterate this process until you run out of alternatives, you are left with a fully transitive social preference established by repeated uses of majority rule.  
  
Black's theorem has considerable real world application, as we have seen, but it of course does not identify necessary and sufficient conditions for majority rule to produce a transitive social preference order. It only identifies a sufficient condition, namely single-peakedness. This means that there are sets of individual preferences that cannot be mapped single-peakedly onto a single graph, and yet which by majority rule produce transitive social preference orders. I leave it to you to construct an example of this.