CONTENTS

Preface ix
Introduction 1
The Concerns of Welfare Economics 1
The Concerns of Social Choice Theory 4
Practical Concerns of Welfare Economics and Social Choice Theory 7

1 Preferences and Utility 9
Fundamental Assumptions 9
Best Alternatives and Utility Functions 11
The Formal Model of Preferences 12
Introduction to Social Preferences 16
Exercises 20
Appendix 21
Selected References 22

2 Barter Exchange 23
Introduction 23
Allocations 24
10 ARROW'S IMPOSSIBILITY THEOREM

INTRODUCTION

We have now looked at several ways to answer the question "When is one alternative socially preferred to another?" Each of the answers has been somewhat disappointing. The Pareto criterion is incomplete; the Kaldor criterion is possibly inconsistent. The Scitovsky criterion reduces to a question of optimality vs. non-optimality. The Samuelson criterion might be completely devoid of content. The fairness and Rawls criteria might be inconsistent with the Pareto criterion. Majority voting might generate cycles, and the single-peakedness condition, which forces transitivity, is quite restrictive.

Our goal throughout has been to discover an unerring rule for generating rational social preferences, rational in the sense that an individual's preferences are rational. That is, we have been looking for a rule that generates complete and transitive social preferences, or, at least, complete and acyclic social preferences. But our series of disappointments raises some questions: Does a foolproof method exist for constructing complete and transitive social preference relations? Does a foolproof method exist for constructing complete and acyclic social preference relations? Does a foolproof method exist for finding best social alternatives? In this chapter we construct a simple model to answer the first question, and we briefly discuss the answer to the second question.

Does a foolproof method exist for constructing complete and transitive social preference relations? The answer to this question clearly depends on what we mean by foolproof. We will impose formal requirements on the method for constructing social preferences, requirements that allow a definite answer. The list of requirements, and the answer to the question, were developed by Kenneth Arrow around 1950, and the answer, to which we shall soon turn, is called Arrow's Impossibility Theorem.

THE MODEL

We now assume for the sake of simplicity that there are only two individuals, and three social alternatives $x$, $y$, and $z$. Also, we suppose that no individual is ever indifferent between any two alternatives. As usual, $xP_iy$ means $i$ prefers $x$ to $y$. Individual $i$'s preference ordering is assumed to be complete and transitive. Since there are only three alternatives and indifference is disallowed, there are only six ways individual 1 can order the alternatives. He can prefer $x$ to $y$ to $z$, or he can prefer $x$ to $z$ to $y$, or he can prefer $y$ to $x$ to $z$, and so on. The same is true of individual 2. Therefore, if any preference ordering for 1 or 2 is allowed, there are exactly $6 \times 6 = 36$ different constellations of individual preferences, or preference profiles, possible in this small society. Table 10–1 includes them all.

Each cell in this table shows a possible pair of rankings of the three alternatives by individuals 1 and 2. On the left side the alternatives are ordered, from top to bottom, according to person 1's preferences, and on the right side they are ordered according to 2’s preferences. For example, the 1st row, 2nd column cell has 1 preferring $x$ to $y$ to $z$, and 2 preferring $x$ to $z$ to $y$.

Our concern here is whether or not there is a foolproof rule to transform any cell in Table 10–1 into a social preference relation. Such a rule is called a collective choice rule. A collective choice rule takes preference profiles and produces social preferences.

Let $R$ stand for a social preference relation, so $xRy$ means $x$ is socially at least as good as $y$. $P$ is the corresponding strict social preference relation: $xPy$ means $x$ is socially preferred to $y$; i.e., $xRy$ and not $yRx$. (Don't confuse this $R$ with the Rawls criterion, or this $P$ with the Pareto criterion.) Finally, $I$ is the social indifference relation: $xIy$ means $x$ and $y$ are socially indifferent; i.e., $xRy$ and $yRx$. 
In the next section we will list five plausible requirements that will be imposed on the collective choice rule. Taken together, these five requirements define what we mean by foolproof.

**REQUIREMENTS ON THE COLLECTIVE CHOICE RULE**

1. **Completeness and transitivity.** The social preference relations generated by a collective choice rule must be complete and transitive. If some preference profile is transformed into a particular $R$, then for any pair of alternatives $x$ and $y$, either $xRy$ or $yRx$ must hold, and for any triple $x, y$, and $z, xRy$ and $yRz$ must imply $xRz$. The requirement says that a collective choice rule must always permit comparisons between two alternatives, and that social preferences must have the nice transitivity property assumed for an individual's preferences.

   We have examined collective choice rules that don't generate complete and transitive social preference relations. The Pareto criterion gives incomplete social rankings: if 1 prefers $x$ to $y$ and 2 prefers $y$ to $x$, the two alternatives are Pareto noncomparable. Majority voting gives nontransitive social rankings. As an example in a two-person case (where the voting cycle of Condorcet cannot be generated), suppose 1 prefers $x$ to $z$ to $y$, and 2 prefers $y$ to $x$ to $z$. If a vote is taken between $x$ and $y$, there is a tie (1 vote for $x$, and 2 votes for $y$). If a vote is taken between $y$ and $z$, there is again a tie. According to majority voting, $x$ and $y$ are socially indifferent, and $y$ and $z$ are socially indifferent. Transitivity would require that $x$ and $z$ be socially indifferent. But in a vote between $x$ and $z$, $x$ gets 2 votes and $z$ none; so $x$ beats $z$, and transitivity fails.

2. **Universality.** A collective choice rule should work no matter what individual preferences happen to be. This means that the rule should give us a social preference ordering for every cell in Table 10–1, not just for the easy ones, like the ones where there is unanimous agreement.

   Universality is a significant requirement. It precludes the assumption of single-peakedness, since it says that the collective choice rule must work for all preference profiles, not just the ones where utility functions have single peaks. Why should it be imposed?

   First, it is difficult to see where to draw the line between permissible and impermissible individual preferences. Which cells in Table 10–1 should be disallowed or ignored? How much diversity can be expected in society? When is there so much conflict that the very idea of social welfare becomes implausible? There are no easy answers to these questions. Second, the theorem we will prove remains valid even when the universality requirement is substantially weakened, and we will indicate how much it can be weakened in a subsequent section.

3. **Pareto consistency.** A collective choice rule should be consistent with the Pareto criterion. For any pair of alternatives $x$ and $y$, if both individuals prefer $x$ to $y$, $x$ must be socially preferred to $y$.

   Pareto consistency is a very mild requirement for a collective choice rule. One would not expect it to hold in societies which are ruled by external forces; in which, for example, everyone prefers lust and gambling, on the one hand, to chastity and frugality on the other; but where, according to a
Holy Book, the social state of chastity and frugality is preferable to the social state of lust and gambling. Economists naturally would recommend lust and gambling.

On a more serious note, let’s recall that the fairness and Rawls criteria could produce results contrary to the Pareto criterion. In our view, this fact is an indictment of fairness and Rawls, not of Pareto. We take Pareto consistency to be fundamental.

4. Non-dictatorship. A collective choice rule must make no one a dictator. Individual i is said to be a dictator if his wishes prevail, no matter how j feels; that is, if xPy,y implies xPy for all x and y, irrespective of Pj. Ruling out dictatorship does not mean that it is never possible to have xPy,y implying xPy for all x and y. Obviously, if both people agree on the rankings of all alternatives (so that P1 = P2, as in the diagonal cells of Table 10–1), then it is perfectly reasonable to have the social preference relation agreeing with 1’s (and 2’s) preference relation, and in fact, the Pareto consistency requirement makes such agreement necessary. Nondictatorship simply says that 1 (or 2) must not always prevail, no matter how 2 (or 1) happens to feel.

5. Independence of irrelevant alternatives. If people’s feelings change about some set of irrelevant alternatives, but do not change about the pair of alternatives x and y, then a collective choice rule must preserve the social ordering of x and y. The social preference between x and y must be independent of individual orderings on other pairs of alternatives. (We should note that this formulation of independence differs slightly from Arrow’s original formulation.)

Independence is the subtlest of the five requirements, and it takes some explanation. Suppose society prefers x to y when z is a third alternative lurking in the wings. Next suppose everyone suddenly changes his mind about the desirability of z, but no one changes his mind about x vs. y. The independence requirement says that, if society is deciding on the relative merits of x and y, and only those two, it must still prefer x to y.

The standard example of an otherwise nice collective choice rule that violates independence is weighted voting. This type of rule was first analyzed in 1781 by Jean-Charles de Borda, in his Mémoire sur les Élections au Scrutin, and it is consequently also called de Borda voting. It works as follows. Each person reports his preference relation, his rank ordering. A first place in a rank ordering is assigned a certain fixed weight, a second place is assigned a (smaller) fixed weight, a third place is assigned a (yet smaller) fixed weight, and so on. (In the two-person, three-alternative model of this chapter we have no ties, no cases of indifference, to worry about.) The weights that each alternative gets from each person are summed, and the social preference relation is derived from the sums of the weights.

For instance, suppose person 1 prefers x to y, while person 2 prefers y to x to z. Suppose a person’s first choice gets a weight of five points, a second choice gets four points, and a third choice gets one point. (The weights are obviously arbitrary. It is common to use equally spaced weights, like 3, 2, and 1, but there is no logically compelling reason to do so. You may construct an example similar to this one using the common weighting scheme.) Now alternative x gets 4 + 4 = 8 points, alternative y gets 1 + 5 = 6 points, and alternative z gets 5 + 1 = 6 points. Therefore, for this preference profile, x is socially preferred to y according to the weighted voting rule.

However, suppose person 1 becomes disillusioned with alternative z, and his preference ordering changes to x over y over z. If the voting is repeated, x gets 5 + 4 = 9 points, y gets 4 + 5 = 9 points, and z gets 1 + 1 = 2 points. Therefore, given this new preference profile, x is socially indifferent to y. Society has become indifferent between x and y, even though neither person has changed his feelings about x and y! Consequently, weighted voting violates the independence requirement.

APPLYING THE REQUIREMENTS

At this stage we shall apply requirements 1, 2, 3 and 5 to Table 10–1. The applications should clarify the meanings of the four requirements. They will also lay the groundwork for the proof of Arrow’s Impossibility Theorem.

The completeness and transitivity and the universality requirements say that, when applied to any cell of Table 10–1, a collective choice rule must generate a complete and transitive social ordering.

The Pareto consistency requirement says a collective choice rule must respect unanimous opinion: If both 1 and 2 prefer one alternative to another, then society must also prefer the one to the other. For example, given the preference profile of the first row, second column cell of Table 10–1, the Pareto requirement says x must be socially preferred to y and x must be socially preferred to z. We must have xPy and xPz. Application of Pareto consistency over the entirety of Table 10–1 gives rise to Table 10–2.

Each cell of this table is produced by applying Pareto consistency to the corresponding cell of Table 10–1, and therefore, any rule for generating social preferences must be entirely consistent with Table 10–2.
TABLE 10-2

<table>
<thead>
<tr>
<th>xPy</th>
<th>xPy</th>
<th>xPy</th>
<th>xPy</th>
<th>xPy</th>
</tr>
</thead>
<tbody>
<tr>
<td>xPz</td>
<td>xPz</td>
<td>yPz</td>
<td>xPy</td>
<td></td>
</tr>
<tr>
<td>yPz</td>
<td>xPy</td>
<td>yPz</td>
<td>yPz</td>
<td></td>
</tr>
<tr>
<td>xPy</td>
<td>xPy</td>
<td>xPz</td>
<td>xPy</td>
<td>zPy</td>
</tr>
<tr>
<td>xPz</td>
<td>zPz</td>
<td>xPz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yPz</td>
<td>yPz</td>
<td>yPz</td>
<td>yPz</td>
<td>yPz</td>
</tr>
<tr>
<td>yPz</td>
<td>yPz</td>
<td>yPz</td>
<td>zPz</td>
<td>yPz</td>
</tr>
<tr>
<td>xPy</td>
<td>xPy</td>
<td>zPz</td>
<td>xPz</td>
<td>zPz</td>
</tr>
<tr>
<td>xPz</td>
<td>xPz</td>
<td>zPz</td>
<td>zPz</td>
<td></td>
</tr>
<tr>
<td>yPz</td>
<td>yPz</td>
<td>yPz</td>
<td>zPz</td>
<td>zPz</td>
</tr>
<tr>
<td>xPy</td>
<td>xPy</td>
<td>zPz</td>
<td>zPz</td>
<td>zPz</td>
</tr>
<tr>
<td>zPz</td>
<td>yPz</td>
<td>yPz</td>
<td>zPz</td>
<td>zPz</td>
</tr>
</tbody>
</table>

Let's be specific about those areas of agreement. Independence requires that all the cells in Table 10-1 where xP1y and yP2x must yield identical social rankings of x and y. Similarly, all the cells where yP1x and xP2y must yield identical social rankings of x and y. There is no presumption, however, that the social ranking of x and y on the xP1y and yP2x cells need be the same as the social ranking of x and y on the yP1x and xP2y cells. Such a neutrality condition is unnecessary for the proof of the Impossibility Theorem, although it is intuitively appealing and useful in other contexts.

Independence also implies these areas of agreement. All the cells of Table 10-1 where zP1z and zP2z must give rise to identical social rankings of x and z; all the cells where zP1x and xP2z must give rise to identical social rankings of x and z; all the cells where yP1z and zP2y must give rise to identical social rankings of y and z; and, finally, all the cells where zP1y and yP2z must give rise to identical social rankings of y and z.

All of this information can be incorporated in a third table. Table 10-3a indicates where the social rankings of x and y must agree because xP1y and yP2x in all the X’d cells, and where the social rankings of x and y must agree because yP1x and xP2y in all the O’d cells. Tables 10-3b and 10-3c show the areas of agreement which arise from applications of the independence requirement to the social preferences between x and z, and y and z, respectively.

Now we turn to the independent requirement. Suppose that when person 1 prefers x to y to z and person 2 prefers y to x to z (the first row, third column cell of Table 10-1) a collective choice rule declares x is socially preferred to y, or xPy. Then independence requires that xPy hold whenever xP1y and yP2x, not matter how 1 and 2 rank alternative z. Similarly, if yPx (or xPy) holds when person 1 prefers x to y to z and person 2 prefers y to x to z, then yPx (or xPy) must hold whenever xP1y and yP2x. In short, the independence requirement forces a collective choice rule to give rise to social preferences that agree over certain preference profiles.

TABLE 10-3a

The crossed cells all produce the same x-y social rankings. The circled cells all produce the same x-y social rankings (which need not be the same as in the crossed cells).
**TABLE 10-3b & c**

The crossed cells all produce the same x-z social rankings. The circled cells all produce the same x-z social rankings (which need not be the same as in the crossed cells).

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

0 0 0

0 0 0

0 0 0

(b)

The crossed cells all produce the same y-z social rankings. The circled cells all produce the same y-z social ranking (which need not be the same as in the crossed cells).

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

0 0 0

0 0 0

(c)

With these preliminaries out of the way, we can turn to a truly remarkable theorem.

**ARROW’S IMPOSSIBILITY THEOREM**

At least since the time of Condorcet and de Borda in the eighteenth century, people have been concerned with the properties of rules for making social choices, election rules in practice, collective choice rules in theory. Does there exist a foolproof rule for discovering, or for defining, social prefer-

ences? Arrow showed that, if foolproof means consistent with the five requirements above, the answer is No.

We now turn to a formal statement and proof of the theorem.

**Arrow’s Impossibility Theorem.** Any collective choice which is consistent with the requirements of (1) completeness and transitivity, (2) universality, (3) Pareto consistency, and (4) independence of irrelevant alternatives, makes one person a dictator. Therefore, there is no rule which satisfies all five requirements.

**Proof.** We start by looking at the preference profile of the first row, second column cell of Table 10-1. For these preferences Pareto consistency requires $xPy$ and $xPz$ (Table 10-2). There are three and only three complete and transitive social preference orderings which satisfy $xPy$ and $xPz$. They are:

1. $xPy$, $xPz$ and $yPz$
2. $xPy$, $xPz$ and $zPy$
3. $xPy$, $xPz$ and $yLz$

Each of these three possibilities will be considered in turn.

**Case 1: $yPz$.** First a word about strategy. The Pareto consistency requirement tells a lot about what social preferences must be, but it leaves a lot unsaid. Table 10-2 is full of blank and partially blank spaces. We will now show how all the blanks can be filled in by repeatedly applying the independence and transitivity requirements.

If $yPz$ holds in the first row, second column cell, then independence (Table 10-3c) requires that $y$ be socially preferred to $z$ whenever individual preferences about $y$ and $z$ are the same as they are in that cell. Therefore, $yPz$ holds in all the cells indicated in Table 10-4a. (The cells that provide crucial steps in the proof are numbered 1 and 2.)

Now consider the first row, fifth column cell, or cell number 2 in Table 10-4a. Pareto consistency (Table 10-2) requires that $xPy$ here, but $xPy$ and $yPz$ implies $xPz$, by transitivity. So in this cell we must also have $xPz$.

But if $xPz$ holds in cell number 2, then independence (Table 10-3b) requires that $x$ be socially preferred to $z$ whenever individual preferences about $x$ and $z$ are the same as they are in that cell. Therefore, $xPz$ holds in all the cells indicated in Table 10-4b. (The cells that provide crucial steps in the proof are numbered 2 and 3.)

Now we have $xPz$ in cell 3. We again invoke Pareto consistency and transitivity to conclude that $xPy$ must hold in cell 3 as well. But this