21 Social choice

Chapter summary

In this chapter we present a model of social choice, which studies how a group of individuals makes a collective choice from among a set of alternatives. The model assumes that each individual in the group holds a preference relation over a given set of alternatives, and the problem is how to aggregate these preferences to one preference relation that is supposed to represent *the preference of the group*. A function that maps each vector of preference relations to a single preference relation is called a *social welfare function*. The main result on this topic is Arrow's Impossibility Theorem, which states that every social welfare function that satisfies two properties, unanimity and independence of irrelevant alternatives, is dictatorial.

This impossibility result is then extended to *social choice functions*. A social choice function assigns to every vector of preference relations of all individuals in the group a single alternative, interpreted as the alternative that is most preferred by the group.

A social choice function is said to be *nonmanipulable* if no individual can manipulate the group's choice and obtain a better outcome by reporting a preference relation that is different from his true preference relation. Using the impossibility result for social choice functions we prove the Gibbard–Satterthwaite Theorem, which states that any nonmanipulable social choice function that satisfies the property of unanimity is dictatorial.

Groups of decision makers are often called upon to choose between several possible alternatives: citizens are called upon to choose between political candidates on election day; union members select committees; family members need to choose which television program they will watch together. In any such example, each individual has his or her own preference ordering with respect to the different alternatives, and all the individuals together need to come to a collective decision that serves as the entire group's "most preferred" alternative. How are such decisions arrived at? In a dictatorship, the choice is made by the "dictator" (the national leader or the head of the family). In democratic countries the choice is usually made by some variation of majority vote. The following example, Example 21.1, shows one way to generalize majority vote to cases in which there are more than two alternatives. Example 21.2 reveals that problems may arise in such generalizations. The following example, due to Borda [1784],[1] points to difficulties in using majority vote for collective decisions.

1 Jean-Charles Chevalier de Borda (1733–99) was a French mathematician, physicist, social scientist and sailor.

Example 21.1 A committee composed of 21 people needs to select one individual from among three candidates named A, B, and C. The committee members' preferences are given in the following table:

No. of committee members	First choice	Second choice	Third choice
1	A	B	C
7	A	C	B
7	B	C	A
6	C	B	A

In words, one committee member ranks the candidates in the order A first, B second, and C third; seven committee members share the preference ordering A first, C second, and B third; another seven share the ordering B first, C second, and A third; and six members rank the candidates by the order C first, B second, and A third. Which candidate will be chosen? The answer, obviously, depends on which voting system is used.

Since we are used to the idea of selecting candidates by majority vote, it seems natural to check whether there is a candidate who would defeat every other candidate by majority vote in a head-to-head competition. Such a candidate is called a "Condorcet winner,"and if one exists, it seems reasonable to choose him as the collective choice of the committee. In our example, candidate C is a Condorcet winner, since he would defeat A by a vote of 13 to 8 if the two of them were the sole candidates, and similarly would win by 13 votes to 8 votes against B. ◀

One problem with this "voting method" is that there might not always be a Condorcet winner, as can be seen in Example 21.2, which is due to Condorcet [1785].[2] In this case, the above voting method will not tell us who to choose.

Example 21.2 A 60-member committee needs to select one individual from candidates A, B, and C. The committee members' preferences are given in the following table:

No. of committee members	First choice	Second choice	Third choice
23	A	B	C
2	B	A	C
17	B	C	A
10	C	A	B
8	C	B	A

In head-to-head pairwise competition, A defeats B by $33 - 27$, B trounces C by $42 - 18$, and C wins against A by $35 - 25$. In other words, there is no Condorcet winner: the preference ordering given by pairwise majority voting is not transitive. This leads to several implications. First of all, the order in which voting between candidates is conducted can affect the result – that is, if we first pit two candidates against each other and then have the winner between them compete against the third candidate, the order in which this is done may be crucial. If A and B compete head-to-head

2 Marie Jean Antoine Nicolas Caritat (Marquis de Condorcet, 1743–94) was a French philosopher and mathematician who wrote about political science.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

with the winner going up against C, then C will be selected. But if we first have A and C compete against each other with the winner squaring off against B, then B ends up being selected. And if B competes against C with the winner between them pitted against A, then A is the ultimate selection. The order of voting is absolutely critical.

Another consequence of the fact that the preference relation is not transitive is that in any voting method that generalizes majority vote between two candidates to a greater number of candidates, the results may depend on the presence or absence of a candidate who is not even the winner! For example, suppose that a certain voting method, which for two candidates chooses the winner by majority vote, leads to the selection of A. If B were to decline to participate, A and C would instead compete directly against each other – and then C would win by majority vote. In other words, B's presence as a candidate can affect the results, even though B does not win when he competes. A similar phenomenon would exist if the voting method were to select B or C.

The condition that the presence or absence of a candidate that is not selected by the procedure should not affect the results of a voting method is called the "independence of irrelevant alternatives." The above example shows that it is not true that every voting system that in the presence of two candidates chooses one of them using majority vote satisfies independence of irrelevant alternatives. This fact forms the basis of the results developed in this chapter. ◀

Another problem with pairwise majority voting is that even when a Condorcet winner exists, it is not always clear that he is the candidate who should be selected (see Exercise 21.5). We can check whether or not several well-known voting methods select the Condorcet winner when such a candidate exists. One popular method is to have each committee member vote for his or her most-preferred candidate, with the candidate receiving the most votes winning.[3] If this method were to be adopted in Example 21.1, then A would receive 8 votes, B would get 7 votes and C only 6 votes, leading to the selection of candidate A and not the Condorcet winner C.

Another method chooses the winning candidate in a two-round process:[4] in round one, every committee member votes for his or her most-preferred candidate. The two candidates who received the greatest number of votes in round one go on to compete against each other in round two, with the candidate garnering the most votes in round two ultimately selected as the winner. In Example 21.1, this method would lead to candidates A and B proceeding to a head-to-head competition in round two, where B would defeat A by 13 votes to 8; once again the Condorcet winner, candidate C, fails to be selected.

Election results, therefore, are extremely sensitive to which voting method is adopted, and as we have seen, two very popular voting methods by-pass the Condorcet winner, when such a candidate exists, and may well end up selecting another candidate.

In addition, the two methods discussed above can be subject to manipulation, in the sense that committee members have incentives to misrepresent their preferences in order to change the results. To see that, note that in Example 21.1, under the voting system in which each committee member votes for only one candidate, and the candidate with the greatest number of votes is chosen, if the committee members who prefer C to B and B to A vote for B instead of C, then B, whom they prefer to A, will win instead of A. In

3 This is a method used by many committees, including committees selecting candidates for public service positions in Britain.

4 This is the method used to elect the President of France.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

the same example with a two-round voting method, if the committee members who rank A over C and C over B vote for C instead of A in the first round, them C and B will be the candidates competing in the second round, with C, whom they prefer to B, ultimately winning. The subject of voting manipulation will be discussed again later in this chapter.

In all the examples so far, the main question was choosing one alternative (one candidate) from a set of alternatives (candidates). We will return to this question later in the chapter, but we will first consider a more general issue. Suppose that every individual in a given population has a preference ordering (or ranking) over a set of alternatives, and society in general seeks to derive, out of all the individual rankings, a single ranking representing society's collective preferences among the alternatives: society's first choice among the alternatives, society's second choice, and so forth. In other words, the question before us is how to "aggregate" all the individual preference rankings into one preference ranking that can be interpreted as that of society's.

For example, suppose several teachers are asked to rank the students in a class by academic achievement. Each teacher ranks the students based on their performance in the subject that he or she teaches, and it is quite reasonable that different teachers will produce different rankings. The teachers may want to find a way to aggregate their rankings into one collective ranking listing the students in order from the "best" student to the "weakest" student.

We will show that, surprisingly, there is no rule producing an aggregate preference ordering that satisfies three very natural-sounding democratic conditions. One of those conditions is that there should be no dictator. If we dispense with this condition, then the only rule that satisfies the other two conditions is dictatorship.

After that we will explore situations in which there is no need to rank all of the possible choices, because it is only necessary to select society's top choice. There are many examples of such situations: selecting a committee chairman, electing a president, the board of directors of a corporation choosing among different investment opportunities, army officers selecting a military course of action, and, in our above example, picking the best student in the class in the collective opinion of the teachers. We will see that in this case as well, there is no selection rule that satisfies three natural democratic conditions, and that the only rule that satisfies two of the conditions is dictatorship.

21.1 Social welfare functions

Let A be a nonempty finite set of *alternatives*, and $N = \{1, 2, \ldots, n\}$ be a finite set of *individuals* ("voters" or "decision makers"). In Chapter 2, which deals with utility theory, we defined a preference relation over the set A as a subset of $A \times A$, and assumed that the preference relations of the players over the set of outcomes are complete, reflexive, and transitive. In this chapter we will consider preference relations that are complete and transitive but not necessarily reflexive. For simplicity, we will include these properties as part of the definition: a preference relation will from here on be a complete, reflexive, and transitive binary relation, and a strict preference relation will be a complete, irreflexive, and transitive binary relation.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

Definition 21.3 *A* preference relation $\succsim_{P_i}$ *of player i over a set A is a binary relation*[5] *satisfying the following properties:*

- *For every pair of elements $a, b \in A$, either $a \succsim_{P_i} b$ or $b \succsim_{P_i} a$ (the relation is* complete*).*
- *$a \succsim_{P_i} a$ (the relation is* reflexive*).*
- *If $a \succsim_{P_i} b$ and $b \succsim_{P_i} c$ then $a \succsim_{P_i} c$ (the relation is* transitive*).*

A strict preference relation $\succ_{P_i}$ *of player i over A is a binary relation satisfying the following properties:*

- *For every pair of distinct elements $a, b \in A$, either $a \succ_{P_i} b$ or $b \succ_{P_i} a$.*
- *$a \not\succ_{P_i} a$ (the relation is* irreflexive*).*
- *If $a \succ_{P_i} b$ and $b \succ_{P_i} c$ then $a \succ_{P_i} c$ (the relation is* transitive*).*

When P^* is a preference relation, if $a \succsim_{P^*} b$ and $b \succsim_{P^*} a$ then we will say that a and b are *equivalent* under the preference relation P^*, and denote this by $a \approx_{P^*} b$. As the following example shows, it is possible for $a \approx_{P^*} b$ even though $a \neq b$.

If the set of alternatives $A = \{-m, -m+1, \cdots, m\}$ is a finite set of natural numbers, the relation $\geq$ is a preference relation, and the relation $>$ is a strict preference relation. The relation $\succsim_{P_i}$ defined by $a \succsim_{P_i} b$ if and only if $|a| \geq |b|$ is also a preference relation. Note that $k \approx_{P_i} -k$. The lexicographic relation $\succsim_L$ defined as follows is also a preference relation: the set A is the following set of pairs of positive integers, $A = \{(n, m) \colon 1 \leq n \leq N, 1 \leq m \leq M, n \in \mathbb{N}, m \in \mathbb{N}\}$, and the relation $\succsim_L$ is defined by $\hat{\imath}(n, m) \succsim_L (\widehat{n}, \widehat{m})$ if and only if $n > \widehat{n}$, or $n = \widehat{n}$ and $m \geq \widehat{m}$.

Denote by $\mathcal{P}^*(A)$ the set of all preference relations over A and by $\mathcal{P}(A)$ the set of all strict preference relations over A.

Definition 21.4 *A* strict preference profile *is a list $P^N = (P_i)_{i \in N}$ of strict preferences, one per individual. The collection of all strict preference profiles is the Cartesian product*

$$(\mathcal{P}(A))^N = \mathcal{P}(A) \times \mathcal{P}(A) \times \cdots \times \mathcal{P}(A). \tag{21.1}$$

A strict preference profile describes how each individual in society ranks all the alternatives. The problem before us is how to "aggregate" all the preferences in a strict preference profile into one preference relation, "the social preference relation."

Definition 21.5 *A* social welfare function *is a function F that maps each strict preference profile $P^N = (P_i)_{i \in N} \in (\mathcal{P}(A))^N$ to a preference relation in $\mathcal{P}^*(A)$ (which is denoted by $F(P^N)$).*

In other words, a social welfare function summarizes the opinions of everyone in society: given the strict preference relations $P^N = (P_i)_{i \in N}$ of all the individuals, society as a collective ranks the alternatives in A by way of the preference relation $F(P^N)$. If society ranks a above b, that is, if $a \succsim_{F(P^N)} b$, we will say that society (weakly) prefers a to b.

5 In the chapter on utility theory (Chapter 2) we studied a preference relation of an individual i and denoted it by $\succsim_i$. In this chapter we may want an individual i to have different preference relations P_i, $\widehat{P}_i$, and so on. We will therefore label a relation $\succsim$ not by the name of the individual but by his preference relation, i.e., $\succsim_{P_i}$, $\succsim_{\widehat{P}_i}$, and so on.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

Note that we are assuming that every individual has a strict preference relation, meaning that no one is indifferent between any pair of alternatives. But the social preference relation, on the other hand, may exhibit indifference. The following example clarifies why we choose this definition. Recall that for every finite set X we denote the number of elements in X by $|X|$.

Example 21.6 **Simple majority rule, $|A| = 2$** Suppose there are only two alternatives $A = \{a, b\}$. For each strict preference profile P^N we will denote the number of individuals who prefer a to b by:

$$m(P^N) = \left|\{i \in N : a \succ_{P_i} b\}\right|. \tag{21.2}$$

The simple majority rule is the social welfare function F defined by:

- If $m(P^N) > \frac{n}{2}$ then society as a whole prefers a to b: $a \succ_{F(P^N)} b$.
- If $m(P^N) < \frac{n}{2}$ then society as a whole prefers b to a: $b \succ_{F(P^N)} a$.
- If $m(P^N) = \frac{n}{2}$ then society as a whole is indifferent between a and b: $a \approx_{F(P^N)} b$.

◀

If we do not permit society to be indifferent between alternatives, then, for this to be a social welfare function, it would need to rank a versus b even when $m(P^N) = \frac{n}{2}$. To avoid a situation in which arbitrary rankings are assigned, we accept indifference in the social preference, even when there is no indifference at the individual level.

Despite this, the theorems presented in this chapter obtain even when the preferences of the individuals are not necessarily strict preferences, but weakening the assumption of strict preference may require using different proofs.

A dictatorship is a simple social welfare function: if the dictator prefers a to b, society must prefer a to b.

Definition 21.7 *A social welfare function F is* dictatorial *if there is an individual $i \in N$ such that $F(P^N) = P_i$ for every profile of strict preferences P^N. In other words, for every pair of alternatives $a, b \in A$, and every strict preference profile P^N*

$$a \succ_{P_i} b \quad \Longrightarrow \quad a \succ_{F(P^N)} b. \tag{21.3}$$

In this case, individual i is called a dictator.

The simple majority rule (see Example 21.6) is not a dictatorial social welfare function, because every individual in society may find himself part of the minority, in which case social preference between a and b will be contrary to his preference.

The approach we adopt for studying social welfare functions is the "normative" (or "axiomatic") approach. This means we ask which "reasonable" properties do we want the social welfare function to satisfy, and which mathematical conclusions can we draw regarding functions satisfying those properties.

A reasonable property one may want a social welfare function to satisfy is that if all individuals in society prefer alternative a to alternative b, society also prefers a to b.

Definition 21.8 *A social welfare function F satisfies the property of* unanimity *if F satisfies the following condition: for every two alternatives $a, b \in A$, and every strict preference profile $P^N = (P_i)_{i \in N}$, if $a \succ_{P_i} b$ for every individual $i \in N$, then $a \succ_{F(P^N)} b$.*

A second property one might wish a social welfare function to satisfy is that the way society determines whether alternative a is preferable to alternative b depends solely on the way the individuals compare a to b.

Definition 21.9 *A social welfare function F satisfies the* independence of irrelevant alternatives *(IIA) Property if for every pair of alternatives $a, b \in A$, and every pair of strict preference profiles P^N and Q^N*

$$a \succ_{P_i} b \iff a \succ_{Q_i} b, \quad \forall i \in N, \tag{21.4}$$

implies that

$$a \succsim_{F(P^N)} b \iff a \succsim_{F(Q^N)} b. \tag{21.5}$$

In other words, if every individual answers identically in both P^N and in Q^N to the question "which do you prefer between a and b?" the social preference between a and b should be identical according to both $F(P^N)$ and $F(Q^N)$.

In the example presented above of the teachers ranking the pupils in a class, if the weighted rankings of all the teachers indicate that Ann is ranked higher than Dan, and then Tanya's grades are changed (because she retook an exam), this should have no effect on (i.e., be irrelevant to) the relative ranking of Ann and Dan: Ann should still be ranked higher than Dan.

A dictatorial social welfare function satisfies the properties of unanimity and independence of irrelevant alternatives (prove this). Similarly, when $|A| = 2$, the simple majority rule (Example 21.6) satisfies the properties of unanimity and independence of irrelevant alternatives (prove this). Can the simple majority rule be extended to any number of alternatives to yield a social welfare function that satisfies unanimity and independence of irrelevant alternatives? As the following surprising theorem shows, the answer to this question is negative.

Theorem 21.10 (Arrow [1951]) *If $|A| \geq 3$, then every social welfare function satisfying the properties of unanimity and independence of irrelevant alternatives is dictatorial.*

An equivalent formulation of the theorem is given by considering nondictatorship to be a desired property.

Definition 21.11 *A social welfare function F satisfies the property of* nondictatorship *if it is not dictatorial.*

Theorem 21.12 *If $|A| \geq 3$, there does not exist a social welfare function satisfying the properties of unanimity, independence of irrelevant alternatives, and nondictatorship.*

Theorem 21.12 is called Arrow's Impossibility Theorem. The significance of the theorem is that when we seek a social welfare function defined over the set of all preference profiles, if dictatorship is not something we desire, we must give up either unanimity or independence of irrelevant alternatives, or restrict the domain of preference profiles over which the function is defined.

Example 21.13 **Three individuals, three alternatives** Suppose there are three individuals $N = \{1, 2, 3\}$, and three alternatives $A = \{a, b, c\}$. The social welfare function F is defined as follows:

- To determine society's most-preferred alternative, first check if there is an alternative that is ranked highest by at least two individuals. If such an alternative exists, this is society's most-preferred alternative.
- If no such alternative exists, check if there is one alternative that is ranked highest or second-highest by all three individuals. If such an alternative exists, this is society's most-preferred alternative.
- If the above two checks fail to determine a most-preferred alternative for society, society is indifferent between the three alternatives.
- After the most-preferred alternative is chosen using the above rules, we use majority vote to determine the ranking of the other two alternatives.

The following table depicts several strict preference profiles, along with the social preference relation corresponding to each preference profile according to the social welfare function F:

	Individual 1	Individual 2	Individual 3	Society
1	$c \succ_{P_1} b \succ_{P_1} a$	$c \succ_{P_2} b \succ_{P_2} a$	$c \succ_{P_1} b \succ_{P_3} a$	$c \succ_{F(P^N)} b \succ_{F(P^N)} a$
2	$c \succ_{P_1} b \succ_{P_1} a$	$c \succ_{P_2} b \succ_{P_2} a$	$b \succ_{P_3} c \succ_{P_3} a$	$c \succ_{F(P^N)} b \succ_{F(P^N)} a$
3	$c \succ_{P_1} b \succ_{P_1} a$	$a \succ_{P_2} b \succ_{P_2} c$	$b \succ_{P_3} c \succ_{P_3} a$	$b \succ_{F(P^N)} c \succ_{F(P^N)} a$
4	$c \succ_{P_1} b \succ_{P_1} a$	$b \succ_{P_2} a \succ_{P_2} c$	$a \succ_{P_3} c \succ_{P_3} b$	$c \approx_{F(P^N)} b \approx_{F(P^N)} a$
5	$c \succ_{P_1} b \succ_{P_1} a$	$b \succ_{P_2} a \succ_{P_2} c$	$c \succ_{P_3} a \succ_{P_3} b$	$c \succ_{F(P^N)} b \succ_{F(P^N)} a$

This social welfare function is not dictatorial, because if two individuals share the same strict preference relation, that preference relation is also society's preference relation (check that this is true). It satisfies the unanimity property: if all the individuals prefer a to b, then either a is society's most-preferred alternative (if at least two individuals rank a highest), or c is ranked first by society and a is ranked second (if at most one individual ranks a highest according to his preference relation). In either case, society prefers a to b.

Theorem 21.10 then implies that this social welfare function cannot satisfy the independence of irrelevant alternatives property. Indeed, if we compare preference profiles 4 and 5 above, we see that in both of them $b \succ_{P_1} a$, $b \succ_{P_2} a$, $a \succ_{P_3} b$, but in the fourth profile $b \approx_{F(P^N)} a$, while in the fifth profile $b \succ_{F(P^N)} a$. ◀

In proving Theorem 21.10 we will make use of several definitions and denotations:

Definition 21.14 *A* coalition is a set of individuals $S \subseteq N$.

Definition 21.15 *Let F be a social welfare function, and let $a, b \in A$ be two different alternatives. A coalition $S \subseteq N$ is called* decisive *for a over b (relative to F) if for every $P^N \in (\mathcal{P}(A))^N$ satisfying:*

1. *$a \succ_{P_i} b$ for every $i \in S$,*
2. *$b \succ_{P_j} a$ for every $j \notin S$,*

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

one has $a \succ_{F(P^N)} b$. The coalition S is called decisive *(relative to F) if there exists a pair of alternatives for which it is decisive.*

In words, a set of individuals S is decisive for a over b if when every member of S prefers a to b, and all the other individuals prefer b to a, society prefers a to b. For example, it is possible for the President, the Secretary of the Treasury, and the Chairman of the Federal Reserve to be a decisive coalition for matters pertaining to economic policy; when issues of national defense require a decision, the President, the Secretary of Defense, and the National Security Adviser may be a decisive coalition.

Before we turn our attention to the characteristics of decisive coalitions, we will check whether there always exists at least one decisive coalition. The definition of the unanimity property leads to the following theorem (Exercise 21.11).

Theorem 21.16 *Let F be a social welfare function satisfying the unanimity property. For every $a, b \in A$, the coalition N is decisive for a over b and the empty coalition $\emptyset$ is not decisive for a over b.*

The next theorem shows that when a social welfare function satisfies the property of independence of irrelevant alternatives, it is easy to check whether a particular coalition is decisive for a given pair of alternatives.

Theorem 21.17 *Let F be a social welfare function satisfying the independence of irrelevant alternatives property, and let $a, b \in A$ be two alternatives. A coalition $S \subseteq N$ is decisive for a over b, if and only if there exists a strict preference profile P^N satisfying:*

(a1) $a \succ_{P_i} b$ for all $i \in S$,
(a2) $b \succ_{P_j} a$ for all $j \notin S$,
(a3) and $a \succ_{F(P^N)} b$.

It follows that if a coalition S is not decisive for a over b, and if a strict preference profile P^N satisfies:

(b1) $a \succ_{P_i} b$ for all $i \in S$,
(b2) $b \succ_{P_j} a$ for all $j \notin S$,

then $b \succsim_{F(P^N)} a$.

In other words, the theorem states that if the function F satisfies the independence of irrelevant alternatives property, then the condition "for all $P^N \in (\mathcal{P}(A))^N \cdots$" in Definition 21.15 can be replaced by the condition "there exists $P^N \in (\mathcal{P}(A))^N \cdots$."

Proof: Start with the first direction: suppose that S is decisive for a over b. Let P^N be a strict preference profile satisfying (a1) and (a2) (give an example of such a preference profile). Since the coalition S is decisive for a over b, (a3) is satisfied, and P^N therefore satisfies (a1)–(a3), as required.

For the other direction, we need to show that if there exists a strict preference profile P^N satisfying the three conditions, then S is decisive for a over b. In other words, we need to show that every strict preference profile Q^N that satisfies the first two conditions also

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

satisfies the third condition. But that follows from the fact that F satisfies the independence of irrelevant alternatives property.

The second part of the statement follows from the first part □

The next theorem states that if a coalition is decisive for a^* over b^*, then it is decisive for all pairs of alternatives.

Theorem 21.18 *Suppose that $|A| \geq 3$ and that F satisfies the unanimity and independence of irrelevant alternatives properties. If coalition V is decisive for a^* over b^*, then V is decisive for any pair of alternatives in A.*

Proof: Let a and b be a pair of alternatives.

Part 1: If V is decisive for a over b then V is decisive for a over c, for any alternative $c \in A \setminus \{a\}$.
If $c = b$ the claim follows by assumption. Otherwise $c \in A \setminus \{a, b\}$. Consider the following strict preference profile P^N:

$$\begin{cases} a \succ_{P_i} b \succ_{P_i} c & i \in V, \\ b \succ_{P_i} c \succ_{P_i} a & i \notin V. \end{cases} \tag{21.6}$$

All the other alternatives in A are ordered by each individual arbitrarily.

As V is decisive for a over b, it follows that $a \succ_{F(P^N)} b$. Since F satisfies the unanimity property, $b \succ_{F(P^N)} c$. Since $F(P^N)$ is a transitive ordering relation, we deduce that $a \succ_{F(P^N)} c$. Theorem 21.17 then implies that V is decisive for a over c.

Part 2: If V is decisive for a over b then V is decisive for b over c, for any $c \in A \setminus \{a, b\}$.
Let $c \in A \setminus \{a, b\}$. Consider the following strict preference profile P^N:

$$\begin{cases} b \succ_{P_i} a \succ_{P_i} c & i \in V, \\ c \succ_{P_i} b \succ_{P_i} a & i \notin V. \end{cases} \tag{21.7}$$

All the other alternatives in A are ordered by each individual arbitrarily.

From Part 1 it follows that V is decisive for a over c, and therefore $a \succ_{F(P^N)} c$. Since F satisfies the unanimity property, $b \succ_{F(P^N)} a$. Since $F(P^N)$ is a transitive ordering relation, we deduce that $b \succ_{F(P^N)} c$. Theorem 21.17 then implies that V is decisive for b over c.

Part 3: The first two parts are sufficient for proving Theorem 21.18.
Let $a \neq b$ be any pair of alternatives in A. We will prove that V is decisive for a over b. Recall that V is decisive for a^* over b^*.

- If $a = a^*$, from Part 1 and the fact that V is decisive for a^* over b^*, we deduce that V is decisive for a over b.
- If $a \neq a^*$ and $b \neq a^*$, from Part 1 and the fact that V is decisive for a^* over b^*, one has that V is decisive for a^* over a. That in turn implies from Part 2 that V is decisive for a over b.
- If $a \neq a^*$ and $b = a^*$, then there exists an alternative $c \in A \setminus \{a, b\}$ since A contains at least three alternatives (it is possible for $c = b^*$). From Part 1, and the fact that V is decisive for a^* over b^*, one has that V is decisive for b over c. From Part 2, this then implies that V is decisive for c over a. Finally, Part 2 implies that V is decisive for a over b.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

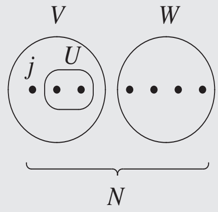

Figure 21.1

This concludes the proof of Theorem 21.18. □

We next prove Theorem 21.10.

Proof of Theorem 21.10 (Arrow's Theorem): Let F be a social welfare function satisfying the properties of unanimity and independence of irrelevant alternatives. We show that there exists a decisive coalition containing a single individual, who is a dictator.

Claim 21.19 *There exists a decisive coalition V containing a single individual.*

Proof: Let V be a nonempty decisive coalition containing a minimal number of individuals. Since by Theorem 21.16 the coalition N is decisive and the empty coalition $\emptyset$ is not decisive, there must exist such a coalition V. If V contains a single individual, there is nothing more to prove. Suppose that V contains at least two individuals; we will show that this leads to a contradiction.

Let $j \in V$. Denote $U = V \setminus \{j\}$ and $W = N \setminus V$ (see Figure 21.1). Since V contains at least two individuals, the coalition U is nonempty. Since by definition, V contains a minimal number of individuals among the nonempty decisive coalitions, the coalition U is nondecisive, and the coalition $\{j\}$ is nondecisive.

Since $|A| \geq 3$, we can choose three distinct alternatives a, b, c. Consider the strict preference profile $P^N = (P_i)_{i \in N}$ defined as follows:

$$\begin{cases} a \succ_{P_i} b \succ_{P_i} c & i = j, \\ c \succ_{P_i} a \succ_{P_i} b & i \in U, \\ b \succ_{P_i} c \succ_{P_i} a & i \in W. \end{cases} \tag{21.8}$$

The other alternatives in A are ordered by each individual arbitrarily.

Since V is decisive, it is decisive for any pair of alternatives (Theorem 21.18). In particular, it is decisive for a over b. Since $a \succ_{P_i} b$, for every individual $i \in V = U \cup \{j\}$ and $b \succ_{P_i} a$ for every individual $i \in N \setminus V = W$, one has $a \succ_{F(P^N)} b$. Since U is not decisive, it is not decisive for c over b. Since $c \succ_{P_i} b$ for every individual $i \in U$ and $b \succ_{P_i} c$ for every individual $i \in N \setminus U = W \cup \{j\}$, Theorem 21.17 implies that $b \succsim_{F(P^N)} c$. Since $F(P^N)$ is a transitive ordering relation, we deduce that $a \succ_{F(P^N)} c$.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

Note that $a \succ_{P_i} c$ for $i = j$, and $c \succ_{P_i} a$ for all $i \neq j$. By Theorem 21.17 we conclude that $\{j\}$ is a decisive coalition for a over c, and it is therefore a decisive coalition, contradicting the assumption that $\{j\}$ is not a decisive coalition. This contradiction establishes that $|V| = 1$. □

Let $V = \{j\}$ be a decisive coalition containing a single individual. We next prove that j is a dictator.

Claim 21.20 *Individual j is a dictator, i.e., $F(P^N) = P_j$ for every $P^N = (P_i)_{i \in N} \in (\mathcal{P}(A))^N$.*

Proof: Let P^N be a strict preference profile, and let $a, b \in A$ be two different alternatives such that $a \succ_{P_j} b$. We wish to show that $a \succ_{F(P^N)} b$.

Since A contains at least three alternatives, there exists an alternative $c \in A \setminus \{a, b\}$. Consider the following strict preference profile Q^N:

$$\begin{cases} a \succ_{Q_i} c \succ_{Q_i} b & i = j, \\ c \succ_{Q_i} a \succ_{Q_i} b & i \neq j, a \succ_{P_i} b, \\ c \succ_{Q_i} b \succ_{Q_i} a & i \neq j, b \succ_{P_i} a. \end{cases} \tag{21.9}$$

The other alternatives in A are ordered by each individual arbitrarily. Since $V = \{j\}$ is decisive for any two alternatives, it is in particular decisive for a over c, and therefore $a \succ_{F(Q^N)} c$. Since F satisfies the unanimity property, it follows that $c \succ_{F(Q^N)} b$. Since $F(Q^N)$ is a transitive ordering relation, we deduce that $a \succ_{F(Q^N)} b$.

Now, individual j prefers a over b, both according to P_j, and according to Q_j, and every individual $i \neq j$ prefers a over b according to P_i if and only if he prefers a over b according to Q_i. Since F satisfies the independence of irrelevant alternatives property, and since $a \succ_{F(Q^N)} b$, we deduce that $a \succ_{F(P^N)} b$, as required. □

We have proved that there exists a decisive coalition containing a single individual, and that this individual is a dictator. The proof of Theorem 21.10 is complete. □

One might imagine that the conclusion of the theorem holds because we asked for too much: we want a social welfare function to rank all the alternatives. The next section, however, shows that a similar negative result holds even if all we ask is for society to choose its most-preferred alternative.

21.2 Social choice functions

In many cases, society is not required to rank all possible alternatives, because it suffices to choose only one alternative. Examples of such situations include the election of a president, congressman, or committee chairman. An additional example is the selection by policymakers of the "best possible" political or economic policy, from a range of alternative policies. In this section, we study the question of associating every strict preference profile with one alternative that is most preferred by society, and striving to ensure that the process of choosing the most-preferred alternative satisfies desirable properties without being dictatorial.

Definition 21.21 *A* social choice function *is a function* $G : (\mathcal{P}(A))^n \to A$.

A social choice function associates every strict preference profile with one alternative, which is called society's "most-preferred alternative." By this definition, society cannot choose two different alternatives as most preferred, and it is not possible to choose the most-preferred alternative by tossing a coin.

It is reasonable to require that a social choice function be monotonic; if P^N and Q^N are two preference profiles in which, for each individual i, the ranking of a in Q^i is not lower that its ranking in P^i, and if $F(P^N) = a$, then $F(Q^N) = a$ as well.

Definition 21.22 *A social choice function G is called* monotonic *if for every pair of strict preference profiles P^N and Q^N satisfying*

$$a \succ_{P_i} c \quad \Longrightarrow a \succ_{Q_i} c, \quad \forall c \neq a, \forall i \in N, \tag{21.10}$$

if $G(P^N) = a$, then $G(Q^N) = a$.

Dictatorship is a simple social choice function:

Definition 21.23 *A social choice function G is* dictatorial *if there is an individual i such that for every strict preference profile P^N, $G(P^N)$ is the preferred alternative of individual i. Such an individual i is called a* dictator.

A dictatorial social choice function is monotonic (prove this).

Example 21.24 **Simple majority rule, $|A| = 2$** Denote $A = \{a, b\}$. When n is odd, the social choice function defined by majority rule is monotonic and nondictatorial (prove this).

When n is even, the social choice function defined by majority rule, where alternative a is chosen in case of a tied vote, is monotonic and nondictatorial. ◀

Example 21.25 Order the alternatives according to: $A = \{1, 2, \ldots, K\}$, where K is the number of alternatives. From among the alternatives that at least one individual most prefers, choose the one of minimal index:

$$F(P^N) = \min\{k \in A : \text{ there exists } i \in N \text{ such that } k \succ_{P_i} b \text{ for all } b \in A \setminus \{k\}\}. \tag{21.11}$$

This social choice function is neither dictatorial nor monotonic (Exercise 21.19). ◀

As for social welfare functions, it is reasonable to require that social choice functions satisfy the property that if every individual prefers a to every other alternative, then society as a whole should prefer alternative a.

Definition 21.26 *A social choice function G satisfies the property of* unanimity *if for every alternative $a \in A$, and every strict preference profile $P^N = (P_i)_{i \in N}$: if $a \succ_{P_i} b$ for every individual $i \in N$ and every alternative $b \neq a$, then $G(P^N) = a$.*

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

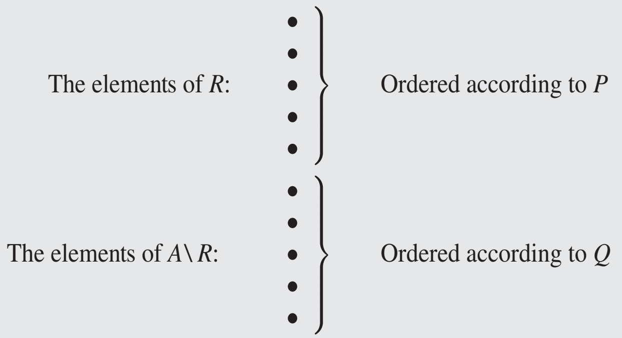

Figure 21.2 The preference relation $Z(P, Q; R)$

If G satisfies unanimity, the range of G is A: for every alternative $a \in A$, there exists a strict preference profile P^N satisfying $G(P^N) = a$ (give an example of such a profile).

Theorem 21.27 *If $|A| \geq 3$, all social choice functions satisfying the properties of unanimity and monotonicity are dictatorial.*

Before proving this theorem, we need several definitions and properties of monotonic social choice functions.

Let P and Q be two strict preference relations over a set of alternatives A, and let R be a subset of the set of alternatives A. Define a strict preference relation over A as follows: place the elements of R prior to the elements that are not in R; the strict preference relation over the elements of R is given by P, and the strict preference relation over the elements that are not in R is given by Q (see Figure 21.2). This strict preference relation is denoted $Z(P, Q; R)$.

The formal definition is as follows:

Definition 21.28 *Let P and Q be strict preference relations, and let $R \subseteq A$ be a subset of alternatives. Denote by $Z(P, Q; R)$ the following strict preference relation:*

- *$Z(P, Q; R)$ is identical with P over R: if $a, b \in R$ then*

$$a \succ_{Z(P,Q;R)} b \iff a \succ_P b. \tag{21.12}$$

- *$Z(P, Q; R)$ is identical with Q over $A \setminus R$: if $a, b \notin R$ then*

$$a \succ_{Z(P,Q;R)} b \iff a \succ_Q b. \tag{21.13}$$

- *The alternatives in R are preferred to the alternatives not in R: if $a \in R$ and $b \notin R$, then $a \succ_{Z(P,Q;R)} b$.*

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

Example 21.29 Suppose that the set of alternatives is $A = \{a_1, a_2, a_3, a_4\}$, and $R = \{a_1, a_4\}$. Furthermore, suppose that two strict preference relations P and Q are given by

$$\begin{aligned} a_1 \succ_P a_2 \succ_P a_3 \succ_P a_4, \\ a_2 \succ_Q a_4 \succ_Q a_1 \succ_Q a_3. \end{aligned} \tag{21.14}$$

The relation $Z(P, Q; R)$ is given by

$$a_1 \succ_{Z(P,Q;R)} a_4 \succ_{Z(P,Q;R)} a_2 \succ_{Z(P,Q;R)} a_3. \tag{21.15}$$

◀

The analogue to Definition 21.28 for strict preference relations is the following:

Definition 21.30 *Let P^N and Q^N be two strict preference profiles, and let $R \subseteq A$ be a subset of alternatives. Denote by $Z(P^N, Q^N; R)$ the strict preference profile in which the strict preference of individual i is $Z(P_i, Q_i; R)$, for each $i \in N$.*

Let $a \in R$ be an alternative in R, and suppose that $G(P^N) = a$. Compared to the strict preference profile P^N, in the strict preference relation $Z(P_i, Q_i; R)$ the ranking of the alternatives in R are improved relative to the alternatives not in R. If the social choice function is monotonic, one would expect that after this improvement, society chooses a. This property is expressed in the next theorem, whose proof is left as an exercise to the reader (Exercise 21.25).

Theorem 21.31 *Let G be a monotonic social choice function, let P^N and Q^N be two strict preference profiles, and let $R \subseteq A$. If $a \in R$, and $G(P^N) = a$, then $G(Z(P^N, Q^N; R)) = a$. Equivalently, if $G(Z(P^N, Q^N; R)) \neq a$ and $a \in R$, then $G(P^N) \neq a$.*

Theorem 21.32 *Let G be a monotonic social choice function satisfying the property of unanimity, let P^N be a strict preference profile, and let $a, b \in A$. If*

$$a \succ_{P_i} b, \quad \forall i \in N, \tag{21.16}$$

then $G(P^N) \neq b$.

In words, under the conditions of the theorem, if there is an alternative b that is ranked by all of the individuals below a, then alternative b is not chosen by society.

Proof: Define $Q^N := Z(P^N, P^N; \{a\})$. Under the preference profile Q^N, all individuals rank a highest. Since G satisfies unanimity, $G(Q^N) = a$. Suppose, by contradiction, that $G(P^N) = b$. Since all the individuals prefer a to b according to both P^N and Q^N, and since the preference relations of all the individuals over the set of alternatives $A \setminus \{a\}$ according to both P^N and Q^N are identical, it follows that for each individual i,

$$b \succ_{P_i} c, \quad \Longleftrightarrow \quad b \succ_{Q_i} c, \quad \forall c \neq b. \tag{21.17}$$

Since G is monotonic, we deduce that $G(Q^N) = b$, contradicting $G(Q^N) = a$. This contradiction shows that the original supposition was wrong, and therefore $G(P^N) \neq b$. $\square$

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

Under the strict preference profile $Z(P^N, Q^N; R)$, every individual prefers every result in R to every result not in R. This leads to the following corollary of Theorem 21.32.

Corollary 21.33 *Let G be a monotonic social choice function satisfying unanimity, let P^N and Q^N be two strict profile preferences, and let $R \subseteq A$. If R is nonempty, then $G(Z(P^N, Q^N; R)) \in R$.*

Proof of Theorem 21.27: We will prove that a monotonic social choice function G satisfying the unanimity property is dictatorial. Towards this end, using the social choice function G we will construct a social welfare function F. We will show that if G is monotonic and satisfies the unanimity property then F satisfies unanimity and independence of irrelevant alternatives. Using Theorem 21.10 (page 859) we will then deduce that F is dictatorial. In conclusion, we will show that the dictator in the social welfare function F is also the dictator in the social choice function G.

Let $W^N \in (\mathcal{P}(A))^N$ be a strict preference profile. Fix this profile throughout the rest of the proof.

Step 1: Defining the function F.
For every strict preference profile P^N, define a binary relation $F(P^N)$ as follows. For every pair of distinct alternatives $a, b \in A$,

$$G(Z(P^N, W^N; \{a,b\})) = a \implies a \succ_{F(P^N)} b, \tag{21.18}$$

$$G(Z(P^N, W^N; \{a,b\})) = b \implies b \succ_{F(P^N)} a. \tag{21.19}$$

For this relation to be reflexive, define, in addition,

$$a \succsim_{F(P^N)} a, \quad \forall a \in A. \tag{21.20}$$

We will prove that F is a social welfare function, by showing that the binary relation $F(P^N)$ is complete and transitive.

By Corollary 21.33, $G(Z(P^N, W^N; \{a,b\})) \in \{a,b\}$. Equations (21.18)–(21.19) then imply that for every pair of distinct alternatives a, b in A, either $a \succ_{F(P^N)} b$, or $b \succ_{F(P^N)} a$, i.e., $F(P^N)$ is a complete preference relation over R. Note that the relation $F(P^N)$ expresses no indifference: either society strictly prefers a to b, or it strictly prefers b to a.

Step 2: $F(P^N)$ is a transitive relation.
Suppose by contradiction that for $a, b, c \in A$ one has $a \succ_{F(P^N)} b, b \succ_{F(P^N)} c$, but $c \succ_{F(P^N)} a$. It is not possible that $a = c$, since if $a = c$ one deduces that both $a \succ_{F(P^N)} b$ and $b \succ_{F(P^N)} a$, which is impossible. It follows that a, b, and c are distinct alternatives.

Since $a \succ_{F(P^N)} b$, one has $G(Z(P^N, W^N; \{a,b\})) = a$. Note that the following identity holds:

$$Z(P^N, W^N; \{a,b\}) = Z(Z(P^N, P^N; \{a,b,c\}), W^N; \{a,b\}). \tag{21.21}$$

To see this, over the complement of $\{a,b\}$ the preference relation is determined in both cases by W^N, while over $\{a,b\}$, the preference relation is determined in both cases by P^N.

Equation (21.21) and $a \succ_{F(P^N)} b$ imply that

$$G(Z(Z(P^N, P^N; \{a,b,c\})), W^N; \{a,b\}) = a. \tag{21.22}$$

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

In particular:

$$G(Z(Z(P^N, P^N; \{a,b,c\})), W^N; \{a,b\}) \neq b. \tag{21.23}$$

Theorem 21.31 implies that

$$G(Z(P^N, P^N; \{a,b,c\})) \neq b. \tag{21.24}$$

We have therefore shown that

$$a \succ_{F(P^N)} b \quad \Longrightarrow \quad G(Z(P^N, P^N; \{a,b,c\})) \neq b. \tag{21.25}$$

Similarly, since $b \succ_{F(P^N)} c$, we deduce that

$$G(Z(P^N, P^N; \{a,b,c\})) \neq c, \tag{21.26}$$

and since $c \succ_{F(P^N)} a$ we get

$$G(Z(P^N, P^N; \{a,b,c\})) \neq a. \tag{21.27}$$

Equations (21.24), (21.26) and (21.27) imply that

$$G(Z(P^N, P^N; \{a,b,c\})) \notin \{a,b,c\}. \tag{21.28}$$

On the other hand, by Corollary 21.33, $G(Z(P^N, P^N; \{a,b,c\})) \in \{a,b,c\}$, contradicting Equation (21.28). We deduce that the assumption that the relation $F(P^N)$ is not transitive is false.

We have shown that $F(P^N)$ is a complete and transitive relation, and therefore F is a social welfare function.

Step 3: The social welfare function F satisfies the unanimity property.
Let $a \neq b$ be two alternatives in A, and let P^N be a strict preference profile satisfying $a \succ_{P_i} b$ for every $i \in N$. This means that for every $i \in N$, the alternative a is most preferred under the strict preference relation $Z(P_i, W_i; \{a,b\})$. Theorem 21.32 implies that $G(Z(P^N, W^N; \{a,b\})) = a$, and by Equation(21.18) one has $a \succ_{F(P^N)} b$.

Step 4: The social welfare function F satisfies the independence of irrelevant alternatives property.
Let $a, b \in A$ be two distinct alternatives, and let P^N, Q^N be two strict preference profiles satisfying

$$a \succ_{P_i} b \quad \Longleftrightarrow \quad a \succ_{Q_i} b, \quad \forall i \in N. \tag{21.29}$$

It follows that for every $i \in N$,

$$Z(P_i, W_i; \{a,b\}) = Z(Q_i, W_i; \{a,b\}). \tag{21.30}$$

To see this, in both strict preference relations, the alternatives $\{a,b\}$ are preferred to all the other alternatives; the strict preference relation over the alternatives that are not in $\{a,b\}$ is determined in both cases by W_i, and by Equation (21.29), the ranking between a and b is identical in both cases.

By the definition of F,

$$a \succ_{F(P^N)} b \quad \Longleftrightarrow \quad a \succ_{F(Q^N)} b, \tag{21.31}$$

and therefore the function F satisfies the property of independence of irrelevant alternatives.

Step 5: Using Theorem 21.10 (page 859).
Since F is a social welfare function satisfying the properties of unanimity and independence of irrelevant alternatives, and since A contains at least three alternatives, we can apply Theorem 21.10. This enables us to deduce that F is dictatorial. In other words, there is an individual i such that

$$F(P^N) = P_i, \quad \forall P^N \in (\mathcal{P}(A))^N. \tag{21.32}$$

We will next show that i is also a dictator in the social choice function G.

Let P^N be a strict preference relation, and suppose that a is the alternative most preferred by individual i. To see that i is a dictator under G, we need to show that $G(P^N) = a$. Let $b \neq a$ be an alternative. Since a is the preferred by individual i, one has $a \succ_{P_i} b$. Since i is a dictator under F, one also has $a \succ_{F(P^N)} b$. The definition of F implies that $G(Z(P^N, W^N; \{a, b\})) = a \neq b$, and Theorem 21.31 implies that $G(P^N) \neq b$. Since this is true for every alternative $b \neq a$, we deduce that $G(P^N) = a$. This establishes that G is a dictatorial social choice function, thus completing the proof of Theorem 21.27. □

Theorem 21.27 states that dictatorial social choice functions are the only monotonic social choice functions satisfying the property of unanimity. As the following example shows, there do exist monotonic social choice functions that are not dictatorial (but they do not satisfy the property of unanimity).

Example 21.34 The set of alternatives is $A = \{a, b, c\}$, and $N = \{1, 2\}$. Consider the social choice function F defined by

$$F(P^N) := \begin{cases} b & b \succ_{P_1} c, \\ c & c \succ_{P_1} b. \end{cases} \tag{21.33}$$

In words, the alternative that is chosen by society is the alternative preferred by Player 1 from the set $\{b, c\}$. This social choice function is not dictatorial since alternative a will not be chosen even if it is the preferred alternative of both players, and hence neither player is a dictator. This also shows that F does not satisfy the property of unanimity. ◀

This example leads to the introduction of the following concept. Denote the image of a social choice function G by range(G):

$$\text{range}(G) = \{a \in A\colon \text{ there exists } P^N \in (\mathcal{P}(A))^N \text{ for which } G(P^N) = a\}. \tag{21.34}$$

That is, range(G) is the set of alternatives that may be chosen by society, if society applies the social choice function G. If a social choice function G satisfies the unanimity property then range(G) $= A$.

Theorem 21.35, whose proof is left to the reader (Exercise 21.27), is the generalization of Theorem 21.27 to the case in which range(G) $\neq A$.

Theorem 21.35 *For any monotonic social choice function G satisfying* $|\text{range}(G)| \geq 3$, *there exists a player i such that for every strict preference profile P^N, the alternative $G(P^N)$ is player i's preferred alternative from among the alternatives in* $\text{range}(G)$.

21.3 Non-manipulability

The model presented in the previous section is a model in which each individual is assumed to report (to, say, a governing body) his or her strict preference relation over the set of alternatives, with the social choice function then choosing an alternative that is declared to be society's most-preferred alternative. The model thus assumes implicitly that each individual reports his or her true preference relation. But why should an individual always report his or her true preference relation? Perhaps there might be cases in which by reporting a preference relation that is different from his true preference, an individual can cause society to choose an alternative that is more preferred to him than the alternative that would be chosen if were to report his true preference relation? If this is possible, we say that the social choice function is manipulable.

Given a strict preference profile P^N, denote by $P_{-i} = (P_j)_{j \neq i}$ the strict preferences of all the individuals who are not individual i. In other words, (P_i, P_{-i}) is an alternative denotation for the profile P^N.

Definition 21.36 *A social choice function G is called* manipulable *if there exist a strict preference profile P^N, an individual $i \in N$, and a strict preference relation Q_i satisfying*

$$G(Q_i, P_{-i}) \succ_{P_i} G(P^N). \tag{21.35}$$

In words, a social choice function G is manipulable if there exists a strict preference profile such that there is an individual who can, by reporting a strict preference relation different from his true one, cause society to choose an alternative that is more preferred by him than the alternative that would be chosen if he were to report his true preference relation. If this is not possible, the social choice function is called *nonmanipulable*. If a social choice function is nonmanipulable, the situation in which every individual reports his or her true strict preference relation is a Nash equilibrium in the game in which the set of strategies of each individual is the set of strict preference relations $\mathcal{P}(A)$, and the outcome is the alternative chosen by society.

Remark 21.37 *Definition 21.36 touches on the possibility that a single individual may influence the alternative that is chosen by reporting a preference relation that differs from his true preference relation. As we saw in Example 21.1, social choice functions may also be manipulable by sets of individuals. We will not expand on this idea in this chapter.* ♦

A dictatorial social choice function is nonmanipulable. The dictator cannot gain by reporting a strict preference relation that is different from his true preference relation, because society's choice is always the most-preferred alternative that he reports. Neither can the other individuals gain by reporting false preference relations, because their reported preference relations have no effect on society's choice, in any event. The next example presents a social choice function that is manipulable by a single voter.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

Example 21.38 **A manipulable social choice function** Lisa (individual 1), Mickey (individual 2), and Emily (individual 3) comprise the membership of a village social committee, charged with choosing the theme of the annual village social event. The committee makes its choice by majority vote. If each committee member votes for a different alternative, the deciding vote is the one cast by Lisa, the committee chairman. Three alternatives have been suggested: a bingo night (B), a dance party (D), or a village singalong (S).

Consider the following strict preference profile P^N in which P_1 is Lisa's preference relation, P_2 is Mickey's preference relation, and P_3 is Emily's preference relation:

$$P^N: \quad \begin{array}{l} B \succ_{P_1} D \succ_{P_1} S, \\ D \succ_{P_2} B \succ_{P_2} S, \\ S \succ_{P_3} D \succ_{P_3} B. \end{array} \tag{21.36}$$

If Lisa, Mickey, and Emily all report their true preference relations, the chosen alternative will be a bingo night. If, however, Emily changes her reported preference relation to

$$D \succ_{Q_3} S \succ_{Q_3} B, \tag{21.37}$$

the chosen alternative will be a dance party, which Emily prefers to bingo. We see that the social choice function in this example, in which the majority rule is applied, with the committee chairman granted the tie-breaking vote, is manipulable. ◄

Gibbard [1973] and Satterthwaite [1975] proved the following theorem.

Theorem 21.39 (Gibbard, Satterthwaite) *Let G be a nonmanipulable social choice function satisfying the unanimity property. If $|A| \geq 3$ then G is dictatorial.*

The practical implication of this theorem is that if we wish to apply a nondictatorial social choice function, there are necessarily situations in which one (or more) of the individuals has an incentive to report a preference relation that is different from his or her true preference relation.

Proof: By Theorem 21.27, it suffices to show that every nonmanipulable social choice function that satisfies the unanimity property is monotonic.

Let G be a nonmanipulable social choice function satisfying the unanimity property. Suppose that G is not monotonic, i.e., that there exist two distinct strict preference profiles P^N and Q^N, and two distinct alternatives a, b such that

$$a \succ_{P_i} c \implies a \succ_{Q_i} c, \quad \forall c \neq a, \forall i \in N, \tag{21.38}$$

while

$$G(P^N) = a, \text{ and } G(Q^N) = b. \tag{21.39}$$

Since P^N and Q^N are distinct strict preference profiles there is at least one individual i for whom $P_i \neq Q_i$. From among all pairs of alternatives a, b and strict preference profiles P^N and Q^N with respect to which the above conditions hold, choose those for which the number of individuals i for which $P_i \neq Q_i$ is minimal.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

For such P^N and Q^N, denote by I the set of individuals i for whom $P_i \neq Q_i$:

$$I = \{i \in N : P_i \neq Q_i\}. \tag{21.40}$$

By assumption, the set I contains at least one individual.

Let j be an individual in I. We claim that $G(P_j, Q_{-j}) = a$. Suppose by contradiction that $G(P_j, Q_{-j}) = c \neq a$ (c may be equal to b). The pair of alternatives a, c, along with the pair of strict preference profiles P^N and (P_j, Q_{-j}) satisfy the above conditions but the number of individuals whose strict preference relations differ in the two profiles is $|I| - 1$ (because the preferences of individual j in the profiles P^N and (P_j, Q_{-j}) are identical), which contradicts the minimality of I. This contradiction establishes that indeed $G(P_j, Q_{-j}) = a$.

In summary, we have deduced that $G(Q^N) = b$ and $G(P_j, Q_{-j}) = a$. Since G is nonmanipulable, when the preference profile is Q^N, individual j has no incentive to report P_j as his preference profile, and therefore

$$b = G(Q^N) \succ_{Q_j} G(P_j, Q_{-j}) = a. \tag{21.41}$$

Similarly, since G is nonmanipulable, when the preference profile is (P_j, Q_{-j}), individual j has not incentive to report Q_j as his preference profile, and therefore:

$$a = G(P_j, Q_{-j}) \succ_{P_j} G(Q^N) = b. \tag{21.42}$$

Equations (21.38) and (21.42) imply that

$$a \succ_{Q_j} b. \tag{21.43}$$

Equation (21.43) contradicts Equation (21.41). This contradiction establishes that G is monotonic. □

21.4 Discussion

In this chapter, we have studied the question of how to aggregate the preferences of a group of individuals into a single "social preference." The approach we have adopted is the normative, or axiomatic, approach. In other words to construct a choice function that associates every strict preference profile of individual preference relations with a social preference relation, we asked what properties such a function should satisfy. Surprisingly, this led us to conclude that if there are at least three alternatives, seemingly natural and reasonable properties cannot hold unless the choice function is dictatorial.

The most fundamental result in this section is Arrow's Impossibility Theorem (Theorem 21.10 on page 859), which states that when there are at least three alternatives that are to be ranked, the only social welfare function satisfying the properties of unanimity and independence of irrelevant alternatives is dictatorial. This implies that the only social choice function (which chooses one alternative, "the best alternative," based on the strict preference profiles of the individuals) satisfying monotonicity or nonmanipulability is again dictatorial.

These results form the foundations of an important branch of study (especially for the disciplines of economics and political science) called "social choice." One way to obtain positive results in this field is to limit the domain of the social welfare (or choice) functions. In other words, if one does not allow individuals to have every possible preference relation, and instead restricts preference relations to a smaller set than the set of all preference relations, it is in some cases possible to obtain positive results. For example, when studying political or economic preferences (from conservative to liberal) it is customary to assume that preferences are single-peaked, meaning that every individual has an "ideal point" along a scale of alternatives that he prefers, with his ranking of other alternatives decreasing the farther those alternatives are from his ideal point. If one assumes that preferences are single-peaked, it is possible to find social choice functions that are not dictatorial and satisfy monotonicity or nonmanipulability.

Another direction of inquiry in social choice theory involves "taking manipulation into account": constructing games whose sole equilibria are "desired outcomes." In other words, one studies implementations of social choice functions by appropriate game mechanisms, taking into account that each individual will do his best to have his preferred alternatives chosen (possibly by not reporting his true preferences). The interested reader is directed to Peleg [1984] for a detailed analysis of this subject.

21.5 Remarks

Example 21.1 was first presented by Borda [1784]. The interested reader may find three simple proofs of Theorem 21.10 in Geanakoplos [2005].

Both the Borda Method (Exercise 21.3) and the Condorcet Method (Exercise 21.4) were suggested as early as the thirteenth century by Ramon Lull. Approval voting (see Exercise 21.28) was used as the voting procedure of the Major Council of the Republic of Venice in the Middle Ages. Exercise 21.5 is based on an example appearing in Balinski and Laraki [2007]. The authors thank Rida Laraki for kindly answering many questions during the composition of this chapter.

21.6 Exercises

21.1 For each of the following relations, determine whether it is complete, reflexive, irreflexive, or transitive, and use this to determine whether it is a preference relation or a strict preference relation.

(a) A is the set of all subsets of some set S, and $a \succsim b$ if and only if b is a subset of a.
(b) A is the set of all natural numbers, and $a \succsim b$ if and only if b is a divisor of a (i.e., $a = bq$ for some integer q).
(c) A is the set of all 26 letters in the Latin alphabet, and $\alpha \succsim \beta$ if and only if $\alpha\beta$ is a standard word in English (where α is the first letter of the word, and β the second letter of the word).

(d) A is the set of all natural numbers, and $a \succsim b$ if and only if $a \times b = 30$.
(e) A is the set of all human beings, past and present, and $b \succsim a$ if and only if a is a descendant of b (meaning a child, grandchild, great-grandchild, etc).
(f) A is the set of people living in a particular neighborhood, and $a \succsim b$ if and only if a likes b.

21.2 Show that if P^* is a strict preference relation then $P := P^* \cup \{(a, a) \colon a \in A\}$ is a preference relation.

21.3 **The Borda Method** The French mathematician Borda proposed the following voting method. Every voter ranks the candidates, from most preferred to least preferred. A candidate receives k points (called Borda points) from a voter if that voter ranks the candidate higher than exactly k other candidates. The Borda ranking of a candidate is given by the total number of Borda points he receives from all the voters. The winning candidate (called the Borda winner) is then the candidate who has amassed the most Borda points.

(a) For every pair of candidates a and b, let $N_{a,b}$ be the number of voters ranking a ahead of b. Show that the Borda ranking of candidate a equals $\sum_{b \neq a} N_{a,b}$.
(b) Compute the Borda ranking, and the Borda winner, from among the three candidates A, B, and C in Example 21.2.

21.4 **The Condorcet Method** The French mathematician Condorcet proposed the following method for determining a social preference order based on the strict preferences of the individuals in society. A voter i with a strict preference order P_i grants k Condorcet points to a strict preference relation P if there are exactly k pairs of alternatives a, b satisfying $b \succ_P a$ and $b \succ_{P_i} a$. The number of Condorcet points amassed by the strict preference relation P is the total sum of the Condorcet points it receives from all the voters. The strict social preference order is the one that has amassed the greatest number of Condorcet points.

(a) Is the strict preference relation amassing the greatest number of Condorcet points unique? If yes, prove this claim. If no, present a counterexample.
(b) Show that if there exists a Condorcet winner, then every strict preference relation receiving the maximal number of Condorcet points ranks the Condorcet winner highest in its preference ordering.
(c) Find the number of Condorcet points that each strict preference relation receives in Examples 21.1 and 21.2, and determine the preference relation that the Condorcet Method chooses.

21.5 **The Borda and Condorcet Methods** The following example was presented by Condorcet, in a critique of the Borda Method. A committee composed of 81 members is to choose a winner from among three candidates, A, B, and C. The rankings of the committee members appear in the following table:

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

No. of voters	First candidate	Second candidate	Third candidate
30	A	B	C
1	A	C	B
29	B	A	C
10	B	C	A
10	C	A	B
1	C	B	A

(a) Is there a Condorcet winner? If yes, who is the Condorcet winner?
(b) What is the Borda ranking, and who is the Borda winner?
(c) Based on your answers above, what is Condorcet's critique of the Borda Method?

A counterclaim to Condorcet's critique might be given by analyzing the preference profile in the following way. A *Condorcet component* is a set of $3n$ individuals whose strict preference relations are as follows:

No. of voters	First candidate	Second candidate	Third candidate
n	a_1	a_2	a_3
n	a_3	a_1	a_2
n	a_2	a_3	a_1

The variables a_1, a_2, a_3 are three distinct alternatives (this example can be generalized to an arbitrary number of alternatives). In a certain sense, these voters "neutralize" each other, and they can therefore be removed from the list of voters.

(a) What are all the Condorcet components in this example?
(b) Remove all the individuals appearing in one of the Condorcet components. From among the remaining individuals, find another Condorcet component and remove all the individuals in that component. Repeat this process until there remain no Condorcet components. Is there a Condorcet winner according to the strict preference profile that remains at the end of this process? If so, which candidate is the Condorcet winner?
(c) Given the above two items, elucidate a counterclaim to Condorcet's critique of the Borda Method.

21.6 Show that there is at most one dictator in every social welfare function F: if i is a dictator in F, and j is also a dictator in F, then $i = j$.

21.7 A committee comprised of 15 members is called upon to rank three colors: red, blue, and yellow, from most preferred to least preferred. The committee members simultaneously announce their strict preference relations over the three colors. If red is the preferred color of at least five members of the committee, red is determined to be the prettiest color. Otherwise, if blue is the preferred color of at least five members of the committee, blue is determined to be the prettiest color. Otherwise, yellow is determined to be the prettiest color. Once the prettiest color is determined, the remaining two colors are then ranked by the simple majority rule.

(a) Is the social welfare function described here dictatorial? Justify your answer.
(b) Does the social welfare function described here satisfy the unanimity property? Justify your answer.
(c) Does the social welfare function described here satisfy the independence of irrelevant alternatives property? Justify your answers.

21.8 Repeat Exercise 21.7 for the following situation. There are two alternatives, $A = \{a, b\}$, and n voters. Let $k \in \{0, 1, 2, \dots, n\}$ and let F_k be the following social welfare function: $a \succsim_{F_k(P^N)} b$ if and only if the number of individuals who prefer a over b is greater than or equal to k.

21.9 Let F be a social welfare function satisfying the independence of irrelevant alternatives property, and let a, b be two distinct alternatives. Let P^N and Q^N be two strict preference profiles satisfying

$$a \succ_{P_i} b \iff a \succ_{Q_i} b, \quad \forall i \in N. \tag{21.44}$$

Prove that $a \approx_{F(P^N)} b$ if and only if $a \approx_{F(Q^N)} b$.

21.10 Denote by $K = |A|$ the number of alternatives in A. For each alternative $a \in A$, denote by $j_a(P_i)$ the ranking of a in the strict preference relation P_i (for example, $j_a(P_i) = 1$ when alternative a is the most-preferred alternative according to P_i). Define a social welfare function as follows. For each alternative a, compute the sum $s_a = \sum_{i \in N} j_a(P_i)$. We say that alternative a is (weakly) preferred to alternative b if and only if $s_a \leq s_b$.

(a) Prove that this defines a social welfare function.
(b) Is this a dictatorial function? Does it satisfy the unanimity property? Does it satisfy the independence of irrelevant alternatives property? Justify your answers.
(c) What is the connection between this voting system and the Borda system appearing in Exercise 21.3?

21.11 Prove Theorem 21.16 (page 861): let F be a social welfare function satisfying the unanimity property. Then for every $a, b \in A$ the coalition N is decisive for a over b, and the empty coalition $\emptyset$ is not decisive for a over b.

21.12 A jury composed of seven members is called upon to find an accused individual either guilty or innocent. Debbie is the jury forewoman, with Bobby and Jack appointed vice-foremen. The jury includes four more jurors, in addition to Debbie, Bobby and Jack. For each of the following cases, describe the set of all decisive coalitions for "guilty" over "innocent," and the set of all minimal decisive coalitions for this pair of alternatives.

(a) The accused is found guilty only if all the jurors unanimously agree that he is guilty.
(b) The accused is found guilty if a majority of jurors declare him to be guilty.
(c) The accused is found guilty if at least four jurors, including Debbie, declare him to be guilty.

(d) The accused is found guilty if at least four jurors, including Debbie, declare him to be guilty, or if at least five jurors, including Bobby and Jack, declare him to be guilty.
(e) The accused is found guilty if at least four jurors declare him to be guilty, or if Debbie, Bobby, and Jack declare him to be guilty.
(f) The accused is found guilty if at least five jurors, including Debbie and at least one of her vice-foremen, declare him to be guilty.

21.13 Suppose that $|A| \geq 3$, and let F be a social welfare function satisfying the properties of unanimity and independence of irrelevant alternatives. What are all the decisive coalitions?

21.14 Show that the assumptions of Theorem 21.18 (page 862) are necessary for its conclusion: if we remove any one of the three assumptions of the theorem, the conclusion does not hold.

21.15 Sam wishes to prove Claim 21.20 (page 864) using the following strict preference profile, instead of the preference profile Q^N appearing in the proof of the claim on page 864. How can you help Sam complete his proof of the claim?

$$\begin{cases} a \succ_{Q_i} c \succ_{Q_i} b & i = j, \\ a \succ_{Q_i} b \succ_{Q_i} c & i \neq j, a \succ_{P_i} b, \\ b \succ_{Q_i} a \succ_{Q_i} c & i \neq j, b \succ_{P_i} a. \end{cases} \tag{21.45}$$

21.16 Ben wishes to prove Claim 21.20 (page 864) using the following strict preference profile, instead of the preference profile Q^N appearing in the proof of the claim on page 864. Why does the proof of the claim fail when using this preference relation?

$$\begin{cases} a \succ_{Q_i} c \succ_{Q_i} b & i = j, \\ a \succ_{Q_i} c \succ_{Q_i} b & i \neq j, a \succ_{P_i} b, \\ c \succ_{Q_i} b \succ_{Q_i} a & i \neq j, b \succ_{P_i} a. \end{cases} \tag{21.46}$$

21.17 A social welfare function F is called *monotonic* if for every alternative $a \in A$, and every pair of strict preference profiles P^N and Q^N satisfying

$$a \succ_{P_i} c \quad \Longrightarrow a \succ_{Q_i} c, \quad \forall c \neq a, \ \forall i \in N, \tag{21.47}$$

the following is also satisfied:

$$a \succ_{F(P)} c \quad \Longrightarrow a \succ_{F(Q)} c, \quad \forall c \neq a. \tag{21.48}$$

In words, if for each individual i the ranking of alternative a relative to the other alternatives is not lowered in moving from P_i to Q_i, then its ranking is not lowered in society's ranking, according to the social welfare function, when moving from profile P^N to profile Q^N.

Answer the following questions:

(a) Are the social welfare functions in Exercises 21.7 and 21.10 monotonic? Justify your answer.
(b) Does every monotonic social welfare function satisfy the unanimity property? Justify your answer.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

(c) Does every monotonic social welfare function satisfy the independence of irrelevant alternatives property? Justify your answer.
(d) Is every social welfare function satisfying the independence of irrelevant alternatives property monotonic? Justify your answer.

21.18 Ron and Veronica need to choose a name for their newborn daughter. After giving the matter much thought, they have narrowed the list of possible names to four: Abigail, Iris, Irene, and Olga. They now must choose one name from this list. Each parent ranks the four names in order of preference. Given each of the following decision rules, determine whether it is a social choice function. If yes, determine whether it is monotonic, and whether it is manipulable. Justify your answers. (If a rule is monotonic, provide a direct proof that it satisfies the property of monotonicity. If a rule is manipulable, provide an example showing how it may be manipulated.)

(a) If both parents select a name as being the most preferred, that name is chosen. Otherwise, if there is only one name that both parents rank within their top two most-preferred names, that name is chosen. Otherwise, Abigail is the chosen name.
(b) If both parents select a name as being the most preferred, that name is chosen. Otherwise, if there is only one name that both parents rank within their top two most-preferred names, that name is chosen. Otherwise, the parents toss a coin to determine the name of their daughter, with Iris chosen if the coin shows heads, and Olga chosen if the coin shows tails.
(c) If Ron most prefers the name Irene, that name is chosen. Otherwise, the name that Veronica most prefers is chosen.
(d) As a first step, the name that each parent ranks last by preference is removed from the list under consideration. This leaves two or three names in contention. In the next step, the name that each parent ranks last by preference from among the two or three remaining names is removed from the list under consideration. This leaves zero, one, or two names. If two names remain, the process of removing the name that each parent least prefers from the list under consideration is repeated again. If only one name remains, that name is chosen. Otherwise, Olga is chosen as the child's name.

21.19 A committee comprised of 15 members is called upon to choose the prettiest color: red, blue, or yellow. The committee members simultaneously announce their strict preference relations among these three colors. If red is the most-preferred color of at least one committee member, red is declared the prettiest color. Otherwise, if blue is the most-preferred color of at least one committee member, blue is declared the prettiest color. Otherwise, yellow is declared the prettiest color.

(a) Is the social choice function described above dictatorial? Justify your answer.
(b) Is the social choice function described above monotonic? Justify your answer.
(c) Is the social choice function described above manipulable? Justify your answer.

21.20 The following electoral method is used to choose the mayor of Whoville: Every resident ranks the candidates from most preferred to least preferred, and places this ranked list in a ballot box. Each candidate receives a number of points equal to the number of residents who prefer him to all the other candidates. The candidate who thus amasses the greatest number of points wins the election. If two or more candidates are tied for first place in the number of points, the winner of the election is the candidate among them whose social security number is smallest. Assume there are at least three candidates.

(a) Show by example that this electoral method is not monotonic.
(b) Show by example that this electoral method is not manipulable.

21.21 Repeat Exercise 21.20, under the following scenario. The following electoral method is used to choose the mayor of Sleepy Hollow: Every resident ranks the candidates from most preferred to least preferred, and places this ranked list in a ballot box. Each candidate receives a number of points equal to the number of residents who rank him or her in the first two positions. The candidate who thus amasses the greatest number of points wins the election. If two or more candidates are tied for first place in the number of points, the winner of the election is the candidate among them whose social security number is smallest. Assume there are at least three candidates.

21.22 The following electoral method is used to choose the mayor of Hobbiton: Every resident ranks the candidates from most preferred to least preferred, and places this ranked list in a ballot box. Each candidate receives a number of points equal to the number of residents who rank him or her least preferred. The candidate who thus amasses the greatest number of points is then removed from the list of candidates. If two or more candidates are tied for first place in the number of points, the candidate among them whose social security number is greatest is removed from the list of candidates. This candidate is then ignored in the strict preference relations submitted by the residents, and the process is repeated as often as is necessary, until only one candidate remains, who is declared the new mayor. Assume there are at least three candidates.

(a) Is it possible for the winner of the election to not be the most preferred candidate of any resident? Justify your answer.
(b) Is it possible for the winner of the election to be ranked least preferred by at least half of the residents? Justify your answer.
(c) Is the social choice function described here dictatorial? Justify your answer.
(d) Is it monotonic? Prove why yes, or show by example that it is not monotonic.
(e) Is it manipulable? If yes, provide an example or otherwise prove that it is not.
(f) After the election of the mayor, using the above method, is completed, the local Elections Board checks which candidate would have won had they implemented instead the electoral method used in Whoville (see Exercise 21.20). Will the same candidate necessarily be chosen under both electoral methods? Justify your answer.

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

21.23 Repeat Exercise 21.20 for the following situation. The following electoral method is used to choose the mayor of Hogsmeade Village: Every resident ranks the candidates from most preferred to least preferred, and places this ranked list in a ballot box. For each candidate a denote by $N_{a,k}$ the ratio of residents (using a number between 0 and 1) who ranked him or her in the top k positions. Further denote:

$$k_0 = \min\left\{k \in \mathbb{N}\colon \text{there is a candidate } a \text{ such that } N_{a,k} > \frac{k}{k+1}\right\}. \tag{21.49}$$

The winning candidate is the one whose value of N_{a,k_0} is maximal (if there are two or more such candidates, the winner is the candidate among them whose social security number is smallest).

Under this method, if there is a candidate who is ranked first by more than half of the population, he or she wins the election. Otherwise, if there are one or more candidates who are ranked first or second by more than two-thirds of the population, the winner of the election is the candidate who is ranked first or second by the greatest number of residents, and so on.

21.24 Let A be a set of alternatives, and let P^N be a strict preference profile. Alternative $a \in A$ is termed the *Condorcet winner* if for every alternative $b \neq a$, more than half of the individuals rank a above b. A social choice function G satisfies the *Condorcet criterion* if for every strict preference profile P^N for which there exists a Condorcet winner a, it chooses the Condorcet winner, i.e., $G(P^N) = a$ holds.

(a) Does a Condorcet winner a exist for every strict preference profile P^N? Justify your answer.
(b) Which of the social choice functions described in Exercises 21.19–21.23 satisfy the Condorcet criterion? Justify your answer.

21.25 Let G be a monotonic social choice function, let P^N and Q^N be two strict preference profiles, and let $R \subseteq A$. Show that if $a \in R$ and $G(P^N) = a$, then $G(Z(P^N, Q^N; R)) = a$. (for the definition of $Z(P^N, Q^N; R)$ see Definition 21.30 on page 867).

21.26 Let G be a monotonic social choice function, let $a \in \text{range}(G)$, and let P^N be a strict preference profile. Show that if

$$a \succ_{P_i} b, \quad \forall i \in N, \tag{21.50}$$

then $G(P^N) \neq b$.

21.27 In this exercise we prove Theorem 21.35 (page 871).

(a) Prove that if G is a monotonic social choice function, then for any pair of strict preference relations P^N and Q^N satisfying

$$a \succ_{P_i} b \iff a \succ_{Q_i} b, \quad \forall a, b \in \text{range}(G), \forall i \in N, \tag{21.51}$$

https://doi.org/10.1017/CBO9780511794216.022 Published online by Cambridge University Press

$G(P^N) = G(Q^N)$. In words, the claim states that the alternatives that are not in range(G) have no effect on the choice of G.

(b) Show that if G is a monotonic social choice function satisfying $|\text{range}(G)| \geq 3$, then there is an individual i such that for every strict preference profile P^N the alternative $G(P^N)$ is the most prefered alternative from i's perspective, from among all the alternatives in range(G).

Guidance: For the first part, choose a strict preference profile W^N, and denote $R = \text{range}(G)$. Show that $Z(P^N, W^N; R) = Z(Q^N, W^N; R)$, and use Theorem 21.31 to derive the claim. For the proof of the second part, use Exercise 21.26.

21.28 **Approval voting** In this question, we consider the case in which the individuals are called upon to choose candidates for a task, by specifying which candidates they most approve for the task. Let A be a nonempty set of alternatives. A binary relation P_i over A is called (at most) *two-leveled* if there exists a set $B(P_i) \subseteq A$ satisfying: (i) $b \approx_{P_i} c$ for every $b, c \in B(P_i)$, (ii) $b \approx_{P_i} c$ for every $b, c \in A \setminus B(P_i)$, (iii) $b \succ_{P_i} c$ for every $b \in B(P_i)$, and $c \in A \setminus B(P_i)$. In words, the individual is indifferent between all the alternatives in $B(P_i)$, he is indifferent between all the alternatives that are not in $B(P_i)$, and he prefers all the elements in $B(P_i)$ to all the elements that are not in $B(P_i)$. The interpretation that we give to such a preference relation is that the individual approves of all the alternatives in $B(P_i)$, and disapproves of all the alternatives not in $B(P_i)$. A two-leveled preference profile P^N is a profile of preference relations all of which are two-leveled.

Consider a choice function H associating every two-leveled profile with a single alternative, which is declared to be society's most-preferred alternative. Such a choice function H is called *monotonic* if for every pair of two-leveled preference profiles P^N and Q^N, if $H(P^N) = a$, and if every individual i satisfies $B(Q_i) = B(P_i)$ or $B(Q_i) = B(P_i) \cup \{a\}$, then $H(Q^N) = a$. In other words, if alternative a is chosen under preference profile P^N, and if Q^N is a preference profile that is identical to P^N except that some individuals have added a to the set of their approved alternatives, then a is also chosen under Q^N.

A choice function H is called *nonmanipulable* if for every two-leveled preference profile P^N, for every individual i, and for every two-leveled preference relation Q_i, it is the case that $H(P^N) \succsim_{P_i} H(Q_i, P_{-i})$.

Define a choice function H^* as follows: a winning alternative is one that is approved by the greatest number of individuals; in other words, a is a winning alternative if it maximizes the value of $|\{i \in N : a \in B(P_i)\}|$. If there are two or more alternatives receiving the greatest number of approval votes, the alternative whose serial number is lowest is chosen.

Is this choice function monotonic? Is it manipulable? Justify your answer.

21.29 This exercise is similar to Exercise 21.28, but we now assume that the individuals are called upon to agree on a set of approved alternatives (as opposed to one most-approved alternative). In other words, the function H associates each two-leveled preference profile P^N with a two-leveled preference profile $H(P^N)$. In this case, a choice function H is called *nonmanipulable* if for every two-leveled preference profile P^N, and every individual i, if $a \in B(P_i)$ and if $a \notin B(H(P^N))$, then $a \notin B(H(Q_i, P_{-i}))$ for every two-leveled preference relation Q_i.

Define a function H^* as follows: the set of approved alternatives, $H^*(P^N)$, is the set of alternatives that are approved by the greatest number of individuals according to P^N. Is this function manipulable? Justify your answer.