9

Aggregating Preferences: Social Choice

In the preceding chapters we adopted what might be called the "agent perspective": we asked what an agent believes or wants, and how an agent should or would act in a given situation. We now adopt a complementary, "designer perspective": we ask what rules should be put in place by the authority (the "designer") orchestrating a set of agents. In this chapter this will take us away from game theory, but before too long (in the next two chapters) it will bring us right back to it.

9.1 Introduction

A simple example of the designer perspective is voting. How should a central authority pool the preferences of different agents so as to best reflect the wishes of the population as a whole? It turns out that voting, the kind familiar from our political and other institutions, is only a special case of the general class of *social choice problems*. Social choice is a motivational but nonstrategic theory—agents have preferences, but do not try to camouflage them in order to manipulate the outcome (of voting, for example) to their personal advantage.[1] This problem is thus analogous to the problem of belief fusion that we present in Section 14.2.1, which is also nonstrategic; here, however, we examine the problem of aggregating preferences rather than beliefs.

social choice problem

We start with a brief and informal discussion of the most familiar voting scheme, plurality. We then give the formal model of social choice theory, consider other voting schemes, and present two seminal results about the sorts of preference aggregation rules that it is possible to construct. Finally, we consider the problem of building ranking systems, where agents rate each other.

9.1.1 *Example: plurality voting*

To get a feel for social choice theory, consider an example in which you are babysitting three children—Will, Liam, Vic—and need to decide on an activity for them. You can choose among going to the video arcade (a), playing basketball (b), and driving around in a car (c). Each kid has a different preference over these

1. Some sources use the term "social choice" to refer to both strategic and nonstrategic theories; we do not follow that usage here.

activities, which is represented as a strict total ordering over the activities and which he reveals to you truthfully. By $a \succ b$ denote the proposition that outcome a is preferred to outcome b.

$$\begin{aligned} \text{Will:} \quad & a \succ b \succ c \\ \text{Liam:} \quad & b \succ c \succ a \\ \text{Vic:} \quad & c \succ b \succ a \end{aligned}$$

What should you do? One straightforward approach would be to ask each kid to vote for his favorite activity and then to pick the activity that received the largest number of votes. This amounts to what is called the *plurality* method. While quite standard, this method is not without problems. For one thing, we need to select a tie-breaking rule (e.g., we could select the candidate ranked first alphabetically). A more disciplined way is to hold a runoff election among the candidates tied at the top.

plurality voting

Even absent a tie, however, the method is vulnerable to the criticism that it does not meet the *Condorcet condition*. This condition states that if there exists a candidate x such that if for all other candidates y at least half the voters prefer x to y, then x must be chosen. If each child votes for his top choice, the plurality method would declare a tie between all three candidates and, in our example, would choose a. However, the Condorcet condition would rule out both a (since in two of the preferences b is preferred to a) and c (since in two of the preference orderings b is preferred to c).

Condorcet condition

Based on this example the Condorcet rule might seem unproblematic (and actually useful since it breaks the tie without resorting to an arbitrary choice such as alphabetical ordering), but now consider a similar example in which the preferences are as follows.

$$\begin{aligned} \text{Will:} \quad & a \succ b \succ c \\ \text{Liam:} \quad & b \succ c \succ a \\ \text{Vic:} \quad & c \succ a \succ b \end{aligned}$$

In this case the Condorcet condition does not tell us what to do, illustrating the fact that it does not tell us how to aggregate arbitrary sets of preferences. We will return to the question of what properties can be guaranteed in social choice settings; for the moment, we aim simply to illustrate that social choice is not a straightforward matter. In order to study it precisely, we must establish a formal model. Our definition will cover voting, but will also handle more general situations in which agents' preferences must be aggregated.

9.2 A formal model

Let $N = \{1, 2, \ldots, n\}$ denote a set of agents, and let O denote a finite set of outcomes (or alternatives, or candidates). Making a multiagent extension to the preference notation introduced in Section 3.1.2, denote the proposition that agent i weakly prefers outcome o_1 to outcome o_2 by $o_1 \succeq_i o_2$. We use the notation $o_1 \succ_i o_2$ to capture strict preference (shorthand for $o_1 \succeq_i o_2$ and not $o_2 \succeq_i o_1$) and $o_1 \sim_i o_2$ to capture indifference (shorthand for $o_1 \succeq_i o_2$ and $o_2 \succeq_i o_1$).

preference ordering

Because preferences are transitive, an agent's preference relation induces a *preference ordering*, a (nonstrict) total ordering on O. Let L_- be the set of nonstrict total orders; we will understand each agent's preference ordering as an element of L_-. Overloading notation, we denote an element of L_- using the same symbol we used for the relational operator: $\succeq_i \in L_-$. Likewise, we define a *preference profile* $[\succeq] \in L_-^n$ as a tuple giving a preference ordering for each agent.

preference profile

Note that the arguments in Section 3.1.2 show that preference orderings and utility functions are tightly related. We can define an ordering $\succeq_i \in L_-$ in terms of a given utility function $u_i : O \mapsto \mathbb{R}$ for an agent i by requiring that o_1 is weakly preferred to o_2 if and only if $u_i(o_1) \geq u_i(o_2)$.

In what follows, we define two kinds of social functions. In both cases, the input is a preference profile. Both classes of functions aggregate these preferences, but in a different way.

Social choice functions simply select one of the alternatives (or, in a more general version, some subset).

social choice function

Definition 9.2.1 (Social choice function) *A* social choice function *(over N and O) is a function $C : L_-^n \mapsto O$.*

social choice correspondence

A *social choice correspondence* differs from a social choice function only in that it can return a set of candidates, instead of just a single one.

social choice correspondence

Definition 9.2.2 (Social choice correspondence) *A* social choice correspondence *(over N and O) is a function $C : L_-^n \mapsto 2^O$.*

In our babysitting example there were three agents (Will, Liam, and Vic) and three possible outcomes (a, b, c). The social choice correspondence defined by plurality voting of course picks the subset of candidates with the most votes; in this example either the subset must be the singleton consisting of one of the candidates or else it must include all candidates. Plurality is turned into a social choice function by any deterministic tie-breaking rule (e.g., alphabetical).[2]

Let $\#(o_i \succ o_j)$ denote the number of agents who prefer outcome o_i to outcome o_j under preference profile $[\succeq] \in L_-^n$. We can now give a formal statement of the Condorcet condition.

Condorcet winner

Definition 9.2.3 (Condorcet winner) *An outcome $o \in O$ is a* Condorcet winner *if $\forall o' \in O$, $\#(o \succ o') \geq \#(o' \succ o)$.*

A social choice function satisfies the *Condorcet condition* if it always picks a Condorcet winner when one exists. We saw earlier that for some sets of preferences there does *not* exist a Condorcet winner. (Indeed, under reasonable conditions the probability that there will exist a Condorcet winner approaches zero as the number of candidates approaches infinity.) Thus, the Condorcet condition does not always tell us anything about which outcome to choose.

2. One can also define probabilistic versions of social choice functions; however, we will focus on the deterministic variety.

An alternative is to find a rule that identifies a *set* of outcomes among which we can choose. Extending on the idea of the Condorcet condition, a variety of other conditions have been proposed that are guaranteed to identify a nonempty set of outcomes. We will not describe such rules in detail; however, we give one prominent example here.

Smith set

Definition 9.2.4 (Smith set) *The* Smith set *is the smallest set $S \subseteq O$ having the property that $\forall o' \notin S$, $\#(o \succ o') \geq \#(o' \succ o)$.*

That is, every outcome *in* the Smith set is preferred by at least half of the agents to every outcome *outside* the set. This set always exists. When there is a Condorcet winner then that candidate is also the only member of the Smith set; otherwise, the Smith set is the set of candidates who participate in a "stalemate" (or "top cycle").

The other important flavor of social function is the *social welfare function.* These are similar to social choice functions, but produce richer objects, total orderings on the set of alternatives.

social welfare function

Definition 9.2.5 (Social welfare function) *A* social welfare function *(over N and O) is a function $W : L_{\text{-}}^n \mapsto L_{\text{-}}$.*

Although the usefulness of these functions is somewhat less intuitive, they are very important to social choice theory. We will discuss them further in Section 9.4.1, in which we present Arrow's famous impossibility theorem.

9.3 Voting

We now survey some important voting methods and discuss their properties. Then we demonstrate that the problem of voting is not as easy as it might appear, showing some counterintuitive ways in which these methods can behave.

9.3.1 *Voting methods*

nonranking voting

The most standard class of voting methods is called *nonranking voting*, in which each agent votes for one of the candidates. We have already discussed plurality voting.

Definition 9.3.1 (Plurality voting) *Each voter casts a single vote. The candidate with the most votes is selected.*

As discussed earlier, ties must be broken according to a tie-breaking rule (e.g., based on a lexicographic ordering of the candidates; through a runoff election between the first-place candidates, etc.). Since the issue arises in the same way for all the voting methods we discuss, we will not belabor it in what follows.

Plurality voting gives each voter a very limited way of expressing his preferences. Various other rules are more generous in this regard. Consider *cumulative voting*.

cumulative voting

Definition 9.3.2 (Cumulative voting) *Each voter is given k votes, which can be cast arbitrarily (e.g., several votes could be cast for one candidate, with the remainder of the votes being distributed across other candidates). The candidate with the most votes is selected.*

approval voting

Approval voting is similar.

Definition 9.3.3 (Approval voting) *Each voter can cast a single vote for as many of the candidates as he wishes; the candidate with the most votes is selected.*

We have presented cumulative voting and approval voting to give a sense of the range of voting methods. We will defer discussion of such rules to Section 9.5, however, since in the (nonstrategic) voting setting as we have defined it so far, it is not clear how agents should choose when to vote for more than one candidate. Furthermore, although it is more expressive than plurality, approval voting still fails to allow voters to express their full preference orderings. Voting methods that do so are called *ranking voting* methods. Among them, one of the best known is *plurality with elimination*; for example, this method is used for some political elections. When preference orderings are elicited from agents before any elimination has occurred, the method is also known as *instant runoff*.

ranking voting

plurality voting with elimination

Definition 9.3.4 (Plurality with elimination) *Each voter casts a single vote for their most-preferred candidate. The candidate with the fewest votes is eliminated. Each voter who cast a vote for the eliminated candidate casts a new vote for the candidate he most prefers among the candidates that have not been eliminated. This process is repeated until only one candidate remains.*

Borda voting

Another method which has been widely studied is *Borda voting*.

Definition 9.3.5 (Borda voting) *Each voter submits a full ordering on the candidates. This ordering contributes points to each candidate; if there are n candidates, it contributes $n-1$ points to the highest ranked candidate, $n-2$ points to the second highest, and so on; it contributes no points to the lowest ranked candidate. The winners are those whose total sum of points from all the voters is maximal.*

Nanson's method

Nanson's method is a variant of Borda that eliminates the candidate with the lowest Borda score, recomputes the remaining candidates' scores, and repeats. This method has the property that it always chooses a member of the Condorcet set if it is nonempty, and otherwise chooses a member of the Smith set.

pairwise elimination

Finally, there is *pairwise elimination*.

Definition 9.3.6 (Pairwise elimination) *In advance, voters are given a schedule for the order in which pairs of candidates will be compared. Given two candidates (and based on each voter's preference ordering) determine the candidate that each voter prefers. The candidate who is preferred by a minority of voters is eliminated, and the next pair of noneliminated candidates in the schedule is considered. Continue until only one candidate remains.*

9.3.2 *Voting paradoxes*

At this point it is reasonable to wonder why so many voting schemes have been invented. What are their strengths and weaknesses? For that matter, is there one voting method that is appropriate for all circumstances? We will give a more formal (and more general) answer to the latter question in Section 9.4. First, however, we will consider the first question by considering some sets of preferences for which our voting methods exhibit undesirable behavior. Our aim is not to point out every problem that exists with every voting method defined above; rather, it is to illustrate the fact that voting schemes that seem reasonable can often fail in surprising ways.

Condorcet condition

Let us start by revisiting the Condorcet condition. Earlier, we saw two examples: one in which plurality voting chose the Condorcet winner, and another in which a Condorcet winner did not exist. Now consider a situation in which there are 1,000 agents with three different sorts of preferences.

$$\begin{aligned} 499 \text{ agents:} \quad & a \succ b \succ c \\ 3 \text{ agents:} \quad & b \succ c \succ a \\ 498 \text{ agents:} \quad & c \succ b \succ a \end{aligned}$$

Observe that 501 people out of 1,000 prefer b to a, and 502 prefer b to c; this makes b the Condorcet winner. However, many of our voting methods would fail to select b as the winner. Plurality would pick a, as a has the largest number of first-place votes. Plurality with elimination would first eliminate b and would subsequently pick c as the winner. In this example Borda does select b, but there are other cases where it fails to select the Condorcet winner—can you construct one?

Sensitivity to a losing candidate

Consider the following preferences by 100 agents.

$$\begin{aligned} 35 \text{ agents:} \quad & a \succ c \succ b \\ 33 \text{ agents:} \quad & b \succ a \succ c \\ 32 \text{ agents:} \quad & c \succ b \succ a \end{aligned}$$

Plurality would pick candidate a as the winner, as would Borda. (To confirm the latter claim, observe that Borda assigns a, b, and c the scores 103, 98, and 99 respectively.) However, if the candidate c did not exist, then plurality would pick b, as would Borda. (With only two candidates, Borda is equivalent to plurality.) A third candidate who stands no chance of being selected can thus act as a "spoiler," changing the selected outcome.

Another example demonstrates that the inclusion of a least-preferred candidate can even cause the Borda method to *reverse* its ordering on the other candidates.

$$\begin{aligned} 3 \text{ agents:} \quad & a \succ b \succ c \succ d \\ 2 \text{ agents:} \quad & b \succ c \succ d \succ a \\ 2 \text{ agents:} \quad & c \succ d \succ a \succ b \end{aligned}$$

Given these preferences, the Borda method ranks the candidates $c \succ b \succ a \succ d$, with scores of 13, 12, 11, and 6 respectively. If the lowest-ranked candidate d is dropped, however, the Borda ranking is $a \succ b \succ c$ with scores of 8, 7, and 6.

Sensitivity to the agenda setter

Finally, we examine the pairwise elimination method, and consider the influence that the agenda setter can have on the selected outcome. Consider the following preferences, which we discussed previously.

$$\begin{aligned} 35 \text{ agents:} &\quad a \succ c \succ b \\ 33 \text{ agents:} &\quad b \succ a \succ c \\ 32 \text{ agents:} &\quad c \succ b \succ a \end{aligned}$$

First, consider the order a, b, c. a is eliminated in the pairing between a and b; then c is chosen in the pairing between b and c. Second, consider the order a, c, b. a is chosen in the pairing between a and c; then b is chosen in the pairing between a and b. Finally, under the order b, c, a, we first eliminate b and ultimately choose a. Thus, given these preferences, the agenda setter can select whichever outcome he wants by selecting the appropriate elimination order!

Next, consider the following preferences.

$$\begin{aligned} 1 \text{ agent:} &\quad b \succ d \succ c \succ a \\ 1 \text{ agent:} &\quad a \succ b \succ d \succ c \\ 1 \text{ agent:} &\quad c \succ a \succ b \succ d \end{aligned}$$

Consider the elimination ordering a, b, c, d. In the pairing between a and b, a is preferred; c is preferred to a and then d is preferred to c, leaving d as the winner. However, *all* of the agents prefer b to d—the selected candidate is Pareto dominated!

Last, we give an example showing that Borda is fundamentally different from pairwise elimination, *regardless* of the elimination ordering. Consider the following preferences.

$$\begin{aligned} 3 \text{ agents:} &\quad a \succ b \succ c \\ 2 \text{ agents:} &\quad b \succ c \succ a \\ 1 \text{ agent:} &\quad b \succ a \succ c \\ 1 \text{ agent:} &\quad c \succ a \succ b \end{aligned}$$

Regardless of the elimination ordering, pairwise elimination will select the candidate a. The Borda method, on the other hand, selects candidate b.

9.4 Existence of social functions

The previous section has illustrated several senses in which some popular voting methods exhibit undesirable or unfair behavior. In this section, we consider this state of affairs from a more formal perspective, examining both social welfare functions and social choice functions.

In this section only, we introduce an additional assumption to simplify the exposition. Specifically, we will assume that all agents' preferences are *strict* total orderings on the outcomes, rather than nonstrict total orders; denote the set of such orders as L, and denote an agent i's preference ordering as $\succ_i \in L$. Denote a preference profile (a tuple giving a preference ordering for each agent) as $[\succ'] \in L^n$, and denote agent i's preferences from preference profile $[\succ']$ as $\succ'_i$. We also redefine social welfare functions to return a strict total ordering over the outcomes, $W : L^n \mapsto L$. In other words, we assume that no agent is ever indifferent between outcomes and that the social welfare function is similarly decisive. We stress that this assumption is *not required* for the results that follow; analysis of the general case can be found in the works cited at the end of the chapter.[3]

Finally, let us introduce some new notation. Social welfare functions take preference profiles as input; denote the preference ordering selected by the social welfare function W, given preference profile $[\succ'] \in L^n$, as $\succ_{W([\succ'])}$. When the input ordering $[\succ']$ is understood from context, we abbreviate our notation for the social ordering as $\succ_W$.

9.4.1 *Social welfare functions*

Arrow's impossibility theorem

In this section we examine *Arrow's impossibility theorem*, without a doubt the most influential result in social choice theory. Its surprising conclusion is that fairness is multifaceted and that it is *impossible* to achieve all of these kinds of fairness simultaneously.

Now, let us review these multifaceted notions of fairness.

Pareto efficiency (PE)

Definition 9.4.1 (Pareto efficiency (PE)) *W is* Pareto efficient *if for any $o_1, o_2 \in O$, $\forall i\ o_1 \succ_i o_2$ implies that $o_1 \succ_W o_2$.*

In words, PE means that when all agents agree on the ordering of two outcomes, the social welfare function must select that ordering. Observe that this definition is effectively the same as *strict Pareto efficiency* as defined in Definition 3.3.2.[4]

independence of irrelevant alternatives (IIA)

Definition 9.4.2 (Independence of irrelevant alternatives (IIA)) *W is* independent of irrelevant alternatives *if, for any $o_1, o_2 \in O$ and any two preference profiles $[\succ'], [\succ''] \in L^n$, $\forall i\ (o_1 \succ'_i o_2$ if and only if $o_1 \succ''_i o_2)$ implies that $(o_1 \succ_{W([\succ'])} o_2$ if and only if $o_1 \succ_{W([\succ''])} o_2)$.*

That is, the selected ordering between two outcomes should depend only on the relative orderings they are given by the agents.

3. Intuitively, because we will be looking for social functions that work given *any* preferences the agents might have, when we show that desirable social welfare and social choice functions cannot exist even when agents are assumed to have strict preferences, we will also have shown that the claim holds when we relax this restriction.

4. One subtle difference does arise from our assumption in this section that all preferences are strict.

nondictatorship **Definition 9.4.3 (Nondictatorship)** *W does not have a* dictator *if* $\neg\exists i\, \forall o_1, o_2 (o_1 \succ_i o_2 \Rightarrow o_1 \succ_W o_2)$.

Nondictatorship means that there does not exist a single agent whose preferences always determine the social ordering. We say that W is *dictatorial* if it fails to satisfy this property.

Surprisingly, it turns out that there exists no social welfare function W that satisfies these three properties for all of its possible inputs. This result relies on our previous assumption that N is finite.

Theorem 9.4.4 (Arrow, 1951) *If $|O| \geq 3$, any social welfare function W that is Pareto efficient and independent of irrelevant alternatives is dictatorial.*

Proof. We will assume that W is both PE and IIA and show that W must be dictatorial. The argument proceeds in four steps.

Step 1: *If every voter puts an outcome b at either the very top or the very bottom of his preference list, b must be at either the very top or very bottom of $\succ_W$ as well.*

Consider an arbitrary preference profile $[\succ]$ in which every voter ranks some $b \in O$ at either the very bottom or very top, and assume for contradiction that the preceding claim is not true. Then, there must exist some pair of distinct outcomes $a, c \in O$ for which $a \succ_W b$ and $b \succ_W c$.

Now let us modify $[\succ]$ so that every voter moves c just above a in his preference ranking, and otherwise leaves the ranking unchanged; let us call this new preference profile $[\succ']$. We know from IIA that for $a \succ_W b$ or $b \succ_W c$ to change, the pairwise relationship between a and b and/or the pairwise relationship between b and c would have to change. However, since b occupies an extremal position for all voters, c can be moved above a without changing either of these pairwise relationships. Thus in profile $[\succ']$ it is also the case that $a \succ_W b$ and $b \succ_W c$. From this fact and from transitivity, we have that $a \succ_W c$. However, in $[\succ']$, every voter ranks c above a and so PE requires that $c \succ_W a$. We have a contradiction.

Step 2: *There is some voter n^* who is* extremely pivotal *in the sense that by changing his vote at some profile, he can move a given outcome b from the bottom of the social ranking to the top.*

Consider a preference profile $[\succ]$ in which every voter ranks b last, and in which preferences are otherwise arbitrary. By PE, W must also rank b last. Now let voters from 1 to n successively modify $[\succ]$ by moving b from the bottom of their rankings to the top, preserving all other relative rankings. Denote as n^* the first voter whose change causes the social ranking of b to change. There must clearly be some such voter: when the voter n moves b to the top of his ranking, PE will require that b be ranked at the top of the social ranking.

Denote by $[\succ^1]$ the preference profile just before n^* moves b, and denote by $[\succ^2]$ the preference profile just after n^* has moved b to the top of his ranking. (These preference profiles are illustrated in Figures 9.1a and 9.1b, with the indicated positions of outcomes a and c in each agent's ranking

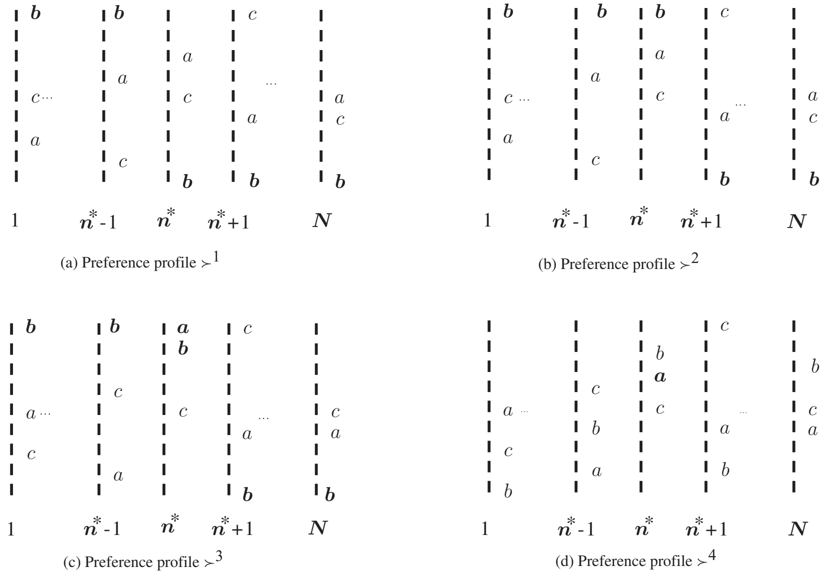

Figure 9.1 The four preference profiles used in the proof of Arrow's theorem. A higher position along the dotted line indicates a higher position in an agent's preference ordering. The outcomes indicated in bold (i.e., b in profiles $[\succ^1]$, $[\succ^2]$, and $[\succ^3]$ and a for voter n^* in profiles $[\succ^3]$ and $[\succ^4]$) must be in the exact positions shown. (In profile $[\succ^4]$, a must simply be ranked above c.) The outcomes not indicated in bold are simply examples and can occur in any relative ordering that is consistent with the placement of the bold outcomes.

serving only as examples.) In $[\succ^1]$, b is at the bottom in $\succ_W$. In $[\succ^2]$, b has changed its position in $\succ_W$, and every voter ranks b at either the top or the bottom. By the argument from Step 1, in $[\succ^2]$ b must be ranked at the top of $\succ_W$.

Step 3: *n^* (the agent who is extremely pivotal on outcome b) is a dictator over any pair ac not involving b.*

We begin by choosing one element from the pair ac; without loss of generality, let us choose a. We will construct a new preference profile $[\succ^3]$ from $[\succ^2]$ by making two changes. (Profile $[\succ^3]$ is illustrated in Figure 9.1c.) First, we move a to the top of n^*'s preference ordering, leaving it otherwise unchanged; thus $a \succ_{n^*} b \succ_{n^*} c$. Second, we arbitrarily rearrange the relative rankings of a and c for all voters other than n^*, while leaving b in its extremal position.

In $[\succ^1]$ we had $a \succ_W b$, as b was at the very bottom of $\succ_W$. When we compare $[\succ^1]$ to $[\succ^3]$, relative rankings between a and b are the same for all voters. Thus, by IIA, we must have $a \succ_W b$ in $[\succ^3]$ as well. In $[\succ^2]$ we had $b \succ_W c$, as b was at the very top of $\succ_W$. Relative rankings between b and c are the same in $[\succ^2]$ and $[\succ^3]$. Thus in $[\succ^3]$, $b \succ_W c$. Using the two aforementioned facts about $[\succ^3]$ and transitivity, we can conclude that $a \succ_W c$ in $[\succ^3]$.

Now construct one more preference profile, $[\succ^4]$, by changing $[\succ^3]$ in two ways. First, arbitrarily change the position of b in each voter's ordering while keeping all other relative preferences the same. Second, move a to an arbitrary position in n^*'s preference ordering, with the constraint that a remains ranked higher than c. (Profile $[\succ^4]$ is illustrated in Figure 9.1d.) Observe that all voters other than n^* have entirely arbitrary preferences in $[\succ^4]$, while n^*'s preferences are arbitrary except that $a \succ_{n^*} c$. In $[\succ^3]$ and $[\succ^4]$, all agents have the same relative preferences between a and c; thus, since $a \succ_W c$ in $[\succ^3]$ and by IIA, $a \succ_W c$ in $[\succ^4]$. Thus we have determined the social preference between a and c without assuming anything except that $a \succ_{n^*} c$.

Step 4: *n^* is a dictator over all pairs ab.*

Consider some third outcome c. By the argument in Step 2, there is a voter n^{**} who is extremely pivotal for c. By the argument in Step 3, n^{**} is a dictator over any pair $\alpha\beta$ not involving c. Of course, ab is such a pair $\alpha\beta$. We have already observed that n^* is able to affect W's ab ranking—for example, when n^* was able to change $a \succ_W b$ in profile $[\succ^1]$ into $b \succ_W a$ in profile $[\succ^2]$. Hence, n^{**} and n^* must be the same agent.

We have now shown that n^* is a dictator over all pairs of outcomes. ■

9.4.2 *Social choice functions*

Arrow's theorem tells us that we cannot hope to find a voting scheme that satisfies all of the notions of fairness that we find desirable. However, maybe the problem is that Arrow's theorem considers too hard a problem—the identification of a social ordering over *all* outcomes. We now consider the setting of social choice functions, which are required only to identify a single top-ranked outcome. First, we define concepts analogous to Pareto efficiency, independence of irrelevant alternatives and nondictatorship for social choice functions.

weak Pareto efficiency

Definition 9.4.5 (Weak Pareto efficiency) *A social choice function C is* weakly Pareto efficient *if, for any preference profile $[\succ] \in L^n$, if there exist a pair of outcomes o_1 and o_2 such that $\forall i \in N$, $o_1 \succ_i o_2$, then $C([\succ]) \neq o_2$.*

This definition prohibits the social choice function from selecting any outcome that is dominated by another alternative for all agents. (That is, if all agents prefer o_1 to o_2, the social choice rule does not have to choose o_1, but it cannot choose o_2.) The definition implies that the social choice rule must respect agents' unanimous choices: if outcome o is the top choice according to each $\succ_i$, then we must have $C([\succ]) = o$. Thus, the definition is less demanding than *strict Pareto efficiency* as defined in Definition 3.3.2—a strictly Pareto efficient choice rule would also always satisfy weak Pareto efficiency, but the reverse is not true.

monotonicity

Definition 9.4.6 (Monotonicity) *C is* monotonic *if, for any $o \in O$ and any preference profile $[\succ] \in L^n$ with $C([\succ]) = o$, then for any other preference profile*

$[\succ']$ *with the property that* $\forall i \in N, \forall o' \in O$, $o \succ'_i o'$ *if* $o \succ_i o'$, *it must be that* $C([\succ']) = o$.

Monotonicity says that when a social choice rule C selects the outcome o for a preference profile $[\succ]$, then for any second preference profile $[\succ']$ in which, for every agent i, the set of outcomes to which o is preferred under $\succ'_i$ is a weak superset of the set of outcomes to which o is preferred under $\succ_i$, the social choice rule must also choose outcome o. Intuitively, monotonicity means that an outcome o must remain the winner whenever the support for it is increased relative to a preference profile under which o was already winning. Observe that the definition imposes no constraint on the relative orderings of outcomes $o_1, o_2 \neq o$ under the two preference profiles; for example, some or all of these relative orderings could be different.

nondictatorship

Definition 9.4.7 (Nondictatorship) *C is* nondictatorial *if there does not exist an agent j such that C always selects the top choice in j's preference ordering.*

Following the pattern we followed for social welfare functions, we can show that no social choice function can satisfy all three of these properties.

Theorem 9.4.8 (Muller–Satterthwaite, 1977) *If $|O| \geq 3$, any social choice function C that is weakly Pareto efficient and monotonic is dictatorial.*

Before giving the proof, we must provide a key definition.

Definition 9.4.9 (Taking O' to the top from $[\succ]$) *Let $O' \subset O$ be a finite subset of the outcomes O, and let $[\succ]$ be a preference profile. Denote the set $O \setminus O'$ as $\overline{O'}$. A second preference profile $[\succ']$* takes O' to the top from $[\succ]$ *if, for all $i \in N$, $o' \succ'_i o$ for all $o' \in O'$ and $o \in \overline{O'}$ and $o'_1 \succ'_i o'_2$ if and only if $o'_1 \succ_i o'_2$.*

That is, $[\succ']$ takes O' to the top from $[\succ]$ when, under $[\succ']$:

- each outcome from O' is preferred by every agent to each outcome from $\overline{O'}$; and
- the relative preferences between pairs of outcomes in O' for every agent are the same as the corresponding relative preferences under $[\succ]$.

Observe that the relative preferences between pairs of outcomes in $\overline{O'}$ are arbitrary: they are not required to bear any relation to the corresponding relative preferences under $[\succ]$.

We can now state the proof. Intuitively, it works by constructing a social welfare function W from the given social choice function C. We show that the facts that C is weakly Pareto efficient and monotonic imply that W must satisfy PE and IIA, allowing us to apply Arrow's theorem.

Proof. We will assume that C satisfies weak Pareto efficiency and monotonicity, and show that it must be dictatorial. The proof proceeds in six steps.

Step 1: *If both* $[\succ']$ *and* $[\succ'']$ *take* $O' \subset O$ *to the top from* $[\succ]$*, then* $C([\succ']) = C([\succ''])$ *and* $C([\succ']) \in O'$.

Under $[\succ']$, for all $i \in N$, $o' \succ'_i o$ for all $o' \in O'$ and all $o \in \overline{O}$. Thus, by weak Pareto efficiency $C([\succ']) \in O'$. For every $i \in N$, every $o' \in O'$ and every $o \neq o' \in O$, $o' \succ'_i o$ if and only if $o' \succ''_i o$. Thus by monotonicity, $C([\succ']) = C([\succ''])$.

Step 2: *We define a social welfare function* W *from* C.

For every pair of outcomes $o_1, o_2 \in O$, construct a preference profile $[\succ^{\{o_1,o_2\}}]$ by taking $\{o_1, o_2\}$ to the top from $[\succ]$. By Step 1, $C([\succ^{\{o_1,o_2\}}])$ will be either o_1 or o_2, and will always be the same regardless of how we choose $[\succ^{\{o_1,o_2\}}]$. Now we will construct a social welfare function W from C. For each pair of outcomes $o_1, o_2 \in O$, let $o_1 \succ_W o_2$ if and only if $C([\succ^{\{o_1,o_2\}}]) = o_1$.

In order to show that W is a social welfare function, we must demonstrate that it establishes a total ordering over the outcomes. Since W is complete, it only remains to show that W is transitive. Suppose that $o_1 \succ_W o_2$ and $o_2 \succ_W o_3$; we must thus show that $o_1 \succ_W o_3$. Let $[\succ']$ be a preference profile that takes $\{o_1, o_2, o_3\}$ to the top from $[\succ]$. By Step 1, $C([\succ']) \in \{o_1, o_2, o_3\}$. We consider each possibility.

Assume for contradiction that $C([\succ']) = o_2$. Let $[\succ'']$ be a profile that takes $\{o_1, o_2\}$ to the top from $[\succ']$. By monotonicity, $C([\succ'']) = o_2$ (o_2 has weakly improved its ranking from $[\succ']$ to $[\succ'']$). Observe that $[\succ'']$ also takes $\{o_1, o_2\}$ to the top from $[\succ]$. Thus by our definition of W, $o_2 \succ_W o_1$. But we already had $o_1 \succ_W o_2$. Thus, $C([\succ']) \neq o_2$. By an analogous argument, we can show that $C([\succ']) \neq o_3$.

Thus, $C([\succ']) = o_1$. Let $[\succ'']$ be a preference profile that takes $\{o_1, o_3\}$ to the top from $[\succ']$. By monotonicity, $C([\succ'']) = o_1$. Observe that $[\succ'']$ also takes $\{o_1, o_3\}$ to the top from $[\succ]$. Thus by our definition of W, $o_1 \succ_W o_3$, and hence we have shown that W is transitive.

Step 3: *The highest-ranked outcome in* $W([\succ])$ *is always* $C([\succ])$.

We have seen that C can be used to construct a social welfare function W. It turns out that C can also be recovered from W, in the sense that the outcome given the highest ranking by $W([\succ])$ will always be $C([\succ])$. Let $C([\succ]) = o_1$, let $o_2 \in O$ be any other outcome, and let $[\succ']$ be a profile that takes $\{o_1, o_2\}$ to the top from $[\succ]$. By monotonicity, $C([\succ']) = o_1$. By the definition of W, $o_1 \succ_W o_2$. Thus, o_1 is the outcome ranked highest by W.

Step 4: *W is Pareto efficient.*

Imagine that $\forall i \in N$, $o_1 \succ o_2$. Let $[\succ']$ take $\{o_1, o_2\}$ to the top from $[\succ]$. Since C is weakly Pareto efficient, $C([\succ']) = o_1$. Thus by the definition of W from Step 2, $o_1 \succ_W o_2$, and so W is Pareto efficient.

Step 5: *W is independent of irrelevant alternatives.*

Let $[\succ^1]$ and $[\succ^2]$ be two preference profiles with the property that for all $i \in N$ and for some pair of outcomes o_1 and $o_2 \in O$, $o_1 \succ^1_i o_2$ if and only if $o_1 \succ^2_i o_2$. We must show that $o_1 \succ_{W([\succ^1])} o_2$ if and only if $o_1 \succ_{W([\succ^2])} o_2$.

Let $[\succ^{1'}]$ take $\{o_1, o_2\}$ to the top from $[\succ^1]$, and let $[\succ^{2'}]$ take $\{o_1, o_2\}$ to the top from $[\succ^2]$. From the definition of W in Step 2, $o_1 \succ_{W([\succ^1])} o_2$ if and only if $C([\succ^{1'}]) = o_1$; likewise, $o_1 \succ_{W([\succ^2])} o_2$ if and only if $C([\succ^{2'}]) = o_1$. Now observe that $[\succ^{1'}]$ also takes $\{o_1, o_2\}$ to the top from $[\succ^2]$, because for all $i \in N$ the relative ranking between o_1 and o_2 is the same under $[\succ^1]$ and $[\succ^2]$. Thus by Part 1, $C([\succ^{1'}]) = C([\succ^{2'}])$, and hence $o_1 \succ_{W([\succ^1])} o_2$ if and only if $o_1 \succ_{W([\succ^2])} o_2$.

Step 6: *C is dictatorial.*

From Steps 4 and 5 and Theorem 9.4.4, W is dictatorial. That is, there must be some agent $i \in N$ such that, regardless of the preference profile $[\succ']$, we always have $o_1 \succ_{W([\succ'])} o_2$ if and only if $o_1 \succ'_i o_2$. Therefore, the highest-ranked outcome in $W([\succ'])$ must also be the outcome ranked highest by i. By Step 3, $C([\succ'])$ is always the outcome ranked highest in $W([\succ'])$. Thus, C is dictatorial. ■

In effect, this theorem tells us that, perhaps contrary to intuition, social choice functions are no simpler than social welfare functions. Intuitively, the proof shows that we can repeatedly "probe" a social choice function to determine the relative social ordering between given pairs of outcomes. Because the function must be defined for all inputs, we can use this technique to construct a full social welfare ordering.

To get a feel for the theorem, consider the social choice function defined by the plurality rule.[5] Clearly, it satisfies weak Pareto efficiency and is not dictatorial. This means it must be nonmonotonic. To see why, consider the following scenario with seven voters.

$$\begin{array}{ll} \text{3 agents:} & a \succ b \succ c \\ \text{2 agents:} & b \succ c \succ a \\ \text{2 agents:} & c \succ b \succ a \end{array}$$

Denote these preferences as $[\succ^1]$. Under $[\succ^1]$ plurality chooses a. Now consider the situation where the final two agents increase their support for a by moving c to the bottom of their rankings as shown below; denote the new preferences as $[\succ^2]$.

$$\begin{array}{ll} \text{3 agents:} & a \succ b \succ c \\ \text{2 agents:} & b \succ c \succ a \\ \text{2 agents:} & b \succ a \succ c \end{array}$$

If plurality were monotonic, it would have to make the same choice under $[\succ^2]$ as under $[\succ^1]$, because for all $i \in N$, $a \succ^2_i b$ if $a \succ^1_i b$ and $a \succ^2_i c$ if $a \succ^1_i c$. However, under $[\succ^2]$ plurality chooses b. Therefore plurality is not monotonic.

5. Formally, we should also specify the tie-breaking rule used by plurality. However, in our example monotonicity fails even when ties never occur, so the tie-breaking rule does not matter here.

9.5 Ranking systems

We now turn to a specialization of the social choice problem that has a computational flavor, and in which some interesting progress can be made. Specifically, consider a setting in which the set of agents is *the same* as the set of outcomes—agents are asked to vote to express their opinions about each other, with the goal of determining a social ranking. Such settings have great practical importance. For example, search engines rank Web pages by considering hyperlinks from one page to another to be votes about the importance of the destination pages. Similarly, online auction sites employ *reputation systems* to provide assessments of agents' trustworthiness based on ratings from past transactions.

Let us formalize this setting, returning to our earlier assumption that agents can be indifferent between outcomes. Our setting is characterized by two assumptions. First, $N = O$: the set of agents is the same as the set of outcomes. Second, agents' preferences are such that each agent divides the other agents into a set that he likes equally, and a set that he dislikes equally (or, equivalently, has no opinion about). Formally, for each $i \in N$ the outcome set O (equivalent to N) is partitioned into two sets $O_{i,1}$ and $O_{i,2}$, with $\forall o_1 \in O_{i,1}, \forall o_2 \in O_{i,2}, o_1 \succ_i o_2$, and with $\forall o, o' \in O_{i,k}$ for $k \in \{1, 2\}$, $o \sim_i o'$. We call this the *ranking systems setting*, and call a social welfare function in this setting a *ranking rule*. Observe that a ranking rule is not required to partition the agents into two sets; it must simply return some total preordering on the agents.

ranking systems setting

ranking rule

Interestingly, Arrow's impossibility system does not hold in the ranking systems setting. The easiest way to see this is to identify a ranking rule that satisfies all of Arrow's axioms.[6]

Proposition 9.5.1 *In the ranking systems setting, approval voting satisfies IIA, PE, and nondictatorship.*

The proof is straightforward and is left as an easy exercise. Intuitively, the fact that agents partition the outcomes into only two sets is crucially important. We would be able to apply Arrow's argument if agents were able to partition the outcomes into as few as three sets. (Recall that the proof of Theorem 9.4.4 requires arbitrary preferences and $|O| \geq 3$.)

Although the possibility of circumventing Arrow's theorem is encouraging, the discussion does not end here. Due to the special nature of the ranking systems setting, there are other properties that we would like a ranking rule to satisfy.

First, consider an example in which Alice votes only for Bob, Will votes only for Liam, and Liam votes only for Vic. These votes are illustrated in Figure 9.2. Who should be ranked highest? Three of the five kids have received votes (Bob, Liam, and Vic); these three should presumably rank higher than the remaining two. But of the three, Vic is special: he is the only one whose voter (Liam) himself received a vote. Thus, intuitively, Vic should receive the highest rank. This intuition is captured by the idea of transitivity.

6. Note that we defined these axioms in terms of strict total orderings; nevertheless, they generalize easily to total preorderings.

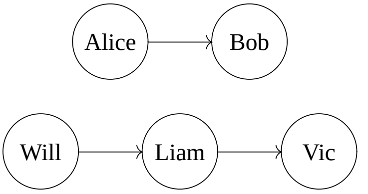

Figure 9.2 Sample preferences in a ranking system, where arrows indicate votes.

First we define *strong transitivity*. We will subsequently relax this definition; however, it is useful for what follows.

Definition 9.5.2 (Strong transitivity) *Consider a preference profile in which outcome o_2 receives at least as many votes as o_1, and it is possible to pair up all the voters for o_1 with voters from o_2 so that each voter for o_2 is weakly preferred by the ranking rule to the corresponding voter for o_1.*[7] *Further assume that o_2 receives more votes than o_1 and/or that there is at least one pair of voters where the ranking rule strictly prefers the voter for o_2 to the voter for o_1. Then the ranking rule satisfies* strong transitivity *if it always strictly prefers o_2 to o_1.*

strong transitivity

Because our transitivity definition will serve as the basis for an impossibility result, we want it to be as weak as possible. One way in which this definition is quite strong is that it does not take into account the *number* of votes that a voting agent places. Consider an example in which Vic votes for almost all the kids, whereas Ray votes only for one. If Vic and Ray are ranked the same by the ranking rule, strong transitivity requires that their votes must count equally. However, we might feel that Ray has been more decisive, and therefore feel that his vote should be counted more strongly than Vic's. We can allow for such rules by weakening the notion of transitivity. The new definition is exactly the same as the old one, except that it is restricted to apply only to settings in which the voters vouch for exactly the same number of candidates.

Definition 9.5.3 (Weak transitivity) *Consider a preference profile in which outcome o_2 receives at least as many votes as o_1, and it is possible to pair up all the voters for o_1 with voters for o_2* who have both voted for exactly the same number of outcomes *so that each voter for o_2 is weakly preferred by the ranking rule to the corresponding voter for o_1. Further assume that o_2 receives more votes than o_1 and/or that there is at least one pair of voters where the ranking rule strictly prefers the voter for o_2 to the voter for o_1. Then the ranking rule satisfies* weak transitivity *if it always strictly prefers o_2 to o_1.*

weak transitivity

Recall the independence of irrelevant alternatives (IIA) property defined earlier in Definition 9.4.2, which said that the ordering of two outcomes should depend only on agents' relative preferences between these outcomes. Such an assumption is inconsistent with even our weak transitivity definitions. However, we can

7. The pairing must use each voter from o_2 at most once; if there are more votes for o_2 than for o_1, there will be agents who voted for o_2 who are not paired. If an agent voted for both o_1 and o_2, it is acceptable for him to be paired with himself.

broaden the scope of IIA to allow for transitive effects, and thereby still express the idea that the ranking rule should rank pairs of outcomes based only on local information.

ranked independence of irrelevant alternatives (RIIA)

Definition 9.5.4 (RIIA, informal) *A ranking rule satisfies* ranked independence of irrelevant alternatives *(RIIA) if the relative rank between pairs of outcomes is always determined according to the same rule, and this rule depends only on:*

1. *the number of votes each outcome received; and*
2. *the relative ranks of these voters.*[8]

Note that this definition prohibits the ranking rule from caring about the *identities* of the voters, which is allowed by IIA.

Despite the fact that Arrow's theorem does not apply in this setting, it turns out that another, very different impossibility result does hold.

Theorem 9.5.5 *There is no ranking system that always satisfies both weak transitivity and RIIA.*

What hope is there then for ranking systems? The obvious way forward is to consider relaxing one axiom and keeping the other. Indeed, progress can be made both by relaxing weak transitivity and by relaxing RIIA. For example, the famous PageRank algorithm (used originally as the basis of the Google search engine) can be understood as a ranking system that satisfies weak transitivity but not RIIA. Unfortunately, an axiomatic treatment of this algorithm is quite involved, so we do not provide it here.

Instead, we will consider relaxations of transitivity. First, what happens if we simply drop the weak transitivity requirement altogether? Let us add the requirements that an agent's rank can improve only when he receives more votes ("positive response") and that the agents' identities are ignored by the ranking function ("anonymity"). Then it can be shown that approval voting, which we have already considered in this setting, is the *only* possible ranking function.

Theorem 9.5.6 *Approval voting is the only ranking rule that satisfies RIIA, positive response, and anonymity.*

Finally, what if we try to modify the transitivity requirement rather than dropping it entirely? It turns out that we can also obtain a positive result here, although this comes at the expense of guaranteeing anonymity. Note that this new transitivity requirement is a different weakening of strong transitivity which does not care about the number of outcomes that agents vote for, but instead requires strict preference only when the ranking rule strictly prefers *every* paired voter for o_2 over the corresponding voter for o_1.

Definition 9.5.7 (Strong quasi-transitivity) *Consider a preference profile in which outcome o_2 receives at least as many votes as o_1, and it is possible to*

8. The formal definition of RIIA is more complicated than Definition 9.5.4 because it must explain precisely what is meant by depending on the relative ranks of the voters. The interested reader is invited to consult the reference cited at the end of the chapter.

forall $i \in N$ **do** $rank(i) \leftarrow 0$
repeat
  forall $i \in N$ **do**
    if $|voters\_for(i)| > 0$ **then**
      $rank(i) \leftarrow \frac{1}{n+1}[|voters\_for(i)| + \max_{j \in voters\_for(i)} rank(j)]$
    else
      $rank(i) \leftarrow 0$
until $rank$ *converges*

Figure 9.3 A ranking algorithm that satisfies strong quasi-transitivity and RIIA.

strong quasi-transitivity

pair up all the voters for o_1 with voters from o_2 so that each voter for o_2 is weakly preferred by the ranking rule to the corresponding voter for o_1. Then the ranking rule satisfies strong quasi-transitivity *if it weakly prefers o_2 to o_1, and strictly prefers o_2 to o_1 if either o_1 received no votes or each paired voter for o_2 is strictly preferred by the ranking rule to the corresponding voter for o_1.*

There exists a family of ranking algorithms that satisfy strong quasi-transitivity and RIIA. These algorithms work by assigning agents numerical ranks that depend on the number of votes they have received, and breaking ties in favor of the agent who received a vote from the highest-ranked voter. If this rule still yields a tie, it is applied recursively; when the recursion follows a cycle, the rank is a periodic rational number with period equal to the length of the cycle. One such algorithm is given in Figure 9.3. This algorithm can be proved to converge in n iterations; as each step takes $O(n^2)$ time (considering all votes for all agents), the worst-case complexity[9] of the algorithm is $O(n^3)$.

9.6 History and references

Social choice theory is covered in many textbooks on microeconomics and game theory, as well as in some specialized texts such as Feldman and Serrano [2006] and Gaertner [2006]. An excellent survey is provided in Moulin [1994].

The seminal individual publication in this area is Arrow [1970], which still remains among the best introductions to the field. The book includes Arrow's famous impossibility result (partly for which he received a 1972 Nobel Prize), though our treatment follows the elegant first proof in Geanakoplos [2005]. Plurality voting is too common and natural (it is used in 43 of the 191 countries in the United Nations for either local or national elections) to have clear origins. Borda invented his system as a fair way to elect members to the French Academy of Sciences in 1770, and first published his method in 1781 as de Borda [1781]. In 1784, Marie Jean Antoine Nicolas Caritat, aka the Marquis de Condorcet, first published his ideas regarding voting [de Condorcet, 1784]. Somewhat later,

9. In fact, the complexity bound on this algorithm can be somewhat improved by more careful analysis; however, the argument here suffices to show that the algorithm runs in polynomial time.

Nanson, a Briton-turned-Australian mathematician and election reformer, published his modification of the Borda count in Nanson [1882]. The Smith set was introduced in Smith [1973]. The Muller–Satterthwaite impossibility result appears in Muller and Satterthwaite [1977]; our proof follows Mas-Colell et al. [1995]. Our section on ranking systems follows Altman and Tennenholtz [2008]. Other interesting directions in ranking systems include developing practical ranking rules and/or axiomatizing such rules (e.g., Page et al. [1998], Kleinberg [1999], Borodin et al. [2005], and Altman and Tennenholtz [2005]), and exploring personalized rankings, in which the ranking function gives a potentially different answer to each agent (e.g., Altman and Tennenholtz [2007a]).