Voting, Arbitration, and Fair Division

The mathematics of social choice

Marcus Pivato
Trent University

April 25, 2004

Copyright

Colophon

All text was prepared using Leslie Lamport's LaTeX2e typesetting language. Pictures were generated using William Chia-Wei Cheng's excellent `TGIF` object-oriented drawing program. This book was prepared entirely on the `RedHat Linux` operating system.

Contents

Preface

Social choice theory is the mathematical study of procedures for finding consensus in a society where people have conflicting goals. These procedures aspire to be egalitarian and consistent, and to maximize the 'collective welfare' of the population. By representing social choice problems mathematically, we can find solutions, or at least determine if they exist. Problems of social choice have traditionally been considered by economists and political scientists, so some parts of social choice theory are reminiscent of mathematical economics, while other parts could properly be called 'mathematical political science'.

Prerequisites: This book has no specific mathematical prerequisites. It does not require knowledge of calculus or linear algebra, and only uses high-school level algebra. Thus, most of the text should be accessible to any intelligent student of economics or politics. However, the book *does* demand considerable 'mathematical maturity', which means an ability to understand abstract definitions and abstract logical arguments, and to follow a mathematical proof from premises to conclusion. Thus, it would be better to have some previous experience with these things, eg. by exposure to a higher level mathematics course in geometry or number theory[1].

The chapters in this book are logically independent, except when otherwise indicated by a list of 'Prerequisites' at the beginning of a chapter or section. Thus, the interested student may skip right to Chapter §5 (*Arbitration & Bargaining*) or Chapter §6 (*Fair Divisions*) without reading the previous chapters. Likewise, a reader may pick out those sections or subsections which interest her, as long as she reads the relevant prerequisites first.

A very brief history: The mathematical theory of social choice began in France and England, near the end of the 18th century. In France, shortly before the Revolution, mathematicians Jean-Antoine-Nicolas Caritat (Marquis de Condorcet) and Jean-Charles Borda proposed several voting systems and analyzed their shortcomings. In England, philosopher and social theorist Jeremy Bentham formulated a quasi-mathematical political philosophy now called *Utilitarianism*, while Adam Smith published the seminal *Wealth of Nations*, which laid the foundations for modern microeconomics.

As economics developed through the 19th century, one of its chief concerns was the maximization of 'social good'. The Italian economist Vilfredo Pareto provided a precise conceptual framework for studying this problem, a framework which is broadly applicable to problems of social choice (even noneconomic ones). A variety of voting systems were contemplated (eg. by Charles Dodgson a.k.a. Lewis Caroll), but there were no major theoretical results.

Starting in the 1920s, mathematicians such as John von Neumann developed *game theory*, a rigorous mathematical analysis of games of strategy and chance, which was broad enough that it could potentially be applied to 'social games' such as economic or political interactions. In the mid 1940s, von Neumann and economist Oskar Morgenstern published *Theory of Games*

[1] Although, of course, the material in this book does not require any specific knowledge of either geometry or number theory.

and Economic Behaviour, which proposed a game-theory approach to economics. In 1950, John Nash applied these ideas to propose an abstract mathematical model of bargaining or arbitration, and to propose a mathematical definition of the 'fair' solution to such a bargaining problem. Meanwhile, the theory of fair divisions began in the late 1940s with the work of Polish mathematicians Bronislaw Knaster, Stefan Banach, and Hugo Steinhaus.

The first major advances in voting theory were in the early 1950s, with the work of Duncan Black and Kenneth May, and of course the famous *Impossibility Theorem* of economist Kenneth Arrow. Since then, voting theory has received important contributions from Stephen J. Brams, Allan Gibbard, Donald G. Saari, Mark Satterthwaite, Amartya K. Sen, Alan D. Taylor, William S. Zwicker, and many others. Meanwhile a variety of fair division schemes were proposed and studied by A.K. Austin, Anatole Beck, Stephen J. Brams, John Horton Conway, Elizabeth Early Cook, Lester E. Dubins, A.M. Fink, Theodore P. Hill, Saul X. Levmore, John L. Selfridge, Edwin H. Spanier, Walter Stromquist, Alan D. Taylor, William S. Zwicker, and many others. Both are still areas of active research.

Chapter 1

Democracy is more complicated than you think

Most people agree that the best form of government is democracy. A *democracy* is a governments where laws are decided by the Will of the People. The 'Will of the People' is usually determined through an election or referendum. However, all of us have seen how the electoral process can fail to adequately measure the Will of the People. A classic example is 'vote splitting'. For instance, suppose that in the hypothetical country of Danaca, the political breakdown of the population is roughly as follows:

Left	15%
Centre	34%
Right	51%

One would expect that Danacians would normally elect a right-wing government. However, suppose that more than three political parties compete in the election. For simplicity, we will assume that each political party falls into neatly one of the three categories 'Left', 'Centre', and 'Right'

Left	*Ultracommunists*	1%	
	New Demagogues	14%	
	Total:		15%
Centre	***Literal Party***		**34%**
Right	*Regressive Coercitives*	26%	
	~~*Deformed*~~ *Behaviour & Compliance*	25%	
	Total:		51%

Clearly, the centrist *Literal* party will win the election, despite the fact that an absolute majority (51%) of Danacians would prefer a right-wing government. The *Literal* party does not have an absolute majority; it only has a *plurality*, ie. the largest (minority) vote of any party. The standard multiparty electoral system is called the *Plurality System*, and the Plurality System has clearly somehow failed in this scenario.

The problem here is that the two right-wing parties have 'split' the right-wing vote between them. One possible solution is to 'unite the right': the *Regressive Coercitives* and the *Behaviour & Compliance* party could unite to form a single *Coercitive* party, which would then win with a 51% majority:

Left	*Ultracommunists*	1%	
	New Demagogues	14%	
	Total:		15%
Centre	*Literal Party*		34%
Right	***Coercitive***		**51%**

If it is not possible to 'unite the right', another option is to split the centrist support of the *Literals*. For example, the *Regressive Coercitives* could covertly support the emergence of 'fringe' centrist parties, fracturing the *Literal* support:

Left	*Ultracommunists*	1%	
	New Demagogues	14%	
	Total:		15%
Centre	*Qubekistan Liberation Front*	2%	
	Earth First! Party	2%	
	Popular Front for the Liberation of Qubekistan	2%	
	Ganja Legalization Party	3%	
	Literal Party	24%	
	People's Democratic Front for Qubekistan Liberation	1%	
	Total		34%
Right	***Regressive Coercitives***	**26%**	
	Behaviour & Compliance	25%	
	Total:		51%

Now the *Regressive Coercitives* will win (barely) with a plurality of 26%. In both cases, the election outcome changed, not because of the 'Will of the People', but because of clever manipulation of the electoral process. This is a simple example of *election manipulation*.

Next, consider a hypothetical gubernatorial election in the state of Kolifönia. The candidates are Ahnold, Bustamante, Carey, and Davis. We will write, for example, $A \succ B$ to mean that a voter prefers Ahnold to Bustamante. The voter's preferences are as follows:

15%	A	$\succ$	D	$\succ$	B	$\succ$	C
15%	A	$\succ$	C	$\succ$	D	$\succ$	B
15%	B	$\succ$	C	$\succ$	D	$\succ$	A
10%	B	$\succ$	D	$\succ$	C	$\succ$	A
25%	C	$\succ$	D	$\succ$	B	$\succ$	A
20%	D	$\succ$	B	$\succ$	C	$\succ$	A

Assuming people vote for their first-place choice, we get the following results:

Ahnold	**30%**
Bustamante	25%
Carey	25%
Davis	20%

Hence, Ahnold will be the new governor of Kolifönia, despite the fact that fully 70% of Kolifönians despise him and ranked him *last* of the four possible candidates. The incumbent, Davis, is defeated by an overwhelming vote of nonconfidence, with the smallest support of any candidate (despite the fact that 70% of Kolifönians prefer him to Ahnold).

Other voting schemes

Because of pathologies like the ones in §1, people have proposed many replacements for the standard "plurality" voting scheme.

1.1 Pairwise elections

In the two examples from §1, the pathology seems to arise from the fact that, with more than two candidates, there is often no candidate who obtains an *absolute majority* (ie. greater than 50% of the vote) so we must instead choose the candidate who achieves a *plurality* (the largest *minority* share, eg. 30%). The obvious solution is to only allow two-candidate elections. With more than two candidates, however, we need more than one such election. For example. we might have the following agenda of elections:

1. First, Ahnold competes with Bustamante.
2. Next, the winner of this election takes on Carey.
3. Finally, the winner of *this* election takes on Davis.

For convenience, we reprint the profile of Kolifönia voter preferences from §1, with extra columns showing how each voter group votes in each pairwise election.

%	Preference	A v B	B v C	B v D	D v C	C v A
15%	$A \succ D \succ B \succ C$	A	B	D	D	A
15%	$A \succ C \succ D \succ B$	A	C	D	C	A
15%	$B \succ C \succ D \succ A$	B	B	B	C	C
10%	$B \succ D \succ C \succ A$	B	B	B	D	C
25%	$C \succ D \succ B \succ A$	B	C	D	C	C
20%	$D \succ B \succ C \succ A$	B	B	D	D	C
	Final tallies:	30/70	60/40	25/75	45/55	70/30

By inspecting the columns ' A v B', 'B v C' and 'B v D' of this table, it is clear the election outcomes will be as follows:

$$\boxed{\begin{array}{ll}\text{A} & 30\% \\ \textbf{B} & \textbf{70\%}\end{array}} \xrightarrow{B} \boxed{\begin{array}{ll}\textbf{B} & \textbf{60\%} \\ \text{C} & 40\%\end{array}} \xrightarrow{B} \boxed{\begin{array}{ll}\text{B} & 25\% \\ \textbf{D} & \textbf{75\%}\end{array}} \xrightarrow{D} \textbf{D}$$

At stage 1, Bustamante defeats Ahnold by a landslide. At stage 2, Bustamante easily defeats Carey. But at stage 3, Bustamante *loses* against Davis. Thus, Davis 'wins' the election, despite the fact that Davis had the *smallest support of any candidate* in the original plurality vote. The People have spoken.

Or have they? The problem here is the election *agenda* —ie. the order in which candidates are compared. If we use a different agenda, we get a different outcome. For example, suppose we use the following agenda:

1. First, Bustamante competes with Davis
2. Next, the winner of this election takes on Carey.
3. Finally, the winner of *this* election takes on Ahnold.

$$\boxed{\begin{array}{ll}\text{B} & 25\% \\ \textbf{D} & \textbf{75\%}\end{array}} \xrightarrow{D} \boxed{\begin{array}{ll}\text{D} & 45\% \\ \textbf{C} & \textbf{55\%}\end{array}} \xrightarrow{C} \boxed{\begin{array}{ll}\textbf{C} & \textbf{70\%} \\ \text{A} & 30\%\end{array}} \xrightarrow{C} \textbf{C}$$

Davis wins the first round against Bustamante, but is then defeated by Carey in the second round. Carey goes on to soundly defeat the much-reviled Ahnold, so now it is *Carey* who wins the election.

Clearly, an electoral scheme isn't very good if the decision varies depending upon the agenda of pairwise elections. The outcome is then not the result of the 'will of the People' but instead an artifact, a consequence of a technicality. This scheme is also vulnerable to manipulation. For example, the incumbent, Davis, can decide the agenda, and he will choose the *first* agenda, which will ensure his re-election.

1.2 The Condorcet Scheme

The problem with pairwise elections is that even if candidate X beats candidate Y, she may lose to candidate Z. Depending on the order of the agenda, a different person may end up winning the last round. The Marquis de Condorcet's response was that someone can only claim legitimate victory if they can beat *every* other candidate. The *Condorcet scheme* works as follows:

- For each possible pair of candidates, determine who wins in an election between that pair.
- The Condorcet winner is the candidate who beats *every other candidate* in a pairwise match.

For example, suppose the profile of voter preferences was as follows:

%	Preference	A v B	A v C	A v D	B v C	B v D	C v D
30%	$A \succ B \succ C \succ D$	A	A	A	B	B	C
15%	$B \succ C \succ D \succ A$	B	C	D	B	B	C
10%	$B \succ D \succ C \succ A$	B	C	D	B	B	D
25%	$C \succ B \succ D \succ A$	B	C	D	C	B	C
20%	$D \succ B \succ C \succ A$	B	C	D	B	D	D
	Final tallies:	30/70	30/70	30/70	75/25	80/20	70/30

The Condorcet scheme would yield the following outcomes:

	vs. A		*vs. B*		*vs. C*		*vs. D*	
A			$_{30}\backslash^{70}$	**B**	$_{30}\backslash^{70}$	C	$_{30}\backslash^{70}$	D
B	$_{70}\backslash^{30}$	**B**			$_{75}\backslash^{25}$	**B**	$_{80}\backslash^{20}$	**B**
C	$_{70}\backslash^{30}$	C	$_{25}\backslash^{75}$	**B**			$_{70}\backslash^{30}$	C
D	$_{70}\backslash^{30}$	D	$_{20}\backslash^{80}$	**B**	$_{30}\backslash^{70}$	C		

Thus, B beats every other individual candidate in a pairwise race, so B is the Condorcet winner. It is easy to prove:

Theorem 1.1 *Suppose X is the Condorcet winner. Then X will be the ultimate victor of a sequence of pairwise elections, no matter what the order of the agenda.*

Proof: **Exercise 1** ______ □

The problem with this method is that there may not *be* a Condorcet winner, in general. Indeed, Theorem 1 immediately implies that there is no Condorcet winner in the Kolifönia election (because otherwise different agendas wouldn't have yielded different outcomes). For an even more extreme example, consider the *Condorcet Paradox:*

%	Preference					A v B	A v C	B v C
33%	A	$\succ$	B	$\succ$	C	A	A	B
33%	B	$\succ$	C	$\succ$	A	B	C	B
34%	C	$\succ$	A	$\succ$	B	A	C	C
					Final tallies:	67/33	33/67	66/34

This yields the following pairwise results:

	vs. A		*vs. B*		*vs. C*	
A			$_{67}\backslash^{33}$	A	$_{33}\backslash^{67}$	C
B	$_{33}\backslash^{67}$	A			$_{66}\backslash^{34}$	B
C	$_{67}\backslash^{33}$	C	$_{34}\backslash^{66}$	B		

Thus, although A beats B, he loses to C. Likewise, B beats C, but loses to A, and C beats A, but loses to B. Like a game of 'Scissors, Rock, Paper', there is no clear winner. This has the following 'paradoxical' consequence:

> No matter which alternative is chosen as leader, this leader will be opposed by a *majority* of voters. Furthermore, this opposing majority can always identify a *specific* alternative they prefer to the current leader.

This clearly has highly undesirable consequences for political stability. A 'Condorcet Paradox' society is a society where a majority of voters are always dissatisfied with the status quo, and constantly seek to replace the existing regime with a new one.

Exercise 2 (a) Show that, in the Condorcet paradox, a sequence of pairwise elections will always elect the *last* candidate named in the agenda. For example, if the agenda is: 'First A vs. B; next the winner takes on C', then the ultimate victor will be C.

(b) Generalize the Condorcet paradox to four or more candidates. Is the analog of part (a) true? Why or why not?

1.3 Borda Count

Perhaps the problem with the traditional plurality vote, pairwise elections, and the Condorcet method is that they all attempt to reduce a complicated piece of information (the complete ordering of the voters' preferences) to a sequence of simplistic binary choices (eg. A vs. B). A more sophisticated method would try to take into account the complete *order structure* of a voter's preferences. One such method is the *Borda count.* Suppose we are choosing amongst N alternatives.

- Assign $(N-1)$ points to each voter's *first* choice, $(N-2)$ points to her *second* choice, and so on, assigning $(N-k)$ points to her kth choice, and 0 points to her *last* choice.
- Add up all the points for each alternative. The winner is the alternative with the highest score.

For example, in the *Condorcet paradox* example (page 6), we get the following scores:

%	Preferences					Points A	Points B	Points C
33%	A	$\succ$	B	$\succ$	C	2	1	0
33%	B	$\succ$	C	$\succ$	A	0	2	1
34%	C	$\succ$	A	$\succ$	B	1	0	2
					Total score:	100	99	101

(1.1)

Thus, C is the winner (barely) with a total Borda count of 101 points.

The Borda count has three shortcomings:

Strategic voting, where voters 'lie' about their preferences to manipulate the outcome.

Failing the Condorcet criterion An alternative can lose in the Borda count, even though it is the Condorcet winner.

Sensitivity to irrelevant alternatives, where a Borda loser can become a Borda winner when extra (losing) alternatives are introduced to the election.

Strategic Voting

To see how vote manipulation can occur, consider the following profile of voter preferences in a competition between Arianne, Bryn, and Chloe:

%	Preferences					Points A	Points B	Points C
45%	A	$\succ$	B	$\succ$	C	2	1	0
45%	B	$\succ$	A	$\succ$	C	1	2	0
10%	C	$\succ$	A	$\succ$	B	1	0	2
					Total score:	145	135	20

In the Borda count election, Arianne narrowly defeats Bryn, and both candidates obliterate the pitifully unpopular Chloe.

However, suppose that a pre-election survey compiles data on voter's preferences, and predicts this outcome. Armed with this knowledge, the 45% who support Bryn decide to manipulate the results. They reason as follows: "We hate Chloe, but there's clearly no danger of her winning. We prefer that Bryn win rather than Arianne, and we are inadvertently contributing to Arianne's victory by ranking her second (rather than third) in our preferences. So, to ensure that Bryn wins, let's *pretend* that we like Chloe more than Arianne, and rank Arianne third."

The new (dishonest) profile of voter preferences is as follows:

%	Preferences					Points		
						A	B	C
45%	A	$\succ$	B	$\succ$	C	2	1	0
45%	B	$\succ$	C	$\succ$	A	0	2	1
10%	C	$\succ$	A	$\succ$	B	1	0	2
					Total score:	100	135	65

If the people vote according to these preferences on election day, then Bryn emerges as the clear winner.

However, the 45% who support Arianne have spies in the Bryn campaign, and they discover this dastardly plot. They plan a counterattack: "We hate Chloe, but there's clearly no danger of her winning. We prefer that Arianne win rather than Bryn, and we are inadvertently contributing to Bryn's illegitimate victory by ranking her second (rather than third) in our preferences. So, to ensure that Arianne wins, let's *pretend* that we like Chloe more than Bryn, and rank Bryn third."

The new (doubly dishonest) profile of voter preferences is as follows:

%	Preferences					Points		
						A	B	C
45%	A	$\succ$	C	$\succ$	B	2	0	1
45%	B	$\succ$	C	$\succ$	A	0	2	1
10%	C	$\succ$	A	$\succ$	B	1	0	2
					Total score:	100	90	110

The machinations of Arianne and Bryn have cancelled out, and annihilated their political advantage; Chloe scrapes by and wins the race!

It should be pointed out that Borda is not the only method vulnerable to strategic voting. Indeed, the traditional Plurality vote is notoriously vulnerable. This is why, during elections, you often hear people talk about voting 'against' one candidate, instead of voting 'for' another candidate, and why you often get the advice, 'Don't waste your vote; vote for X'; which is essentially asking you to vote strategically.

Failing the Condorcet Criterion

Consider the following profile of voter preferences in a competition between Arianne, Bryn, and Chloe:

Arianne vs. Bryn vs. Chloe								
%	Preferences					Points		
						A	**B**	C
60%	A	$\succ$	B	$\succ$	C	2	**1**	0
40%	B	$\succ$	C	$\succ$	A	0	**2**	1
					Total score:	120	**140**	40

(1.2)

Clearly Bryn wins the Borda count, with a score of 140. However, observe that Arianne is the Condorcet winner: she beats both Bryn and Chloe in pairwise races. Thus, the Borda Count does *not* satisfy Condorcet's criterion.

This isn't necessarily a fatal flaw, but it will certainly cause political instability if the winner of the Borda count is a Condorcet loser: this guarantees that a *strict majority* of voters will react to the outcome by saying, 'Why did Bryn win? I preferred Arianne.'

Sensitivity to irrelevant alternatives

The Borda count also yields scenarios where the electoral outcome can change when a (losing) candidate is added or removed. Imagine that Arianne, Bryn, and Chloe are mathematicians shortlisted for the Fields Medal. A committee has been convened to compare the candidates and decide the winner. At the press conference, the committee chair stands up and begins, "We have decided to award the Fields Medal to Bryn..." At that moment, an aide bursts into the room and announces that Chloe has withdrawn from the competition because a subtle but fatal error was found in her proof of the Beiberbach Conjecture. "Ah," says the committee chair. "In that case, the winner is Arianne." You can imagine that the assembled dignitaries would find this somewhat peculiar.

And yet, this is exactly the outcome of the following scenario. Suppose that at the beginning, the profile of voter preferences is as in table (1.2) from the previous section. Thus, even at the beginning, Chloe is not a serious contender; the race is basically between Arianne and Bryn, and Bryn is the overall winner, with a Borda count of 140.

Now, with the news of Chloe's withdrawal, she drops to bottom place in everyone's preferences, effectively out of the running. This yields the following profile:

Arianne vs. Bryn vs. Chloe								
%	Preferences					Points		
						A	B	C
60%	A	$\succ$	B	$\succ$	C	**2**	1	0
40%	B	$\succ$	A	$\succ$	C	**1**	2	0
					Total score:	**160**	140	0

(1.3)

Now Arianne wins the Borda count! What's going on?

Perhaps the introduction of Chloe has revealed information about the 'intensity' of support for Arianne and Bryn. Arianne's 60% like her only 'slightly' more than Bryn (which is why they rank Bryn over Chloe). However, Bryn's 40% like her a *lot* more than Arianne (so they even rank Chloe ahead of Arianne), as illustrated in the following figure:

$$\begin{array}{ll} & \text{Good} \longleftarrow ---(\text{Cardinal utility}) -- \longrightarrow \text{Bad} \\ 60\% & A - B ----------------------C \\ 40\% & B ------C --------------- A \end{array} \tag{1.4}$$

Thus, Bryn's supporters prefer her to Arianne much more 'intensely' than Arianne's supporters prefer her to Bryn, and this tips the balance in Bryn's favour. However, it takes the presence of a third candidate (even a losing candidate) to reveal this 'intensity'. If we apply this reasoning to the 'Fields Medal' parable, then the last minute withdrawal of Chloe should *not* change the outcome, because her presence in the competition has already revealed this 'intensity' information, and that information is still valid even after she withdraws. Hence, the award should *still* go to Bryn.

However, this 'intensity' defence of the Borda count is debatable. According the 'intensity' defence, the positional ranking of the alternatives acts as a crude proxy for *cardinal utility*; hence, the Borda count approximates Bentham's *Utilitarian* system (see §4). But it's easy to construct scenarios where an alternative's positional ranking is a *very* poor proxy for its cardinal utility; in these situations, the presence of a losing 'Chloe alternative' really *is* irrelevant, or worse, actually misleading.

For example, perhaps Arianne's 60% prefer Arianne to Bryn to exactly the same degree that Bryn's 40% prefer Bryn to Arianne. It's just that Arianne's 60% really despise Chloe, whereas Bryn's 40% are indifferent to Chloe, as illustrated in the following figure:

$$\begin{array}{ll} & \text{Good} \longleftarrow ---(\text{Cardinal utility}) -- \longrightarrow \text{Bad} \\ 60\% & A -----------B ----------C \\ 40\% & B --- C -------A ----------- \end{array} \tag{1.5}$$

In this case, it seems wrong that the presence/absence of Chloe in the election can affect the choice between Arianne and Bryn.

We will return to the issue of 'irrelevant alternatives' when we discuss Arrow's Impossibility Theorem (§2.4).

1.4 Approval voting

The problems with the Borda Count seem to arise from the assignment of variable numbers of points to candidates depending on their rank. The introduction of a new candidate (even an unpopular one) changes the ranks of all other candidates, thereby differentially inflating their scores, and possibly changing the outcome. Perhaps a solution is to only assign *one* point to each 'preferred' candidate. This is the rationale behind *Approval voting*. Approval voting is similar to the Borda count, except that each voter can only assign 1 or 0 points to each candidate. However, the voter can give a point to *several different* candidates. The candidate with the highest score wins.

There are two versions of approval voting:

Fixed allotment: Each voter is given a fixed number of points, which she must spend. For example, there might be 4 candidates, and each voter must vote for exactly 2 of them.

In the limit case, when each voter is given exactly *one* point, this is just the traditional Plurality vote.

Variable allotment: A voter can vote for any number of candidates (including none of them or all of them).

The analysis of *fixed allotment* approval voting is simpler, so that is what we'll consider here. Suppose that there are four candidates, Arianne, Bryn, Chloe, and Dominique, and thirteen voters, with the following preferences:

#	Preferences
4	$A \succ D \succ C \succ B$
1	$B \succ A \succ D \succ C$
2	$B \succ A \succ C \succ D$
3	$C \succ B \succ D \succ A$
2	$D \succ B \succ C \succ A$
1	$D \succ C \succ B \succ A$

This example (from Riker [32, §4E]) shows how different voting procedures produce different winners. First if each voter casts a *single* approval vote, so that we basically have a traditional plurality competition, then Arianne wins, with 4 votes:

#	Preferences	Approval Points			
		A	B	C	D
4	$A \succ D \succ C \succ B$	1	0	0	0
1	$B \succ A \succ D \succ C$	0	1	0	0
2	$B \succ A \succ C \succ D$	0	1	0	0
3	$C \succ B \succ D \succ A$	0	0	1	0
2	$D \succ B \succ C \succ A$	0	0	0	1
1	$D \succ C \succ B \succ A$	0	0	0	1
	Total score:	**4**	3	3	3

Next, if each voter casts *two* approval vote, then Bryn wins, with 8 points:

#	Preferences	Approval Points			
		A	B	C	D
4	$A \succ D \succ C \succ B$	1	0	0	1
1	$B \succ A \succ D \succ C$	1	1	0	0
2	$B \succ A \succ C \succ D$	1	1	0	0
3	$C \succ B \succ D \succ A$	0	1	1	0
2	$D \succ B \succ C \succ A$	0	1	0	1
1	$D \succ C \succ B \succ A$	0	0	1	1
	Total score:	7	**8**	4	7

However, if each voter casts *three* approval vote, then Chloe wins, with 12 points:

#	Preferences	Approval Points			
		A	B	C	D
4	$A \succ D \succ C \succ B$	1	0	1	1
1	$B \succ A \succ D \succ C$	1	1	0	1
2	$B \succ A \succ C \succ D$	1	1	1	0
3	$C \succ B \succ D \succ A$	0	1	1	1
2	$D \succ B \succ C \succ A$	0	1	1	1
1	$D \succ C \succ B \succ A$	0	1	1	1
	Total score:	7	9	**12**	11

Finally, if we use the Borda count method, then Dominique wins with 21 points:

%	Preferences	Borda Points			
		A	B	C	D
4	$A \succ D \succ C \succ B$	3	0	1	2
1	$B \succ A \succ D \succ C$	2	3	0	1
2	$B \succ A \succ C \succ D$	2	3	1	0
3	$C \succ B \succ D \succ A$	0	2	3	1
2	$D \succ B \succ C \succ A$	0	2	1	3
1	$D \succ C \succ B \succ A$	0	1	2	3
	Total Score:	18	20	19	**21**

So the question is, who is *really* the democratically legitimate choice of the People?

Further reading: This section contains only a few of the multitude of voting systems which have been proposed. Others include: the Copeland and Black rules (Examples (3) and (4) on page 24); the Gocha, Dodgson, Peleg, and Lexmin rules [20, §4.1]; the Nash rule [32, §2B], the Schwartz and Kemeny rules [32, §4C], and a plethora of quasi-economic methods.

An inventory of these and other voting procedures can be found in any book on voting theory, such as [9, 13, 47, 50].

The Condorcet method was first proposed [25] by the French mathematician and revolutionary political theorist Jean-Antoine-Nicolas Caritat, the Marquis de Condorcet (1743-1794). The Borda count was originally proposed [10] by the French mathematical physicist, political theorist, and naval captain Jean-Charles Borda (1733-1799). An extensive discussion of the merits of the Borda count can be found in Saari [33], which also contains a nice account of Borda's life and achievements [33, §1.3.3]. The pros and cons of approval voting are examined in [39] and [20, §4.6].

Chapter 2

Voting Procedures

> *Democracy is the worst form of government except all those other forms that have been tried from time to time.* —Winston Churchill

The examples of Chapter 1 leave us with the question: is there *any* voting procedure which will not produce the 'wrong' answer in certain circumstances? To answer this question, we must mathematically define what we mean by a *voting procedure* and precisely specify the sorts of 'wrong' answers we want to avoid.

2.1 Preferences, Profiles, and Procedures

Recommended: §1

Preference orderings: In the examples of Chapter 1, we began with a hypothetical population of *voters*, $\mathcal{V}$, and a collection of *alternatives*, $\mathcal{A}$, and we assumed that each voter is capable of ranking the various alternatives in $\mathcal{A}$ in some 'linear' way, eg.

$$A \succ B \succ C \succ D.$$

In other words, each voter's preferences determine a **preference ordering**: a relation '$\succeq$' on $\mathcal{A}$ which satisfies three axioms:

Completeness: For any pair of alternatives X and Y, either $X \succeq Y$ or $Y \succeq X$.

Reflexiveness: For any alternative $X \in \mathcal{A}$, $\quad X \succeq X$.

Transitivity: For any alternatives X, Y and Z,

$$\Big(X \succeq Y \text{ and } Y \succeq Z \Big) \Longrightarrow \Big(X \succeq Z \Big).$$

If $X \succeq Y$, then we say the voter **prefers** X to Y. Note that it is possible to have $X \succeq Y$ and $X \preceq Y$. In this case, we say the voter is **indifferent** between X and Y, and write $X \approx Y$. We then say that $\{X, Y\}$ is an **indifferent pair** of alternatives. If $X \succeq Y$ but $X \not\approx Y$, then the voter **strictly prefers** X to Y, and we write $X \succ Y$.

In some situations, however, it may be necessary for the voter to make a choice; she cannot be indifferent between two alternatives. You can either have your cake later or eat it now; you can't do both. A **strict preference ordering** is a relation ' $\succ$ ' on $\mathcal{A}$ which satisfies three axioms:

Completeness: For any pair of alternatives X and Y, either $X \succ Y$ or $Y \succ X$.

Antisymmetry: For any pair of alternatives X and Y, we *cannot* have both $X \succ Y$ and $Y \succ X$.

Transitivity: For any alternatives X, Y and Z,

$$\Big(X \succ Y \text{ and } Y \succ Z \Big) \Longrightarrow \Big(X \succ Z \Big).$$

Thus, strict preference is like preference, except that we replace the *Reflexiveness* axiom with an *Antisymmetry* axiom. Clearly, any strict preference ordering ' $\succ$ ' can be expanded to a (nonstrict) preference ordering ' $\succeq$ ' by defining

$$\Big(X \succeq Y \Big) \iff \Big(X \succ Y \text{ or } X = Y \Big)$$

The converse is half true: A (nonstrict) preference ordering ' $\succeq$ ' can be reduced to a strict preference ordering ' $\succ$ *if and only if* there exist no indifferent pairs of alternatives. In this case, we can define a strict preference ' $\succ$ ' by:

$$\Big(X \succ Y \Big) \iff \Big(X \succeq Y \text{ and } X \neq Y \Big).$$

Voting procedures: Intuitively, a *voting procedure* is some method which takes a collection of voters (each with some preference ordering), and chooses a single alternative in $\mathcal{A}$ as the 'collective choice' of the group. Actually, a voting procedure must be more complicated than this. Suppose the group elects alternative A as president, but A then dies in a mysterious accident. There must be someone else who is 'next in line' for the the presidency. To put it another way: if A withdrew from the election at the last moment, after all the voters had finalized their preference orderings, then who would they elect instead? Suppose it was B, and suppose that B also withdrew. Who would the *third* choice be?

Reasoning in this manner, it is clear that a voting procedure doesn't just pick a single 'first' choice, it actually implicitly defines a *preference order* on the set of alternatives; a preference order which supposedly reflects the 'collective will of the People'.

Thus, we could define a voting procedure as a function which takes a *collection* of preference orders as input, and produces a *single* preference order as output. To be more precise, let $\mathcal{P}(\mathcal{A})$ be the set of all possible preference orderings on the set of alternatives $\mathcal{A}$. For example, if $\mathcal{A} = \{A, B, C\}$ is a set of three alternatives, then $\mathcal{P}(\mathcal{A})$ has thirteen elements:

$$\begin{aligned}
\mathcal{P}(\mathcal{A}) \quad = \quad \{ & A \succ B \succ C, \quad B \succ C \succ A, \quad C \succ A \succ B, \\
& A \succ C \succ B, \quad C \succ B \succ A, \quad B \succ A \succ C \\
& A \succ B \approx C, \quad B \succ C \approx A, \quad C \succ A \approx B, \\
& A \approx C \succ B, \quad C \approx B \succ A, \quad B \approx A \succ C \\
& A \approx B \approx C \}.
\end{aligned}$$

Let $\mathcal{V}$ be a collection of voters. A **profile** is a function $\rho : \mathcal{V} \longrightarrow \mathcal{P}(\mathcal{A})$ assigning a preference ordering to each voter. Let $\mathfrak{R}(\mathcal{V}, \mathcal{A})$ be the set of all profiles for the voters in $\mathcal{V}$ and alternatives in $\mathcal{A}$. A **voting procedure**[1] is a function

$$\Pi : \mathfrak{R}(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}(\mathcal{A}).$$

In other words, Π is a function which takes any profile as input, and produces a single **collective** (or **social**) preference ordering $\Pi(\rho)$ as output. We will indicate the preference ordering $\Pi(\rho)$ with the relation '$\overset{\rho}{\sqsupseteq}$' (or simply '$\sqsupseteq$', when ρ is clear from context). Thus, $B \overset{\rho}{\sqsupseteq} C$ means that, given profile ρ, the voting procedure has ranked alternative B over alternative C.

Example 2.1:

⟨a⟩ **Plurality Vote:** Let $\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A})$ be a profile. For each $A \in \mathcal{A}$, let

$$N(A) \quad = \quad \#\left\{ v \in \mathcal{V} \; ; \; A \underset{v}{\overset{\rho}{\succeq}} B, \text{ for all } B \in \mathcal{A} \right\}$$

be the number of voters who rank A 'first' in their preferences. Define ranking $\overset{\rho}{\sqsupseteq}$ by:

$$\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(N(A) \geq N(B) \right).$$

Thus, the winner is the alternative which is ranked 'first' by the most voters.

[1] Sometimes called a **social choice function** or a **social welfare function.**

⟨b⟩ **Borda Count:** Let $\rho \in \mathfrak{R}(\mathcal{V},\mathcal{A})$ be a profile. For each $v \in \mathcal{A}$, define $U_v^\rho : \mathcal{A} \longrightarrow \mathbb{N}$ as follows: for any $A \in \mathcal{A}$, $U_v^\rho(A) \quad = \quad \#\left\{B \in \mathcal{A}\,;\ A \underset{v}{\overset{\rho}{\succeq}} B\right\} - 1$ is the number of alternatives which voter v deems 'no better' than A (not including A itself).

Then define $U^\rho : \mathcal{A} \longrightarrow \mathbb{N}$ by $\quad U^\rho(A) \quad = \quad \sum_{v\in\mathcal{V}} U_v^\rho(A) \quad$ (the 'Borda Count' of A).

Then define ranking $\overset{\rho}{\sqsupseteq}$ by: $\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(U^\rho(A) \geq U^\rho(B) \right)$. Thus, the winner is the alternative with the highest Borda Count.

⟨c⟩ **Approval Voting:** Suppose $\#(\mathcal{A}) = N$ and $M < N$, and suppose each voter must vote for exactly M out of N alternatives. Let $\rho \in \mathfrak{R}(\mathcal{V},\mathcal{A})$ be a profile, and for each $v \in \mathcal{A}$, define $T_v^\rho : \mathcal{A} \longrightarrow \mathbb{N}$ as follows: for any $A \in \mathcal{A}$, let

$$T_v^\rho(A) \quad = \quad \#\left\{B \in \mathcal{A}\,;\ B \underset{v}{\overset{\rho}{\succeq}} A\right\}.$$

be the number of alternatives which voter v prefers to A.

Define $f : \mathbb{N} \longrightarrow \{0,1\}$ by $\quad f(t) \quad = \quad \begin{cases} 1 & \text{if} \quad t \leq M \\ 0 & \text{if} \quad t > M \end{cases}.$

Thus, $f\left(T_v^\rho(A)\right) = 1$ if and only if alternative A is in the 'top M' alternatives for voter v. Now define $T^\rho : \mathcal{A} \longrightarrow \mathbb{N}$ as follows: for any $A \in \mathcal{A}$, let

$$T^\rho(A) \quad = \quad \sum_{v\in\mathcal{V}} f\left(T_v^\rho(A)\right),$$

be the total number of 'approval votes' for A.

Define ranking $\overset{\rho}{\sqsupseteq}$ by: $\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(T^\rho(A) \geq T^\rho(B) \right)$. Thus, the winner is the alternative with the most approval votes.

⟨d⟩ **Condorcet:** Suppose $\mathcal{A} = \{A,B,C,D\}$. If there *is* a Condorcet winner (say A), then we have preference order $A \overset{\rho}{\sqsupset} B \overset{\rho}{\approx} C \overset{\rho}{\approx} D$. If there is *no* Condorcet winner, then we have preference order $A \overset{\rho}{\approx} B \overset{\rho}{\approx} C \overset{\rho}{\approx} D$.

Remarks: **(a)** The four voting procedures described here are not the *only* ways to implement Plurality Vote, Borda Count, Approval Voting, and the Condorcet method (but they are arguably the most 'natural'). For example, there are many voting procedures which will identify a Condorcet winner (if one exists); see if you can invent another one.

(b) Observe that a voting procedure can produce an outcome which is *indifferent* between two or more alternatives. For example, suppose the Viking Longboat Society is using the Condorcet vote to decide whether to serve *Ale*, *Beer*, or *Cider* at their annual fundraiser, but the outcome is the *Condorcet Paradox* (page 6). The only reasonable response is to serve all three beverages!

Strict voting procedures: However, sometimes we cannot accept 'indifferent' outcomes; sometimes our procedure *must* give a strict ordering with a maximal element. For example, in a presidential election, there must be a *unique* choice; we can't have a scenario where three people tie for first place and share the job. Hence, the output must not only be a preference ordering, but a *strict* preference ordering. However, we cannot expect a strict preference as output if we do not provide strict preferences as input. Let $\mathcal{P}^*(\mathcal{A})$ be the set of strict preference orderings on $\mathcal{A}$. For example, if $\mathcal{A} = \{A, B, C\}$, then $\mathcal{P}(\mathcal{A})$ has six elements:

$$\begin{aligned}\mathcal{P}(\mathcal{A}) \;=\; \{&A \succ B \succ C, \quad B \succ C \succ A, \quad C \succ A \succ B, \\ &A \succ C \succ B, \quad C \succ B \succ A, \quad B \succ A \succ C\}.\end{aligned}$$

A **strict profile** is a function $\rho : \mathcal{V} \longrightarrow \mathcal{P}^*(\mathcal{A})$ assigning a strict preference ordering to each voter. Let $\mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ be the set of all strict profiles for the voters in $\mathcal{V}$ and alternatives in $\mathcal{A}$. A **strict voting procedure** is a function $\Pi : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}^*(\mathcal{A})$. We will indicate the preference ordering $\Pi(\rho)$ with the relation '$\overset{\rho}{\sqsupset}$' (or simply '$\sqsupset$', when ρ is clear from context). Thus, $B \overset{\rho}{\sqsupset} C$ means that, given strict profile ρ, the strict voting procedure strictly prefers alternative B to alternative C.

Example 2.2: Agenda of Pairwise Elections

Suppose that $\#(\mathcal{V})$ is *odd*, so that it is impossible for a pairwise election to result in a tie (assuming all voters have *strict* preferences). Given a particular agenda of pairwise elections, we define a strict preference ordering on $\mathcal{A}$ as follows:

1. Find the ultimate winner of the agenda of pairwise elections; rank this candidate *first.*
2. Eliminate this candidate from $\mathcal{A}$. From the *remaining* alternatives, find the ultimate winner of the agenda of pairwise elections; rank this candidate *second.*
3. Eliminate the *second* candidate from $\mathcal{A}$. From the *remaining* alternatives, find the ultimate winner of the agenda of pairwise elections; rank this candidate *third.*
4. Proceed in this fashion until you run out of alternatives.

We illustrate with an example. Suppose that $\mathcal{A} = \{A, B, C, D\}$. Let $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$, and suppose that A is the ultimate winner of the pairwise election agenda with profile ρ. Let $\mathcal{A}_1 = \mathcal{A} \setminus \{A\} = \{B, C, D\}$, and let ρ_1 be the profile we get by restricting each voter's ρ-preferences to $\mathcal{A}_1$. For example if voter v had ρ-preferences $B \underset{v}{\overset{\rho}{\succ}} A \underset{v}{\overset{\rho}{\succ}} C \underset{v}{\overset{\rho}{\succ}} D$, then v would have ρ_1-preferences $B \underset{v}{\overset{\rho_1}{\succ}} C \underset{v}{\overset{\rho_1}{\succ}} D$.

Suppose that B is the ultimate winner of the pairwise voting agenda with profile ρ_1 (we skip the election where A would have been introduced). Let $\mathcal{A}_2 = \mathcal{A}_1 \setminus \{B\} = \{C, D\}$, and let ρ_2 be the profile we get by restricting each voter's ρ_1-preferences to $\mathcal{A}_2$.

Suppose that C is the ultimate winner of the pairwise voting agenda with profile ρ_2 (we skip the elections where A and B would have been introduced). Then $\mathcal{A}_3 = \mathcal{A}_2 \setminus \{C\} = \{D\}$. We define order $\overset{\rho}{\sqsupset}$ by: $A \overset{\rho}{\sqsupset} B \overset{\rho}{\sqsupset} C \overset{\rho}{\sqsupset} D$.

Strict vs. nonstrict procedures: A **tiebreaker rule** is a function $\tau : \mathcal{P}(\mathcal{A}) \longrightarrow \mathcal{P}^*(\mathcal{A})$ which converts any nonstrict preference ordering into a strict ordering in an order-preserving way. That is, if $A, B \in \mathcal{A}$ are two alternatives and $A \succ B$, then τ preserves this. However, if $A \approx B$, then τ forces either $A \succ B$ or $B \succ A$.

In general, the tiebreaker rule can be totally arbitrary (eg. flipping a coin, trial by combat, putting alternatives in alphabetical order, etc.), because if a voter is *indifferent* about two alternative then *by definition* it doesn't matter which one we put first.

Given τ, any *nonstrict* voting procedure Π can be turned into a *strict* voting procedure Π^* as follows:

1. Apply Π to a (strict) voter profile.
2. Use τ to convert the resulting (nonstrict) preference ordering to a strict ordering.

Conversely, any *strict* voting procedure Π^* can be extended to a *nonstrict* voting procedure Π as follows:

1. Apply τ to convert each voter's (nonstrict) preference ordering into a strict preference ordering, thereby converting society's nonstrict voter profile into a strict voter profile.
2. Now apply Π^* to the 'strictified' profile.

We formalize this:

Proposition 2.3 *Let $\tau : \mathcal{P}(\mathcal{A}) \longrightarrow \mathcal{P}^*(\mathcal{A})$ be a tie-breaker rule.*

(a) *Suppose $\Pi : \mathfrak{R}(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}(\mathcal{A})$ is a (nonstrict) voting procedure. Let $\Pi^* := \tau \circ \Pi$. Then $\Pi^* : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}^*(\mathcal{A})$ is a strict voting procedure.*

(b) *Suppose $\Pi^* : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}^*(\mathcal{A})$ is a strict voting procedure. Define $\tau : \mathfrak{R}(\mathcal{V}, \mathcal{A}) \longrightarrow \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ by applying τ separately to each voter's (nonstrict) preference ordering. Now define $\Pi = \Pi^* \circ \tau$. Then Π is a (nonstrict) voting procedure.*

Proof: <u>**Exercise 3**</u> $\square$

However, it is not always desirable to cavalierly convert nonstrict procedures into strict ones. For one thing, by using an arbitrary tie-breaker rule, we will likely violate desirable properties such as the *Neutrality* axiom **(N)** (see §2.2). Also, if the nonstrict procedure *too often* produces indifferent outcomes, then we will *too often* end up making decisions based on some arbitrary rule. For example, the (nonstrict) Condorcet procedure *rarely* yields a strict order. If τ is the 'alphabetical order' rule, and Π is the Condorcet rule, then $\Pi^* = \tau \circ \Pi$ will, in practice, end up meaning 'arrange candidates in alphabetical order' 90% of the time. This is hardly a good way to pick the president.

2.2 Desiderata

Prerequisites: §2.1

We began by asking: are there any voting procedures which produce sensible results? We will formalize what we mean by a 'sensible result' by requiring the voting procedure to satisfy certain axioms.

Pareto (Unanimity): The *Pareto* (or *Unanimity*) axiom is the following:

(P) If $B, C \in \mathcal{A}$, and ρ is a profile where *all* voters prefer B to C (ie. for all $v \in \mathcal{V}$, $B \overset{\rho}{\underset{v}{\succeq}} C$), then $B \overset{\rho}{\sqsupseteq} C$.

This seems imminently sensible. Any voting scheme which chose a unanimously unpopular alternative over a unanimously popular alternative would be highly undemocratic!

Exercise 4 Check that the following voting procedures satisfy axiom **(P)**:

1. Plurality Vote (Example 2.1(a)).
2. Borda Count (Example 2.1(b)).
3. Approval Voting Example 2.1(c).
4. Agenda of Pairwise votes (Example 2.2).

Monotonicity: If the voting procedure selects a certain alternative C as the 'collective choice' of the society, and some voter changes his preferences to become *more* favourable toward C, then surely C should *remain* the collective choice of the society. This is the content of the *Monotonicity* axiom:

(M) Let $B, C \in \mathcal{A}$, and let ρ be a profile such that $B \overset{\rho}{\sqsubseteq} C$. Let $v \in \mathcal{V}$ be some voter such that $C \overset{\rho}{\underset{v}{\preceq}} B$, and let δ be the profile obtained from ρ by giving v a new preference ordering $\overset{\delta}{\underset{v}{\succeq}}$, such that $C \overset{\delta}{\underset{v}{\succeq}} B$ (all *other* voters keep the same preferences). Then $B \overset{\delta}{\sqsubseteq} C$.

Exercise 5 Check that the following voting procedures satisfy axiom **(M)**:

1. Plurality Vote (Example 2.1(a)).
2. Borda Count (Example 2.1(b)).
3. Approval Voting Example 2.1(c).
4. Agenda of Pairwise votes (Example 2.2).

Anonymity: A basic democratic principle is *political equality:* all voters have the same degree of influence over the outcome of a vote. To put it another way, the voting procedure is incapable of distinguishing one voter from another, and therefor treats all their opinions equally. In other words, the voters are *anonymous*. To mathematically encode this, we imagine that all the voters exchange identities (ie. are *permuted*). A truly 'anonymous' voting procedure should be unable to tell the difference...

(A) Let $\sigma : \mathcal{V} \longrightarrow \mathcal{V}$ be a permutation of the voters. Let ρ be a profile, and let δ be the profile obtained from ρ by permuting the voters with σ. In other words, for any $v \in \mathcal{V}$, $\delta(v) = \rho\Big(\sigma(v)\Big)$. Then ρ and δ yield identical collective preference orderings. In other words, for any alternatives $B, C \in \mathcal{A}$,

$$\left(B \overset{\rho}{\sqsupseteq} C \right) \iff \left(B \overset{\delta}{\sqsupseteq} C \right).$$

<u>**Exercise 6**</u> Check that the following voting procedures satisfy axiom **(A)**:

1. Plurality Vote (Example 2.1(a)).
2. Borda Count (Example 2.1(b)).
3. Approval Voting Example 2.1(c).
4. Agenda of Pairwise votes (Example 2.2).

If we wish to impose axiom **(A)** as a blanket assumption, then we can restrict our attention to *anonymous voting procedures*. An **anonymous profile** is a function $\alpha : \mathcal{P}(\mathcal{A}) \longrightarrow \mathbb{N}$. For example, any profile $\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A})$ yields an anonymous profile $\alpha_\rho : \mathcal{P}(\mathcal{A}) \longrightarrow \mathbb{N}$, where for any $P \in \mathcal{P}(\mathcal{A})$,

$$\alpha(P) \quad = \quad \#\{v \in \mathcal{V} \,;\, \rho(v) = P\}$$

is the number of voters with preference ordering P. Let $\widetilde{\mathfrak{R}}(\mathcal{A})$ be the set of anonymous profiles. An **anonymous voting procedure** is a function $\widetilde{\Pi} : \widetilde{\mathfrak{R}}(\mathcal{A}) \longrightarrow \mathcal{P}(\mathcal{A})$, which takes an anonymous profile as input, and yields a preference ordering on $\mathcal{A}$ as output.

<u>**Exercise 7**</u> Let $\Pi : \mathfrak{R}(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}(\mathcal{A})$ be a voting procedure. Show that $\Big(\Pi$ satisfies axiom **(A)** $\Big)$ $\iff \left(\begin{array}{l} \text{There is some anonymous procedure } \widetilde{\Pi} \\ \text{so that } \Pi(\rho) \ = \ \widetilde{\Pi}(\alpha_\rho) \text{ for any } \rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A}) \end{array} \right)$.

Neutrality: Just as a voting procedure should be impartial amongst voters, it should be impartial amongst alternatives. An 'unbiased' procedure does not favour the 'first' alternative over the 'second' alternative, and so on. In other words, if we permute the alternatives, we should get the same outcome.

(N) Let $\sigma : \mathcal{A} \longrightarrow \mathcal{A}$ be a permutation of the alternatives. Let ρ be a profile, and let δ be the profile obtained from ρ by permuting the alternatives with σ. In other words, for any $B, C \in \mathcal{A}$, and any $v \in \mathcal{V}$,

$$\left(B \underset{v}{\overset{\rho}{\succeq}} C \right) \iff \left(\sigma(B) \underset{v}{\overset{\delta}{\succeq}} \sigma(C) \right).$$

Then the preference ordering produced by δ is obtained by similarly permuting the preference ordering produced from ρ. That, is, for any $B, C \in \mathcal{A}$,

$$\left(B \overset{\rho}{\sqsupseteq} C \right) \iff \left(\sigma(B) \overset{\delta}{\sqsupseteq} \sigma(C) \right).$$

__**Exercise 8**__ Show that an *agenda* of pairwise elections (Example 2.2) between three or more alternatives does *not* satisfy the axiom **(N)**.

__**Exercise 9**__ Check that the following voting procedures *do* satisfy axiom **(N)**:

1. Plurality Vote (Example 2.1(a)).
2. Borda Count (Example 2.1(b)).
3. Approval Voting Example 2.1(c).

__**Exercise 10**__ Show that, if a voting procedure satisfies the *Monotonicity* axiom **(M)** and the *Neutrality* axiom **(N)**, then it must satisfy the *Pareto* axiom **(P)**.

__**Exercise 11**__ Suppose a *nonstrict* voting procedure Π is transformed into a *strict* voting procedure Π^* using Proposition 2.3(b) on page 20. Show that Π^* might not satisfy axiom **(N)**, even if Π did.

Condorcet: As we saw in §1.2, the Condorcet scheme, while laudable, is somewhat unsatisfactory as a voting procedure because it usually doesn't produce a clear winner. Nevertheless, it's based on a good idea, the *Condorcet criterion*, which we might desire in other voting methods.

If $A, B \in \mathcal{A}$ are two alternatives, let $\#[A \succ B]$ be the number of voters who strictly prefer A to B:

$$\#[A \succ B] \quad = \quad \#\left\{ v \in \mathcal{V} \; ; \; A \underset{v}{\succ} B \right\}.$$

Let's write "$A \gg B$" if A defeats B in a pairwise vote:

$$\left(A \gg B \right) \iff \left(\#[A \succ B] \; > \; \#[B \succ A] \right).$$

The Condorcet Criterion states:

(C) If $A \in \mathcal{A}$ is an alternative who beats every other alternative in a pairwise vote, then A is the top-ranked element of the collective preference ordering. That is

$$\left(\forall \; B \in \mathcal{A}, \quad A \gg B \right) \quad \Longrightarrow \quad \left(\forall \; B \in \mathcal{A}, \quad A \sqsupset B \right).$$

Procedures satisfying axiom **(C)** are sometimes called *Condorcet extensions*, because they often take the form, "If there is a clear Condorcet winner, then choose her. If not, then choose a winner using the following method instead...'. Thus, Condorcet extensions reduce (but usually do not eliminate) the possibility of a tie or ambiguous outcome.

Example 2.4:

⟨a⟩ The Borda Count does *not* satisfy axiom **(C)**. See §1.3.

⟨b⟩ **Sequence of pairwise votes:** Theorem 1.1 says that any sequence of pairwise votes will choose a Condorcet winner, if one exists. Thus, all such sequences are Condorcet extensions. The problem is, as we saw in §1.1, different sequences can produce outcomes; hence a pairwise vote sequence violates the *Neutrality* axiom **(N)**.

⟨c⟩ **Copeland Rule:** The **Copeland index** of A is the number of alternatives A defeats in pairwise votes, minus the number which defeat A:

$$i(A) \quad = \quad \#\{B \in \mathcal{A}\,;\ A \gg B\} \ - \ \#\{B \in \mathcal{A}\,;\ B \gg A\}$$

Thus, if A is the Condorcet winner, then $i(A) = \#(\mathcal{A}) - 1$.

The **Copeland rule** tells us to rank the alternatives in decreasing order of their Copeland indices; thus, the Copeland 'winner' is the alternative with the highest Copeland index. The Copeland rule satisfies the Condorcet criterion **(C)** because, if a Condorcet winner exists, he is automatically the Copeland winner.

⟨d⟩ **Black Rule:** The Black Rule is very simple: if a Condorcet winner A exists, then choose A. If there is no Condorcet winner, then use the Borda Count method to order the alternatives.

The Copeland and Black rules are still not *strict* voting procedures, because ties are still possible; they are simply less likely than in the original Condorcet rule.

2.3 Sen and (Minimal) Liberalism

Prerequisites: §2.2

Surely in a democratic society, there are certain decisions, concerning your person, over which only you should have control. For example, society can impose some constraints on your actions (eg. *Thou shalt not steal*), but only you should be able to decide what you wear, what you say, and who you choose to associate with.

The idea that individuals have certain inalienable rights is called *liberalism*, and the most minimal form of liberalism is one where a particular voter has control over *one* decision in society. If $v \in \mathcal{V}$ is a voter, and $B, C \in \mathcal{A}$ are alternatives, then we say that v is **decisive** over the pair $\{B, C\}$ if, for any profile ρ,

$$\left(B \overset{\rho}{\sqsupseteq} C \right) \iff \left(B \underset{v}{\overset{\rho}{\succeq}} C \right)$$

The axiom of *Minimal Liberalism* requires:

(ML) There are two voters $v_1, v_2 \in \mathcal{V}$, and four alternatives $B_1, C_1,\ B_2, C_2 \in \mathcal{A}$, such that v_1 is decisive over $\{B_1, C_1\}$ and v_2 is decisive over $\{B_2, C_2\}$.

All this says is that there are at least two individuals in the society who have some minimal degree of control over some single aspect of their lives —a very minimal form of liberalism indeed! Nevertheless, we are confronted with...

Sen's Impossibility Theorem: *Suppose there are at least three alternatives in $\mathcal{A}$, and at least two voters in $\mathcal{V}$. Then there is no strict voting procedure which satisfies axioms* **(P)** *and* **(ML)**.

Proof: Suppose that, as in axiom **(ML)**, voter v_1 is decisive over the pair $\{B_1, C_1\}$ and voter v_2 is decisive over the pair $\{B_2, C_2\}$. Given *any* strict profile as input, the voting procedure should produce a strict preference ordering as output. But suppose that voters have the following preferences:

v_1	C_2	$\succ$	$\mathbf{B_1}$	$\succ$	$\mathbf{C_1}$	$\succ$	B_2
		$\diamond$		$*$		$\star$	
v_2	C_1	$\succ$	$\mathbf{B_2}$	$\succ$	$\mathbf{C_2}$	$\succ$	B_1
		$\star$		$\dagger$		$\diamond$	
Others	C_1	$\succ$	B_2	and	C_2	$\succ$	B_1
		$\star$				$\diamond$	

(we have accented in bold the choices over which each voter is decisive). Now,

($*$) v_1 is decisive over $\{B_1, C_1\}$, so we must have $B_1 \sqsupset C_1$.

($\star$) Society is unanimous that $C_1 \succ B_2$, so we must have $C_1 \sqsupset B_2$, by axiom **(P)**.

($\dagger$) v_2 is decisive over $\{B_2, C_2\}$, so we must have $B_2 \sqsupset C_2$.

($\diamond$) Society is unanimous that $C_2 \succ B_1$, so we must have $C_1 \sqsupset B_1$, by axiom **(P)**.

We conclude that

$$B_1 \sqsupset C_1 \sqsupset B_2 \sqsupset C_2 \sqsupset B_1.$$

By *transitivity*, it follows that $C_1 \sqsupset B_1$. But since we also have $B_1 \sqsupset C_1$, this contradicts the requirement of *antisymmetry*. We conclude that $\sqsupset$ cannot be a strict preference ordering. $\square$

Remark: The astute reader will notice that the previous proof seems to assume the existence of *four* alternatives (B_1, C_1, B_2 and C_2), despite the fact that Sen's Theorem only hypothesizes *three*. This apparent inconsistency is reconciled by recognizing that the pairs $\{B_1, C_1\}$ and $\{B_2, C_2\}$ may not be *distinct*. For example, we could set $B_1 = C_2$; and then rework the proof of Sen's Theorem without needing to include the 'unanimous' decision that $B_1 \sqsupset C_2$. The details are **Exercise 12** .

Exercise 13 (a) Show that no voting method can satisfy both the *Anonymity* axiom **(A)** and the *Minimal Liberalism* axiom **(ML)**.

(b) Likewise, show that no voting method can satisfy both the *Neutrality* axiom **(N)** and the *Minimal Liberalism* axiom **(ML)**.

(c) Suggest how you might replace both **(A)** and **(N)** with a modified axiom which allows for **(ML)**, while still encoding the idea that society gives equal political rights to all voters, and decides between conflicting alternatives in an 'unbiased' manner.

Further reading: The original references to Sen's theorem are Sen [38, 37]. An elementary discussion and proof are given in Saari [34]; other references are Saari [33, §3.4.1] or Kim and Roush [20, Thm 4.4.1, p.81]

2.4 Arrow and Independence of Irrelevant Alternatives

Prerequisites: §2.2

Recall the Danacian election example of §1, where the introduction of extra 'fringe' parties into an election 'split' the vote of the ruling *Literal* party, allowing the *Regressive Coercitives* to seize office. Although both the *Literal* and *Coercitive* parties are much more popular than these fringe groups, the existence of fringe parties changes the outcome. The *Plurality* voting procedure is sensitive to the 'irrelevant alternative' of the fringe parties.

Likewise, in §1.3 we were able to change the outcome of a Borda count election between Arianne and Bryn by introducing a third alternative, Chloe. Despite the fact that *everyone* prefers one of the other alternatives to Chloe, her presence in the race still tips the balance. The choice between Arianne and Bryn is sensitive to the 'irrelevant alternative' of Chloe.

We saw how this sensitivity to irrelevant alternatives makes a process vulnerable to manipulation. The *Coercitives* can manipulate the outcome by covertly supporting the fringe parties. Likewise, the friends of Bryn may encourage Chloe to participate in the election, even though she has no chance of winning, simply to manipulate the results in their favour. We want a procedure which is immune to these machinations. We say that a voting procedure is *Independent of Irrelevant Alternatives* if the following is true.

(IIA) Let $A, B \in \mathcal{A}$ be two alternatives. Suppose ρ, δ are two profiles, such that each voter's ρ-preference concerning the pair $\{A, B\}$ is identical to his δ-preferences concerning $\{A, B\}$. That is, for every $v \in \mathcal{V}$,

$$\left(A \underset{v}{\overset{\rho}{\succeq}} B \right) \iff \left(A \underset{v}{\overset{\delta}{\succeq}} B \right)$$

Then the collective ρ-preference concerning $\{A, B\}$ will be identical to the collective δ-preference concerning $\{A, B\}$. That is: $\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(A \overset{\delta}{\sqsupseteq} B \right)$.

To translate this into English, suppose that $\mathcal{A} = \{\text{Arianne}, \text{Bryn}, \text{Chloe}\}$. Let ρ be the profile of table (1.2) on page 9 of §1.3, and let δ be the profile of table (1.3) on page 9. Then we see that the Borda count does *not* satisfy **(IIA)**, because it says $B \overset{\rho}{\sqsupset} A$ but $A \overset{\delta}{\sqsupset} B$, despite the fact that all voters order A and B the *same* in both profiles. In §1.3 we showed how this led to an undesirable scenario, where the 'winner' of the Fields Medal changed because a losing candidate (Chloe) dropped out of the race. This is the reason why **(IIA)** is considered a desirable property.

Dictatorship: A **dictatorship** is a voting procedure where one voter makes all the decisions. In other words, there is some voter $v \in \mathcal{V}$ (the **dictator**) so that, for any $B, C \in \mathcal{A}$,

$$\left(B \sqsupseteq C \right) \iff \left(B \underset{v}{\succeq} C \right).$$

We now come to the most famous result in mathematical political science:

Arrow's Impossibility Theorem: *Suppose that $\mathcal{A}$ has at least three alternatives, and $\mathcal{V}$ has at least two voters. Then the only voting procedure which satisfies axioms* **(P)** *and* **(IIA)** *is a dictatorship.*

Proof: First we will show that any procedure satisfying axioms **(P)** and **(IIA)** must also satisfy a version of the *Neutrality* axiom **(N)**.

Claim 1: *Let $A_1, B_1, A_2, B_2 \in \mathcal{A}$ be four alternatives*[2]*, and suppose ρ is a profile such that every voter's preference ordering of the pair $\{A_1, B_1\}$ is identical to her ordering of $\{A_2, B_2\}$. In other words, for every voter $v \in \mathcal{V}$,*

$$\left(A_1 \underset{v}{\succeq} B_1 \right) \iff \left(A_2 \underset{v}{\succeq} B_2 \right).$$

Then the voting procedure will yield a preference order which also assigns the same order to the pair $\{A_1, B_1\}$ as to the pair $\{A_2, B_2\}$. That is: $\left(A_1 \overset{\rho}{\sqsupseteq} B_1 \right) \iff \left(A_2 \overset{\rho}{\sqsupseteq} B_2 \right)$.

Proof: Assume WOLOG that $A_1 \overset{\rho}{\sqsupseteq} B_1$. We want to show that $A_2 \overset{\rho}{\sqsupseteq} B_2$. To do this, we will create a new profile δ such that:

(a) Every voter's ordering of $\{A_2, B_2\}$ is identical in δ and ρ. Hence, by **(IIA)**, we have

$$\left(A_2 \overset{\rho}{\sqsupseteq} B_2 \right) \iff \left(A_2 \overset{\delta}{\sqsupseteq} B_2 \right).$$

[2] Here, we assume $A_1 \neq B_1$ and $A_2 \neq B_2$; however, the sets $\{A_1, B_1\}$ and $\{A_2, B_2\}$ might not be disjoint (eg. if $\mathcal{A}$ only has three alternatives in total).

(b) δ is structured so that it is clear that $A_2 \overset{\delta}{\sqsupseteq} B_2$.

Combining facts **(a)** and **(b)** yields $A_2 \overset{\rho}{\sqsupseteq} B_2$, as desired.

To obtain δ, take each voter and change her ρ-ranking of A_2 so that it is just above A_1. Likewise, change her ranking of B_2 so that it is just below B_1. We can always do this in a manner which preserves her ordering of the pairs $\{A_1, B_1\}$ and $\{A_2, B_2\}$, as shown by the diagram below:

Before (in ρ)	After (in δ)
$A_1 \overset{\rho}{\succeq} B_1$ and $A_2 \overset{\rho}{\succeq} B_2$	$A_2 \overset{\delta}{\succeq} A_1 \overset{\delta}{\succeq} \ldots \overset{\delta}{\succeq} B_1 \overset{\delta}{\succeq} B_2$
$B_1 \overset{\rho}{\succeq} A_1$ and $B_2 \overset{\rho}{\succeq} A_2$	$B_1 \overset{\delta}{\succeq} B_2 \overset{\delta}{\succeq} \ldots \overset{\delta}{\succeq} A_2 \overset{\delta}{\succeq} A_1$

Now, in δ,

- Every voter prefers A_2 to A_1, so we have $A_2 \overset{\delta}{\sqsupseteq} A_1$ by the Pareto axiom **(P)**.
- Every voter prefers B_1 to B_2, so we have $B_1 \overset{\delta}{\sqsupseteq} B_2$ by the Pareto axiom **(P)**.
- Every voter's ordering of $\{A_1, B_1\}$ is identical in δ and ρ. Hence, $A_1 \overset{\delta}{\sqsupseteq} B_1$, by **(IIA)**.

We now have $A_2 \overset{\delta}{\sqsupseteq} A_1 \overset{\delta}{\sqsupseteq} B_1 \overset{\delta}{\sqsupseteq} B_2$. Because $\overset{\delta}{\sqsupseteq}$ is a transitive ordering, it follows that $A_2 \overset{\delta}{\sqsupseteq} B_2$.

However, every voter's ordering of $\{A_2, B_2\}$ is identical in δ and ρ. Hence, by **(IIA)**, we conclude that $A_2 \overset{\rho}{\sqsupseteq} B_2$, as desired.

We have now shown that

$$\left(A_1 \overset{\rho}{\sqsupseteq} B_1 \right) \Longrightarrow \left(A_2 \overset{\rho}{\sqsupseteq} B_2 \right).$$

By switching A_1 with A_2 and switching B_1 with B_2 throughout the proof, we can likewise show that

$$\left(A_2 \overset{\rho}{\sqsupseteq} B_2 \right) \Longrightarrow \left(A_1 \overset{\rho}{\sqsupseteq} B_1 \right).$$

This completes the proof. .... □ `[Claim 1]`

Now, suppose we number the voters $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ in some arbitrary way. If $A, B \in \mathcal{A}$ are any two alternatives, we'll write '$A \underset{n}{\overset{\rho}{\succeq}} B$' to mean 'voter v_n prefers A to B in profile ρ'

Claim 2: *There is a 'swing voter' v_m with the following property: Suppose $X, Y \in \mathcal{A}$ are any two alternatives, and ρ is any profile such that*

$$X \underset{n}{\overset{\rho}{\succeq}} Y, \text{ for all } n < m \quad \text{and} \quad X \underset{n}{\overset{\rho}{\preceq}} Y, \text{for all } n > m. \tag{2.1}$$

Then

$$\left(X \overset{\rho}{\sqsupseteq} Y \right) \iff \left(X \underset{m}{\overset{\rho}{\succeq}} Y \right). \tag{2.2}$$

Thus, v_m can 'tip the balance' between X and Y in any profile satisfying eqn.(2.1).

Proof: Fix $A, B \in \mathcal{A}$, and consider the following sets of profiles:

$$\begin{array}{rcl}
\mathfrak{R}_0 & = & \{\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A}) \ ; \ \text{every voter prefers } B \text{ to } A\}; \\
\mathfrak{R}_1 & = & \left\{\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A}) \ ; \ \mathcal{A} \underset{1}{\overset{\rho}{\succeq}} \mathcal{B}, \text{ but for all } n > 1, \ \mathcal{A} \underset{n}{\overset{\rho}{\preceq}} \mathcal{B}\right\}; \\
\vdots & \vdots & \vdots \\
\mathfrak{R}_m & = & \left\{\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A}) \ ; \ \begin{array}{c} A \underset{n}{\overset{\rho}{\succeq}} B, \text{ for all } n \leq m \\ \text{but } A \underset{n}{\overset{\rho}{\preceq}} B, \text{ for all } n > m \end{array}\right\}; \\
\vdots & \vdots & \vdots \\
\mathfrak{R}_N & = & \{\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A}) \ ; \ \text{every voter prefers } A \text{ to } B\}.
\end{array}$$

Claim 2.1: *The voting procedure assigns a preference of A vs. B which is <u>constant</u> on each of the sets $\mathfrak{R}_0, \mathfrak{R}_1, \ldots, \mathfrak{R}_N$. In other words, for any $n \in [0..N]$, if there is one $\rho \in \mathfrak{R}_n$ so that $A \overset{\rho}{\sqsupseteq} B$, then it must be the case for <u>all</u> $\delta \in \mathfrak{R}_n$ that $A \overset{\delta}{\sqsupseteq} B$.*

Proof: Axiom **(IIA)** says that the voter's preferences over *other* pairs of alternatives have no effect on the collective preference of A vs. B. Thus, the collective preference of A vs. B is determined only by the pattern of voters' $\{A, B\}$-preferences, and this pattern is constant across each set $\mathfrak{R}_n$, by definition. .... □ `[Claim 2]`

We will write "$A \overset{\mathfrak{R}_n}{\sqsupseteq} B$" to mean that $A \overset{\rho}{\sqsupseteq} B$ for all $\rho \in \mathfrak{R}_n$. By the Pareto axiom **(P)**, we know that

$$A \overset{\mathfrak{R}_0}{\sqsubseteq} B \quad \text{and} \quad A \overset{\mathfrak{R}_N}{\sqsupseteq} B.$$

Thus, there must be some $m \in [1..N]$ such that $A \overset{\mathfrak{R}_{m-1}}{\sqsubseteq} B$, but $A \overset{\mathfrak{R}_m}{\sqsupseteq} B$. Hence eqn.(2.2) is true for $X = A$ and $Y = B$.

But the claim says that eqn.(2.2) will be true for *any* alternatives X and Y. To see this, we apply Claim 1, which says, in effect, that if we take any other alternatives A_1 and B_1, and substitute A_1 for A and B_1 for B_0 everywhere in the above construction, we will reach the same result, namely that eqn.(2.2) is true for $X = A_1$ and $Y = B_1$. . □ `[Claim 2]`

Claim 3: *v_m is a dictator.*

Proof: Let $A, B \in \mathcal{A}$. We want to show that $\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(A \underset{m}{\overset{\rho}{\succeq}} B \right)$ (regardless of what the other voters think).

Suppose that ρ is some profile. We will first show that

$$\left(A \underset{m}{\overset{\rho}{\succeq}} B \right) \Longrightarrow \left(A \overset{\rho}{\sqsupseteq} B \right) \tag{2.3}$$

To do this, we will construct a new profile δ, so that each voter's ρ-preferences concerning $\{A, B\}$ are identical to her δ-preferences. Thus, by **(IIA)**,

$$\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(A \overset{\delta}{\sqsupseteq} B \right).$$

We will build δ so that it is clear that $A \overset{\delta}{\sqsupseteq} B$. To do this, we introduce a third alternative, $C \in \mathcal{A}$. By axiom **(IIA)**, the position of C in the δ-preferences of voters $v_1, \ldots, v_n$ has no effect on whether $A \overset{\delta}{\sqsupseteq} B$ or $B \overset{\delta}{\sqsupseteq} A$. Hence, we can build δ so that:

- For all $n < m$, $\quad C \underset{n}{\overset{\delta}{\succeq}} A$ and $C \underset{n}{\overset{\delta}{\succeq}} B$.
- For all $n > m$, $\quad A \underset{n}{\overset{\delta}{\succeq}} C$ and $B \underset{n}{\overset{\delta}{\succeq}} C$.
- $A \underset{m}{\overset{\delta}{\succeq}} C \underset{m}{\overset{\delta}{\succeq}} B$.
- For every $n \in [1..N]$, $\quad \left(A \underset{n}{\overset{\delta}{\succeq}} B \right) \iff \left(A \underset{n}{\overset{\rho}{\succeq}} B \right)$.

We portray this schematically:

v_1	v_2	v_3	$\ldots$	v_{m-1}	v_m	v_{m+1}	v_{m+2}	v_{m+3}	$\ldots$	v_n
C	C	C	$\ldots$	C	A	A	B	A	$\ldots$	B
A	B	A	$\ldots$	B	C	B	A	B	$\ldots$	A
B	A	B	$\ldots$	A	B	C	C	C	$\ldots$	C

By setting $X = C$ and $Y = A$ in Claim 2, we get $A \overset{\delta}{\sqsupseteq} C$.

By setting $X = C$ and $Y = B$ in Claim 2, we get $C \overset{\delta}{\sqsupseteq} B$.

Hence, by transitivity, we conclude that $A \overset{\delta}{\sqsupseteq} B$. Then by **(IIA)** we also have $A \overset{\rho}{\sqsupseteq} B$.

Now, we can do this in any profile ρ where $A \underset{m}{\overset{\rho}{\succeq}} B$; hence we have shown (2.3).

By reversing the roles of A and B throughout the whole argument, we can likewise show that:

$$\left(B \underset{m}{\overset{\rho}{\succeq}} A \right) \Longrightarrow \left(B \overset{\rho}{\sqsupseteq} A \right).$$

Thus, v_m is a dictator. ________________ $\square$ `[Claim 3 & Theorem]`

Exercise 14 The conclusion of Claim 1 seems superficially different than the *Neutrality* axiom **(N)**, but in fact they are the same.

(a) Show that axiom **(N)** implies Claim 1.

(b) Show that Claim 1, together with axiom **(IIA)**, implies axiom **(N)**.

Discussion: Arrow's Impossibility Theorem says that no 'democratic' procedure can be constructed which is immune to distortion through the introduction of additional alternatives. Arrow's Theorem does *not* say 'democracy is impossible'; it merely says that any democracy will be inherently flawed by 'sensitivity to irrelevant alternatives'.

Indeed, it's not even clear that this sensitivity is a 'flaw'. Despite the terminology of 'impossibility' and 'dictators', Saari [33, §3.4.5,§3.4.9] interprets Arrow's Theorem in a very benign way, as simply saying that **(IIA)** is an unreasonable requirement for a voting procedure; furthermore it isn't even clear that **(IIA)** is desirable. Recall that, in the 'Arianne, Bryn, and Chloe' example on page 9 of §1.3, the 'irrelevant alternative' Chloe is perhaps not *really* irrelevant, because perhaps her presence in the competition provides information about the 'intensity' with which the voters support Arianne or Bryn (see §1.3).

Exercise 15 Which of the following voting procedures satisfy **(IIA)**? Which don't? Provide a proof/counterexample in each case.

1. Plurality Vote (Example 2.1(a)).

2. Borda Count (Example 2.1(b)).

3. Approval Voting Example 2.1(c).

4. Agenda of Pairwise votes (Example 2.2).

Further reading: The original exposition of Arrow's Theorem is [4], but it is discussed in pretty much any book on voting theory or social choice, eg. Sen [37, Chap.3] or Fishburn [13]. Taylor [50, 10.5] contains a very readable and elementary proof. Kim and Roush [20, 4.3] contains a proof using boolean matrix theory. Saari [33, §3.4] contains a proof by convex geometry methods, while Saari [34] contains an elementary introduction to the theorem. Luce and Raiffa [30, §14.4] gives a classical presentation in the context of game theory. Riker [32, §5A] discusses the theorem with application to political science.

The proof I've given here is adapted from the third of three short, elegant proofs by John Geanokoplos [15]; Geanokoplos also mentions that Luis Ubeda-Rives has a similar (unpublished) proof.

2.5 Strategic Voting: Gibbard & Satterthwaite

Prerequisites: §2.4

Recall from §1.3 that *strategic voting* means voting *contrary* to your true desires, because you believe that doing so will actually yield a *more* desirable outcome. To be precise, we must introduce some additional terminology.

Social choice functions: Suppose we have a strict voting procedure Π. For any strict profile $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$, the **leader** picked by ρ is the unique maximal element of of the ordering $\overset{\rho}{\sqsupset}$. We indicate the leader by $\chi_\Pi(\rho)$. Thus, if ρ describes the electorate's preferences in a presidential election, then $\chi_\Pi(\rho)$ is the person actually elected president.

The function χ_Π is an example of a social choice function. A **social choice function** is any function $\chi : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{A}$. In other words, a social choice function takes a profile of voter preferences as input, and yields a single alternative as output.

Any strict voting procedure yields a social choice function. If $\Pi : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}^*(\mathcal{A})$ is a strict voting procedure, then the **leadership function** $\chi_\Pi : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{A}$ (described in the first paragraph) is a social choice function. Conversely, given a social choice function χ, we can construct many strict voting procedures Π so that $\chi = \chi_\Pi$ (**Exercise 16**).

Strategic voting: Suppose $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ is some strict profile, Let $P \in \mathcal{P}^*(\mathcal{A})$ be a strict preference ordering, and let $v \in \mathcal{V}$ be a voter such that $\rho(v) = P$. Let $P' \in \mathcal{P}^*(\mathcal{A})$ be another strict preference ordering, and let δ be the profile we get if voter v 'pretends' to have preference order P'. In other words, for any $w \in \mathcal{V}$,

$$\delta(w) = \begin{cases} \rho(w) & \text{if} \quad v \neq w; \\ P' & \text{if} \quad v = w. \end{cases}$$

We say that P' is **strategic vote** for v if $\chi(\delta) \underset{P}{\succ} \chi(\rho)$. In other words, the alternative $\chi(\delta)$ (obtained if v *pretends* to have preference P') is *preferable* for v to the alternative $\chi(\rho)$ (obtained if v is honest).

We say that a social choice function is **nonmanipulable** (or **strategy proof**) if, for any profile $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$, no voter $v \in \mathcal{V}$ has a strategic vote. For example, the Borda Count is *not* strategy proof, as we saw in §1.3. This fact has often been used to discredit the Borda Count. However, the Gibbard-Satterthwaite Theorem (below) basically says that *any* democratic system is susceptible to strategic voting. Hence, the Borda Count is no worse than any other democratic procedure.

Dictatorship If χ is a social choice function, then a **dictator** for χ is a voter $v \in \mathcal{V}$ so that χ always picks v's favourite choice. In other words, for any profile $\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A})$, and any $A \in \mathcal{A}$,

$$\Big(\chi(\rho) = A \Big) \iff \Big(A \overset{\rho}{\underset{v}{\succ}} B, \text{ for all } B \in \mathcal{A} \Big).$$

We then say that χ is a **dictatorship**. It is easy to show:

Lemma 2.5 *Let Π be a strict voting procedure, with leadership function χ_Π. If v is the dictator of Π, then v is also the dictator of χ_Π.* □

Surjectivity The social choice function χ is **surjective** if, for any $A \in \mathcal{A}$, there is some profile ρ so that $A = \chi(\rho)$.

In a sense, surjectivity is a 'nontriviality' requirement: clearly, if the alternative A can never win, under any conditions, then why is A even in the competition? We can thus always assume that χ is surjective, because if it is not, we should simply remove any candidates who are never chosen.

Gibbard-Satterthwaite Impossibility Theorem: *The only surjective, nonmanipulable social choice function is a dictatorship.*

Proof: Suppose $\chi : \mathfrak{R}(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{A}$ is nonmanipulable and surjective. We will define a strict voting procedure Π so that $\chi = \chi_\Pi$. We will show that Π satisfies the axioms **(P)** (from §2.2) and **(IIA)** (from §2.4). We will then invoke Arrow's Impossibility Theorem (page 27).

To do this, we first need two technical lemmas. If $P \in \mathcal{P}^*(\mathcal{A})$ is a strict ordering on $\mathcal{A}$, and $A, B \in \mathcal{A}$, with $A \underset{P}{\succ} B$, then we say that A and B are P**-adjacent** if there is no alternative C with $A \underset{P}{\succ} C \underset{P}{\succ} B$. The $\{A, B\}$**-transposition** of P is the new ordering P' obtained from P by simply exchanging A and B in the ordering of P, while keeping all other alternatives the same:

$$\begin{array}{rl} P: & C_1 \underset{P}{\succ} C_2 \underset{P}{\succ} \cdots \underset{P}{\succ} C_j \underset{P}{\succ} \quad A \underset{P}{\succ} B \quad \underset{P}{\succ} D_1 \underset{P}{\succ} \cdots \underset{P}{\succ} D_k \\ P': & C_1 \underset{P'}{\succ} C_2 \underset{P'}{\succ} \cdots \underset{P'}{\succ} C_j \underset{P'}{\succ} \quad B \underset{P'}{\succ} A \quad \underset{P'}{\succ} D_1 \underset{P'}{\succ} \cdots \underset{P'}{\succ} D_k \end{array}$$

Let $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ be some profile, and let $v \in \mathcal{V}$ be a voter with $\rho(v) = P$. Define ρ' by:

$$\rho'(w) \;=\; \begin{cases} \rho(w) & \text{if} \quad v \neq w; \\ P' & \text{if} \quad v = w. \end{cases}$$

We say that ρ' is a **transposition** of ρ, which **promotes** B, **demotes** A, and **fixes** all other alternatives.

Claim 1: *Let $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$. Let $A, B \in \mathcal{A}$, and let ρ' be a transposition of ρ which promotes B and demotes A.*

(a) *If $\chi(\rho) = A$, then $\chi(\rho') = A$ or B.*

(b) *If $\chi(\rho) \neq A$ (in particular, if $\chi(\rho) = B$) then $\chi(\rho') = \chi(\rho)$.*

Proof: **(a)** By contradiction, suppose $\chi(\rho') = C$, where $C \notin \{A, B\}$. Thus, either $C \underset{P}{\succ} A$ or $C \underset{P}{\prec} A$.

If $C \underset{P}{\succ} A$, then P' is a strategic vote for the voter v (in profile ρ), contradicting nonmanipulability of χ.

If $C \underset{P}{\prec} A$, then also $C \underset{P}{\prec} B$ (because A and B are adjacent) and then $C \underset{P'}{\prec} A$ (by definition of P'). Thus, P is a strategic vote for the voter v (in profile ρ'), again contradicting nonmanipulability of χ.

(b) Suppose $\chi(\rho) = C \neq A$ and $\chi(\rho') = C'$. We want to show that $C = C'$. Suppose by contradiction that $C \neq C'$. Then either $C \underset{P}{\prec} C'$ or $C \underset{P}{\succ} C'$.

If $C \underset{P}{\prec} C'$, then P' is a strategic vote for voter v in profile ρ.

If $A \neq C \underset{P}{\succ} C'$, then also $C \underset{P'}{\succ} C'$ (**Exercise 17** Hint: A and B are adjacent); hence P is a strategic vote for voter v in profile ρ'. .... □ `[Claim 1]`

Now, if $\rho, \delta \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ are two strict profiles, and $B \in \mathcal{A}$, then a B-**promoting walk** from δ to ρ is a sequence of strict profiles

$$\delta \;=\; \delta_0,\ \delta_1,\ \delta_2, \ldots, \delta_K \;=\; \rho$$

where, for each $k \in [1...K]$, δ_k is a transposition of δ_{k-1} which either promotes B or fixes B.

Claim 2: *Let $\rho, \delta \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ be two strict profiles.*

(a) *Suppose $\chi(\delta) = B$. If there is a B-promoting walk from δ to ρ, then $\chi(\rho) = B$, also.*

(b) *If B is the maximal element in every voter's ρ-preference, then there is a B-promoting walk from δ to ρ.*

Proof: **(a)** By repeated application of Claim 1(b), we have

$$B \;=\; \chi(\delta_0) \;=\; \chi(\delta_1) \;=\; \cdots \;=\; \chi(\delta_K) \;=\; \chi(\rho).$$

(b) follows from the definition of a B-promoting walk. .... □ `[Claim 2]`

We use Claim 2 to prove that χ satisfies a kind of 'Pareto' property.

Claim 3: *Let $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ be a strict profile.*

(a) *If B is the maximal element in every voter's ρ-preference, then $\chi(\rho) = B$.*

(b) *If A and B are the 'top two' alternatives in every voter's ρ-preference, then either $\chi(\rho) = A$ or $\chi(\rho) = B$.*

Proof: **(a)** By hypothesis, χ is *surjective*, so there is *some* profile δ such that $\chi(\delta) = B$. Now, Claim 2(b) yields a B-promoting walk from δ to ρ; then Claim 2(a) says $\chi(\rho) = B$.

(b) Let ρ be a profile so that every voter ranks A and B as her 'top two' alternatives. Let δ be the modified profile where every voter ranks B *first* and ranks A *second*, and ranks all other alternatives the same as in ρ. Thus, B is maximal in every voter's δ-preference, so part **(a)** implies that $\chi(\delta) = B$. But there is an A-promoting walk to get from δ to ρ, so Claim 1(b) implies that either $\chi(\rho) = A$ or $\chi(\rho) = B$. .... □ `[Claim 3]`

Now, suppose $\mathcal{A} = \{A_1, \ldots, A_N\}$, where the alternatives are numbered in some entirely arbitrary order. Given any strict profile $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$, and any pair of alternatives $\{B, C\}$, we define the profile $\rho_{B,C}$ as follows:

- Each voter $v \in \mathcal{V}$ places B and C as her top two alternatives, ranked in the same order as she ranked them in ρ.
- Each voter then ranks all remaining candidates in decreasing numerical order.

To illustrate this, suppose $\mathcal{A} = \{A_1, A_2, \ldots, A_7\}$, and that $B = A_3$ and $C = A_6$. Then we have the following:

Before (in ρ)	After (in $\rho_{B,C}$)
$\cdots \succ B \succ \cdots \succ C \succ \cdots$	$B \succ C \succ A_1 \succ A_2 \succ A_4 \succ A_5 \succ A_7$
$\cdots \succ C \succ \cdots \succ B \succ \cdots$	$C \succ B \succ A_1 \succ A_2 \succ A_4 \succ A_5 \succ A_7$

Now, we define a strict voting procedure $\Pi : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}^*(\mathcal{A})$ as follows. For any strict profile ρ, we define the relation $\stackrel{\rho}{\sqsupset}$ by

$$\left(A \stackrel{\rho}{\sqsupset} B \right) \iff \left(\chi(\rho_{A,B}) = A \right)$$

We don't yet know that $\stackrel{\rho}{\sqsupset}$ is even a preference ordering. However, we can already show:

Claim 4: *The procedure* Π *satisfies axiom* **(IIA)**. *In other words, if $A, B \in \mathcal{A}$, and ρ and δ are two profiles such that*

$$\text{For all } v \in \mathcal{V}, \quad \left(A \underset{v}{\stackrel{\rho}{\succ}} B \right) \iff \left(A \underset{v}{\stackrel{\delta}{\succ}} B \right), \tag{2.4}$$

then we have: $\left(A \stackrel{\rho}{\sqsupset} B \right) \iff \left(A \stackrel{\delta}{\sqsupset} B \right)$.

Proof: Suppose ρ and δ satisfy eqn.(2.4). Then, by definition, $\rho_{A,B} = \delta_{A,B}$. Hence, $\chi(\rho_{A,B}) = \chi(\delta_{A,B})$. Hence $\left(A \stackrel{\rho}{\sqsupset} B \right) \iff \left(A \stackrel{\delta}{\sqsupset} B \right)$. . . . . $\Box$ `[Claim 4]`

To show that Π is a strict voting procedure, we must show:

Claim 5: *For any $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$, the relation $\stackrel{\rho}{\sqsupset}$ is a strict preference ordering.*

Proof: *Complete:* By Claim 3(b), either $\chi(\rho_{A,B}) = A$ or $\chi(\rho_{A,B}) = B$. Hence, either $A \stackrel{\rho}{\sqsupset} B$ or $B \stackrel{\rho}{\sqsupset} A$.

Antisymmetric: Clearly, it is impossible to have both $\chi(\rho_{A,B}) = A$ and $\chi(\rho_{A,B}) = B$. Thus, it is impossible to have both $A \stackrel{\rho}{\sqsupset} B$ and $B \stackrel{\rho}{\sqsupset} A$.

Transitive: Suppose $A \stackrel{\rho}{\sqsupset} B \stackrel{\rho}{\sqsupset} C$. We must show that $A \stackrel{\rho}{\sqsupset} C$. Suppose by contradiction that $C \stackrel{\rho}{\sqsupset} A$, so that we have a cycle

$$A \stackrel{\rho}{\sqsupset} B \stackrel{\rho}{\sqsupset} C \stackrel{\rho}{\sqsupset} A. \tag{2.5}$$

Define the strict profile $\delta \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ such that:

- Each voter $v \in \mathcal{V}$ places A, B and C as her top three alternatives, ranked in the same order as she ranked them in ρ.
- Each voter then ranks all remaining candidates in decreasing numerical order.

Then Claim 4 and eqn.(2.5) imply that $A \overset{\delta}{\sqsupset} B \overset{\delta}{\sqsupset} C \overset{\delta}{\sqsupset} A$.

Claim 5.1: $\chi(\delta) \in \{A, B, C\}$.

Proof: Suppose, by way of contradiction, that $\chi(\delta) = D$, where $D \notin \{A, B, C\}$. We define a D-promoting walk from δ to the profile $\rho_{A,B}$, by simply 'walking' C down each voter's list of preferences, one transposition at a time. Each transposition fixes D, except for at most one transposition which switches C and D (thereby promoting D). We conclude, from Claim 2(a), that $\chi(\rho_{A,B}) = \chi(\delta) = D$. But $A \overset{\delta}{\sqsupset} B$, which means, by definition, that $\chi(\rho_{A,B}) = A$. Contradiction. .... □ `[Claim 5.1]`

Thus, $\chi(\delta) \in \{A, B, C\}$. Assume $\chi(\delta) = A$. (Since the cycle $A \overset{\delta}{\sqsupset} B \overset{\delta}{\sqsupset} C \overset{\delta}{\sqsupset} A$ is symmetrical, the following argument will also work if $\chi(\delta) = B$ or C).

Claim 5.2: $A \overset{\rho}{\sqsupset} C$.

Proof: We define an A-promoting walk from δ to the profile $\rho_{A,C}$, by simply 'walking' B down each voter's list of preferences, one transposition at a time. Each transposition fixes A, except for at most one transposition, which switches A and B (thereby promoting A). We conclude, from Claim 2(a), that $\chi(\rho_{A,C}) = \chi(\delta) = A$. But, by definition of Π, this means that $A \overset{\rho}{\sqsupset} C$. .... □ `[Claim 5]`

Claim 5.2 contradicts the assumption that $C \overset{\rho}{\sqsupset} A$. By contradiction, the cyclical ordering in eqn.(2.5) is impossible. Thus, $\overset{\rho}{\sqsupset}$ must be transitive. .... □ `[Claim 5]`

Thus, $\overset{\rho}{\sqsupset}$ is always a strict preference ordering, for any $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$. Hence, Π is a strict voting procedure.

Claim 6: *The procedure* Π *satisfies the* Pareto *axiom* **(P)**.

Proof: Let $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ and suppose $A, B \in \mathcal{A}$, are such that $A \underset{v}{\overset{\rho}{\succ}} B$ for all $v \in \mathcal{V}$. We must show that $A \overset{\rho}{\sqsupset} B$ —ie. we must show that $\chi(\rho_{A,B}) = A$

But $A \underset{v}{\overset{\rho}{\succ}} B$ for all $v \in \mathcal{V}$, so it follows that, also, $A \underset{v}{\overset{\rho_{A,B}}{\succ}} B$ for all $v \in \mathcal{V}$. Hence, A is the *maximal* element in every voter's preferences under $\rho_{A,B}$. Thus, Claim 3**(a)** says that $\chi(\rho_{A,B}) = A$. Hence $A \overset{\rho}{\sqsupset} B$. .... □ `[Claim 6]`

By combining Claims 4 and 6 with Arrow's Impossibility Theorem, we conclude that the procedure Π is a dictatorship.

Claim 7: χ *is the leadership function for* Π.

Proof: Let $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ and suppose $\chi(\rho) = A$. We must show that A is the maximal element of the ordering $\overset{\rho}{\sqsupset}$ generated by Π. In other words, for any $B \in \mathcal{A}$, we must show that $A \overset{\rho}{\sqsupset} B$, or, equivalently, that $\chi(\rho_{A,B}) = A$. To see this, we define an A-promoting walk from ρ to $\rho_{A,B}$ as follows:

1. For any $v \in \mathcal{V}$ such that $B \underset{v}{\overset{\rho}{\succ}} A$, we 'walk' B to the top of v's preference ordering with a sequence of transpositions that fix A.
2. For any $v \in \mathcal{V}$ such that $A \underset{v}{\overset{\rho}{\succ}} B$, we 'walk' A to the top of v's preference ordering with a sequence of transpositions that promote A.
3. Finally, for every $v \in \mathcal{V}$, we rearrange all remaining alternatives in decreasing numerical order, with a sequence of transpositions that fix both A and B.

Thus, Claim 2(B) implies that $\chi(\rho_{A,B}) = A$. Hence $A \overset{\rho}{\sqsupset} B$. This is true for any $B \in \mathcal{A}$, so A is maximal. .... $\Box$ `[Claim 7]`

Let $v \in \mathcal{V}$ be the dictator of Π. Then Lemma 2.5 says v is also the dictator of χ. $\Box$

One can actually conduct a more sophisticated analysis, and measure how 'susceptible' various voting procedures are to manipulation. Intuitively, the 'susceptibility' of a voting procedure is measured by the probability of a scenario where a small number of voters can change the outcome through strategic voting. Let $\mathfrak{R}(\mathcal{V}, \mathcal{A})$ be the 'space' of all possible profiles; then any voting procedure partitions $\mathfrak{R}(\mathcal{V}, \mathcal{A})$ into regions corresponding to different outcomes. The strategic voting opportunities occur along the *boundaries* between these regions. By defining the 'susceptibility' of a voting procedure in terms of the size of these boundaries, one can actually prove:

Theorem 2.6 [33, §5.3] *The voting procedure* <u>*least*</u> *susceptible to manipulation is the Borda count. The voting procedure* <u>*most*</u> *susceptible to manipulation is the plurality vote.* $\Box$

Further reading: The Gibbard-Satterthwaite theorem was proved independently by Gibbard [16] and Satterthwaite [35]. Other discussions are Gärdenfors [14], Saari [33, §5.1], and Kim and Roush [20, §4.4].

The proof of Gibbard-Satterthwaite given here was adapted from the proof by Sonnenschein and Schmeidler (1974) as transmitted by Kim and Roush [20, Thm 4.4.3]. The original proof doesn't require χ to be surjective, but (at the expense of additional technicalities) uses the weaker assumption that χ takes on at least three distinct values.

Chapter 3

Binary Voting Procedures

Democracy is a device that insures we shall be governed no better than we deserve.
—George Bernard Shaw

Strictly speaking, democracy only insures that the *majority* will be governed no better than they deserve; the rest of us will also be governed no better than they deserve. Even this is only true when a true majority has chosen the government or policies in question, which rarely occurs when there are three or more alternatives. We can only ensure that the democratic choice is the *majority* choice when there are only two alternatives to choose from. In this chapter, we study such *binary* voting procedures.

A *binary choice* is a choice between two alternatives (eg. 'yes' or 'no'). Thus, we say that a social choice function $\Pi : \mathfrak{R}(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{A}$ is **binary** if $\#(\mathcal{A}) = 2$. Likewise, a voting procedure $\Pi : \mathfrak{R}(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}(\mathcal{A})$ is **binary** if $\#(\mathcal{A}) = 2$. Note that any *strict* binary voting procedure determines a binary social choice function and vice versa; the two concepts are semantically equivalent.

All of the pathological scenarios from Chapter 1 (eg. the Condorcet paradox, strategic voting opportunities) require at least three alternatives. So do the impossibility theorems of Sen, Arrow, Gibbard-Satterthwaite (§2.3 §2.4, and §2.5). Hence, if there are only two alternatives, we can hope for a satisfactory voting theory which is not full of depressing negative results.

3.1 Simple Majority Voting: May's Theorem

Prerequisites: §2.1 **Recommended:** §2.2

The most obvious social choice function for two alternatives is the simple majority vote. The more elaborate voting procedures (Borda count, pairwise votes, approval voting, etc.) all reduce to the majority vote when $|\mathcal{A}| = 2$. Indeed, the conventional wisdom says that majority vote is the 'only' sensible democratic procedure for choosing between two alternatives. The good news is that, for once, the conventional wisdom is right.

Suppose that $\mathcal{A} = \{A, B\}$. In §2.2 we introduced three desiderata which any 'reasonable' voting procedure should satisfy:

(M) (*Monotonicity*) Let ρ be a profile such that $A \overset{\rho}{\sqsubseteq} B$. Let $v \in \mathcal{V}$ be some voter such that $B \underset{v}{\overset{\rho}{\preceq}} A$, and let δ be the profile obtained from ρ by giving v a new preference ordering $\underset{v}{\overset{\delta}{\succeq}}$, such that $A \underset{v}{\overset{\delta}{\succeq}} B$ (all *other* voters keep the same preferences). Then $A \overset{\delta}{\sqsubseteq} B$.

(A) (*Anonymity*) Let $\sigma : \mathcal{V} \longrightarrow \mathcal{V}$ be a permutation of the voters. Let ρ be a profile, and let δ be the profile obtained from ρ by permuting the voters with σ. In other words, for any $v \in \mathcal{V}$, $\delta(v) = \rho\Big(\sigma(v)\Big)$. Then $\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(A \overset{\delta}{\sqsupseteq} B \right)$.

(N) (*Neutrality*) Let ρ be a profile, and let δ be the profile obtained from ρ by reversing the positions of A and B for each voter. In other words, for any $v \in \mathcal{V}$,

$$\left(A \underset{v}{\overset{\rho}{\succeq}} B \right) \iff \left(B \underset{v}{\overset{\delta}{\succeq}} A \right).$$

Then the outcome of δ is the reverse of the outcome of ρ. That, is, for any $B, C \in \mathcal{A}$,

$$\left(A \overset{\rho}{\sqsupseteq} B \right) \iff \left(B \overset{\delta}{\sqsupseteq} A \right).$$

A *semistrict* voting procedure is a function $\Pi : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}(\mathcal{A})$. Thus, the voters must provide *strict* preferences as input, but the output might have ties. Let $V := \#(\mathcal{V})$. We say Π is a **quota system** if there is some $Q \in [0..V]$ so that, for any $\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A})$,

(Qa) If $\#\left\{ v \in \mathcal{V} \,;\, A \underset{v}{\overset{\rho}{\succ}} B \right\} > Q$, then $A \overset{\rho}{\sqsupset} B$.

(Qb) If $\#\left\{ v \in \mathcal{V} \,;\, B \underset{v}{\overset{\rho}{\succ}} A \right\} > Q$, then $B \overset{\rho}{\sqsupset} A$.

(Qc) If neither of these is true, then $A \overset{\rho}{\approx} B$.

For example:

- The **simple majority vote** is a quota system where $Q = V/2$.
- The **two thirds majority vote** is a quota system where $Q = \frac{2}{3}V$. If an alternative does not obtain at least two thirds support from the populace, then it is not chosen. If *neither* alternative gets two thirds support, then *neither* is chosen; the result is a 'tie'.
- The **unanimous vote** is a quota system where $Q = V - 1$. Thus, an alternative must be receive unanimous support to be chosen. This is the system used in courtroom juries.

- The **totally indecisive** system is one where $Q = V$; hence condition **(Qc)** is always true, and we always have a tie.

Note that the quota Q must be no less than $V/2$. If $Q < V/2$, then it is theoretically possible to satisfy conditions **(Qa)** and **(Qb)** simultaneously, which would be a contradiction (since **(Qa)** implies a *strict* preference of A over B, and **(Qb)** implies the opposite).

If V is odd, and we set $Q = V/2$ (the simple majority system) then condition **(c)** is never satisfied. In other words, ties never occur, so we get a *strict* voting procedure.

Theorem 3.1 *Any semistrict binary voting system satisfying* **(A)**, **(N)** *and* **(M)** *is a quota system.*

Proof: For any $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$, let

$$\alpha(\rho) \quad = \quad \#\left\{v \in \mathcal{V}\ ;\ A \underset{v}{\overset{\rho}{\succ}} B\right\} \qquad \text{and} \qquad \beta(\rho) \quad = \quad \#\left\{v \in \mathcal{V}\ ;\ B \underset{v}{\overset{\rho}{\succ}} A\right\}.$$

Now, suppose $\Pi : \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \longrightarrow \mathcal{P}(\mathcal{V}, \mathcal{A})$ satisfies **(A)**, **(N)** and **(M)**. Since Π is anonymous, the outcome is determined entirely by the number of voters who prefer A to B, and the number who prefer B to A. In other words, $\Pi(\rho)$ is determined entirely by $\alpha(\rho)$ and $\beta(\rho)$.

However, the voters must provide *strict* preferences, so we also know that $\beta(\rho) = V - \alpha(\rho)$. Thus, $\Pi(\rho)$ is is really determined by $\alpha(\rho)$. Hence, there is some function $\widetilde{\Pi} : \mathbb{N} \longrightarrow \mathcal{P}(\mathcal{A})$ such that $\Pi(\rho) \ = \ \widetilde{\Pi}\left(\alpha(\rho)\right)$, for any $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$.

Claim 1: *Suppose $\rho, \delta \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ are two profiles, such that $\alpha(\delta) \geq \alpha(\rho)$. Then*

$$\left(A \overset{\rho}{\sqsupset} B \right) \Longrightarrow \left(A \overset{\delta}{\sqsupset} B \right).$$

Proof: <u>**Exercise 18**</u> Hint: use the Monotonicity axiom **(M)**. .... □ `[Claim 1]`

Let $Q \quad = \quad \min\left\{q \in \mathbb{N}\ ;\ \text{there exists some } \rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A}) \text{ such that } \alpha(\rho) = q \text{ and } A \overset{\rho}{\sqsupset} B\right\}.$

Claim 2: *If $\delta \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ is any profile, then* $\left(A \overset{\delta}{\sqsupset} B \right) \iff \left(\alpha(\delta) \geq Q \right).$

Proof: '$\Longrightarrow$' is true by definition of Q.

'$\Longleftarrow$': By definition, there is some profile $\rho \in \mathfrak{R}^*(\mathcal{V}, \mathcal{A})$ such that $\alpha(\rho) = Q$ and $A \overset{\rho}{\sqsupset} B$. Thus, if $\alpha(\delta) \geq Q = \alpha(\rho)$, then Claim 1 implies that $A \overset{\delta}{\sqsupset} B$. .... □ `[Claim 2]`

Claim 3: *If $\delta \in \mathfrak{R}(\mathcal{V}, \mathcal{A})$ is any profile, then* $\left(B \overset{\delta}{\sqsupset} A \right) \iff \left(\beta(\delta) \geq Q \right).$

Proof: <u>**Exercise 19**</u> Hint: use Claim 2 and the Neutrality axiom **(N)**. ... □ `[Claim 3]`

Claim 2 says Q satisfies property **(Qa)** of a quota system. Claim 3 says Q satisfies property **(Qb)** of a quota system. Claims 2 and 3 together imply that Q satisfies property **(Qc)** of a quota system. □

Corollary 3.2 May's Theorem (1952) [28]

Suppose $V = \#(\mathcal{V})$ is odd. Then the only strict binary voting system satisfying **(A)**, **(N)** *and* **(M)** *is the simple majority vote.*

Proof: The previous theorem says that any binary voting system satisfying **(A)**, **(N)** and **(M)** must be a quota system. To ensure that this is a *strict* voting system, we must guarantee that ties are impossible; ie. that condition **(Qc)** never occurs. This can only happen if V is odd and we set $Q = V/2$ —ie. the we have the simple majority vote. □

Further reading: May's theorem first appeared in [28]. A good discussion is in Taylor [50, §10.3].

3.2 Weighted Voting Systems

Prerequisites: §2.1 **Recommended:** §2.2

Suppose $\mathcal{A} = \{\mathit{Yes}, \mathit{No}\}$, where we imagine *Yes* to be some proposal (eg. new legislation) and *No* to be the negation of this proposal (eg. the status quo). We can assume that any a binary voting procedure to decide between *Yes* and *No* must be *strict*, because in a tie between *Yes* and *No*, the 'default' choice will be *No*. Observe that such a procedure will generally *not* satisfy neutrality axiom **(N)**, because the status quo (*No*) is favoured over novelty (*Yes*).

A **weighted** voting system is a strict binary voting procedure where the votes of different voters have different 'weights'. To be precise, there is a **weight function** $\omega : \mathcal{V} \longrightarrow \mathbb{N}$ so that, for any $\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A})$, the *total support* for alternative *Yes* is defined:

$$\Upsilon(\rho) \quad = \quad \sum_{y \in \mathcal{Y}(\rho)} \omega(y) \qquad \text{where} \quad \mathcal{Y}(\rho) \;=\; \left\{ v \in \mathcal{V}\ ;\ \mathit{Yes} \underset{v}{\overset{\rho}{\succ}} \mathit{No} \right\}.$$

The **total weight** of the system is $W = \sum_{v \in \mathcal{V}} \omega(v)$. Finally, we set a **quota** $Q \in [0...W]$ so that

$$\left(\mathit{Yes} \overset{\rho}{\sqsupset} \mathit{No} \right) \iff \left(\Upsilon(\rho) \geq Q \right).$$

Most systems favour the 'status quo' alternative (*No*), which means that $Q \geq \frac{1}{2} W$.

Example 3.3: The European Economic Community

The Treaty of Rome (1958) defined a voting procedure with six voters:

$$\mathcal{V} \;=\; \{\text{France}, \text{Germany}, \text{Italy}, \text{Belgium}, \text{Netherlands}, \text{Luxembourg}\}$$

and the following weights:

$$\begin{array}{rcrcrcl} \omega(\text{France}) & = & \omega(\text{Germany}) & = & \omega(\text{Italy}) & = & 4 \\ & & \omega(\text{Belgium}) & = & \omega(\text{Netherlands}) & = & 2 \\ & & & & \omega(\text{Luxembourg}) & = & 2 \end{array}$$

Thus, $W \;=\; 4+4+4+2+2+1 \;=\; 17$. The quota was set at $Q = 12$. Thus, for example, Yes would be chosen if it had the support of France, Germany, and Italy because

$$\Big(\ \mathcal{Y}(\rho) \;=\; \{\text{France, Germany, Italy}\}\ \Big) \Longrightarrow \Big(\ \Upsilon(\rho) \;=\; 4+4+4 \;=\; 12 \;\geq\; Q\ \Big) \Longrightarrow \Big(\ \mathit{Yes} \overset{\rho}{\sqsupseteq} \mathit{No}\ \Big).$$

However, Yes would not be chosen if it only had the support of France, Germany, Belgium, and Luxembourg, because

$$\Big(\ \mathcal{Y}(\rho) \;=\; \{\text{France, Germany, Belgium, Luxembourg}\}\ \Big) \Longrightarrow$$
$$\Big(\ \Upsilon(\rho) \;=\; 4+4+2+1 \;=\; 11 \;<\; Q\ \Big) \Longrightarrow \Big(\ \mathit{Yes} \overset{\rho}{\sqsubseteq} \mathit{No}\ \Big).$$

Example 3.4: United Nations Security Council

According to the Charter of the United Nations, the UN Security Council consists of five *permanent members*

$$\mathcal{V}^* \quad = \quad \{\text{U.S.A., U.K., France, Russia, China}\}$$

along with ten *nonpermanent members* (positions which rotate amongst all member nations). Thus, the set $\mathcal{V}$ has fifteen members in total. Approval of a resolution requires two conditions:

(a) The support of at least 9 out of 15 Security Council members.

(b) The support of *all five* permanent members (any permanent member has a veto).

The 'veto' clause in condition **(b)** suggests that the Security Council is *not* a weighted voting system, but it actually is. First, note that conditions **(a)** and **(b)** can be combined as follows:

(c) A resolution is approved if and only if it has the support of *all five* permanent members, and *at least four* nonpermanent members.

We assign *seven* votes to every permanent member, and *one* vote to every nonpermanent member, and set the quota at 39. That is:

$$\begin{array}{rcll} \omega(v) & = & 7, & \text{for all } v \in \mathcal{V}^* \\ \omega(v) & = & 1, & \text{for all } v \in \mathcal{V} \setminus \mathcal{V}^*. \\ Q & = & 39. & \end{array}$$

Now let $\mathcal{Y}(\rho)$ be the set of voters supporting proposal Yes. Suppose $\mathcal{Y}(\rho) \not\supset \mathcal{V}^*$ (ie. at least one permanent member is 'vetoing'). Even if *every* other nation supports the resolution, we still have

$$\Upsilon(\rho) \ \leq \ 4 \times 7 \ + \ 10 \times 1 \ = \ 28 + 10 \ = \ 38 \ < \ 39 \ = \ Q.$$

Hence the resolution is not approved.

Now suppose $\mathcal{Y}(\rho) \ \subset \ \mathcal{V}^*$. Then

$$\Upsilon(\rho) \ = \ 5 \times 7 \ + \ N \times 1 \ = \ 35 + N,$$

where N is the number of nonpermanent members supporting the resolution. Thus,

$$\Big(\ \Upsilon(\rho) \geq Q \ \Big) \iff \Big(\ 35 + N \geq 39 \ \Big) \iff \Big(\ N \geq 4 \ \Big)$$

Thus, once all five permanent members support the resolution, it will be approved if and only if it also has the support of at least four nonpermanent members, exactly as required by **(c)**.

Example 3.5: Factionalism: Block voting and party discipline

Sometimes even supposedly 'nonweighted' voting systems can behave like weighted systems, because the voters organize themselves into factions, which synchronize their votes on particular issues. In this case, it is no longer correct to model the electorate as a large population of individual voters with equal weight; instead, we must model the electorate as a small number of competing factions, whose votes are 'weighted' in proportion to the size of their membership. Two examples of this phenomenon are *block voting* and *party discipline.*

Block Voting: Ideological groups (eg. labour unions, religious organizations, etc.) with a highly dedicated membership often dictate to their members how to vote in particular issues. Assuming that the group members are mostly obedient to the voting instructions of their leadership, the entire group can be treated as a single voting block.

Party Discipline: Modern elections involve hugely expensive advertising campaigns, and it is difficult to be elected without access to a powerful campaign finance machine. Thus, an individual politician must affiliate herself to some political *party*, which she depends upon to bankroll her campaigns. Political parties thus effectively 'own' their member politicians, and can dictate how their members vote on particular issues. A politician who defies her party might be denied access to crucial campaign funding. In a parliamentary system, the currently governing party can also reward 'loyalty' through prestigious cabinet positions and patronage appointments. These mechanisms guarantee that the party can generally be treated as a unified voting block.

Theorem 3.6 *Let Π be a weighted voting system.*

(a) Π *always satisfies the* **monotonicity** *axiom* **(M)** *and* **Pareto** *axiom* **(P)**.

(b) Π *satisfies* **anonymity** *axiom* **(A)** *if and only if* ω *is a constant (ie. all voters have the same weight).*

(c) Π *satisfies* **neutrality** *axiom* **(N)** *if and only if* $Q = \frac{1}{2}W$.

(d) Π *is a* **dictatorship** *if and only if there is some* $v \in \mathcal{V}$ *whose vote 'outweighs' everyone else combined; ie.* $\rho(v) > \sum_{w \neq v} \omega(w)$.

Proof: **Exercise 20** $\square$

In any binary voting system (weighted or not), a **winning coalition** is a collection of voters $\mathcal{W} \subset \mathcal{V}$ so that

$$\Big(\mathcal{Y}(\rho) \;=\; \mathcal{W} \Big) \Longrightarrow \Big(\mathcal{Y}\!es \;\overset{\rho}{\sqsupset}\; \mathcal{N}\!o \Big).$$

Thus, for example, in a weighted voting system, $\mathcal{W}$ is a winning coalition iff $\sum_{w \in \mathcal{W}} \omega(w) \;\geq\; Q$.

A voting system is called **trade robust** if the following is true: If $\mathcal{W}_1$ and $\mathcal{W}_2$ are two winning coalitions, and they 'swap' some of their members, then at least one of the new coalitions will still be winning. In other words, given any subsets $\mathcal{U}_1 \subset \mathcal{W}_1$ and $\mathcal{U}_2 \subset \mathcal{W}_2$ (where $\mathcal{U}_1$ is disjoint from $\mathcal{W}_2$, while $\mathcal{U}_2$ is disjoint from $\mathcal{W}_1$) if we define

$$\mathcal{W}_1' \;=\; \mathcal{U}_2 \sqcup \mathcal{W}_1 \setminus \mathcal{U}_2 \quad \text{and} \quad \mathcal{W}_2' \;=\; \mathcal{U}_2 \sqcup \mathcal{W}_2 \setminus \mathcal{U}_1$$

...then at least one of $\mathcal{W}_1'$ or $\mathcal{W}_2'$ is *also* a winning coalition.

Theorem 3.7 (Taylor & Zwicker, 1992) [2]
Let Π *be a binary voting system. Then* $\Big(\Pi$ *is a weighted system* $\Big) \iff \Big(\Pi$ *is trade robust* $\Big)$.

Proof: '$\Longrightarrow$' If $\mathcal{W}_1$ and $\mathcal{W}_2$ are winning coalitions, then $\sum_{w \in \mathcal{W}_1} \omega(w) \;\geq\; Q$ and $\sum_{w \in \mathcal{W}_2} \omega(w) \;\geq\; Q$.

Let $\mathcal{U}_1 \subset \mathcal{W}_1$ and $\mathcal{U}_2 \subset \mathcal{W}_2$ be arbitrary subsets. Observe that

$$\sum_{w \in \mathcal{W}_1'} \omega(w) \;=\; \sum_{u \in \mathcal{U}_2} \omega(u) \;+\; \sum_{w \in \mathcal{W}_1} \omega(w) \;-\; \sum_{u \in \mathcal{U}_1} \omega(u),$$

$$\text{and} \quad \sum_{w \in \mathcal{W}_2'} \omega(w) \;=\; \sum_{u \in \mathcal{U}_1} \omega(u) \;+\; \sum_{w \in \mathcal{W}_2} \omega(w) \;-\; \sum_{u \in \mathcal{U}_2} \omega(u)$$

Suppose that $\sum_{u \in \mathcal{U}_1} \omega(u) \;\geq\; \sum_{u \in \mathcal{U}_2} \omega(u)$. Then $\sum_{u \in \mathcal{U}_1} \omega(u) \;-\; \sum_{u \in \mathcal{U}_2} \omega(u) \geq 0$. Thus,

$$\sum_{w \in \mathcal{W}_2'} \omega(w) \;=\; \left(\sum_{u \in \mathcal{U}_1} \omega(u) - \sum_{u \in \mathcal{U}_2} \omega(u) \right) \;+\; \sum_{w \in \mathcal{W}_2} \omega(w) \;\geq\; \sum_{w \in \mathcal{W}_2} \omega(w) \;\geq\; Q.$$

so $\mathcal{W}_2'$ is still a winning coalition. On the other hand, if $\sum_{u\in\mathcal{U}_2}\omega(u) \geq \sum_{u\in\mathcal{U}_1}\omega(u)$, then symmetric reasoning shows that $\mathcal{W}_1'$ is still a winning coalition.

'$\Longleftarrow$' See Taylor and Zwicker [2]. □

Not all binary voting procedures are weighted systems, because not all systems are trade robust.

Example 3.8: Amendment Formula for Canadian Constitution

To approve an amendment to the Canadian Constitution, the amendment must have the support of at least seven out of ten provinces, which together must represent at least 50% of the Canadian population. For the sake of argument, say the populations are as follows:

Ontario: 30%

Quebec: 30%

B.C.: 10%

Alberta, Manitoba & Saskatchewan: 15%

New Brunswick & Nova Scotia: 10%

P.E.I. & Newfoundland: 5%

Now consider the following coalitions:

$$\begin{aligned}
\mathcal{W}_1 &= \{\text{Ontario, B.C., Alberta, Manitoba, Saskatchewan, P.E.I., Newfoundland }\} \\
&\quad \text{7 members,} \quad \textbf{Total weight:} \quad 30+10+15+5 \;=\; 60\% \\
\mathcal{W}_2 &= \{\text{Quebec, B.C., Alberta, Manitoba, Saskatchewan, New Brunswick, Nova Scotia}\} \\
&\quad \text{7 members,} \quad \textbf{Total weight:} \quad 30+10+15+10 \;=\; 65\%.
\end{aligned}$$

Now, let $\mathcal{U}_1 = \{\text{P.E.I., Newfoundland}\}$ and $\mathcal{U}_2 = \{\text{Quebec}\}$, so that

$$\begin{aligned}
\mathcal{W}_1' &= \{\text{Ontario, Quebec, B.C., Alberta, Manitoba, Saskatchewan}\} \\
&\quad \text{6 members,} \quad \textbf{Total weight:} \quad 30+30+10+15 \;=\; 85\%. \\
\mathcal{W}_2' &= \{\text{ B.C., Alberta, Manitoba, Saskatchewan, New Brunswick, Nova Scotia,} \\
&\qquad \text{P.E.I., Newfoundland}\} \\
&\quad \text{8 members,} \quad \textbf{Total weight:} \quad 10+15+10+5 = 40\%
\end{aligned}$$

Now, $\mathcal{W}_1'$ is losing because it only has six members, while $\mathcal{W}_2'$ is losing because it only comprises 40% of the population.

Thus, the Canadian constitutional amendment formula is *not* trade robust, so Theorem 3.7 says that it is *not* a weighted voting system.

3.3 Vector-Weighted Voting Systems

Prerequisites: §3.2

A **vector weighted** voting system is a strict binary voting procedure where the votes of different voters have vector-valued 'weights'. To be precise, there is a **vector weight function** $\boldsymbol{\omega} : \mathcal{V} \longrightarrow \mathbb{N}^D$ (for some $D \geq 2$) so that, for any $\rho \in \mathfrak{R}(\mathcal{V}, \mathcal{A})$, the *total support* for alternative $\mathcal{Y}\!es$ is the vector $\boldsymbol{\Upsilon}(\rho) \in \mathbb{N}^D$ defined:

$$\boldsymbol{\Upsilon}(\rho) \quad = \quad \sum_{y \in \mathcal{Y}(\rho)} \boldsymbol{\omega}(y) \qquad \text{where} \quad \mathcal{Y}(\rho) \;=\; \left\{ v \in \mathcal{V} \; ; \; \mathcal{Y}\!es \underset{v}{\overset{\rho}{\succ}} \mathcal{N}\!o \right\}$$

The **total weight** of the system is the vector $\mathbf{W} = \sum_{v \in \mathcal{V}} \boldsymbol{\omega}(v)$. Finally, we set a vector-valued **quota** $\mathbf{Q} \in \mathbb{N}^D$ so that

$$\left(\mathcal{Y}\!es \overset{\rho}{\sqsupset} \mathcal{N}\!o \right) \iff \left(\boldsymbol{\Upsilon}(\rho) \geq \mathbf{Q} \right), \tag{3.1}$$

where "$\boldsymbol{\Upsilon} \geq \mathbf{Q}$" means $\Upsilon_1 \geq Q_1, \;\; \Upsilon_2 \geq Q_2, \; \ldots \; , \Upsilon_D \geq Q_D$.

Most systems favour the 'status quo' alternative ($\mathcal{N}\!o$), which means that $\mathbf{Q} \geq \frac{1}{2}\mathbf{W}$.

The **dimension** of a vector-weighted voting system Π is the smallest D so that we can represent Π using D-dimensional weight vectors so that eqn.(3.1) is satisfied. Clearly, Π has dimension 1 if and only if Π is a 'scalar' weighted voting system, as described in §3.2.

Example 3.9: Canadian Constitutional Amendment Formula

The Canadian Constitutional Amendment Formula (Example 3.8) is a vector-weighted voting system of dimension 2. To see this, let $\mathcal{V}$ be the ten provinces of Canada. For any $v \in \mathcal{V}$, let P_v be the population of province v. Define $\boldsymbol{\omega} : \mathcal{V} \longrightarrow \mathbb{N}^2$ so that, for any $v \in \mathcal{V}$, $\boldsymbol{\omega}(v) = (1, P_v)$. Now let $\mathbf{Q} = (7, H)$, where H is half the population of Canada. Thus, for any coalition $\mathcal{U} \subset \mathcal{V}$,

$$\sum_{u \in \mathcal{U}} \boldsymbol{\omega}(u) \quad = \quad \left(\#(\mathcal{U}), \;\; \sum_{u \in \mathcal{U}} P_u \right)$$

Hence, $\left(\sum_{u \in \mathcal{U}} \boldsymbol{\omega}(u) \;\; \geq \;\; \mathbf{Q} \right) \iff \left(\#(\mathcal{U}) \geq 7, \;\; \text{and} \;\; \sum_{u \in \mathcal{U}} P_u \;\; \geq \;\; H \right)$.

In other words, $\mathcal{U}$ is a winning coalition if and only if $\mathcal{U}$ consists of at least seven provinces, together comprising at least half of Canada's population.

Thus, we can represent the Constitutional Amendment Formula using 2-dimensional vectors. We already know that we can't use 1-dimensional vectors (because Example 3.8 shows that the Constitutional Amendment Formula is *not* a 'scalar' weighted voting system). Thus, the Amendment Formula has dimension 2.

Theorem 3.10 *Every monotone binary voting procedure is a vector-weighted procedure.*

Proof: A **losing coalition** is a collection of voters $\mathcal{L} \subset \mathcal{V}$ so that

$$\Big(\mathcal{Y}(\rho) \ = \ \mathcal{L} \Big) \Longrightarrow \Big(\mathit{Yes} \ \overset{\rho}{\sqsubset} \ \mathit{No} \Big).$$

Thus, for example, in a weighted voting system, $\mathcal{L}$ is a losing coalition iff $\sum_{\ell \in \mathcal{L}} \omega(\ell) \ < \ Q$.

Let $= \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_D\}$ be the set of all losing coalitions. (We know is finite because $\mathcal{V}$ is finite, and a finite set only has finitely many subsets). For each $\mathcal{L}_d \in$, we define a weight function $\omega_d : \mathcal{V} \longrightarrow \mathbb{N}$ by

$$\omega_d(v) \quad = \quad \begin{cases} 1 & \text{if} \quad v \notin \mathcal{L}_d \\ 0 & \text{if} \quad v \in \mathcal{L}_d \end{cases}$$

Claim 1: *If $\mathcal{U} \subset \mathcal{V}$ is some coalition, then*

$$\Big(\mathcal{U} \textit{ is a losing coalition} \Big) \quad \Longleftrightarrow \quad \left(\sum_{u \in \mathcal{U}} \omega_d(u) \ = \ 0 \textit{ for some } \mathcal{L}_d \in \ \right)$$

Proof: '$\Longrightarrow$' Clearly, if $\mathcal{U}$ is losing, then $\mathcal{U} \in$. Suppose $\mathcal{U} = \mathcal{L}_d$; then $\sum_{u \in \mathcal{U}} \omega_d(u) \ = \ 0$.

'$\Longleftarrow$' Suppose $\sum_{u \in \mathcal{U}} \omega_d(u) \ = \ 0$, for some $\mathcal{L}_d \in$. Then $\mathcal{U} \subset \mathcal{L}_d$. Thus, $\mathcal{U}$ is also a losing coalition, because Π is monotone, by hypothesis. .... □ `[Claim 1]`

Now define $\boldsymbol{\omega} : \mathcal{V} \longrightarrow \mathbb{N}^D$ by

$$\boldsymbol{\omega}(v) \ = \ (\omega_1(v), \omega_2(v), \ldots, \omega_D(v)).$$

Now let $\mathbf{Q} = (1, 1, \ldots, 1)$. Then for any $\mathcal{U} \subset \mathcal{V}$,

$$\begin{aligned} \Big(\mathcal{U} \text{ is a losing coalition} \Big) \quad & \Longleftrightarrow \quad \left(\sum_{u \in \mathcal{U}} \omega_{\mathcal{L}_d}(u) \ = \ 0 \text{ for some } \mathcal{L}_d \in \ \right) \\ & \Longleftrightarrow \quad \left(\sum_{u \in \mathcal{U}} \omega_d(u) \ = \ 0 \text{ for some } d \in [1..D] \right) \\ & \Longleftrightarrow \quad \left(\sum_{u \in \mathcal{U}} \boldsymbol{\omega}(u) \ \not\geq \ \mathbf{Q} \right). \end{aligned}$$

Thus, $\Big(\mathcal{U} \text{ is a winning coalition} \Big) \iff \left(\sum_{u \in \mathcal{U}} \boldsymbol{\omega}(u) \ \geq \ \mathbf{Q} \right)$, as desired. ______________ □

Further reading: Much of the material in this section was drawn from Chapters 3 and 8 of Taylor [50], which, in turn, are based mainly on papers by Taylor and Zwicker [2, 3]. The Canadian constitutional amendment formula was first studied by Kilgour [21]; see also [48].

3.4 Voting Power Indices

Prerequisites: §3.2

The weighted and vector-weighted voting schemes of §3.2 show how different voters can wield different amounts of 'power' in a voting system. Naively, we would expect the 'power' of a particular voter to be proportional to the 'weight' of her vote, but this is not true. For example, consider a weighted voting system with four voters, Ulrich, Veronique, Waldemar, and Xavier, with the following weights:

$$\omega(u) \;=\; \omega(v) \;=\; \omega(w) \;=\; 3; \qquad \omega(x) \;=\; 2.$$

Thus, $W \;=\; 3+3+3+2 = 11$. Let $Q = 6$; hence the alternative *Yes* will be chosen if and only if *Yes* recieves 6 or more votes. Naively, we would say that Ulrich, Veronique, and Waldemar each have 3/11 of the total power, while Xavier has 2/11. But in fact, Xavier has *no* power, because his vote is irrelevant. In a binary vote of *Yes* vs. *No*, the alternative *Yes* will be chosen if and only if *Yes* recieves the support of at least two out of three members of the set {Ulrich, Veronique, Waldemar}, *regardless* of Xavier's vote. To see this, consider the following table of possible profiles and outcomes:

Individual Votes				Total Score		Outcome
u	v	w	x	*Yes*	*No*	
No	*No*	*No*	*No*	0	11	*No* wins
No	*No*	*No*	*Yes*	2	9	*No* wins
Yes	*No*	*No*	*No*	3	8	*No* wins
Yes	*No*	*No*	*Yes*	5	6	*No* wins
Yes	*Yes*	*No*	*No*	6	5	*Yes* wins
Yes	*Yes*	*No*	*Yes*	8	3	*Yes* wins
Yes	*Yes*	*Yes*	*No*	9	2	*Yes* wins
Yes	*Yes*	*Yes*	*Yes*	11	0	*Yes* wins

(Ulrich, Veronique, and Waldemar have identical weights, so the same outcomes will occur if we permute the 'u', 'v', and 'w' columns of this table.)

This example tells us that 'voting weight' is not the correct measure of 'voting power'. Instead, 'voting power' should answer the question

(P) *How often does Xavier's vote actually make a difference to the outcome?*

In other words, assuming all other voters have already decided their votes, how often will it be the case that *Yes* will be chosen over *No* if and only if Xavier votes for *Yes*? This is the motivation behind the definition of various *Voting Power Indices*. In the previous example, the answer to question **(P)** is 'never'; hence, by any measure, the "voting power" of Xavier is *zero*.

The Shapley-Shubik Index: On any particular issue, we can arrange all the voters in some linear order, from those most in favour of a particular policy to those most opposed to that policy. We can imagine arranging the voters in an *ideological spectrum* from left to right[1], eg:

$$a \quad b \quad c \quad \ldots\ldots \quad m \mid \quad n \quad o \quad \ldots\ldots \quad x \quad y \quad z$$

Those at the right end of the spectrum (eg x, y, z) strongly support the policy; those at the left end (eg. a, b, c) are strongly opposed, and those in the middle (eg. m, n, o) are more or less neutral, perhaps with some slight bias one way or the other.

We can then draw a line so thate the voters *right* of this line vote 'yes' and the voters *left* of this line vote 'no'. In this case, the line is between m and n. If the voters to the *right* of the line form a winning coalition, then the policy will be chosen. Otherwise it will be rejected.

We say that voter n is a **critical** if the set $\{n, o, p, \ldots, x, y, z\}$ *is* a winning coalition, but the set $\{o, p, \ldots, x, y, z\}$ is *not* a winning coalition. In other words, *Yes* will be chosen if and only if n votes for *Yes*.

For example, in a *weighted* voting system, the policy will be chosen if and only if $\omega(n) + \omega(o) + \cdots + \omega(x) + \omega(y) + \omega(z) \ \geq\ Q$, where Q is the quota and $\omega(m)$ is the weight of voter m, etc. Thus, the voter n is critical if and only if

$$\omega(o)+\omega(p)+\cdots+\omega(x)+\omega(y)+\omega(z) \quad < \quad Q \quad \leq \quad \omega(n)+\omega(o)+\omega(p)+\cdots+\omega(x)+\omega(y)+\omega(z)$$

which is equivalent to saying

$$0 \quad < \quad Q - \Big(\omega(o) + \omega(p) + \cdots + \omega(x) + \omega(y) + \omega(z)\Big) \quad \leq \quad \omega(n).$$

Clearly, for a particular ordering of voters into an ideological spectrum, there will be exactly *one* critical voter. The Shapley-Shubik index of voter n is a crude estimate of the probability that voter n will be this unique critical voter. Formally, define

$$\mathcal{O} \ := \ \{\text{All orderings of the voters}\} \quad \text{and} \quad \mathcal{O}_n \ := \ \{\text{All orderings where } n \text{ is critical}\}.$$

Then we define $\quad \mathtt{SSI}(n) \quad := \quad \dfrac{\#\mathcal{O}_n}{\#\mathcal{O}} \quad = \quad \dfrac{\#\mathcal{O}_n}{N!}, \quad$ where N is the number of voters.

Note that this definition assumes that *all orderings of the voters are equally likely*. In the present abstract context, this *a priori* assumption is clearly no worse than any other. However, in a specific concrete setting, it may be clear that some orderings are much more likely than others. (For example, on economic issues, we can expect that the voters will always arrange themselves in roughly the same order, forming a spectrum from socialism to laissez-faire capitalism).

Observe that the Shapley-Shubik indices of all players add up to 1:

$$\mathtt{SSI}(a) + \mathtt{SSI}(b) + \cdots + \mathtt{SSI}(z) \quad = \quad 1. \tag{3.2}$$

This follows from the fact that $\mathcal{O}_a \sqcup \mathcal{O}_b \sqcup \cdots \sqcup \mathcal{O}_z \ = \ \mathcal{O}$.

[1]'Left' and 'right' here do not necessarily correspond to the vague political categories of 'left-wing' vs. 'right-wing', although of course they might.

Example 3.11: United Nations Security Council

Recall that the U.N. Security Council has five 'permanent' members and ten 'nonpermanent' members. A resolution passes if and only if it is supported by *all five* permanent members, and, in addition, by *at least four* nonpermanent members.

Let $\mathcal{P} = \{p_1, p_2, p_3, p_4, p_5\}$ be the set of permanent members, and let $\mathcal{N} = \{n_1, \ldots, n_{10}\}$ be the set of nonpermanent members, so that $\mathcal{C} = \mathcal{P} \sqcup \mathcal{N}$ is the entire Council.

Let's compute the Shapley-Shubik Index of a 'permanent' member, say, p_1. First, note that, if $\mathcal{W} \subset \mathcal{C}$ is *any* winning coalition, then p_1 is critical for $\mathcal{W}$, because any permanent member has a veto. Thus, to count the orderings where p_1 is critical, we must simply count all orderings of all possible winning coalitions, such that p_1 is the 'borderline' member in the ordering.

Now, $\mathcal{W}$ is a winning coalition only if $\mathcal{P} \subset \mathcal{W}$ and $\mathcal{W}$ also contains at least four nonpermanent members. Hence if $\widetilde{\mathcal{W}} = \mathcal{W} \setminus \mathcal{P}$, then $\widetilde{\mathcal{W}}$ must have $m \geq 4$ elements.

Suppose $\widetilde{\mathcal{W}}$ is chosen, and let $\widehat{\mathcal{W}} = \mathcal{W} \setminus \{p_1\}$; then $\#(\widehat{\mathcal{W}}) = 4 + m$, so there are $(4+m)!$ ways that $\widehat{\mathcal{W}}$ can be ordered. If $\mathcal{V} = \mathcal{C} \setminus \mathcal{W}$ is the set of remaining Council members (those 'opposed'), then $\#(\mathcal{V}) = 10 - m$, so there are $(10-m)!$ ways to order $\mathcal{V}$. Thus, having fixed $\widetilde{\mathcal{W}}$, there are a total of $(10-m)! \cdot (4+m)!$ orderings of $\mathcal{C}$ where p_1 is the 'boundary' element between $\widehat{\mathcal{W}}$ and $\mathcal{V}$.

There are $\begin{pmatrix} 10 \\ m \end{pmatrix}$ subsets of $\mathcal{N}$ which contain m elements. Hence, there are $\begin{pmatrix} 10 \\ m \end{pmatrix} = \dfrac{10!}{m!(10-m)!}$ choices for $\widetilde{\mathcal{W}}$. This gives a total of

$$\frac{10!}{m!(10-m)!} \cdot (10-m)! \cdot (4+m)! \quad = \quad \frac{10!(4+m)!}{m!}$$

possible orderings involving m nonpermanent members where p_1 is critical. Thus, the total number of orderings where p_1 is critical is given:

$$\#(\mathcal{O}_{p_1}) \quad = \quad \sum_{m=4}^{10} \frac{10!(4+m)!}{m!}$$

Now, $\#(\mathcal{C}) = 15$, so there are $15!$ orderings in total. Thus, the Shapley-Shubik Index for p_1 is given:

$$\mathtt{SSI}(p_1) \quad = \quad \frac{10!}{15!} \sum_{m=4}^{10} \frac{(4+m)!}{m!} \quad = \quad \left(\frac{70728}{360360}\right) \quad = \quad \left(\frac{421}{2145}\right) \quad \approx \quad 0.19627\ldots$$

which is roughly 19.6%.

To compute the Shapley-Shubik index of a *nonpermanent* member, we could go through a similar combinatorial argument. However, there is a simpler way, using eqn.(3.2). First, note that, by symmetry, all permanent members have the same SSI. Let P denotes the SSI of a permanent member; then we've just shown that $P = \frac{421}{2145}$. Likewise, by symmetry, all nonpermanent members have the same SSI, say N. Then eqn.(3.2) implies that

$$5 \cdot P + 10 \cdot N \quad = \quad 1$$

Substituting $P = \frac{421}{2145}$ and simplifying, we have:

$$\begin{aligned} N &= \frac{1}{10}(1-5P) \quad = \quad \frac{1}{10}\left(1 - 5 \cdot \frac{421}{2145}\right) \quad = \quad \frac{1}{10}\left(1 - \frac{421}{429}\right) \quad = \quad \frac{1}{10}\left(\frac{8}{429}\right) \\ &= \frac{8}{4290} \quad = \quad \frac{4}{2145}. \quad \approx \quad 0.001865\ldots \end{aligned}$$

Hence, each of the five permanent members has about 19.6% of the total voting power, while each of the ten nonpermanent members has about 0.187% of the total voting power. ____

The Shapley-Shubik index is only one of many 'voting power indices' which have been developed; others include the Banzhaf index, the Johnston index, and the Deegan-Packel index. All the indices measure the power of a voter v by counting the scenarios where v is 'critical'; the indices differ in how they define and enumerate these scenarios. For example, the Shapley-Shubik index defines 'scenarios' as linear orderings of the voters; hence counting critical scenarios means counting orderings. The Banzhaf index, on the other hand, simply counts the total number of *unordered* winning coalitions where v is a critical member; hence, one 'critical' coalition in the Banzhaf model corresponds to many 'critical' coalitions in the Shapley-Shubik model (because it could have many orderings). This yields a different formula, with a different numerical value.

In some cases, the various voting power indices roughly agree, while in others, they wildly disagree, suggesting that at least one of them must be wrong. Like Shapley-Shubik, the other indices are based on some *a priori* estimate of the probability of various voting scenarios (eg. Shapley-Shubik assumes that all $N!$ orderings are equally probable). To the extent that these *a priori* estimates are unrealistic (because some orderings are much more ideologically plausible than others), *none* of the indices will be perfectly accurate. Nevertheless, they are valuable as a rough estimate of how (non)egalitarian a voting system is. For example, our analysis of the U.N. Security Council confirms the conventional wisdom that the ten nonpermanent members have virtually no power.

Further reading: A good overview of the various power indices is in Chapters 4 and 9 of Taylor [50]. The Shapley-Shubik index first appeared in [27].

Chapter 4

Bentham's Utilitarianism

Prerequisites: None **Recommended:** §2.2

All the voting methods we've examined have a common shortcoming: they only use information about the *order* in which voters rank the various alternatives, and they do *not* use information about *how much* a voter prefers one alternative to another. As we saw in *Arianne, Bryn and Chloe* example on page 9 of §1.3, and in the discussion of Arrow's Theorem (page 31), we can justify the Borda count's 'sensitivity to irrelevant alternatives' by arguing that the Borda count detects the 'intensity' of voter's preferences. However, by comparing figures (1.4) and (1.5) on page 10, one can see that the *same* Borda count profile can result from very different intensities. If we are to take the 'intensity' idea seriously, we need a better way of measuring intensity.

For a more extreme example, reconsider the *Condorcet Paradox* profile (page 6). Table (1.1) on page 7 shows how the Borda count 'resolves' this paradox, in favour of candidate C. However, suppose that the 34% who prefer C only like him slightly more than B, and only like B slightly more than C. Likewise, the 33% who prefer B only like him slightly more than C, and only like C slightly more than A. However, the 33% who prefer A absolutely despise C, and rank him a distant third behind B (for concreteness, suppose C is a genocidal maniac, who proposes extermination of this 33% of the population). Wouldn't it make more sense to assign points as follows?

	Points		
	A	B	C
33%	2	1	-10
33%	0	2	1
34%	1	0	2
Total:	100	99	-229

In this case, A is the Borda winner.

However, once we're allowed to manipulate the point assignments, what is the 'right' point assignment? Why -10, and not -20? If 33% of the population *really* likes A, can they give him *more* than 2 points? How much more? If people are allowed to assign as many points as they want, then everyone will simply assign astronomically large figures to their favourite candidate, and astronomically large negative figures to their least favourite, and the election will become a farce.

4.1 Utility functions

We need an objective way to measure the *utility* (ie. degree of happiness or unhappiness) which each voter assigns to each alternative. To be precise, we first distinguish between *ordinal* and *cardinal* utility.

Let $v \in \mathcal{V}$ be a voter (say, *Veronique*) who assigns some preference ordering '$\underset{v}{\succeq}$' to the alternatives in $\mathcal{A}$. An **ordinal utility function** for Veronique is a function $U : \mathcal{A} \longrightarrow \mathbb{R}$ such that for any $A, B \in \mathcal{A}$,

$$\left(A \underset{v}{\succeq} B \right) \iff \left(U(A) \geq U(B) \right)$$

Thus, U preserves the ordering of the alternatives. Notice that U says nothing about *how much* Veronique prefers A to B.

For example the *Borda count* method (§1.3) implicitly uses an ordinal utility function, where we assign $N-1$ points to Veronique's highest ranked alternative, and 0 points to her lowest ranked alternative.

A **cardinal utility function** for Veronique is an ordinal utility function $U : \mathcal{A} \longrightarrow \mathbb{R}$ with an additional property. Suppose A_1, B_1 and A_2, B_2 are two pairs of alternatives such that $U(A_1) > U(B_1)$ and $U(A_2) > U(B_2)$. Let $\Delta U_1 = U(A_1) - U(B_1)$ and let $\Delta U_2 = U(A_2) - U(B_2)$. If $\Delta U_1 = r \cdot \Delta U_2$ for some $r > 0$, then we interpret this to mean that Veronique's preference for A_1 over B_1 is 'r times as great' as her preference for A_2 over B_2. Our problem is to make sense of this notion.

We do this using a hypothetical gambling game or *lottery*. First we introduce the concept of *expected utility*. Suppose Veronique is playing a lottery where she will randomly win one of the alternatives in $\mathcal{A} = \{A, B, C\}$. Suppose the probabilities of these outcomes are $\mathbf{p}(A)$, $\mathbf{p}(B)$ and $\mathbf{p}(C)$ respectively, and suppose Veronique assigns them (cardinal) utilities $U(A)$, $U(B)$ and $U(C)$. Then Veronique's *expected utility* in this lottery is quantity:

$$\mathbb{E}(U, \mathbf{p}) \quad = \quad \mathbf{p}(A) \cdot U(A) \; + \; \mathbf{p}(B) \cdot U(B) \; + \; \mathbf{p}(C) \cdot U(C)$$

One can interpret this as the *average utility she can expect to gain by playing the lottery many times over.*

More generally, in a lottery with some set of alternatives $\mathcal{A}$, let $\mathbf{p} : \mathcal{A} \longrightarrow [0, 1]$ be a probability distribution (a function assigning a probability to each alternative), and let $U : \mathcal{A} \longrightarrow \mathbb{R}$ be a (cardinal) utility function. Veronique's **expected utility** is defined:

$$\mathbb{E}(U, \mathbf{p}) \quad = \quad \sum_{A \in \mathcal{A}} \mathbf{p}(A) \cdot U(A). \qquad (4.1)$$

Now we come to the key idea. Given a choice between several lotteries (ie. several probability distributions over $\mathcal{A}$), *Veronique will always pick the lottery which maximizes her expected utility.*

This intuitively plausible observation can then be flipped around, to provide the *definition* of the cardinal utility function. To be precise, a **cardinal utility function** for Veronique is a function $U : \mathcal{A} \longrightarrow \mathbb{R}$ so that, if Veronique is allowed to choose between various lotteries (with different probability distributions), she will always choose the lottery which yields the highest expected utility, as defined by eqn.(4.1).

To guarantee that such a utility function exists, we must assume that Veronique is 'rational' in her choices amongst various lotteries. Let $\mathbb{P}(\mathcal{A})$ be the space of all possible probability distributions on the set of alternatives $\mathcal{A}$. For example, if $\mathcal{A} = \{A_1, \ldots, A_N\}$, then we can identify $\mathbb{P}(\mathcal{A})$ with the N-simplex:

$$\mathbb{P}(\mathcal{A}) \quad = \quad \left\{ \mathbf{p} = (p_1, \ldots, p_N) \in [0,1]^N \; ; \; \sum_{n=1}^{N} p_n \; = \; 1 \right\}.$$

Thus, $\mathbb{P}(\mathcal{A})$ represents the space of all possible lotteries on these alternatives. When choosing between lotteries, Veronique must satisfy the following three 'rationality' axioms:

(T) (*Transitivity*) There is a preference ordering $\underset{v}{\succeq}$ on $\mathbb{P}(\mathcal{A})$ so that, for any $\mathbf{p}, \mathbf{q} \in \mathbb{P}(\mathcal{A})$,

$$\Big(\text{Veronique picks lottery } \mathbf{p} \text{ over } \mathbf{q} \Big) \iff \Big(\mathbf{p} \underset{v}{\succeq} \mathbf{q} \Big).$$

In particular, axiom **(T)** yields a preference ordering on $\mathcal{A}$. For any alternative $A \in \mathcal{A}$, let $\mathbb{1}_A$ be the lottery which gives probability 1 (ie. certainty) to outcome A. Then

$$\Big(A \underset{v}{\succeq} B \Big) \iff \Big(\mathbb{1}_A \underset{v}{\succeq} \mathbb{1}_B \Big).$$

In other words, if choosing between lotteries with 'guaranteed' outcomes, Veronique will pick the outcome she prefers.

(L) (*Linearity*) Suppose $\mathbf{p}_0, \mathbf{p}_1 \in \mathbb{P}(\mathcal{A})$ are two lotteries. For any $r \in [0,1]$, let $\mathbf{p}_r$ be the lottery obtained by convex-combining $\mathbf{p}_0$ and $\mathbf{p}_1$:

$$\mathbf{p}_r(A) \quad = \quad r \cdot \mathbf{p}_1(A) \; + \; (1-r) \cdot \mathbf{p}_0(A), \quad \text{for all } A \in \mathcal{A}.$$

Likewise, let $\mathbf{q}_0, \mathbf{q}_1 \in \mathbb{P}(\mathcal{A})$ be two lotteries, and for any $r \in [0,1]$, define $\mathbf{q}_r$ by

$$\mathbf{q}_r(A) \quad = \quad r \cdot \mathbf{q}_1(A) \; + \; (1-r) \cdot \mathbf{q}_0(A), \quad \text{for all } A \in \mathcal{A}.$$

Then

$$\Big(\mathbf{p}_0 \underset{v}{\succeq} \mathbf{q}_0 \text{ and } \mathbf{p}_1 \underset{v}{\succeq} \mathbf{q}_1 \Big) \Longrightarrow \Big(\forall\, r \in [0,1], \quad \mathbf{p}_r \underset{v}{\succeq} \mathbf{q}_r \Big).$$

To understand axiom **(L)**, think of $\mathbf{p}_r$ as representing a 'two-stage lottery', where the prize at the *first* stage is a ticket to a *second* lottery held at the *second* stage. At the first stage, there is probability r of winning a ticket to lottery $\mathbf{p}_1$, and probability $(1-r)$ of winning a prize to lottery $\mathbf{p}_0$. The net effect is as if Veronique was competing in a *single* lottery with probability distribution $\mathbf{p}_r$. The axiom **(L)** says: if Veronique prefers $\mathbf{p}_0$ to $\mathbf{q}_0$, and prefers $\mathbf{p}_1$ to $\mathbf{q}_1$, then she will surely prefer a lottery between $\mathbf{p}_0$ and $\mathbf{p}_1$ to a similar lottery between $\mathbf{q}_0$ and $\mathbf{q}_1$.

(C) (*Continuity*[1]) Suppose $A_0, B, A_1 \in \mathcal{A}$, and $A_0 \underset{v}{\preceq} B \underset{v}{\preceq} A_1$. For every $r \in [0,1]$, let $\mathbf{p}_r \in \mathbb{P}(\mathcal{A})$ be the lottery giving probability r to A_1 and probability $(1-r)$ to A_0. Then there exists a value $r_0 \in [0,1]$ such that

$$\begin{aligned} \Big(r \leq r_0 \Big) &\implies \Big(\mathbf{p}_r \underset{v}{\preceq} \mathbb{1}_B \Big); \\ \Big(r = r_0 \Big) &\implies \Big(\mathbf{p}_r \underset{v}{\approx} \mathbb{1}_B \Big); \\ \text{and} \quad \Big(r \geq r_0 \Big) &\implies \Big(\mathbf{p}_r \underset{v}{\succeq} \mathbb{1}_B \Big). \end{aligned}$$

For example, suppose alternative A_0 represents having \$99.00, alternative B represents having \$100.00, and alternative A_1 represents having \$1,000,099.00. Suppose Veronique has \$100.00, and is considering buying a \$1.00 ticket to a lottery with a \$1,000,000 jackpot. The three alternatives then mean the following:

- A_0: Buy a ticket and lose the lottery; Veronique is left with \$99.00.
- B: Don't buy a ticket; Veronique is left with \$100.00.
- A_1: Buy a ticket and win; Veronique is left with \$1,000,099.00.

If the odds of winning are too low (ie $r < r_0$), then Veronique considers it a waste of money to buy a ticket. If the odds of winning are high enough (ie. $r > r_0$), then she considers it a good bet, so she will by a ticket. In between these extremes, there is a critical probability (ie. $r = r_0$), where Veronique can't decide whether or not to buy a ticket. The exact value of r_0 depends upon how much utility Veronique assigns to having different amounts of money (ie. how much she fears bankruptcy, how greedy she is to be rich, etc.)

Clearly, axioms **(T)**, **(L)** and **(C)** are reasonable to expect from any rational person. The next theorem says, loosely speaking, *Any rational gambler has a utility function.*

Theorem 4.1 (von Neumann and Morgenstern) *Suppose* $\underset{v}{\succeq}$ *is a relation on* $\mathbb{P}(\mathcal{A})$ *satisfying axioms* **(T)**, **(L)**, *and* **(C)**. *Then:*

[1] This is sometimes called the *Archimedean* axiom, or the axiom of *Substitutability*.

(a) *There exists a cardinal utility function $U : \mathcal{A} \longrightarrow \mathbb{R}$ so that, if $\mathbf{p}, \mathbf{q} \in \mathbb{P}(\mathcal{A})$ are two lotteries, then*

$$\left(\mathbf{p} \underset{v}{\succeq} \mathbf{q} \right) \iff \left(\mathbb{E}(U, \mathbf{p}) \; \geq \; \mathbb{E}(U, \mathbf{q}) \right), \tag{4.2}$$

where $\mathbb{E}(U, \mathbf{p})$ is the expected utility defined by eqn.(4.1).

(b) *Furthermore, the function U is unique up to affine transformation. That is, if $U_1, U_2 : \mathcal{A} \longrightarrow \mathbb{R}$ are two functions satisfying eqn.(4.2), then there exist constants $m > 0$ and $b \in \mathbb{R}$ so that $U_1 \;=\; m \cdot U_2 + b$.* □

Proof: Fix a three-element subset $\{A, B, C\} \subset \mathcal{A}$. Assume WOLOG that $A \underset{v}{\succeq} B \underset{v}{\succeq} C$, and use axiom **(C)** to find a value r_0 so that, in the language of axiom **(C)**,

$$\left(r = r_0 \right) \Longrightarrow \left(\mathbf{p}_r \underset{v}{\approx} \mathbb{1}_B \right) \tag{4.3}$$

Assuming that a cardinal utility function U exists, the definition (4.1) of *expected utility* says:

$$\mathbb{E}(U, \mathbf{p}_r) \quad = \quad r \cdot U(A) + (1 - r) \cdot U(C) \tag{4.4}$$

Assume U satisfies eqn.(4.2); then eqns.(4.3) and (4.4) imply:

$$r \cdot U(A) + (1 - r) \cdot U(C) \quad = \quad U(B). \tag{4.5}$$

If we do this for every three-element subset of $\mathcal{A}$, we end up with a system linear equations like eqn.(4.5) in the unknown variables $\{U(A)\}_{A \in \mathcal{A}}$. We then solve this system to get values for $U(A_1), \ldots, U(A_N)$.

(For example, if $\mathcal{A} = \{A, B, C, D\}$, then we get four equations:
$$\begin{aligned} r \cdot U(A) + (1 - r) \cdot U(C) &= U(B), \\ s \cdot U(A) + (1 - s) \cdot U(D) &= U(B), \\ t \cdot U(A) + (1 - t) \cdot U(D) &= U(C), \\ u \cdot U(B) + (1 - u) \cdot U(C) &= U(C), \end{aligned}$$
for some constants r, s, t and u. We then solve for $U(A)$, $U(B)$, $U(C)$, and $U(D)$.)

Now there are two problems:

1. Show that this system of equations is actually *consistent* (ie. has a solution).
2. Show that the resulting values for $U(A_1), \ldots, U(A_N)$ actually satisfy eqn.(4.2)

Both problems are solved by the *linearity* axiom **(L)**. The details are **Exercise 21** . □

Exercise 22 One advantage of the 'gambling' interpretation of utility is that it provides an empirical method to *measure* the utility functions of a real person, by performing an experiment where she chooses between various lotteries. Each choice she makes is a data point, and given sufficient data, we can solve a system of linear equations to figure out what her utility function is.

1. Design an experimental protocol to measure the cardinal utilities of three alternatives A, B, C. For simplicity, assume $A \succeq B \succeq C$ and set $U(A) = 1$ and $U(C) = 0$. The problem is to find the value of $U(B) \in [0, 1]$. Your experiment should involve two stages:

(a) First, use a series of choices between lotteries to find a value of r_0 so that the experimental subject is indifferent between lottery $\mathbb{1}_B$ and the lottery $\mathbf{p}_{r_0}$, as described by axiom **(C)**. (In principle, it will take an infinite sequence of choices to find the exact value of r_0. In practice, of course, we stop when we get a good enough estimate).

(b) Assume you know the value of r_0. Set up and solve a linear equation like (4.5) to determine the value of $U(B)$.

2. Use this protocol in a real experiment with a friend. See if you can measure her cardinal utilities for three alternatives (eg. three restaurants, three movies, etc.)

4.2 The problem of interpersonal comparisons

Now that we have a clear notion of utility, it seems clear that true 'democracy' means that collective social decisions should be made so as to maximize the utilities of all voters. Let $\mathcal{A}$ be a set of alternatives and let $\mathcal{V}$ be a set of voters. Suppose each voter $v \in \mathcal{V}$ has some cardinal utility function $U_v : \mathcal{A} \longrightarrow \mathbb{R}$. We define the **collective utility function** $\overline{U} : \mathcal{A} \longrightarrow \mathbb{R}$ by:

$$\overline{U}(A) \quad = \quad \sum_{v \in \mathcal{V}} U_v(A), \quad \text{for all } A \in \mathcal{A}. \tag{4.6}$$

The **Utilitarian Procedure** then states

(U) *Society should choose the alternative which maximizes the collective utility function $\overline{U}$.*

This beautiful idea suffers from a critical flaw: the *Problem of interpersonal comparisons of utility*. To illustrate the problem, suppose there are two voters, Veronique and Waldemar, with utility functions U_v and U_w. Then the collective utility function is defined:

$$\overline{U}(A) \quad = \quad U_v(A) \; + \; U_w(A), \quad \text{for all } A \in \mathcal{A}.$$

Suppose $\mathcal{A} = \{A, B, C\}$, and we have the following utilities:

	U_v	U_w	$\overline{U}$
A	0	1	**1**
B	1	-1	0
C	-1	0	-1

Clearly, society should choose alternative A, which maximizes the collective utility with $\overline{U}(A) = 1$. Now suppose we have the following utilities:

	U'_v	U_w	$\overline{U}$
A	0	1	1
B	10	-1	**9**
C	-10	0	-10

In this case, society should chose alternative B. However, the von Neumann-Morgenstern Theorem (Thm.4.1) says that the utility function U_v is only well-defined up to affine transformation, and it is clear that $U'_v = 10 \cdot U_v$. Thus, using the von Neumann-Morgenstern 'gambling' definition of utility, *there is no way to determine whether Veronique has utility function U_v or U'_v.* The two utility functions will produce identical 'gambling' behaviour in Veronique, but they clearly yield different outcomes in the collective social choice.

The problem is that there is no natural 'unit' we can use to compare Veronique's utility measurements to Waldemar's. Put another way: there is no way we can empirically determine that Veronique prefers alternative B to A 'ten times as much' as Waldemar prefers A to B. Indeed, in terms of the von Neumann-Morgenstern definition of utility, it doesn't even *make sense* to say such a thing.

This ambiguity makes the 'collective utility' concept meaningless, and worse, subject to manipulation through exaggeration. For example, Waldemar can regain control of the social choice by exaggerating his preferences as follows:

	U_v	U_w	$\overline{U}$
A	0	1000	**1000**
B	10	-1000	-900
C	-10	0	-10

Veronique can then retaliate, and so on. Clearly this rapidly gets ridiculous.

4.3 Normalization & Voting Dollars

The solution to this 'inflationary' scenario is to *normalize* the cardinal utility functions, by forcing all voters to assign $U = 1$ to their *most preferred* alternative, and assign $U = -1$ to their *least preferred* alternative. We will call this a **normalized cardinal utility function.** It follows from the von Neumann-Morgenstern theorem that every voter has a unique normalized utility function, which we will write as U^*.

Voting dollars: However, normalizing utility functions on the scale from -1 to 1 may not always be appropriate. Sometimes it might *really be true* that some voters have stronger preferences than others. For example:

- In a referendum concerning public schools, parents will likely have stronger preferences than nonparents, so perhaps their opinions should have more weight. But how much more? Should a parent of two children get more weight than a parent with one child?

- In a referendum about where to build a new freeway, people who drive a lot (eg. taxi drivers) will have stronger preferences than people who don't. Should their opinion get extra weight? How much?

To resolve this, we must contextualize each particular referendum within a larger series of referenda. Different voters will care about different issues, and thus, will have stronger preferences in different referenda. For example, Waldemar may have stronger preferences about the school referendum, while Veronique has stronger feelings about the freeway issue. It is okay to give someone more voting power in *one* referendum, and less voting power in *another*, so long as both Veronique and Waldemar have the same *average* amount of voting power in the long run.

To achieve this, we provide each voter with a steady allowance of 'voting dollars' (represented by the symbol $\check{\$}$). By spending D voting dollars, Veronique can multiply the effect of her vote by a factor of D. Thus, although her normalized utility function has a range of -1 to 1, if she spends $\check{\$}5.00$, it will be *as if* her utility function has a range of -5 to 5. Thus, the *collective utility* is computed:

$$\overline{U}(A) \quad = \quad \sum_{v \in \mathcal{V}} D_v \cdot U_v^*(A), \quad \text{for all } A \in \mathcal{A}. \tag{4.7}$$

Here, U_v^* represents voter v's *normalized* utility function (ranging from -1 to 1), and D_v is the number of voting dollars which v has chosen to spend on this particular referendum. We then choose the alternative A which maximizes the expression (4.7).

To minimize vote manipulation and ensure political equality, we need three conditions:

- *Each voter receives the same allowance of voting dollars* (eg. $\check{\$}1.00$ per day). Thus, in the long run, all voters have equal political influence (although different voters may choose to exert more or less influence on different issues).

  Since you have an extremely finite supply of dollars, you have an incentive not to exaggerate the intensity of your feelings. You aren't going to spend $\check{\$}10.00$ on a particular referendum unless the issue *really* matters 10 times as much to you as a $\check{\$}1.00$ issue.

- *Voting dollars are nontransferable.* You cannot sell, loan, bequeath, or give your voting dollars to another person. This is very important to prevent transactions where people exchange voting dollars for physical goods, services, or money. Otherwise, wealthy people can control the political process by buying voting dollars.

  To achieve nontransferability, voting dollars must not be a physical currency. Instead, they must only exist as numerical records kept in a central bank. Thus, you can't give someone your voting dollars for the same reason you can't give someone your social insurance number. (Of course, you could still make a secret agreement to spend your voting dollars according to some wealthy patron's instructions (in exchange for money). But voting is anonymous, so it is impossible for the patron to ever know that you are voting the way he told you to. Hence, this sort of arrangement wouldn't be very attractive for him.)

- *Voting dollars cannot be hoarded.* At any time, you can have a maximum of $\check{\$}100.00$. Once your supply reaches $\check{\$}100.00$, your $\check{\$}1.00$/day allowance ceases. This prevents 'single-issue extremists' from hoarding huge quantities of dollars and then spending them all at once to 'swing' a particular vote.

- *Upcoming referendum horizon exceeds 100 days.* Voters must receive at least one hundred days of advanced notice on any issue to be put to referendum. This allows all voters to build up their reserves to the maximum amount of $\check{\$}$100.00. (It thereby prevents a 'surprise attack', where Waldemar lures Veronique into depleting her voting power in a contentious 'decoy' referendum, and then he immediately follows this up with a second referendum, where he easily wins his *real* objectives.)

Exercise 23 A **utility profile** is a function $U : \mathcal{V} \times \mathcal{A} \longrightarrow \mathbb{R}$, so that $U(v, A)$ is the utility which voter v assigns to alternative A. A **utilitarian voting procedure** is a function which takes a utility profile as input and produces a preference ordering on $\mathcal{A}$ as output.

(a) Explain how rule **(U)** is a utilitarian voting procedure.

(b) Explain how an ordinary voting procedure (as defined in §2) can be extended to a utilitarian procedure.

(c) Suggest appropriate formulations of the axioms **(P)**, **(M)**, **(A)**, **(N)**, **(C)**, **(ML)** and **(IIA)** for utilitarian procedures. In each case justify why your formulation is appropriate.

(d) Does the procedure **(U)** satisfy these axioms? Why or why not?

(e) How is the procedure **(U)** is susceptible to strategic voting?

Further reading: Utilitarianism was first articulated by British political philosopher Jeremy Bentham (1748-1832), and was later elaborated by other utilitarian philosophers, most notably John Stuart Mill (1806-1873). These early approaches suffered from a fundamental problem: they took for granted that happiness (or 'utility') could be treated as a mathematical quantity, but they gave no concrete way of *quantifying* it. This problem was resolved by game theorists John von Neumann and Oskar Morgenstern [19], who showed how to quantify utility using Theorem 4.1. Good introductions to the von Neumann-Morgenstern approach are Luce and Raiffa [30, §2.4] or Riker [32, §4F]. The 'voting dollar' proposal is, to my knowledge, new.

Other solutions to the problem of interpersonal comparison of utility are Raiffa [31] and Braithwaite [11], Goodman-Markowitz [24], Hildreth [17]; see sections 6.11 and 14.6 of Luce & Raiffa [30] for a summary.

Chapter 5

Arbitration & Bargaining

Prerequisites: None

Bargaining is another kind of group decision making, which differs from voting in two ways:

- Each participant begins with an initial 'endowment' or 'bargaining position' which we call the *status quo*. If a mutually satisfactory agreement cannot be reached, than any participant may terminate negotiations and maintain the status quo. (However, if an agreement *is* reached, then all participants must honour it. In other words, contracts can be enforced.)
- Rather than a finite ballot of alternatives, there exists a *continuum* of possible outcomes (normally represented by a subset of $\mathbb{R}^N$). For example, when trading 'infinitely divisible' commodities such as money and oil, there exist a continuum of possible trading positions, each representing a particular quantity of oil purchased at a particular price.

In a *Bargaining* scenario, we regard the participants as playing an unrefereed 'game', and we then apply game theory to predict what sorts of outcomes are (im)possible or (un)likely, assuming all players are rational strategists. It is not important whether the outcome is 'fair'; only whether it is strategically realistic. In an *Arbitration* scenario, however, we assume the existence of an 'arbiter', who tries to propose a 'fair' outcome. The question then becomes: what is fair?

5.1 The von Neumann-Morgenstern Model

Let $\mathcal{V} = \{v, w\}$ be a set of two bargainers. Let $\mathcal{A}$ be a (possibly infinite) set of alternatives which they can jointly choose amongst (Figure 5.1A). Each element of $\mathcal{A}$ represents an outcome for v and a corresponding outcome for w. For example:

- Suppose v and w are negotiating an exchange of several commodities (say, *Ale*, *Beer*, and *Cider*), then each element of $\mathcal{A}$ allocates a specific amount of Ale, Beer, and Cider to v, and a complementary amount of Ale, Beer, and Cider to w.

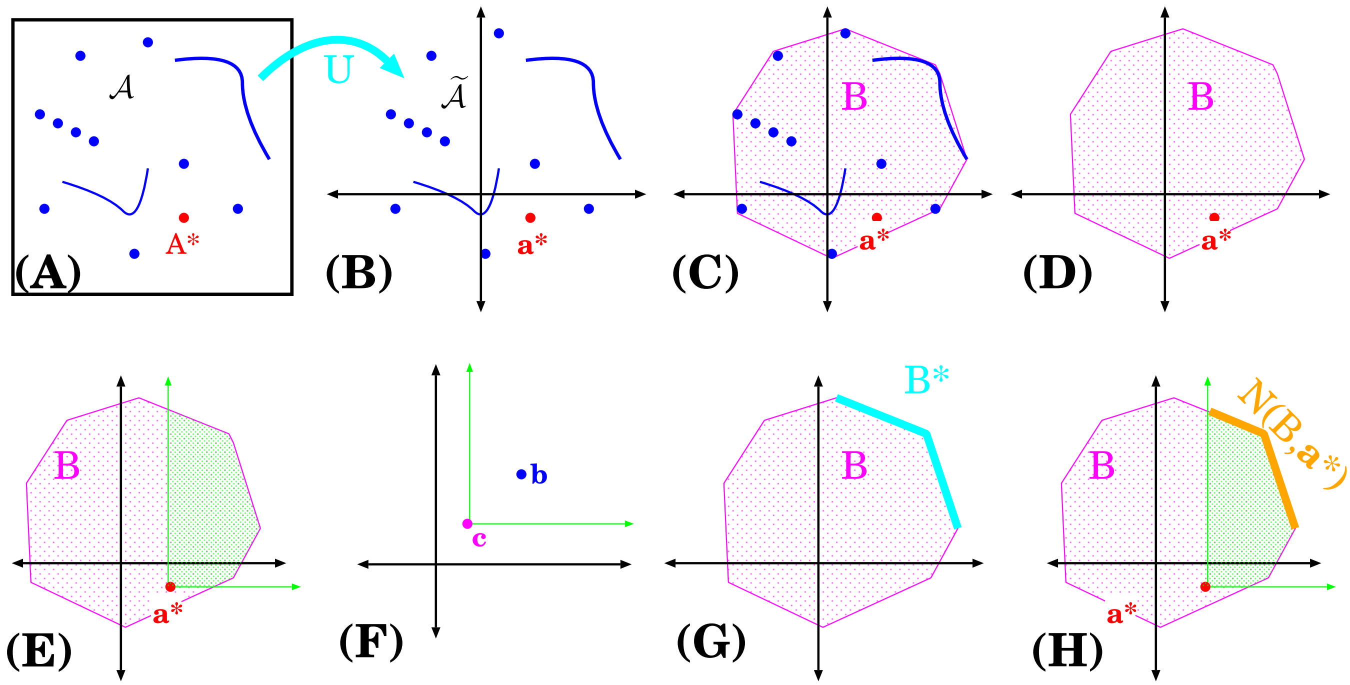

Figure 5.1: The von Neumann-Morgenstern bargaining model. **(A)** $\mathcal{A}$ is an abstract set of 'alternatives', and A^* is the 'status quo' **(B)** $\widetilde{\mathcal{A}} \subset \mathbb{R}^2$ is the image of $\mathcal{A}$ under $U : \mathcal{A} \longrightarrow \mathbb{R}^2$, and $\mathbf{a}^* = U(A^*)$. **(C)** $\mathcal{B}$ is the convex closure of $\widetilde{\mathcal{A}}$; we can therefore assume axiom **(C)**. **(D)** We forget about $\widetilde{\mathcal{A}}$; the bargaining problem is determined by $\mathcal{B}$ and $\mathbf{a}^*$. **(E)** The axiom **(MB)** says any viable bargain must be Pareto-preferred to the status quo. This restricts us to the green-shaded 'northeast' corner. **(F)** $\mathbf{b}$ is Pareto-preferred to $\mathbf{c}$ if $b_v \geq c_v$ and $b_w \geq c_w$. **(G)** The axiom **(P)** says any viable bargain must be Pareto-optimal. This restricts us to the *Pareto frontier* $\mathcal{B}^*$. **(H)** Axioms **(MB)** and **(P)** together restrict us to the *negotiating set* $\mathcal{N}(\mathcal{B}, \mathbf{a}^*)$.

- If v is a labour union and w is management, then each element of $\mathcal{A}$ corresponds to a labour contract with particular terms concerning wages, benefits, holiday time, job security, etc.

- If v and w are roommates, then each element of $\mathcal{A}$ corresponds to some agreement about how to split the housework responsibilities (eg. dishwashing) and how to peacefully coexist (eg. who gets to play music, entertain guests, or watch television at particular times).

We assume that v and w have cardinal utility functions, $U_v : \mathcal{A} \longrightarrow \mathbb{R}$ and $U_w : \mathcal{A} \longrightarrow \mathbb{R}$. Together, these determine a *joint utility function* $U : \mathcal{A} \longrightarrow \mathbb{R}^2$, where $U(A) = \Big(U_v(A),\ U_w(A)\Big)$ for each $A \in \mathcal{A}$. For the purposes of bargaining, the details of each alternative in $\mathcal{A}$ are unimportant; all that is really important is how much utility each alternative has for each bargainer. Thus, we can forget $\mathcal{A}$ and instead consider the image set $\widetilde{\mathcal{A}} = U(\mathcal{A}) \subset \mathbb{R}^2$ (Figure 5.1B). We treat the bargainers as negotiating over elements of $\widetilde{\mathcal{A}}$.

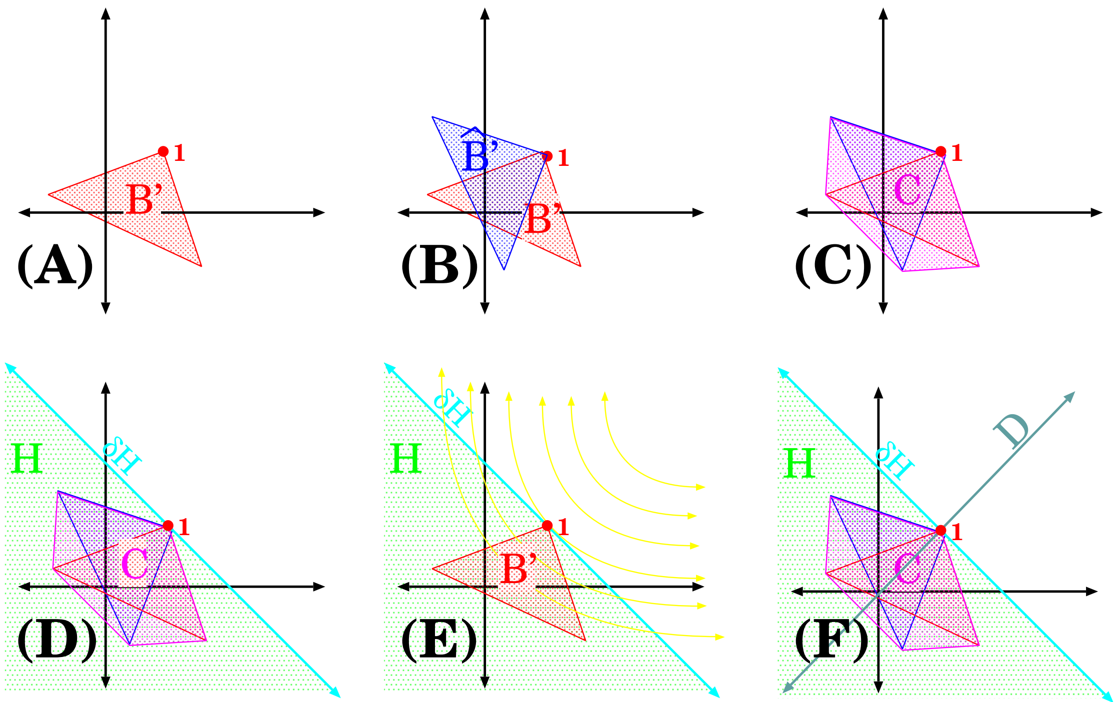

Figure 5.2: **(A)** The bargaining set $\mathcal{B}'$. **(B)** $\widehat{\mathcal{B}}' = \{(b_w, b_v) \in \mathbb{R}^2 \; ; \; (b_v, b_w) \in \mathcal{B}'\}$. **(C)** $\mathcal{C}$ is the convex closure of $\mathcal{B}' \cup \widehat{\mathcal{B}}'$. **(D)** $\mathcal{H} = \{\mathbf{h} \in \mathbb{R}^2 \; ; \; h_1 + h_2 < 2\}$. **(E)** The level curve of $\widetilde{U}$ must be tangent to the boundary of $\mathcal{B}'$ at $\mathbb{1}$. **(F)** $\mathcal{D} = \{\mathbf{d} \in \mathbb{R}^2 \; ; \; d_v = d_w\}$, and $\mathcal{C}^* \cap \mathcal{D} = \{\mathbb{1}\}$.

A **convex combination** of elements in $\widetilde{\mathcal{A}}$ is any linear combination:

$$\sum_{j=1}^{J} c_j \mathbf{a}_j$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_J \in \widetilde{\mathcal{A}}$, and where $c_1, \ldots, c_j \in [0,1]$ are coefficients such that $\sum_{j=1}^{J} c_j = 1$. This convex combination represents the 'average utility' obtained by 'mixing' several alternatives together. There are several ways in which we could 'mix' alternatives:

- Suppose $\mathcal{A}$ represents alternatives which could be shared over time. Then a convex combination represents a 'time-sharing' agreement.

  For example, suppose Veronique and Waldemar are roommates, and each element of $\mathcal{A}$ represents a division of household chores. Suppose A and B are two such divisions:

  A: Veronique washes dishes, Waldemar washes floors.

  B: Veronique washes floors, Waldemar washes dishes.

If $\mathbf{a} = U(A)$ and $\mathbf{b} = U(B)$, then the convex combination $\frac{1}{3}\mathbf{a} + \frac{2}{3}\mathbf{b}$ represents the 'time shared' outcome, 'one third of the time, Veronique washes the dishes and Waldemar washes the floor; two thirds of the time, Veronique washes the floor and Waldemar washes the dishes'.

- Suppose $\mathcal{A}$ is a set of *quantitative* alternatives (eg. a division of goods or money). Then convex combinations represent compromises obtained by 'splitting the difference' between alternatives in $\mathcal{A}$.

  For example, suppose v is the *Vintner's Union*, and w is the management of *Wilson Wine Works*. Suppose A and B are two possible contracts:

  A: Wage of \$10.00/hour, and 21 days paid vacation

  B: Wage of \$12.00/hour, and only 7 days paid vacation.

  If $\mathbf{a} = U(A)$ and $\mathbf{b} = U(B)$, then the convex combination $\frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{b}$ might[1] represent a compromise contract with a wage of \$11.00/hour and 14 days paid vacation.

- If v and w are willing to gamble over the outcome, then a convex combination represents a lottery. If $\mathbf{a}_j = U(A_j)$ for all $j \in [1..J]$, then the convex combination $\sum_{j=1}^{J} c_j \mathbf{a}_j$ represents the *expected utility* of the lottery where alternative A_j has probability c_j of occurring.

As shown in Figure 5.1(C), the **bargaining set** $\mathcal{B}$ is the *closed convex hull* of $\widetilde{\mathcal{A}}$. That is,

$$\mathcal{B} \quad = \quad \overline{\left\{ \sum_{j=1}^{J} c_j \mathbf{a}_j \ ; \ \mathbf{a}_1, \ldots, \mathbf{a}_J \in \widetilde{\mathcal{A}}; \ c_1, \ldots, c_j \in [0,1]; \ \sum_{j=1}^{J} c_j \ = \ 1 \right\}}$$

Thus, $\mathcal{B}$ is the set of all joint utilities which are realizable through some kind of timesharing, compromise, or lottery. We will therefore assume that the bargainers can choose any point in $\mathcal{B}$. In other words, we will assume the axiom of *Convexity:*

(C) *The bargaining set $\mathcal{B}$ is always a closed, convex subset of $\mathbb{R}^2$.*

Next, we assume that each player begins with an initial *endowment* or *bargaining position*; these endowments determine the **status quo** alternative $A^* \in \mathcal{A}$. If the players cannot come to a mutually agreeable arrangement, then either one can terminate negotiations and both will receive the 'status quo' outcome. For example

- If Veronique and Waldemar are potential roommates, then A^* means they cannot come to a satisfactory cohabitation agreement, and thus, they do not become roommates.

[1] I say 'might', because this assumes that the utilities of workers and management are 'linear' in the variables 'wage' and 'holiday time', and this assumption is generally not true.

- If the Vintner's Union and Wilson Wine Works are negotiating a new contract, then A^* means 'no contract'. If the Vintners choose A^*, this is tantamount to a strike; if Wilson chooses A^*, this is tantamount to a lock-out.

Let $\mathbf{a}^* = U(A^*) \in \mathbb{R}^2$, and suppose $\mathbf{a}^* = (a_v^*, a_w^*)$, where a_v^* is the utility of A^* for v, and a_w^* is the utility of A^* for w. If $\mathbf{b} \in \mathcal{B}$, and $\mathbf{b} = (b_v, b_w)$, then clearly, $\mathbf{b}$ is acceptable to v only if $b_v \geq a_v^*$. Likewise, $\mathbf{b}$ is acceptable to w only if $b_w \geq a_w^*$. In other words, a bargain will only occur if it is mutually beneficial. We therefore assume the axiom of *Mutual Benefit*:

(MB) *An element* $\mathbf{b} \in \mathcal{B}$ *is an acceptable bargain only if* $b_v \geq a_v^*$ *and* $b_w \geq a_w^*$. (Figure 5.1E).

If $\mathbf{b}, \mathbf{c} \in \mathcal{B}$, then we say $\mathbf{b}$ is **Pareto preferred** to $\mathbf{c}$ if $b_v \geq c_v$ and $b_w \geq c_w$ —in other words, both v and w agree that $\mathbf{b}$ is 'better' than $\mathbf{c}$ (Figure 5.1F). We then write $\mathbf{b} \underset{P}{\succeq} \mathbf{c}$. Thus, the axiom **(MB)** can be rephrased: $\mathbf{b}$ *is an acceptable bargain only if* $\mathbf{b} \underset{P}{\succeq} \mathbf{a}^*$.

Clearly, if $\mathbf{b}$ is Pareto-preferred to $\mathbf{c}$, then v and w will never choose $\mathbf{c}$ if they could instead choose $\mathbf{b}$. We say that $\mathbf{b}$ is a **Pareto optimal** element of $\mathcal{B}$ if there exists no $\mathbf{c} \in \mathcal{B}$ with $\mathbf{c} \underset{P}{\succeq} \mathbf{b}$ (except $\mathbf{b}$ itself). The **Pareto frontier** of $\mathcal{B}$ is the set $\mathcal{B}^*$ of Pareto-optimal points in $\mathcal{B}$ (Figure 5.1G). If $\mathcal{B}$ is a bounded domain in $\mathbb{R}^2$, then $\mathcal{B}^*$ is the 'northeast' frontier of this domain. Clearly, v and w will only agree to a Pareto-optimal bargain. We therefore assume *Pareto Optimality:*

(P) *An element* $\mathbf{b} \in \mathcal{B}$ *is an acceptable bargain only if* $\mathbf{b} \in \mathcal{B}^*$ *—ie.* $\mathbf{b}$ *is Pareto-optimal.*

As we have seen, any bargaining problem can be represented by a pair $(\mathcal{B}, \mathbf{a}^*)$, where $\mathcal{B} \subset \mathbb{R}^2$ is some convex *bargaining set*, and $\mathbf{a}^* \in \mathcal{B}$ is some *status quo* point. We will thus refer to the ordered pair $(\mathcal{B}, \mathbf{a}^*)$ as a **bargaining problem** (Figure 5.1D). Given a bargaining problem $(\mathcal{B}, \mathbf{a}^*)$, the von Neumann-Morgenstern **negotiating set** is the set of bargains satisfying axioms **(MB)** and **(P)**:

$$\mathcal{N}(\mathcal{B}, \mathbf{a}^*) \quad = \quad \left\{ \mathbf{b} \in \mathcal{B}^* \ ; \ \mathbf{b} \underset{P}{\succeq} \mathbf{a}^* \right\} \qquad \text{(Figure 5.1H).}$$

The above reasoning implies that any mutually agreeable outcome will always be an element of $\mathcal{N}(\mathcal{B}, \mathbf{a}^*)$. There are now three cases:

(0) If $\mathcal{N}(\mathcal{B}, \mathbf{a}^*)$ is empty, then there is no mutually agreeable bargain, and v and w will chose the status quo $\mathbf{a}^*$.

(1) If $\mathcal{N}(\mathcal{B}, \mathbf{a}^*) = \{\mathbf{b}\}$ is a singleton, then $\mathbf{b}$ represents the unique mutually agreeable bargain for v and w.

(2) If $\mathcal{N}(\mathcal{B}, \mathbf{a}^*)$ contains many points (the usual case), then we must impose further constraints.

Case **(2)** is the most common, and we will focus on this case. Clearly, v prefers the easternmost point on $\mathcal{N}(\mathcal{B}, \mathbf{a}^*)$, while w prefers the northernmost point in $\mathcal{N}(\mathcal{B}, \mathbf{a}^*)$. These goals are incompatible, and the bargaining problem comes down to finding a 'fair' compromise between these extremes.

5.2 The Nash Arbitration Scheme

An **arbitration scheme** is a function which takes any bargaining problem $(\mathcal{B}, \mathbf{a}^*)$ as input, and yields, as output, a unique point $\mathbf{b} \in \mathcal{N}(\mathcal{B}, \mathbf{a}^*)$ which is deemed the 'fair' outcome of this bargaining problem. Nash [29][30, §6.5] proposes that any 'fair' arbitration scheme must satisfy three axioms. Each axiom constrains what we consider an 'fair' outcome of arbitration. We will discover that the three axioms together restrict the set of 'fair' outcomes to a unique point.

Invariance under rescaling of utility functions: Recall that, in the von Neumann-Morgenstern definition of cardinal utility, the cardinal utility functions of v and w are only well-defined up to *rescaling*. Thus, if v has utility function $U_v : \mathcal{A} \longrightarrow \mathbb{R}$, and we define $U'_v : \mathcal{A} \longrightarrow \mathbb{R}$ by $U'_v(A) = k \cdot U_v(A) + j$ for some constants k and j, then U'_v and U_v are both equally valid as utility functions for v (ie. both satisfy the conditions of the von Neumann Morgenstern theorem).

If v asserts that his utility function is U'_v, then a 'fair' arbitration scheme should produce the same outcome as if v had said his utility function was U_v. If the arbitration scheme was sensitive to a rescaling of v's utility function, then he could manipulate the outcome by choosing the constants k and j so as to skew results in his favour.

Suppose $U'_v : \mathcal{A} \longrightarrow \mathbb{R}$ and $U'_w : \mathcal{A} \longrightarrow \mathbb{R}$ were rescaled versions of U_v and U_w, and let $U' : \mathcal{A} \longrightarrow \mathbb{R}^2$ be the resulting joint utility function. Let $\widetilde{\mathcal{A}}' = U'(\mathcal{A})$; let $\mathcal{B}'$ be the convex closure of $\widetilde{\mathcal{A}}'$ and let $\mathbf{a}' = U'(A^*)$ be the 'status quo'. We say that the bargaining problem $(\mathcal{B}', \mathbf{a}')$ is a **rescaling** of the bargaining problem $(\mathcal{B}, \mathbf{a}^*)$.

To be precise, suppose $U'_v(A) = k_v \cdot U_v(A) + j_v$ and $U'_w(A) = k_w \cdot U_w(A) + j_w$ for some constants $k_v, k_w, j_v, j_w \in \mathbb{R}$. If we define the **rescaling function** $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ by $F(b_v, b_w) = (k_v b_v + j_v,\ k_w b_w + j_w)$, then clearly $\mathcal{B}' = F(\mathcal{B})$ and $\mathbf{a}' = F(\mathbf{a}^*)$.

Clearly, $(\mathcal{B}', \mathbf{a}')$ and $(\mathcal{B}, \mathbf{a}^*)$ represent the *same* bargaining problem, only with the utility functions for each player 'rescaled' by some amount. A 'fair' arbitration scheme should therefore yield the same outcome for $(\mathcal{B}', \mathbf{a}')$ and $(\mathcal{B}, \mathbf{a}^*)$. Thus, we require the axiom of *Rescaling Invariance:*

(RI) *Suppose $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is a rescaling function, and that $(\mathcal{B}, \mathbf{a}^*)$ is rescaled to $(\mathcal{B}', \mathbf{a}')$ via F. If $\mathbf{b} \in \mathcal{B}$, then* $\Big($ $\mathbf{b}$ *is a fair outcome of* $(\mathcal{B}, \mathbf{a}^*)$ $\Big) \iff \Big($ $F(\mathbf{b})$ *is a fair outcome of* $(\mathcal{B}', \mathbf{a}')$. $\Big)$

Symmetry: In a 'fair' arbitration scheme, the two parties should be treated equally. Thus, if v and w switch places, then the arbitration scheme should respond by switching the outcomes.

If $(\mathcal{B}, \mathbf{a}^*)$ is a bargaining problem, then the *dual* bargaining problem $(\widehat{\mathcal{B}}, \widehat{\mathbf{a}}^*)$ is obtained by switching coordinates. That is,

$$\widehat{\mathcal{B}} \;=\; \big\{(b_w, b_v) \in \mathbb{R}^2\ ;\ (b_v, b_w) \in \mathcal{B}\big\}$$

and, if $\mathbf{a}^* = (a^*_v, a^*_w)$, then $\widehat{\mathbf{a}}^* = (a^*_w, a^*_v)$. The axiom of *Symmetry* states:

(S) *Let* $\mathbf{b} = (b_v, b_w) \in \mathcal{B}$*, and let* $\widehat{\mathbf{b}} = (b_w, b_v)$*. Then*

$$\Big(\ \mathbf{b} \text{ is a fair outcome of } (\mathcal{B}, \mathbf{a}^*)\ \Big) \iff \Big(\ \widehat{\mathbf{b}} \text{ is a fair outcome of } (\widehat{\mathcal{B}}, \widehat{\mathbf{a}}^*)\ \Big).$$

Lemma 5.1 *Suppose* $\mathcal{B}$ *is a* symmetric *set (ie.* $\widehat{\mathcal{B}} = \mathcal{B}$*) and* $a_v^* = a_w^*$ *(ie. the bargainers have identical status quo positions). If* $\mathbf{b}$ *is any fair outcome satisfying* **(S)***, then* $b_v = b_w$ *(ie. the outcome is identical for each bargainer).*

Proof: <u>**Exercise 24**</u> □

(In fact, Lemma 5.1 is really the only consequence of **(S)** we will need, so in Nash's original work, he *defined* axiom **(S)** to be the statement of Lemma 5.1).

Independence of Irrelevant Alternatives: Suppose the bargainers are initially negotiating over some bargaining problem $(\mathcal{B}, \mathbf{a}^*)$, and they agree on outcome $\mathbf{b}$. Suppose that they then discover that, *actually*, their range of alternatives was more restricted than they thought; so that the *real* bargaining set was some subset $\mathcal{B}' \subset \mathcal{B}$. The good news, however, is that $\mathbf{b}$ is *still* an admissible outcome; ie. $\mathbf{b} \in \mathcal{B}'$. Clearly, if $\mathbf{b}$ was a mutually agreeable outcome for the bargaining set $\mathcal{B}$, it should *still* be mutually agreeable for $\mathcal{B}'$.

We can reverse this scenario: suppose the bargainers initially agree on an outcome $\mathbf{b}$ for the bargaining problem $(\mathcal{B}', \mathbf{a}^*)$. Suppose now that their alternatives are *enhanced*, so that the bargaining set is *expanded* to some superset $\mathcal{B} \supset \mathcal{B}'$. The outcome of the new bargaining problem $(\mathcal{B}, \mathbf{a}^*)$ should either *remain* $\mathbf{b}$, or, if it changes, it should change to some previously inaccessible outcome —ie. an element of $\mathcal{B} \setminus \mathcal{B}'$. It certainly makes no sense for the players to change from $\mathbf{b}$ to another element of $\mathcal{B}'$, when confronted with the richer possibilities of $\mathcal{B}$. We thus arrive at the axiom of *Independence of Irrelevant Alternatives*:

(IIA) *Suppose* $\mathcal{B}' \subset \mathcal{B}$*. If the bargain* $\mathbf{b}$ *is a fair outcome of* $(\mathcal{B}, \mathbf{a}^*)$*, and* $\mathbf{b} \in \mathcal{B}'$*, then* $\mathbf{b}$ *is also a fair outcome of* $(\mathcal{B}', \mathbf{a}^*)$*.*

The Nash Solution: If $(\mathcal{B}, \mathbf{a}^*)$ is a bargaining problem, then the **Nash utility function** is the function $\widetilde{U} : \mathcal{B} \longrightarrow \mathbb{R}$ defined:

$$\widetilde{U}(b_v, b_w) \quad = \quad (b_v - a_v^*) \cdot (b_w - a_w^*)$$

In other words, for any $\mathbf{b} \in \mathcal{B}$, we first compute the *net change* in utility from the status quo for each bargainer. Then we multiply these net changes. The **Nash solution** is the (unique) point in $\mathcal{B}$ which maximizes the value of $\widetilde{U}$. The **Nash arbitration scheme** says:

- *Given any bargaining problem* $(\mathcal{B}, \mathbf{a}^*)$*, choose the Nash solution as the outcome.*

This may seem like a strange arbitration scheme, but we have the following:

Nash's Theorem: *There exists a unique Nash solution to any bargaining problem.*

Furthermore, the Nash arbitration scheme is the unique arbitration scheme satisfying axioms **(RI)**, **(S)** *and* **(IIA)**.

Proof: *Existence:* To see that a Nash solution exists, we must show that the function $\widetilde{U}$ takes a maximum on $\mathcal{B}$. This follows from the fact that $\widetilde{U}$ is continuous, and that $\mathcal{B}$ is a compact subset of $\mathbb{R}^2$.

Uniqueness: We must show that $\widetilde{U}$ cannot have *two* maxima in $\mathcal{B}$. We'll use two facts:

- $\mathcal{B}$ is a convex set.
- $\widetilde{U}$ is a *strictly convex* function. That is, for any $\mathbf{b}_1 \neq \mathbf{b}_2 \in \mathbb{R}^2$, and any $c_1, c_2 \in (0,1)$ such that $c_1+c_2 = 1$, we have $\widetilde{U}(c_1\mathbf{b}_1+c_2\mathbf{b}_2) \quad > \quad c_1\widetilde{U}(\mathbf{b}_1)+c_2\widetilde{U}(\mathbf{b}_2)$. (**Exercise 25**)

Suppose M was the maximal value of $\widetilde{U}$ in $\mathcal{B}$, and suppose $\mathbf{b}_1$ and $\mathbf{b}_2$ were both maxima for $\widetilde{U}$, so that $\widetilde{U}(\mathbf{b}_1) = M = \widetilde{U}(\mathbf{b}_2)$. Let $\mathbf{b} = \frac{1}{2}\mathbf{b}_1 + \frac{1}{2}\mathbf{b}_2$. Then $\mathbf{b} \in \mathcal{B}$ (because $\mathcal{B}$ is convex), and $\widetilde{U}(\mathbf{b}) > \frac{1}{2}M + \frac{1}{2}M = M$, (because $\widetilde{U}$ is strictly convex), thereby contradicting the maximality of M. By contradiction, the maximum of F in $\mathcal{B}$ must be *unique.*

The Nash solution satisfies **(RI)***:* Let $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a rescaling function defined: $F(b_v, b_w) = (k_v b_v + j_v,\ k_w b_w + j_w)$. Let $\mathcal{B}' = F(\mathcal{B})$ and let $\mathbf{a}' = F(\mathbf{a}^*)$. Let $\widetilde{U}' : \mathcal{B}' \longrightarrow \mathbb{R}$ be the Nash function for the bargaining problem $(\mathcal{B}', \mathbf{a}')$.

Claim 1: **(a)** *For any $\mathbf{b} \in \mathcal{B}$, if $\mathbf{b}' = F(\mathbf{b})$, then $\widetilde{U}'(\mathbf{b}') \; = \; k_v k_w \cdot \widetilde{U}(\mathbf{b})$.*

(b) *Thus,* $\Big($ $\mathbf{b}$ *is the maximum of* $\widetilde{U}$ *on* $\mathcal{B}$ $\Big) \iff \Big($ $\widehat{\mathbf{b}}$ *is the maximum of* $\widehat{\widetilde{U}}'$ *on* $\widehat{\mathcal{B}}$ $\Big)$.

Proof: If $\mathbf{b} = (b_v, b_w)$, then $\mathbf{b}' = (k_v b_v + j_v,\ k_w b_w + j_w)$. Also, $\mathbf{a}' = (k_v a^*_v + j_v,\ k_w a^*_w + j_w)$. Thus,

$$
\begin{aligned}
\widetilde{U}'(\mathbf{b}') &= (b'_v - a'_v) \cdot (b'_w - a'_w) \\
&= \Big((k_v b_v + j_v) - (k_v a^*_v + j_v)\Big) \cdot \Big((k_w b_w + j_w) - (k_w a^*_w + j_w)\Big) \\
&= (k_v b_v - k_v a^*_v) \cdot (k_w b_w - k_w a^*_w) \\
&= k_v \cdot k_w \cdot (b_v - a^*_v) \cdot (b_w - a^*_w) \quad = \quad k_v k_w \cdot \widetilde{U}(\mathbf{b}).
\end{aligned}
$$

.... □ [Claim 1]

The Nash solution satisfies **(S)***:* Let $\widehat{\mathcal{B}}$ and $\widehat{\mathbf{a}}^*$ be as in the definition of **(S)**. Let $\widehat{\widetilde{U}} : \widehat{\mathcal{B}} \longrightarrow \mathbb{R}$ be the Nash function for the bargaining problem $(\widehat{\mathcal{B}}, \widehat{\mathbf{a}}^*)$.

Claim 2: **(a)** *For any $\mathbf{b} = (b_v, b_w) \in \mathcal{B}$, if $\widehat{\mathbf{b}} = (b_w, b_v) \in \widehat{\mathcal{B}}$, then $\widehat{\widetilde{U}}(\widehat{\mathbf{b}}) \; = \; \widetilde{U}(\mathbf{b})$.*

(b) *Thus,* $\Big($ $\mathbf{b}$ *is the maximum of* $\widetilde{U}$ *on* $\mathcal{B}$ $\Big) \iff \Big($ $\widehat{\mathbf{b}}$ *is the maximum of* $\widehat{\widetilde{U}}$ *on* $\widehat{\mathcal{B}}$ $\Big)$.

Proof: $\widehat{\widetilde{U}}(\widehat{\mathbf{b}}') \; = \; (\widehat{b}_v - \widehat{a}^*_v) \cdot (\widehat{b}_w - \widehat{a}^*_w) \; = \; (b_w - a^*_w) \cdot (b_v - a^*_v) \; = \; (b_v - a^*_v) \cdot (b_w - a^*_w) \; = \; \widetilde{U}(\mathbf{b})$.
□ [Claim 2]

The Nash arbitration scheme satisfies **(IIA)***:* **<u>Exercise 26</u>** .

Nash is the unique scheme satisfying axioms **(RI)***,* **(S)** *and* **(IIA)***:* Suppose was some arbitration scheme satisfying **(RI)**, **(S)** and **(IIA)**. Let $(\mathcal{B}, \mathbf{a}^*)$ be a bargaining problem, and let $\mathbf{b}$ be the Nash solution. We will show that $(\mathcal{B}, \mathbf{a}^*) = \mathbf{b}$.

Let $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be the rescaling defined: $F(c_v, c_w) = \left(\dfrac{c_v - a_v^*}{b_v}, \dfrac{c_w - a_w^*}{b_w} \right)$, for any $\mathbf{c} = (c_v, c_w) \in \mathcal{B}$.

Thus, $F(\mathbf{a}^*) = \mathbf{o} := (0,0)$ and $F(\mathbf{b}) = \mathbb{1} := (1,1)$. Let $\mathcal{B}' = F(\mathcal{B})$ (Figure 5.2A). Thus, the axiom **(AI)** implies that

$$\Big(\ (\mathcal{B}, \mathbf{a}^*) = \mathbf{b} \ \Big) \iff \Big(\ (\mathcal{B}', \mathbf{o}) = \mathbb{1} \ \Big). \tag{5.1}$$

Hence we will prove that $(\mathcal{B}', \mathbf{o}) = \mathbb{1}$. To do this, let $\widehat{\mathcal{B}'} = \{(b_w, b_v) \in \mathbb{R}^2 \ ; \ (b_v, b_w) \in \mathcal{B}'\}$ (Figure 5.2B), and let $\mathcal{C}$ be the convex closure of $\mathcal{B}' \cup \widehat{\mathcal{B}'}$ (Figure 5.2C). Then $\mathcal{C}$ is a convex, symmetric set containing $\mathcal{B}'$.

Consider the halfspace $\mathcal{H} = \{\mathbf{h} \in \mathbb{R}^2 \ ; \ h_1 + h_2 < 2\}$ (Figure 5.2D).

Claim 3: $\mathcal{C} \subset \mathcal{H}$.

Proof: The Nash utility function for $(\mathcal{B}', \mathbf{o})$ is just the function $\widetilde{U}(b_v, b_w) = b_v \cdot b_w$. Claim 1(b) implies that $\mathbb{1}$ is the maximum of $\widetilde{U}$ in $\mathcal{B}'$. Thus, the level curve of $\widetilde{U}$ must be tangent to the boundary of $\mathcal{B}'$ at $\mathbb{1}$ (Figure 5.2E). But the level curve of $\widetilde{U}$ at $\mathbb{1}$ has slope -1 (**<u>Exercise 27</u>**); in other words, it is tangent to the line $\partial\mathcal{H} = \{\mathbf{h} \in \mathbb{R}^2 \ ; \ h_1 + h_2 = 2\}$.
It follows that the boundary of $\mathcal{B}'$ is tangent to $\partial\mathcal{H}$ at $\mathbb{1}$. But $\mathcal{B}'$ is convex, so it follows that *all* of $\mathcal{B}'$ must be below $\partial\mathcal{H}$ (**<u>Exercise 28</u>**); in other words, $\mathcal{B}' \subset \mathcal{H}$. Thus, $\widehat{\mathcal{B}'} \subset \widehat{\mathcal{H}}$. But $\mathcal{H}$ is symmetric under exchange of coordinates; hence $\widehat{\mathcal{H}} = \mathcal{H}$, so we have $\widehat{\mathcal{B}'} \subset \mathcal{H}$. Thus, $\mathcal{B}' \cup \widehat{\mathcal{B}'} \subset \mathcal{H}$. Since $\mathcal{H}$ is convex, we conclude that the convex closure $\mathcal{C}$ must also be a subset of $\mathcal{H}$. .... □ `[Claim 3]`

Let $\mathcal{C}^*$ be the Pareto frontier of $\mathcal{C}$.

Claim 4: $\mathbb{1} \in \mathcal{C}^*$.

Proof: We know that $\mathbb{1} \in \mathcal{C}$ because $\mathbb{1} \in \mathcal{B}$. Suppose $\mathbf{c} \in \mathcal{C}$ was strictly Pareto preferred to $\mathbb{1}$. Then $c_v \geq 1$ and $c_w \geq 1$, and at least one of these is a strict inequality. Hence $c_v + c_w > 1 + 1 = 2$. Hence $\mathbf{c} \notin \mathcal{H}$, contradicting Claim 3. .... □ `[Claim 4]`

Claim 5: $(\mathcal{C}, \mathbf{o}) = \mathbb{1}$.

Proof: Let $(\mathcal{C}, \mathbf{o}) = \mathbf{c}$. Since $\mathcal{C}$ and $\mathbf{o}$ are both symmetric, the symmetry axiom **(S)** implies that $\mathbf{c}$ must also be symmetric —ie. $c_v = c_w$. Thus, if $\mathcal{D} = \{\mathbf{d} \in \mathbb{R}^2 \ ; \ d_v = d_w\}$ is the diagonal line in Figure 5.2(F), then we know $\mathbf{c} \in \mathcal{D}$. However, the Pareto axiom **(P)** also requires that $\mathbf{c}$ be an element of the Pareto frontier $\mathcal{C}^*$. Thus, $\mathbf{c} \in \mathcal{C}^* \cap \mathcal{D}$.

Now Claim 4 says $\mathbb{1} \in \mathcal{C}^*$, and clearly, $\mathbb{1} \in \mathcal{D}$; hence $\mathbb{1} \in \mathcal{C}^* \cap \mathcal{D}$; However, $\mathcal{C}^*$ can never be tangent to $\mathcal{D}$ (**Exercise 29**), so we know that $\mathcal{C}^*$ and $\mathcal{D}$ can intersect in at most one place; hence $\mathcal{C}^* \cap \mathcal{D} = \{\mathbb{1}\}$. Thus, we must have $\mathbf{c} = \mathbb{1}$. .... □ `[Claim 5]`

Since $\mathbb{1} \in \mathcal{B}' \subset \mathcal{C}$, it follows from Claim 5 and axiom **(IIA)** that $(\mathcal{B}', \mathbf{o}) = \mathbb{1}$. Thus, eqn.(5.1) implies that $(\mathcal{B}, \mathbf{a}^*) = \mathbf{b}$. □

Discussion: The Nash arbitration scheme is applicable when the axioms **(RA)**, **(S)**, and **(IIA)** are appropriate. However, the Nash scheme has been rejected by some who argue that the axioms **(RA)** and **(S)** are often inappropriate.

For example, the 'Rescaling' axiom **(RA)** explicitly disallows any judgements about the 'intensity' of the utility valuations of the bargainers. In some cases, this may be appropriate, because we have limited information, and because subjective testimonials by the bargainers concerning the 'intensity' of their feelings will be at best imprecise, and at worst, deliberately manipulative.

However, in some scenarios, it is *clear* that one bargainer has much stronger preferences than the other. Suppose that Veronique is starving, and trying to beg food from Waldemar the grocer. We would all agree that Veronique's utility valuations regarding food will be much more 'intense' than Waldemar's. However, the axiom **(RA)** does not allow us to include this information in our model.

The 'Symmetry' axiom **(S)** is also contentious. **(S)** is appropriate when dealing with two individual persons, if we assume that all persons must be treated equally. However, in many bargaining situations, the bargainers are not *people* but *groups*. For example, v might be a union representing the interests of 5000 workers, while w is the management of a company, representing the interests of 3000 shareholders. It is not clear how we should weigh the 'moral importance' of 5000 workers against 3000 shareholders, but it is clear that an *a priori* insistence on exactly symmetric treatment is simpleminded at best.

In some cases, both **(S)** and **(RA)** are questionable. For example, suppose that the poor developing country of Voderstan is negotiating a multimillion dollar development loan with the World Bank. The Voderstan negotiators represent of millions of peasants, for whom the terms of the loan may literally make the difference between survival and starvation. The World Bank represents the interests of wealthy industrialized nations, and in particular, the interests of a few thousand wealthy financiers in these nations, for whom the terms of the loan mean a 1% difference in their annual profit. In this case, great asymmetry exists, both in the nature of the parties, and in the intensity of their preferences.

To obviate the objections to **(S)**, we can insist that the Nash solution only be applied to bargaining between individuals. A bargaining scenario between two *groups* can then be treated as a *multiparty* bargain between all the individuals comprising these groups. For example, arbitration between a labour union (representing 5000 workers) and management (representing 3000 shareholders) can be treated as a multiparty arbitration involving 8000 individuals. The von Neumann-Morgenstern model and the Nash arbitration scheme generalize to multiparty bargaining in the obvious way, and Nash's Theorem still holds in this context.

To obviate the objections to **(RI)**, we can insist that all individuals must assign utilities to the same collection of alternatives. For example, if starving Veronique is begging food from Waldemar the grocer, the axiom **(S)** is inappropriate, because Veronique's alternatives are eating versus starving, while Waldemar's alternatives are more versus less profit. To apply **(S)**, we should ask Waldemar to also assign utility to the (imaginary) scenario where he is a starving person, and we should ask Veronique to assign a utility to the (imaginary) scenario where she is a struggling grocer trying to remain in business (even though these imaginary scenarios are not part of the bargaining set). The problem is that we may not be able to obtain accurate utility estimates for such far-fetched imaginary scenarios. For example, it is probably impossible for a World Bank executive to imagine the life of a malnourished Voderstanian peasant who spends her days wading through a rice paddy under a subtropical sun. It is equally impossible for the peasant to imagine the life of the executive. So how can we expect them to accurately assign 'utilities' to one another's circumstances?

Chapter 6

Fair Divisions

All bad precedents began as justifiable measures. —Gaius Julius Caesar

Prerequisites: None.

Suppose two or more people are dividing a cake. What is a procedure we can use to guarantee that each person gets a 'fair' portion? We assume that the participants are selfish and do not trust one another, and perhaps even dislike one another. Nevertheless, we seek a procedure, which the people can execute themselves (ie. without the intervention of some arbiter), which ensures that each person will come away feeling that she has a fair share. Think of this procedure as a 'game' such that, if each person plays 'rationally', then all people will be ensured a fair outcome.

Example 6.1: I cut, you choose

The classic cake division procedure for two people is well-known. If Veronique and Waldemar are trying to split a cake, then one of them (say, Veronique) divides the cake into two parts, and the other (Waldemar) chooses which half he likes better. Veronique then takes the remaining half. Veronique therefore has an incentive to make the division as even as possible, to avoid the risk of getting a smaller piece.

To see this, note that Veronique's worst-case scenario is that Waldemar takes the larger portion, thereby leaving her with the smaller portion. In the language of game theory, the smaller portion is her *minimum payoff*. Hence, she seeks a *maximin* strategy: the cake division which *maximizes* the *minimum* portion which she could receive. Clearly, her unique maximin strategy is to make sure the cake is divided into two exactly equal portions.

Since Waldemar always picks what he perceives as the larger portion, he will always perceive the outcome as fair. If Veronique plays rationally (ie. according to her maximin strategy), then she will also see the outcome as 'fair' no matter which piece Waldemar chooses, because she has ensured that both portions are equal (in her perception).

The beauty of this procedure is that it does not require any 'referee' or 'arbiter' to decide the fair outcome; the fair outcome arises naturally from the rational choices of the players. Can this elegant solution to the 'cake-cutting problem' be generalized to three or more people? The problem becomes more complicated if the players actually have different preferences (eg. Veronique likes the vanilla part more, but Waldemar likes chocolate more) or if the cake has 'indivisible' components (ie. nuts or cherries which cannot easily be divided).

Of course, our real goal is not to prevent fist-fights at birthday parties. 'Cake-cutting' is a metaphor for a lot of contentious real-life problems, including:

- Resolving a border dispute between two or more warring states.
- Dividing an inheritance amongst squabbling heirs.
- Splitting the property in a divorce settlement.
- Allocating important government positions amongst the various factions in a coalition government.
- Defining 'fair use' of intrinsically common property (eg. resource rights in international waters).

Fair division procedures can be used to divide up 'bads' as well as 'goods'. In this case, each participant seeks to *minimize* their portion, rather than *maximizing* it. Some examples include:

- Partitioning chores amongst the members of a household.
- Allocating military responsibilities to different member states in an alliance.

Fair division is generally more complicated than the simple and elegant 'I cut, you choose' algorithm for two persons, because of the following factors:

- There are generally more than two participants.
- The participants may have different preferences (eg. Veronique likes vanilla, Waldemar likes chocolate), or at the very least, different perceptions of the situation. Hence, what looks like a 'fair' division to one person may appear 'unfair' to the other. To mathematically represent this, we endow each person with a *utility measure*, which encodes how *she* values different parts of the cake. See §6.1.
- In dividing a piece of physical territory, there are military and economic reasons why the portions should be *connected*. Thus, we cannot achieve 'fairness' by giving each party a chopped up collection of tiny bits. See §6.2.5.
- The participants may be actively hostile and distrusting of one another. Thus, each of four participants may not only demand that she receives *at least* one quarter, but she may also require that (in her perception) no *other* participant receives *more* than she does. We call this an *envy-free* partition. See §6.3.

- Ideally, we'd like a partition which maximizes the happiness of the participants. For example, if Veronique likes vanilla and Waldemar likes chocolate, it makes sense to give her more of the vanilla, and him more of the chocolate, even if this does *not* result in a strict 50/50 division of the cake. See §6.4 and §6.4.4.

- Some components are indivisible. For example, an inheritance may involve single, high-value items (eg. a car, a painting) which cannot easily be split or shared amongst two heirs. See §6.6.2.

- Some participants may be 'entitled' to a larger share than others. For example, in dividing an inheritance, the spouse of the deceased may be entitled to one half the estate, while each of the three children is entitled to only one sixth. See §6.6.1.

- If the participants have knowledge of one another's preferences, they can cooperate to maximize their common well-being. However, one person can also use this knowledge to *manipulate* the procedure, obtaining a disproportionately large share at someone else's expense. See §6.6.4.

6.1 Partitions, Procedures, and Games

Let $\mathbf{X}$ be a set which represents the cake (or the inheritance, or the disputed territory, etc.). A **portion** is some subset $\mathbf{P} \subset \mathbf{X}$. Each person assigns some utility $\mu(\mathbf{P})$ to the portion $\mathbf{P}$. This defines a function μ from the collection of all subsets of $\mathbf{X}$ to the set of real numbers. We assume that μ satisfies the following axioms:

(U0) $\mu[\emptyset] \;=\; 0$. In other words, the value of an empty portion is zero.

(U1) $\mu[\mathbf{X}] \;=\; 1$. The value of the entire cake is one.

(UA) For any disjoint subsets $\mathbf{P}, \mathbf{Q} \subset \mathbf{X}$, $\quad \mu[\mathbf{P} \sqcup \mathbf{Q}] \;=\; \mu[\mathbf{P}] + \mu[\mathbf{Q}]$. (We say that μ is *additive*.)

More generally, in some procedures involving an sequence of approximations, we require:

(UA$^\infty$) For any infinite sequence of disjoint subsets $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \ldots \subset \mathbf{X}$,

$$\mu[\mathbf{P}_! \sqcup \mathbf{P}_2 \sqcup \mathbf{P}_3 \sqcup \cdots] \quad = \quad \mu[\mathbf{P}_1] \;+\; \mu[\mathbf{P}_2] \;+\; \mu[\mathbf{P}_2] \;+\; \cdots$$

(We say that μ is *sigma-additive*.)

A function μ satisfying properties **(U0)**, **(U1)**, and **(UA)** is called a **utility measure**[1]. We assume that seeks a piece $\mathbf{P}$ which maximizes $\mu(\mathbf{P})$.

[1]Actually μ is a special case of a mathematical object called a *signed measure*. Signed measures cannot, in general, be well-defined on *every* subset of $\mathbf{X}$. Instead, we must choose a collection of 'measurable' subsets called a *sigma-algebra*, which is closed under countable intersections and unions. This is a technicality which we will neglect in this discussion.

Remark on Additivity: The axiom **(UA)** assumes that utilities are *additive*, but in many real situations, this is not true, because it neglects the phenomena of *complementarity* and *substitutability* between various parts of $\mathbf{X}$. For example suppose $\mathbf{X}$ is a collection of food, and $\mathcal{V}$ is a bunch of hungry people, who seek to divide the food in a manner such that each person gets a decent meal. Suppose that $\mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{S} \subset \mathbf{X}$ are four disjoint subsets of $\mathbf{X}$. $\mathbf{P}$ is a piece of *pumpernickel* bread, $\mathbf{Q}$ is *quark* cheese, $\mathbf{R}$ is *rye* bread, and $\mathbf{S}$ is *salami*. For a well-balanced meal, each person wants to make a sandwich with some bread and some filling. Clearly, if I already have **P**umpernickel bread, then what I want next is either **Q**uark or **S**alami. I *don't* want **R**ye bread. Conversely, if I have the **S**alami, then I want bread, not cheese.

In economics jargon, the items $\mathbf{P}$ and $\mathbf{R}$ are **substitutes**; they are both bread, and thus, if you have one, then you no longer desire the other. In other words, the utility of having both $\mathbf{P}$ and $\mathbf{R}$ is *less* than the sum of their separate utilities:

$$\mu[\mathbf{P} \sqcup \mathbf{R}] \quad < \quad \mu[\mathbf{P}] + \mu[\mathbf{R}].$$

On the other hand, the items $\mathbf{P}$ and $\mathbf{Q}$ are **complements**. By themselves, neither is worth very much (a piece of bread by itself is a pretty poor lunch). But together, they make a tasty sandwich. Thus, the value of the combination $\mathbf{P} \sqcup \mathbf{Q}$ is *greater* than the sum of the separate parts:

$$\mu[\mathbf{P} \sqcup \mathbf{Q}] \quad > \quad \mu[\mathbf{P}] + \mu[\mathbf{Q}].$$

There are other more complex ways in which different subsets of $\mathbf{X}$ can combine to have nonadditive utility. For example, suppose $\mathbf{X}$ is a disputed territory and $\mathcal{V}$ is a collection of hostile military powers trying to divide $\mathbf{X}$ between them. Clearly, a *connected* piece of territory is much more valuable to any party than several *disconnected* components. So, suppose $\mathbf{P} \subset \mathbf{X}$ is a disconnected piece of territory, and $\mathbf{Q} \subset \mathbf{X}$ is another disconnected piece of territory, but the combination $\mathbf{P} \sqcup \mathbf{Q}$ is connected (eg. $\mathbf{Q}$ forms a 'bridge' between the two parts of $\mathbf{P}$). Then from a strategic point of view, the unified territory $\mathbf{P} \sqcup \mathbf{Q}$ is worth much more than the sum of the two separate pieces. (See §6.2.5 for a discussion of connectivity in partitions).

Notwithstanding this objection, we will keep axiom **(UA)** because it is a good approximation in many situations, and because it would be too complicated to mathematically represent complementarity and substitutability[2].

Let $\mathcal{V} = \{v_1, \ldots, v_N\}$ be the set of people dividing $\mathbf{X}$; each person v_n has a utility measure μ_n. A **partition** is a collection of disjoint portions $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ so that $\mathbf{X} = \mathbf{P}_1 \sqcup \cdots \sqcup \mathbf{P}_N$. We assume portion $\mathbf{P}_n$ goes to person v_n; this person then assigns the partition a utility of $\mu_n(\mathbf{P}_n)$. Clearly, each person v_n seeks a partition $\mathcal{P}$ which maximizes the value of $\mu_n(\mathbf{P}_n)$.

A *partition procedure* is some function which takes information about the utility measures of the various parties, and yields a partition which (we hope) will satisfy each party. Formally, let be the set of all utility measures on $\mathbf{X}$. Then a complete description of the preferences of

[2]It is perhaps possible to mathematically represent complementarity and substitutability using a measure μ defined on $\mathbf{X} \times \mathbf{X}$ (so that $\mu[\mathbf{P} \times \mathbf{Q}] > 0$ if $\mathbf{P}$ and $\mathbf{Q}$ are complementary, and $\mu[\mathbf{P} \times \mathbf{Q}] < 0$ if $\mathbf{P}$ and $\mathbf{Q}$ are substitutes. We will not pursue this here.

all parties is given by an N-tuple $(\mu_1, \mu_2, \ldots, \mu_N) \in {}^N = \underbrace{\times \; \times \cdots \times}_{N}$. Let ${}_N$ be the set of all possible partitions of $\mathbf{X}$ into N portions. Then an **N-partition procedure** is a function $\Pi : {}^N \longrightarrow {}_N$.

Partition procedures involve 'dividing the value' of the cake, which requires that the value be divisible. Indivisible components of value are called *atoms*, and present obstructions to partition. To be precise, if μ is a utility measure on $\mathbf{X}$, then an **atom** of μ is a point $x \in \mathbf{X}$ such that $\mu\{x\} > 0$. Intuitively, an atom represents a valuable but indivisible item, a 'diamond in the cake' (Akin [1]). We say μ is **nonatomic** if it has no atoms. Failing that, we say that μ is **at most $\frac{1}{N}$ atomic** if the total mass of all atoms of μ is less than $\frac{1}{N}$. In other words, there is a subset $\mathbf{Y} \subset \mathbf{X}$ which contains *no* atoms, such that $\mu[\mathbf{Y}] > 1 - \frac{1}{N}$. The consequence is that any portion of size $\frac{1}{N}$ cannot be *entirely* made of atoms, and hence, is divisible.

Procedure 6.2: I cut, you choose

Let $\mathbf{X} = [0, 1]$ be the unit interval (representing a one-dimensional cake). Let μ_1 and μ_2 be utility measures on $\mathbf{X}$. Assume μ_1 is at most $\frac{1}{2}$ atomic.

(1) Let $r \in [0, 1]$ be such that $\mu_1[0, r) = \frac{1}{2} = \mu_1[r, 1]$ (ie. player v_1 cuts the cake into two pieces which she perceives have equal size; this is possible because μ_1 is at most $\frac{1}{2}$ atomic)

(2a) If $\mu_2[0, r) \geq \mu_2[r, 1]$, then define $\mathbf{P}_2 = [0, r)$ and $\mathbf{P}_1 = [r, 1]$. (If v_2 thinks that $[0, r)$ is larger, then he takes this piece, and v_1 takes the other one).

(2b) Otherwise, if $\mu_2[0, r) < \mu_2[r, 1]$, then define $\mathbf{P}_1 = [0, r)$ and $\mathbf{P}_2 = [r, 1]$. (If v_2 thinks that $[r, 1]$ is larger, then he takes *this* piece, and v_1 takes the other one).

Now let $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2\}$.

In general, of course, we do *not* have complete information about every person's preferences. Instead, each person provides a small amount of information (eg. by proposing a certain portion as 'fair' or by rejecting a proposed portion as 'too small'). We must use this limited information to *interpolate* her true desires. Also, in general, we cannot assume that an 'objective arbiter' will be present to implement the partition procedure; it must be something which the participants can do themselves, even if they are *not* friends and do *not* trust one another. Thus, for practical purposes, we seek not a procedure, but a *game*, so that, if all participants play 'rationally', then the outcome will be *as if* we had implemented some partition procedure.

An N-person **partition game** is a structure $\Gamma := (\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_N;\ \gamma)$, where $\mathcal{S}_n$ is some set of 'strategies' (ie. 'moves') for player v_n, and where $\gamma : \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_N \longrightarrow {}_N$. An N-tuple of strategies $\mathbf{s} := (s_1, \ldots, s_N)$ is called a **play** of the game (representing a specific choice of strategy by each player). Thus, γ is a function which takes any play $\mathbf{s}$, and produces an **outcome** $\gamma(\mathbf{s}) = \mathcal{P} \in {}_N$ which is an N-partition of $\mathbf{X}$. Each player might perform a different 'role' in the game, and thus, different players may have different strategy-sets to choose from. We assume each player picks the strategy which she believes will maximize her portion.

Game 6.3: I cut, you choose

Let $\mathbf{X} = [0, 1]$ be the unit interval (representing a one-dimensional cake). Let μ_1 and μ_2 be two utility measures on $\mathbf{X}$ (with μ_1 being at most $\frac{1}{2}$ atomic). We define the sequential game Γ as follows:

1. First, player v_1 chooses a number $r \in [0, 1]$ (ie. v_1 'cuts the cake'.)
2. Next, player v_2 choses between partitions **(a)** and **(b)**:
   (a) $\mathbf{P}_1 = [0, r)$ and $\mathbf{P}_2 = [r, 1]$.
   (b) $\mathbf{P}_2 = [0, r)$ and $\mathbf{P}_1 = [r, 1]$.

Thus, player v_1's strategy is a point $r \in [0, 1]$, so we can define $\mathcal{S}_1 = [0, 1]$. Player v_2's strategy is to then choose either the left portion or the right portion, so we'll say $\mathcal{S}_2 = \{L, R\}$. The function $\gamma : \mathcal{S}_1 \times \mathcal{S}_2 \longrightarrow_2$ is then defined by $\gamma(r, s) = \{\mathbf{P}_1, \mathbf{P}_2\}$, where

$$\mathbf{P}_1 = [0, r) \text{ and } \mathbf{P}_2 = [r, 1] \text{ if } s = L,$$

and $\mathbf{P}_2 = [0, r)$ and $\mathbf{P}_1 = [r, 1]$ if $s = R$.

Suppose $\mathbf{s} \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_N$ is play of game Γ and $\gamma(\mathbf{s}) = \mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ is the resulting partition. We define $\mu_n(\mathbf{s}) := \mu_n(\mathbf{P}_n)$; this is called the Γ-**payoff** for player v_n in the play $\mathbf{s}$.

A *dominant* strategy for player v_1 is a strategy $s_1^* \in \mathcal{S}_1$ which yields a maximal payoff for v_1, no matter *what* the other players do. Formally, we say that $s_1^* \in \mathcal{S}_1$ is **dominant** if, for *any* counterstrategies $s_2 \in \mathcal{S}_2, \ldots, s_N \in \mathcal{S}_N$, the strategy s_1^* is best for v_1: for any other $s_1 \in \mathcal{S}_1$,

$$\mu_1(s_1^*, s_2, \ldots, s_N) \quad \geq \quad \mu_1(s_1, s_2, \ldots, s_N).$$

Clearly, it is irrational for v_1 to choose anything but a dominant strategy, if she has one. Of course, in general, v_1 may not have a dominant strategy. In this case, v_1 can evaluate the worth of any strategy $s_1 \in \mathcal{S}_1$ by considering its 'worst case scenario'. The **minimum payoff** for s_1 is defined:

$$\underline{\mu}_1(s_1) \quad = \quad \min_{s_2 \in \mathcal{S}_2, \ldots, s_N \in \mathcal{S}_N} \mu_1(s_1, s_2, \ldots, s_N).$$

In other words, $\underline{\mu}_1(s_1)$ is the *worst* payoff which v_1 can expect from s_1 under *any* circumstances. A **maximin strategy** is a strategy $s_1^\dagger \in \mathcal{S}_1$ which *maximizes* her worst-case scenario payoff:

$$\underline{\mu}_1(s_1^\dagger) \quad = \quad \max_{s_1 \in \mathcal{S}_1} \underline{\mu}_1(s_1).$$

The value of $\underline{\mu}_1(s_1^\dagger)$ is then the **maximin payoff** for v_1. This is the *worst* she ever expect to do in the game, if she plays according to her maximin strategy.

If $s_1^* \in \mathcal{S}_1$ is dominant for v_1, then s_1^* is automatically a maximin strategy (**Exercise 30**). However, a maximin strategy may exist even when a dominant strategy does not. We deem it irrational for v_1 to chose anything but a maximin strategy, if one exists.

'I cut, you choose' is a **sequential** game, meaning that the players play one at a time, in numerical order. First v_1 plays, then v_2 plays, and so on. Thus, v_2 *knows* the strategy of v_1, and thus, he can choose dominant/maximin strategies *given* this information. More generally, in a sequential game, player v_n already *knows* the strategies of players $v_1, \ldots, v_{n-1}$, and thus, she can choose dominant/maximin strategies *given* this information.

Example 6.4: I cut, you choose (maximin strategies)

In Game 6.3, Player v_1 has no idea which piece v_2 will think is better (she has no idea what his utility measure is). However, she doesn't want to *risk* getting a small piece. Hence, to *maximize* the utility of the worst-case scenario, her *maximin* strategy is to chose r so that $\mu_1[0, r) = \frac{1}{2} = \mu_1[r, 1]$. In other words, she effectively implements step **(1)**. of the 'I cut, you choose' *procedure* (Procedure 6.2).

Player v_2 plays after v_1, and *given* a strategy $r \in \mathcal{S}_1$ by player v_1, his dominant strategy is clearly to pick the piece he thinks is better. Thus, he will effectively implement steps **(2a)** and **(2b)** of the 'I cut, you choose' *procedure* (Procedure 6.2).

The outcome of a game appears unpredictable. However, we can often predict the outcome if we make four assumptions about the 'psychology' of the players:

(Ψ1) Each player has **complete self-awareness** about her own preferences. In other words, she has 'perfect knowledge' of her own utility measure.

(Ψ2) Each player has **complete ignorance** of other players' preferences. Thus, she cannot in any way predict or anticipate their behaviour.

(Ψ3) Each player is **rational**, in the sense that she carefully considers all of her strategies and the possible counterstrategies of other players. For each possible play $\mathbf{s} \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_N$, she determines what her payoff would be. She thereby determines her *minimum payoff* for each of her possible strategies, and thereby determines her *maximin strategy.*

(Ψ4) Each player is **conservative** (or **risk-averse**), in the sense that she wants to minimize personal risk. She will *not* choose 'risky' strategies which threaten low minimum payoffs (even if they also tempt her with high *maximum* payoffs). Instead, she will 'play safe' and choose the strategy with the best minimum payoff: her *maximin* strategy.

Being *psychological* assertions, we can provide no mathematical justification for these axioms[3]. Nevertheless, we must postulate **(Ψ1)-(Ψ4)** to develop a predictive theory of partition games.

We can translate a partition *game* into a partition *procedure* if, using axioms **(Ψ1)-(Ψ4)**, we can *predict* that the game players will act *as if* they were executing that procedure. To be precise, suppose Π is an N-partition procedure, and Γ is an N-person partition game. We say that Γ **yields** Π if:

1. Each player in Γ has a unique pure maximin strategy, and

[3]Indeed, axioms **(Ψ1)-(Ψ4)** are questionable on purely psychological grounds. Nevertheless, we can argue that, even if they are false, these axioms are at least 'reasonable approximations' which are 'good enough' for practical purposes. See [30, Chapt. 1 & 2] for more discussion of this.

2. If all players play their maximin strategies, then the outcome of Γ will be the same partition as produced by Π.

Thus, Example 6.4 shows that the 'I cut, you choose' *game* (Game 6.3) yields the 'I cut, you choose' *procedure* (Procedure 6.2).

6.2 Proportional Partitions

6.2.1 Introduction

Prerequisites: §6.1

We say that a partition $\mathcal{P}$ is **proportional** if $\mu_n(\mathbf{P}_n) \geq \frac{1}{N}$ for all $n \in [1..N]$. For example, if $\mathcal{V} = \{v_1, v_2\}$, then the partition $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2\}$ is proportional if $\mu_1(\mathbf{P}_1) \geq \frac{1}{2}$ and also $\mu_2(\mathbf{P}_2) \geq \frac{1}{2}$. In other words, each person feels that (in their estimation), they received at *least* half the value of the cake. A partition procedure is **proportional** if it *always* produces proportional partitions.

Example 6.5: 'I cut, you choose' is proportional.

Recall the 'I cut, you choose' procedure (Procedure 6.2). Notice that $\mu_2(\mathbf{P}_2) \geq \frac{1}{2}$ by definition (steps **(2a)** and **(2b)**) and $\mu_1(\mathbf{P}_1) = \frac{1}{2}$ by step **(1)**. Thus, $\mathcal{P}$ will be a proportional partition.

6.2.2 Banach and Knaster's 'Last Diminisher' game

Prerequisites: §6.2.1

Is there a proportional partition procedure for more than two players? Yes.

Procedure 6.6: (Banach & Knaster) [22, 40, 41]

Let $\mathbf{X} = [0, 1]$ be the unit interval. Suppose $\mathcal{V} = \{v_1, \ldots, v_N\}$, and, for each $n \in [1..N]$, let v_n have a utility measure μ_n that is at most $\frac{1}{N}$ atomic. The Banach-Knaster procedure is defined recursively as follows:

1. If $\mathcal{V} = \{v_1, v_2\}$ has only two players, then play the 'I cut, you choose' game (Game 6.3)

2. Suppose $\mathcal{V}$ has $N \geq 3$ players. For each $n \in [1..N]$, let r_n be the largest value such that $\mu_n[1, r_n) = \frac{1}{N}$. In other words, $[1, r_n)$ is the largest piece of cake that v_n thinks is worth $\frac{1}{N}$ of the entire cake.

Claim 1: *Such an r_n exists.*

Proof: Let $R_n \in [0, 1]$ be the smallest value such that $\mu_n[1, R_n) \geq \frac{1}{N}$. If $\mu_n[1, R_n) = \frac{1}{N}$, then we are done. If not, then $\mu_n[1, R_n) > \frac{1}{N}$. Recall that μ_n is at most $\frac{1}{N}$ atomic. Thus $[0, R_n]$ cannot be entirely atomic. We can assume that any atoms in $[0, R_n]$ are clustered near 0, and not near R_n. (This can be achieved if necessary, but cutting $[0, R_n]$ into pieces and reordering them.). Hence we can assume that there is some

$S_n < R_n$ so that $\mu_n[0, S_n] < \frac{1}{N}$, and so that all the atoms in $[0, R_n]$ are actually in $[0, S_n]$. Thus, $(S_n, R_n]$ contains *no* atoms. Now define $f : [S_n, R_n] \longrightarrow \mathbb{R}$ by $f(r) = \mu_n[0, r]$. Then f is continuous on $[S_n, R_n]$ (because there are no atoms). But $f(S_n) < \frac{1}{N} < f(R_n)$; hence the Intermediate Value Theorem yields some $r_n \in (S_n, R_n)$ with $f(r_n) = \frac{1}{N}$. .... ☐ `[Claim 1]`

Let v_n be the player with the smallest value of r_n (if two players are tied, then choose the smaller value of n). We define $\mathbf{P}_n = [0, r_n)$. Observe that $\mu_n(\mathbf{P}_n) = \frac{1}{N}$. (In other words, player v_n thinks he got $\frac{1}{N}$ of the cake.) Let $\mathbf{X}_1 = [r_n, 1]$ (ie. $\mathbf{X}_1$ is the remaining cake).

Claim 2: *For every $m \neq n$, $\mu_m[\mathbf{X}_1] \geq \frac{N-1}{N}$.*

Proof: By hypothesis, $r_n \leq r_m$. Thus, $\mu_m[1, r_n) \leq \mu_m[1, r_m) = \frac{1}{N}$. Thus,

$$\mu_m[r_n, 1] \underset{\text{(UA)}}{=\!=} 1 - \mu_m[1, r_n) \geq 1 - \frac{1}{N} = \frac{N-1}{N}.$$

Here, **(UA)** is by axiom **(UA)** on page 77. .... ☐ `[Claim 2]`

Thus, each of the remaining players thinks that at least $\frac{N-1}{N}$ of the cake remains to be divided.

3. Now let $\mathcal{V}_1 = \mathcal{V} \setminus \{v_n\}$ (the remaining players). We apply the Banach-Knaster procedure recursively to divide $\mathbf{X}_1$ into $N-1$ slices such that each of the players in $v_m \in \mathcal{V}_1$ thinks he got a portion $\mathbf{P}_m$ such that

$$\mu_m[\mathbf{P}_m] \underset{\text{(S2)}}{\geq} \frac{1}{N-1} \cdot \mu_m[\mathbf{X}_1] \underset{\text{(C1)}}{\geq} \left(\frac{1}{N-1}\right) \cdot \left(\frac{N-1}{N}\right) = \frac{1}{N}. \tag{6.1}$$

where **(S2)** follows from step 2 of the procedure, and **(C1)** follows from Claim 1. ________

The Banach-Knaster partition is proportional, because of equation (6.1). Is there a game which yields the Banach-Knaster procedure? Yes.

Game 6.7: 'Last Diminisher' (Banach & Knaster)

Let $\mathbf{X} = [0, 1]$ be the unit interval, and let $\mathcal{V} = \{v_1, \ldots, v_N\}$.

1. Player v_1 cuts a portion from the cake. In other words, v_1 chooses some $r_1 \in [0, 1]$ (the position of the 'cut').
2. Player v_2 then has the option (but is not obliged) to 'trim' this portion; ie. to cut off a small slice and return it to the original cake. In other words, v_2 chooses some $r_2 \in [0, 1]$; if $r_2 < r_1$, then v_2 is 'trims' the portion; if $r_2 \geq r_1$, then he leaves it alone.
3. Player v_3 then has the option (but is not obliged) to 'trim' this new portion; ie. to cut off a small slice and return it to the original cake. In other words, v_3 chooses some $r_3 \in [0, 1]$; if $r_3 < \min\{r_1, r_2\}$, then v_2 is 'trims' the portion; otherwise she leaves it alone.
4. The portion passes by each successive player in turn. Each has the option of trimming off a further slice

5. Once all N players have inspected the portion, the 'Last Diminisher' is the *last player who trimmed the portion* (or player v_1, if no one else touched it).

   The 'Last Diminisher' receives this portion as *his* portion, and leaves the game.

6. The remaining $(N-1)$ players then repeat the game to divide up the remaining cake.

(Observe that, if there are only two players, then the 'Last Diminisher' game is equivalent to 'I cut, you choose'). ______________________

Strictly speaking, the 'Last Diminisher' game is not a partition game. Instead of yielding an entire partition all at once, this game consists of a sequence of *apportionment games*, each of which yields a single portion for a single player. We need some machinery to make this precise.

Apportionment games: let $\mathcal{B}$ be the set of all subsets of $\mathbf{X}$ which *could* be a portion for some player[4]. An N-player **apportionment game** is a structure $\Gamma_N := (\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_N;\ \gamma)$, where $\mathcal{S}_n$ is some set of 'strategies' (ie. 'moves') for player v_n, and where $\gamma : \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_N \longrightarrow \mathcal{B} \times [1..N]$.

Let $\mathbf{s} := (s_1, \ldots, s_N)$ be an N-tuple of strategies, and suppose $\gamma(\mathbf{s}) = (\mathbf{P}, n)$ for some $\mathbf{P} \subset \mathbf{X}$ and $n \in [1..N]$. This means that player v_n gets the portion $\mathbf{P}$, and the other $(N-1)$ players are left to divide the remaining cake $\mathbf{X} \setminus \mathbf{P}$.

Apportionment cascades: To complete the description, we must describe how the remaining players divide the remaining cake. An **apportionment cascade** is a sequence of apportionment games $\mathbf{\Gamma} := (\Gamma_N, \Gamma_{N-1}, \ldots, \Gamma_3, \Gamma_2)$, each of which awards a slice to *one* player, who then leaves the game (technically Γ_2 is a true *partition* game, since it only has two players, and implicitly assigns a portions to each of them). At the end of this sequence, each player has a portion, so, taken as a totality, the apportionment cascade $\mathbf{\Gamma}$ yields a partition game.

Payoffs and Strategies: To evaluate strategies in the apportionment cascade $\mathbf{\Gamma}$, we must define the $\mathbf{\Gamma}$-payoffs for each players. We do this inductively. First, we define the Γ_2-payoffs by treating it as standard partition game. We can then compute the maximin strategies and maximin payoffs for Γ_2.

Next, we move onto Γ_3. Clearly, if player v_n receives the portion $\mathbf{P}$ in the game Γ_3, then his Γ_3-payoff is just $\mu_n(\mathbf{P})$. We define the Γ_3-payoffs of all other players to be their *maximin payoffs* in the game Γ_2. Having defined the Γ_3-payoffs for all players, we can then compute the maximin strategies and maximin payoffs for Γ_3.

Inductively, assume we've computed the maximin payoffs for Γ_{N-1}. Consider Γ_N. Clearly, if player v_n receives the portion $\mathbf{P}$ in the game Γ_N, then his Γ_N-payoff is just $\mu_n(\mathbf{P})$. We define the Γ_N-payoffs of all other players to be their *maximin payoffs* in the game Γ_{N-1}.

Proposition 6.8 *If all players use their maximin strategies, then the Last Diminisher game yields the Banach-Knaster procedure.*

[4] Technically, $\mathcal{B}$ is a sigma-algebra of *measurable* subsets, but we're neglecting this technicality

Proof: Let Γ_N be the 'Last Diminisher' game with N players.

Claim 1: *Player v_1's maximin Γ_N-strategy is to cut portion which she believes is exactly $\frac{1}{N}$ of the whole cake. In other words, v_1 will choose $r_1 \in [0,1]$ so that $\mu_1[0,r_1] = \frac{1}{N}$.*

Her maximin Γ_N-payoff with this strategy is $\frac{1}{N}$.

Proof: (by induction on N)

Base case ($N = 2$): In this case, we're playing 'I cut, you choose', and Example 6.4 shows that v_1's maximin strategy is to choose r_1 so that $\mu_1[0,r_1] = \frac{1}{2}$. Regardless of how v_2 plays, we know that v_1 will end up with a portion $\mathbf{P}_1$ such that $\mu_1[\mathbf{P}_1] = \frac{1}{2}$.

Induction: Suppose the claim is true for $N-1$ players. We first consider three strategies for v_1

`Strategy I:` (v_1 *chooses* r_1 *so that* $\mu_1[0,r_1] \;=\; \frac{1}{N}$.)

In this case, either v_1 gets this portion (a payoff of $\frac{1}{N}$), or someone else gets gets a 'trimmed' version of it. If someone else gets a trimmed version, then this recipient got a piece *smaller* than $\frac{1}{N}$. Hence, *more* than $\frac{N-1}{N}$ cake remains to be divided in Γ_{N-1}. In other words, $\mu_1[\mathbf{X}_1] > \frac{N-1}{N}$.

Now, v_1 enters game Γ_{N-1} with $(N-1)$ other players. By induction hypothesis, v_1's maximin Γ_{N-1}-payoff will be:

$$\frac{1}{N-1}\,\mu_1[\mathbf{X}_1] \quad > \quad \left(\frac{1}{N-1}\right)\cdot\left(\frac{N-1}{N}\right) \quad = \quad \frac{1}{N}.$$

Thus, v_1's minimum Γ_N-payoff under `Strategy I` is $\frac{1}{N}$.

`Strategy II:` (v_1 *chooses* r_1 *so that* $\mu_1[0,r_1] \;>\; \frac{1}{N}$.)

In this case, v_1 runs the risk that someone else will receive this 'oversized' portion. But then the recipient gets *more* than $\frac{1}{N}$ of the cake, which means that *less* than $\frac{N-1}{N}$ cake remains during the next round of play, which will be played amongst $N-1$ players. In other words, $\mu_1[\mathbf{X}_1] < \frac{N-1}{N}$.

Now, v_1 enters game Γ_{N-1}. By induction hypothesis, her maximin Γ_{N-1}-payoff will be:

$$\frac{1}{N-1}\,\mu_1[\mathbf{X}_1] \quad < \quad \left(\frac{1}{N-1}\right)\cdot\left(\frac{N-1}{N}\right) \quad = \quad \frac{1}{N}.$$

Thus, v_1's minimum Γ_N-payoff under `Strategy II` is less than $\frac{1}{N}$.

`Strategy III:` (v_1 *chooses* r_1 *so that* $\mu_1[0,r_1] \;<\; \frac{1}{N}$)

In this case, v_1 runs the risk of getting this 'undersized' portion $[0,r_1]$ if no one else trims it. Hence her minimum Γ_N-payoff is less than $\frac{1}{N}$.

Clearly, `Strategy I` yields the *best* minimum payoff, so this will be v_1's maximin strategy.
$\Box$ `[Claim 1]`

Claim 2: *For any $n > 1$, Player v_n's maximin Γ_N-strategy is to trim the portion if and only if he thinks it is too large, and if so, to trim the portion until he believes it is exactly $\frac{1}{N}$ of the whole cake. In other words, v_n will choose $r_n \in [0,1]$ so that $\mu_1[0, r_n] = \frac{1}{N}$.*

His maximin Γ_N-payoff with this strategy is $\frac{1}{N}$.

Proof: **Exercise 31** The proof is by induction on N, and is similar to Claim 1; the only difference is that, if $N = 2$, then player v_2 takes the role of the 'chooser' in 'I cut, you choose.' □ `[Claim 2]`

Thus, assuming each player follows their maximin strategy, we can redescribe the Last Diminisher Game as follows:

1. Player v_1 cuts from the cake a portion she believes to be of size $\frac{1}{N}$. In other words, v_1 chooses some $r_1 \in [0,1]$ so that $\mu_1[0, r_1] = \frac{1}{N}$.
2. If player v_2 thinks that this portion is too large, then he can 'trim' off a small slice (which is returned to the rest of the cake) so that he believes the new trimmed piece is exactly $\frac{1}{N}$ of the whole cake. If player v_2 thinks the portion is not too large (or possibly to small), then he leaves it alone.

   That is, v_2 chooses some $r_2 \in [0,1]$ so that $\mu_2[0, r_2] = \frac{1}{N}$. If $r_2 < r_1$ he 'trims' the cake; otherwise he leaves it alone.
3. If player v_3 thinks that this new portion is too large, then she can 'trim' off a small slice (which is returned to the rest of the cake) so that she believes the new trimmed piece is exactly $\frac{1}{N}$ of the whole cake. If player v_3 thinks the portion is not too large (or possibly to small), then she leaves it alone.

   That is, v_3 chooses some $r_3 \in [0,1]$ so that $\mu_3[0, r_3] = \frac{1}{N}$. If $r_3 < \min\{r_1, r_2\}$, then v_3 'trims' the cake; otherwise she leaves it alone.
4. The portion passes by each successive player in turn. Each has the option of trimming off a further slice, if he thinks the portion is too large.
5. The last person to trim the cake (ie. the player v_n with the smallest value of r_n) is the person who gets this portion. That is, $\mathbf{P}_n = [1, r_n)$.
6. The rest of the players then play the game with the remaining cake.

Notice how, once we describe the player's maximin strategies, it is clear that the Last Diminisher Game yields the Banach-Knaster Procedure. □

6.2.3 The Latecomer Problem: Fink's 'Lone Chooser' game

Prerequisites: §6.1, §6.2.1

What if three people have just finished dividing a cake into three fair portions, and suddenly a fourth person shows up and wants a fair share? In 1964 A.M. Fink [12] devised a proportional

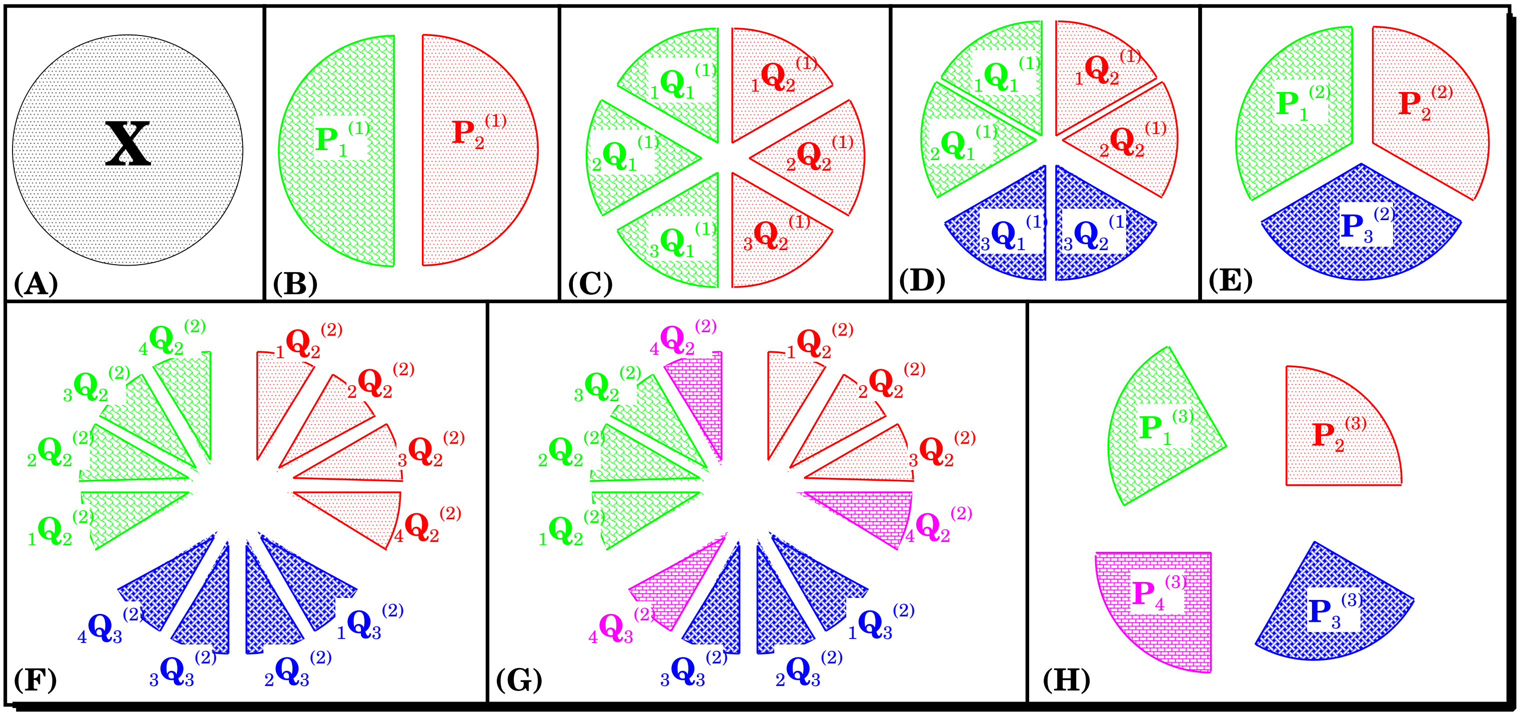

Figure 6.1: The Fink 'Lone Chooser' game.

partition game which can easily accommodate 'late comers'. We will describe the game rules in normal font, and the maximin strategy in *italics*; the rules and the maximin strategies together yield the partition procedure.

Game 6.9: 'Lone Chooser' (A.M. Fink)

Suppose $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$. Refer to Figure 6.1 for each step of the algorithm:

(A) We begin with a cake $\mathbf{X}$. It is convenient (but not essential) to imagine $\mathbf{X}$ as a disk.

(B) Players v_1 and v_2 use 'I cut, you choose' to split the cake into $\mathbf{P}_1^{(1)}$ and $\mathbf{P}_2^{(1)}$.
Maximin strategy yields $\mu_1[\mathbf{P}_1^{(1)}] = \frac{1}{2}$ *and* $\mu_2[\mathbf{P}_2^{(1)}] \geq \frac{1}{2}$.

(C) Player v_1 trisects her portion $\mathbf{P}_1^{(1)}$ into three equal parts ${}_1\mathbf{Q}_1^{(1)}$, ${}_2\mathbf{Q}_1^{(1)}$, and ${}_3\mathbf{Q}_1^{(1)}$.
Maximin strategy: $\mu_1[{}_1\mathbf{Q}_1^{(1)}] = \mu_1[{}_2\mathbf{Q}_1^{(1)}] = \mu_1[{}_3\mathbf{Q}_1^{(1)}] = \frac{1}{6}$.
Likewise v_2 trisects his portion $\mathbf{P}_2^{(1)}$ into three equal parts ${}_1\mathbf{Q}_2^{(1)}$, ${}_2\mathbf{Q}_2^{(1)}$, and ${}_3\mathbf{Q}_2^{(1)}$.
Maximin strategy: $\mu_2[{}_1\mathbf{Q}_2^{(1)}] = \mu_2[{}_2\mathbf{Q}_2^{(1)}] = \mu_2[{}_3\mathbf{Q}_2^{(1)}] = \frac{1}{3}\mu_2[\mathbf{P}_2^{(1)}] \geq \frac{1}{6}$.

(D) Player v_3 chooses one of ${}_1\mathbf{Q}_1^{(1)}$, ${}_2\mathbf{Q}_1^{(1)}$, or ${}_3\mathbf{Q}_1^{(1)}$, and one of ${}_1\mathbf{Q}_2^{(1)}$, ${}_2\mathbf{Q}_2^{(1)}$, or ${}_3\mathbf{Q}_2^{(1)}$.
Maximin strategy: Choose ${}_j\mathbf{Q}_1^{(1)}$, *and* ${}_k\mathbf{Q}_2^{(1)}$ *so as to maximize* $\mu_3[{}_j\mathbf{Q}_1^{(1)} \sqcup {}_k\mathbf{Q}_2^{(1)}]$.

(E) Assume without loss of generality that v_3 chooses ${}_3\mathbf{Q}_1^{(1)}$, and ${}_3\mathbf{Q}_2^{(1)}$. At this point,

- v_1 has $\mathbf{P}_1^{(2)} = {}_1\mathbf{Q}_1^{(1)} \sqcup {}_2\mathbf{Q}_1^{(1)}$.
- v_2 has $\mathbf{P}_2^{(2)} = {}_1\mathbf{Q}_2^{(1)} \sqcup {}_2\mathbf{Q}_2^{(1)}$.
- v_3 has $\mathbf{P}_3^{(2)} = {}_3\mathbf{Q}_1^{(1)} \sqcup {}_3\mathbf{Q}_2^{(1)}$.

Maximin outcome: $\mu_1[\mathbf{P}_1^{(2)}] = \frac{1}{3}$, $\mu_2[\mathbf{P}_2^{(2)}] \geq \frac{1}{3}$, *and* $\mu_3[\mathbf{P}_3^{(2)}] \geq \frac{1}{3}$. *Thus, each player has (in his estimation) at least one third of the cake.*

Now we introduce player v_4.

(F) For each $n \in [1..3]$, player v_n quadrisects his/her portion $\mathbf{P}_n^{(2)}$ into four equal parts ${}_1\mathbf{Q}_n^{(2)}, \ldots, {}_4\mathbf{Q}_n^{(2)}$.
Maximin strategy: $\mu_n[{}_1\mathbf{Q}_n^{(2)}] = \cdots = \mu_1[{}_4\mathbf{Q}_n^{(2)}] = \frac{1}{4}\mu_n[\mathbf{P}_n^{(2)}]$.

(G) Player v_4 then chooses one of ${}_1\mathbf{Q}_1^{(2)}, \ldots, {}_4\mathbf{Q}_1^{(2)}$, one of ${}_1\mathbf{Q}_2^{(2)}, \ldots, {}_4\mathbf{Q}_2^{(2)}$, and one of ${}_1\mathbf{Q}_3^{(2)}, \ldots, {}_4\mathbf{Q}_3^{(2)}$.
Maximin strategy: For each $n \in [1..3]$, *choose* ${}_{j_n}\mathbf{Q}_n^{(2)}$ *so as to maximize* $\mu_4[{}_{j_n}\mathbf{Q}_n^{(2)}]$.

(H) Assume without loss of generality that v_4 chooses ${}_4\mathbf{Q}_1^{(2)}$, ${}_4\mathbf{Q}_2^{(2)}$, and ${}_4\mathbf{Q}_3^{(2)}$. At this point,

- v_1 has $\mathbf{P}_1^{(3)} = {}_1\mathbf{Q}_1^{(2)} \sqcup {}_2\mathbf{Q}_1^{(2)} \sqcup {}_3\mathbf{Q}_1^{(2)}$.
- v_2 has $\mathbf{P}_2^{(3)} = {}_1\mathbf{Q}_2^{(2)} \sqcup {}_2\mathbf{Q}_2^{(2)} \sqcup {}_3\mathbf{Q}_2^{(2)}$.
- v_3 has $\mathbf{P}_3^{(3)} = {}_3\mathbf{Q}_3^{(2)} \sqcup {}_3\mathbf{Q}_3^{(2)} \sqcup {}_3\mathbf{Q}_3^{(2)}$.
- v_4 has $\mathbf{P}_4^{(3)} = {}_4\mathbf{Q}_1^{(2)} \sqcup {}_4\mathbf{Q}_2^{(2)} \sqcup {}_4\mathbf{Q}_3^{(2)}$.

Maximin outcome: $\mu_1[\mathbf{P}_1^{(3)}] = \frac{1}{4}$, $\mu_2[\mathbf{P}_2^{(2)}] \geq \frac{1}{4}$, $\mu_3[\mathbf{P}_3^{(2)}] \geq \frac{1}{4}$, *and* $\mu_4[\mathbf{P}_4^{(2)}] \geq \frac{1}{4}$. *Thus, each player has (in his estimation) at least one fourth of the cake.*

A fifth player is dealt with similarly. Proceed inductively. ______

<u>**Exercise 32**</u> Verify that the maximin strategies and payoffs for 'Lone Chooser' are as described in above.

<u>**Exercise 33**</u> One issue with partitioning games is the number of cuts.

1. Show that the N-player Banach-Knaster 'Last Diminisher' game requires *at most* $\frac{N(N-1)}{2}$ cuts.
2. Show that Fink's 'Lone Chooser' game for N players *always* requires $N!$ cuts.

Further reading: The problem of fairly dividing a cake amongst three people was first considered by Steinhaus in 1943, who developed a three-person game called 'Lone Divider' [40, 41, 23]. Steinhaus was unable to extend his solution to more than three players; this problem was solved by his students, Stefan Banach and Bronislaw Knaster, with their 'Last Diminisher' game [22, 40, 41]. Woodall [52] has modified Fink's scheme so that each of N players is guaranteed *strictly more* than $\frac{1}{N}$ of the cake (in his estimation). Austin [5] has modified Fink's scheme so that each player thinks he gets *exactly* $\frac{1}{N}$ of the cake. A thorough discussion of all these proportional partition procedures is given in Chapters 2 and 3 of Brams and Taylor [43].

6.2.4 Symmetry: Dubins and Spanier's 'Moving Knife' game

Prerequisites: §6.2.2

One objection to 'I cut, you choose' (Game 6.3) is that the 'chooser' player has a clear advantage. If 'cutter' plays her maximin strategy, she will be guaranteed *exactly* half the cake (in her perception). However, 'chooser' will be guaranteed *at least* half the cake (in his perception), and possibly much more, if his perception differs sufficiently from that of 'cutter'.

The same objection applies to the Banach-Knaster 'Last Diminisher' game (Game 6.7), because 'Last Diminisher' reduces to 'I cut, you choose' once the number of players is reduced to two. For example, if five people play 'Last Diminisher', then the person who ends up being 'chooser' will again be guaranteed *at least* one fifth of the cake, which gives him an advantage over not only 'cutter', but also the three 'diminishers' who have already left the game, each of whom received *exactly* one fifth (in their estimations).

To remove this asymmetry, Dubins and Spanier proposed a 'symmetric' form of the Banach-Knaster procedure, where no specific player gets to be 'chooser', or is forced into the role of 'cutter'. The Dubins-Spanier procedure looks very similar to Banach-Knaster, except that we do *not* resort to 'I cut, you choose' in the base case.

Procedure 6.10: Dubins & Spanier [26]

Again suppose $\mathbf{X} = [0,1]$ is the unit interval. Suppose $\mathcal{V} = \{v_1, \ldots, v_N\}$, and let v_n have nonatomic utility measure μ_n. The Dubins-Spanier procedure is defined recursively as follows:

1. Suppose $N \geq 2$. For each $n \in [1..N]$, let r_n be the largest value such that $\mu_n[1, r_n) = \frac{1}{N}$. In other words, $[1, r_n)$ is the largest piece of cake that v_n thinks is $\frac{1}{N}$ of the entire cake (such an r_n exists because μ_n is nonatomic).

Let v_n be the player with the smallest value of r_n (if two players are tied, then choose the smaller value of n). We define $\mathbf{P}_n = [0, r_n)$. Observe that $\mu_n(\mathbf{P}_n) = \frac{1}{N}$. (In other words, player v_n thinks he got $\frac{1}{N}$ of the cake.) Let $\mathbf{X}_1 = [r_n, 1]$ (ie. $\mathbf{X}_1$ is the remaining cake).

As in the Banach-Knaster Procedure (Procedure 6.6), we have:

Claim 1: *For every $m \neq n$,* $\mu_m[\mathbf{X}_1] \geq \frac{N-1}{N}$.

Thus, each of the remaining players thinks that at least $\frac{N-1}{N}$ of the cake remains to be divided.

2. Now let $\mathcal{V}_1 = \mathcal{V} \setminus \{v_n\}$ (the remaining players). We apply the Dubins-Spanier procedure recursively to divide $\mathbf{X}_1$ into $N-1$ slices such that each of the players in $v_m \in \mathcal{V}_1$ thinks he got a portion $\mathbf{P}_m$ such that

$$\mu_m[\mathbf{P}_m] \underset{\text{(S1)}}{\geq} \frac{1}{N-1} \cdot \mu_m[\mathbf{X}_1] \underset{\text{(C1)}}{\geq} \left(\frac{1}{N-1}\right) \cdot \left(\frac{N-1}{N}\right) = \frac{1}{N}.$$

where **(S1)** follows from step 1 of the procedure, and **(C1)** follows from Claim 1.

In what sense is the Dubins-Spanier procedure symmetric? Let $\sigma : \mathcal{V} \longrightarrow \mathcal{V}$ be a permutation. If $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N) \in {}^N$ is an N-tuple of utility measures, let $\sigma(\boldsymbol{\mu}) = (\mu_{\sigma(1)}, \ldots, \mu_{\sigma(N)})$.

(Heuristically speaking, the players are rearranged in a different order). We say that a partition procedure Π is **symmetric** if the following is true: Let $\boldsymbol{\mu} \in {}^N$ and let $\Pi(\boldsymbol{\mu}) = \mathcal{P}$. Let $\sigma : \mathcal{V} \longrightarrow \mathcal{V}$ be any permutation, and let $\Pi(\sigma(\boldsymbol{\mu})) = \mathcal{Q}$. Then for all $n \in [1..N]$, $\mu_n(\mathbf{P}_n) = \mu_{\sigma(n)}(\mathbf{Q}_{\sigma(n)})$. In other words, if we reorder the players and then apply the procedure, then each player receives a portion which he thinks is the same *size* (although it might not actually be the same *portion*) as he received before reordering.

Lemma 6.11 *The Dubins-Spanier procedure is symmetric*

Proof: Exercise 34 □

In this sense, the Dubins-Spanier procedure is more 'fair' than Banach-Knaster, because the 'last' player has no advantage over the 'first' player. Is there a partition game which yields the Dubins-Spanier procedure? Yes.

Game 6.12: 'Moving Knife' (Dubins & Spanier)

Suppose $\mathbf{X} = [0, 1]$ is the unit interval.

1. A 'referee' (who could be one of the players) takes a knife, and places it at the left end of the cake (ie. at 0).
2. The referee then very slowly moves the knife rightwards across the cake. (from 0 to 1).
3. At any moment, any player can say 'cut'. Say player v says 'cut', then she receives the piece to the left of the knife and exits the game (ie. if the knife is at $k \in [0, 1]$, then v gets the portion $[0, k]$.).
4. The rest of the players then continue playing with the remaining cake.

Notice that 'Moving Knife' is an *apportionment game*, like 'Last Diminisher'. Thus, to create a real *partitioning game*, we actually need to arrange a sequence of 'Moving Knife' games in an *apportionment cascade*. However, 'Moving Knife' is different from 'Last Diminisher' in one important respect. 'Last Diminisher' was a *sequential* game, where the players played one at a time. In contrast, 'Moving Knife' is a **simultaneous** game, where all players must choose and execute their strategies *simultaneously*, each in ignorance of the choices made by the others. In 'Moving Knife', each player's 'strategy' is his choice of the moment when he will say 'cut' (assuming no one else has said it first).

In a simultaneous game, *all* players are in the same strategic position that player v_1 had in a sequential game. We can thus define 'dominant' and 'maximin' strategies for all players of a simultaneous game in a fashion analogous to the dominant and maximin strategies for player v_1 in a sequential game (which simplifies analysis).

Recall: in an apportionment game Γ_N, the Γ_N-payoff of the 'winner' v_n of portion $\mathbf{P}$ is the value of $\mu_n(\mathbf{P})$, while the Γ_N-payoffs of all the *other* players are their maximin payoffs in Γ_{N-1}.

Proposition 6.13 *The Moving Knife game yields the Dubins-Spanier procedure.*

Proof: Let Γ_N be the 'Moving Knife' game with N players. It suffices to show:

Claim 1: *For any player v_n, let $r_n \in [0,1]$ be the largest value such that $\mu_n[1, r_n) \leq \frac{1}{N}$. Then v_n's maximin Γ_N-strategy is to say 'cut' exactly when the knife is at r_n (unless some other player says 'cut' first).*

His maximin Γ_N-payoff under this strategy is $\frac{1}{N}$.

Proof:

We proceed by induction on N. The logic is very similar to the proof of Proposition 6.8 (concerning the Banach-Knaster game). Hence, here we will be more sketchy.

Base Case: ($N = 2$) **Exercise 35** .

Induction: If v_n says 'cut' *before* the knife reaches r_n, then he will end up with a piece which is he *definitely* thinks is less than $\frac{1}{N}$.

If v_n waits until *after* the knife reaches r_n, then another player might say 'cut' first. Then this *other* player will then get a portion which v_n believes is *more* than $\frac{1}{N}$ of the cake. Then v_n enters game Γ_{N-1} with $(N-1)$ other players. By induction, his maximin payoff in Γ_{N-1} now $\frac{1}{N-1}$ of the remaining cake, which he believes is strictly *less* than $\frac{N-1}{N}$. Thus, his maximin Γ_{N-1}-payoff is *less* than $\left(\frac{1}{N-1}\right)\left(\frac{N-1}{N}\right) = \frac{1}{N}$. Hence, his minimum Γ_N-payoff is less than $\frac{1}{N}$.

If v_n says cut *exactly* at r_n, then he gets a piece of size $\frac{1}{N}$, so his minimum payoff is exactly $\frac{1}{N}$. This strategy yields a *higher* minimum payoff then either of the other two strategies, so this is v_n's *maximin* strategy in Γ_N. .... □ [Claim 1]

It follows from Claim 1 that we can expect the player with the smallest value of r_n to be 'honest', and say 'cut' exactly when the knife is at r_n. This player receives the portion $[0, r_n)$, and we continue playing. Thus, step **1** of the Dubins-Spanier procedure is implemented.

The rest of the players now play the game Γ_{N-1}; this implements step **2**. □

If $N = 2$, then the Dubins-Spanier game is called **Austin's single moving knife procedure**, and works as follows: the referee moves the knife across the cake until either of the two players says 'cut'. The player who calls 'cut' gets the left-hand portion, and the other player gets the right-hand portion. By the **Base Case** of Claim 1 above, we see that this game yields a proportional partition.

6.2.5 Connectivity: Hill and Beck's Fair Border Procedure

Prerequisites: §6.1, §6.2.1; Basic Topology (see appendix)

The partitioning procedures of Banach-Knaster (§6.2.2) and Fink (§6.2.3) made no reference to the *topology* of the portions. We *represented* the 'cake' $\mathbf{X}$ with the unit interval $[0,1]$, but the two procedures don't really depend on the linear topology of $[0,1]$, and are equally applicable to subsets of $\mathbb{R}^N$ (or even abstract measure spaces). This is 'topological insensitivity' is acceptable and even helpful, if the real 'cake' does not have any topological structure (eg. the 'cake' is a set of items to be divided in an inheritance). If the cake is a dessert, these procedures may leave each player with a plate full of chopped up bits, which is suboptimal but at least edible. However, if the 'cake' is a physical piece of land (eg. a contested property in an inheritance; a disputed territory in a military conflict), then disconnected portions will be unacceptable.

The Dubins-Spanier 'Moving Knife' procedure (§6.2.4) *does* use the topology of $[0, 1]$, and guarantees each player a connected subinterval as a portion. The Dubins-Spanier procedure can easily be adapted to higher dimensional spaces such as a two-dimensional territory (we can imagine a metaphorical knife sweeping slowly across a map of the territory until some player says 'stop'). However, the resulting portions may not be connected; if the territory has a complicated boundary (eg. a coastline), then the 'cuts' made by the knife may intersect this boundary in several places, leaving each player with several disconnected components.

There are economic, logistical, and strategic reasons for desiring *connected* territory. To ensure future peaceful coexistence, we need a procedure which yields a proportional partition with connected portions. In 1983, Theodore P. Hill [18] proved that such a partition existed, but he gave no construction procedure. In 1987, Anatole Beck [8] provided a partition procedure which yields a stronger version of Hill's theorem. The Hill-Beck solution involves some basic planar topology, which is reviewed in the *Appendix on Topology* at the end of this section (page 96).

Theorem 6.14 (Hill & Beck)

Let $\mathbf{X} \subset \mathbb{R}^2$ *be an open connected region in the plane* (the disputed territory). *Let* $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_N \subset \partial\mathbf{X}$ *be disjoint connected subsets of the boundary such that*

$$\partial\mathbf{X} \quad = \quad \mathbf{B}_1 \sqcup \mathbf{B}_2 \sqcup \cdots \sqcup \mathbf{B}_N \qquad \text{(Figure 6.2A)}$$

($\mathbf{B}_n$ is the 'border' between $\mathbf{X}$ and some adjacent country v_n which wants a piece of $\mathbf{X}$).

Let $\mu_1, \ldots, \mu_N$ *be nonatomic utility measures on* $\mathbf{X}$. *Then there exists a proportional partition* $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N\}$ *of* $\mathbf{X}$ *such that (as shown in Figure 6.2B), for each* $n \in [1..N]$

- $\mathbf{P}_n$ *is connected*
- $\mathbf{B}_n \subset \mathbf{P}_n$ (ie. $\mathbf{P}_n$ is connected to the country v_n).
- $\mu_n[\mathbf{P}_n] \ \geq \ \dfrac{1}{N}$. □

Beck provides a partition game which yields the partition described in this theorem. The partition game is a cascade of apportionment games, each of which gives a portion of territory to one player, who then leaves the game. First we need some notation.

$$\begin{array}{rrcll}
 & \text{Let} \quad \mathbf{D} & := & \left\{\mathbf{x} \in \mathbb{R}^2 \ ; \ |\mathbf{x}| \leq 1\right\} & \text{be the unit disk. (Fig.6.3A)} \\
 & \text{Let} \quad \mathbf{C} & := & \left\{\mathbf{x} \in \mathbb{R}^2 \ ; \ |\mathbf{x}| = 1\right\} & \text{be the unit circle. (Fig.6.3A)} \\
\text{For each } r \in [0,1], \text{ let} & \mathbf{D}(r) & := & \{\mathbf{x} \in \mathbf{D} \ ; \ |\mathbf{x}| \leq r\} & \text{be the disk of radius } r. \text{ (Fig.6.3B)}
\end{array}$$

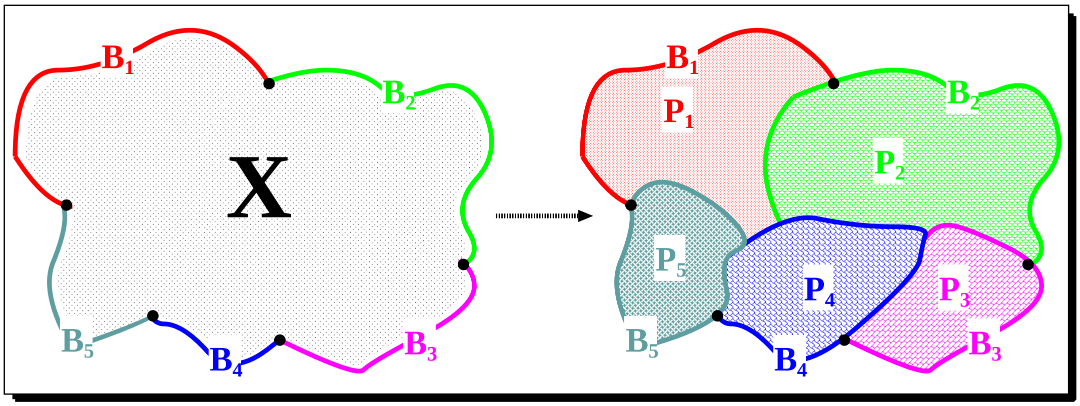

Figure 6.2: The Hill-Beck Theorem

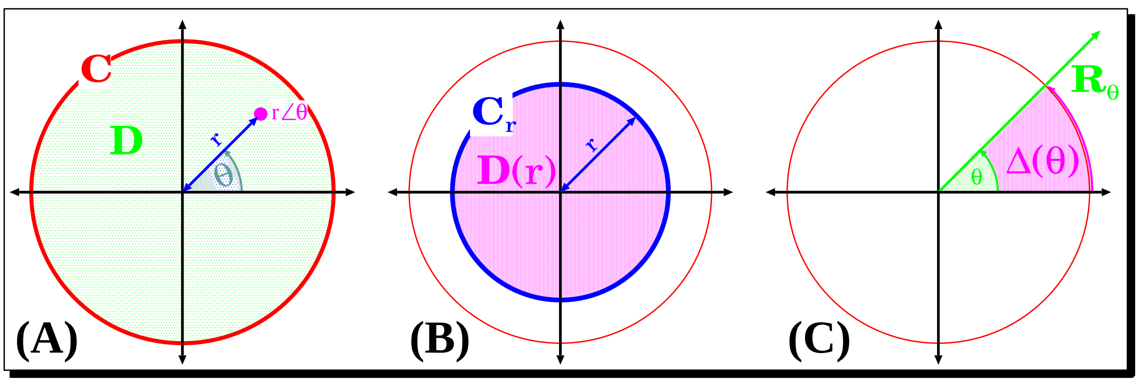

Figure 6.3: The unit disk **D** and subsets.

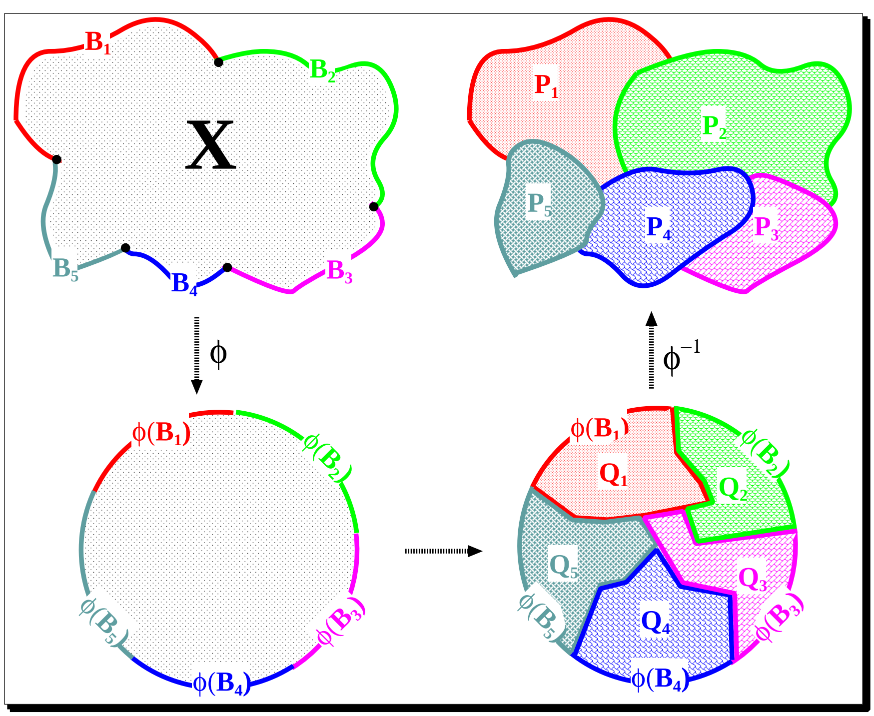

Figure 6.4: Representing the partition problem on the unit disk.

and let $\mathbf{C}_r := \{\mathbf{x} \in \mathbf{D} \,;\, |\mathbf{x}| = r\}$ be the circle of radius r. (Fig.6.3B)

If $r \geq 0$ and $\theta \in [0, 2\pi]$, let $r\angle\theta := (r\cos(\theta), r\sin(\theta))$ be the point with polar coordinates r and θ. (Fig.6.3A)

For each $\theta \in [0, 2\pi]$, let $\mathbf{R}_\theta := \{r\angle\theta \,;\, r \in [0, 1]\}$ be the ray at angle θ. (Fig.6.3C)

For any $\Theta \in [0, 2\pi]$, let $\Delta(\Theta) := \{r\angle\theta \,;\, r \in [0, 1],\ \theta \in [0, \Theta)\}$ be the wedge between angles 0 and Θ. (Fig.6.3C)

Beck's apportionment game assumes that the territory $\mathbf{X}$ is the unit disk $\mathbf{D}$, so that $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_N$ are subsets of the unit circle $\mathbf{C}$ (because $\partial\mathbf{D} = \mathbf{C}$). This assumption causes no loss of generality, because Beck shows that we can accurately represent any territorial dispute using an analogous territorial dispute on the unit disk, as shown in Figure 6.4, and described by the next lemma:

Lemma 6.15 *Let $\mathbf{X} \subset \mathbb{R}^2$ be an open, simply connected region, and let $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_N \subset \partial\mathbf{X}$ be disjoint connected sets such that $\partial\mathbf{X} = \mathbf{B}_1 \sqcup \mathbf{B}_2 \sqcup \cdots \sqcup \mathbf{B}_N$. Let $\mu_1, \ldots, \mu_N$ be nonatomic utility measures on $\mathbf{X}$.*

There exists a homeomorphism $\phi : \mathbf{X} \longrightarrow \mathbf{D}$ such that, if $\nu_1 = \phi(\mu_1), \ldots, \nu_N = \phi(\mu_N)$, then for all $n \in [1..N]$,

$\nu_n(\mathbf{R}_\theta) = 0$, *for all* $\theta \in [0, 2\pi]$, *and* $\nu_n(\mathbf{C}_1) = 0$, *for all* $r \in [0, 1]$. □

Now, if we can achieve a partition $\mathcal{Q} = \{\mathbf{Q}_1, \ldots, \mathbf{Q}_n\}$ of $\mathbf{D}$ which is connected and proportional with respect to the measures $\nu_1, \ldots, \nu_N$, and we define $\mathbf{P}_n = \phi^{-1}(\mathbf{Q}_n)$ for all $n \in [1..N]$ (see Fig.6.4), then Proposition 6.18 of the *Appendix* says that $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_n\}$ is a partition of $\mathbf{X}$ which is connected and proportional with respect to the measures $\mu_1, \ldots, \mu_N$. Thus, it suffices to find a proportional, connected partition of the disk $\mathbf{D}$.

To do this, we use a procedure similar to the Dubins-Spanier Moving Knife, only now the 'knife blade' is a *circle* of varying radius.

1. Each player submits a 'bid' for the largest radius r such that the disk $\mathbf{D}_r$ has mass $\frac{1}{N}$ in that player's estimation. In other words, for all $n \in [1..N]$, we define r_n by

$$r_n \quad = \quad \max\left\{r \in [0, 1] \,;\, \mu_n\left[\mathbf{D}(r)\right] = \frac{1}{N}\right\}.$$

   Claim 1: *Such an r_n exists.*

   Proof: Define $f_n : [0, 1] \longrightarrow [0, 1]$ by $f_n(r) = \mu_n[\mathbf{D}(r)]$. Then f_n is a *continuous* function of r, because Lemma 6.15 says $\mu_n[\mathbf{C}_r] = 0$ for all r. Observe that $f_n(0) = 0$ and $f_n(1) = 1$. Thus, the Intermediate Value Theorem yields some r with $f_n(r) = \frac{1}{N}$. □ `[Claim 1]`

2. Consider the player(s) whose value(s) of r_n are minimal. There are two cases:

Case 1: There is a *unique* player whose value of r_n is minimal (highly probable)

Case 2: There are several players whose values of r_n are equal and all minimal (highly improbable, but still theoretically possible).

We deal with these cases separately.

Case 1: Assume without loss of generality that r_1 is minimal (if necessary, permute the players to achieve this). Also assume without loss of generality that $\mathbf{B}_1$ is the arc of the circle $\mathbf{C}$ between angles 0 and Θ^*:

$$\mathbf{B}_1 \quad = \quad \{1\angle\theta \;;\; \theta \in [0, \Theta^*)\}.$$

(if necessary, rotate the disk to achieve this). Thus, to connect the disk $\mathbf{D}(r_1)$ to his territory, player v_1 must define some sort of 'corridor' from $\mathbf{D}(r_1)$ to $\mathbf{B}_1$. This corridor will take the form of a 'wedge'. For each $n \in [2..N]$, let v_n propose some $\Theta_n \in [0, \Theta^*]$ so that

$$\mu_n \Big[\mathbf{D}(r_1) \cup \Delta(\Theta_n)\Big] \quad = \quad \frac{1}{N}. \tag{6.2}$$

Claim 2: *Such a Θ_n exists.*

Proof: Define $f_n : [0, 2\pi] \longrightarrow [0, 1]$ by $f_n(\theta) = \mu_n [\mathbf{D}(r) \cup \Delta(\theta)]$. Then f_n is *continuous* as a function of θ, because Lemma 6.15 says $\mu_n[\mathbf{R}_\theta] = 0$ for all θ. Now, $f_n(0) = \mu_n [\mathbf{D}(r)] < \frac{1}{N}$ (by hypothesis), while $f_n(2\pi) = 1$. Thus, the Intermediate Value Theorem yields some θ such that $f_n(\theta) = \frac{1}{N}$. .... ☐ `[Claim 2]`

Now let $\Theta = \min_{n\in[2..N]} \Theta_n$. Define $\mathbf{P}_1 \;=\; \mathbf{D}(r_1) \cup \Delta(\Theta)$, as shown in Figure 6.5.

Case 2: This case is resolved through a sequence of 'subsidiary auctions'. We'll only sketch the idea here.

Assume without loss of generality that there is some $M \leq N$ so that players $v_1, \ldots, v_M$ all tied for the minimum bid in the first auction. Then we hold a second auction, where each of $v_1, \ldots, v_N$ submits a 'bid' for the *smallest* radius r such that the disk $\mathbf{D}(r)$ has mass $\frac{1}{N}$ in that player's estimation. In other words, for all $m \in [1..M]$, we define $r'_m <$ by

$$r'_m \quad = \quad \min\left\{r \in [0, 1] \;;\; \mu_m [\mathbf{D}(r)] = \frac{1}{N}\right\}.$$

Now, if there is a unique minimal bid in this auction (say, r'_1), then we move onto the 'wedge' procedure from **Case 1**, and give player v_1 the disk $\mathbf{D}(r'_1)$ plus a wedge-shaped corridor connecting $\mathbf{D}(r'_1)$ to his territory. If there is also a tie in the second auction, then Beck introduces a complex procedure to resolve the tie; the details are in [8].

3. Assume without loss of generality that player v_1 won the auction(s), and was awarded portion $\mathbf{P}_1$. Player v_1 exits the game. Let $\mathbf{X}_1 = \mathbf{D} \setminus \mathbf{P}_1$, as shown in Figure 6.5. Then $\mathbf{X}_1$ is open and simply connected, and $\mu_n[\mathbf{X}_1] \geq \frac{N-1}{N}$ for each $n \in [2..N]$. The remaining players repeat the game to divide $\mathbf{X}_1$.

The Beck procedure provides an elegant constructive proof of Hill's 'existence' theorem. However, it is not clear that Beck's procedure could resolve a real-world territorial dispute between two or more adversarial parties, for two reasons:

- The Hill-Beck theorem specifically requires that the utility measures $\mu_1, \ldots, \mu_N$ be *non-atomic* (this is necessary for the proof of Lemma 6.15). But in real-world territorial conflicts, nontrivial value is often concentrated at a single indivisible entity (eg. a city, an oil well, a gold mine).

- Beck's partition *procedure* does not lend itself to a partition *game* because there is no incentive for players to bid honestly. It's true that the players will bid honestly for the 'minimal radius disk' $\mathbf{D}(r)$ in **Step 1** (for the same reason that we expect honest bidding in Banach-Knaster or in Dubins-Spanier). However, consider the construction of the 'wedge' corridor $\Delta(\Theta)$ in **Case 1** of **Step 2**. In this stage, players $v_2, \ldots, v_n$ have no incentive to provide honest bids satisfying equation (6.2); instead, they all have an incentive to make the wedge *as small as possible*. It's certainly true that, as long the wedge $\Delta(\Theta)$ has nonzero width, v_1 will think that he has obtained 'more' than his fair share, because $\mu_1[\mathbf{P}_1] > \frac{1}{N}$. However, rival countries can 'shave' this wedge to be so thin that it is useless for practical purposes. If $\Delta(\Theta)$ cuts across rugged terrain, it can so thin that it becomes impassable. If $\Delta(\Theta)$ is flanked by enemies, it can be made so thin that it is indefensible.

Appendix on Topology

Let $\mathbf{X} \subset \mathbb{R}^2$ and $y \in \mathbb{R}^2$, we say that y is **adjacent** to $\mathbf{X}$ if there is a sequence of points $\{x_n\}_{n=1}^{\infty} \subset \mathbf{X}$ such that $\lim_{n\to\infty} x_n = y$ (Figure 6.6A) . If $\mathbf{P}, \mathbf{Q} \subset \mathbb{R}^2$ are disjoint subsets, we say that $\mathbf{P}$ and $\mathbf{Q}$ are **adjacent** if either $\mathbf{P}$ contains a point adjacent to $\mathbf{Q}$ or $\mathbf{Q}$ contains a point adjacent to $\mathbf{P}$ (Figure 6.6B). A set $\mathbf{X} \subset \mathbb{R}^2$ is **connected** if, whenever we partition $\mathbf{X}$ into two disjoint subsets $\mathbf{P}$ and $\mathbf{Q}$ so that $\mathbf{X} = \mathbf{P} \sqcup \mathbf{Q}$, we find that $\mathbf{P}$ and $\mathbf{Q}$ are adjacent (Figure 6.6C and 6.6D).

If $\mathbf{X} \subset \mathbb{R}^2$, then the **complement** of $\mathbf{X}$ is the set $\mathbf{X}^{\complement} := \mathbb{R}^2 \setminus \mathbf{X}$ (Figure 6.6E). The **boundary** of $\mathbf{X}$ is the set $\partial\mathbf{X}$ of all points which are both adjacent to $\mathbf{X}$ and adjacent to $\mathbf{X}^{\complement}$ (Figure 6.6F). For example, if $\mathbf{D} := \{\mathbf{x} \in \mathbb{R}^2 \,;\, |\mathbf{x}| \leq 1\}$ is the **unit disk** (Figure 6.6G) , then $\mathbf{D}^{\complement} = \{\mathbf{x} \in \mathbb{R}^2 \,;\, |\mathbf{x}| > 1\}$. If $\mathbf{C} := \{\mathbf{x} \in \mathbb{R}^2 \,;\, |\mathbf{x}| = 1\}$ is the **unit circle** (Figure 6.6H), then $\partial\mathbf{D} = \mathbf{C}$.

A subset $\mathbf{X} \subset \mathbb{R}^2$ is **open** if no point in $\mathbf{X}$ is adjacent to $\mathbf{X}^{\complement}$. Equivalently, $\mathbf{X}$ is open if $\partial\mathbf{X} \subset \mathbf{X}^{\complement}$. For example, the disk $\mathbf{D}$ is *not* open, because $\partial\mathbf{D} = \mathbf{C} \subset \mathbf{D}$. However, the set $\mathbf{O} = \{\mathbf{x} \in \mathbb{R}^2 \,;\, |\mathbf{x}| < 1\}$ *is* open, because $\partial\mathbf{O} = \mathbf{C} \subset \mathbf{O}^{\complement}$.

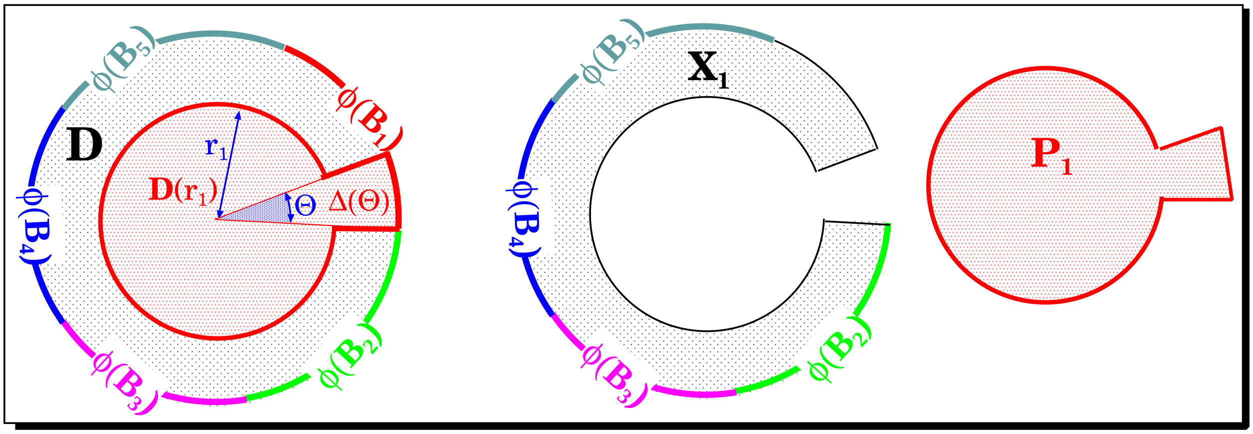

Figure 6.5: The Beck procedure.

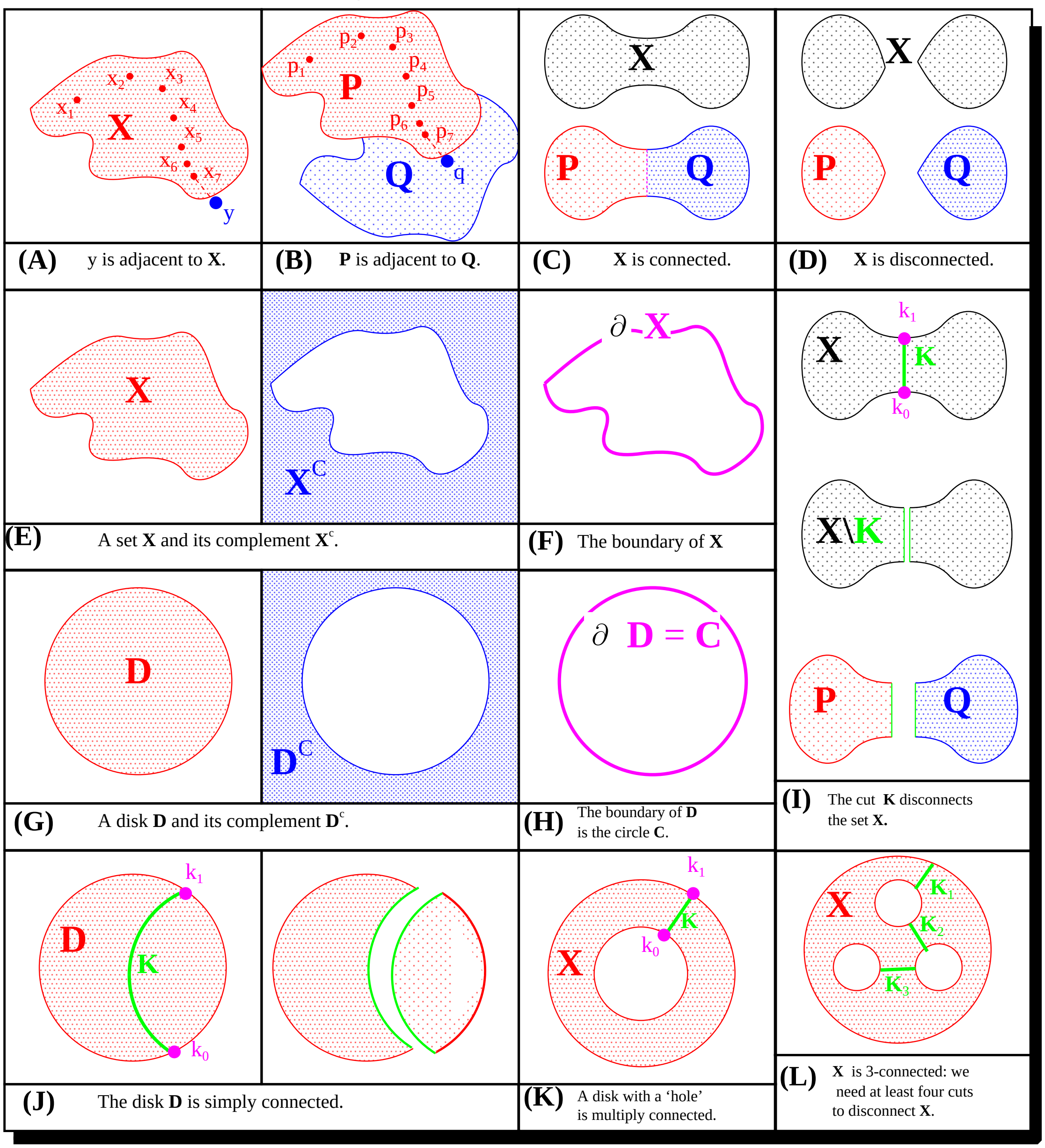

Figure 6.6: Some concepts in planar topology

Suppose $\mathbf{X} \subset \mathbb{R}^2$ is a connected domain. A **cut** is a curve $\mathbf{K} \subset \mathbf{X}$ which goes from one boundary point $k_0 \in \partial\mathbf{X}$ to another boundary point $k_1 \in \partial\mathbf{X}$, as in Figure 6.6I. We say that **$\mathbf{K}$ disconnects $\mathbf{X}$** if $\mathbf{X} \setminus \mathbf{K}$ is disconnected, as in Figure 6.6I. In other words, 'cutting' $\mathbf{X}$ along the curve $\mathbf{K}$ splits $\mathbf{X}$ into two pieces. We say that $\mathbf{X}$ is **simply connected** if *any* cut $\mathbf{K}$ disconnects $\mathbf{X}$. For example, the unit disk $\mathbf{D}$ is simply connected (Figure 6.6J).

However, suppose $\mathbf{X}$ has a 'hole', as in Figure 6.6(K). If k_0 is a point on the 'exterior' boundary of $\mathbf{X}$, and k_1 is a point on the 'hole' boundary, then a cut from k_0 to k_1 will *not* disconnect $\mathbf{X}$. Thus, $\mathbf{X}$ is *not* simply connected. We say $\mathbf{X}$ is **multiply connected**, meaning that there is some cut $\mathbf{K}$ which does *not* disconnect $\mathbf{X}$ (ie. $\mathbf{X} \setminus \mathbf{K}$ is still connected). More generally, we say that $\mathbf{X}$ is **N-connected** if there are N cuts $\mathbf{K}_1, \ldots, \mathbf{K}_N \subset \mathbf{X}$ such that the set $\mathbf{X} \setminus (\mathbf{K}_1 \cup \cdots \cup \mathbf{K}_N)$ is simply connected, as in Figure 6.6(L). (hence, at this point, one more cut will disconnect $\mathbf{X}$). Loosely speaking, $\mathbf{X}$ is N-connected if $\mathbf{X}$ has N distinct 'holes' in its interior.

If $\mathbf{X}, \mathbf{Y} \subset \mathbb{R}^2$, then a **homeomorphism** is a function $\phi : \mathbf{X} \longrightarrow \mathbf{Y}$ such that:

- ϕ is bijective [and thus, has a well-defined inverse function $\phi^{-1} : \mathbf{Y} \longrightarrow \mathbf{X}$].
- ϕ is *continuous* [ie. if $\{x_1, x_2, \ldots\} \subset \mathbf{X}$, and $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} \phi(x_n) = \phi(x)$].
- ϕ^{-1} is also continuous.

Heuristically, ϕ provides a method to 'represent' $\mathbf{X}$ using the region $\mathbf{Y}$; all topological phenomena (eg. adjacency, connectivity) on $\mathbf{X}$ are transformed by ϕ into analogous phenomena on $\mathbf{Y}$, as follows:

Lemma 6.16 *Suppose $\phi : \mathbf{X} \longrightarrow \mathbf{Y}$ is a homeomorphism.*

(a) *If $\mathbf{P}, \mathbf{Q} \subset \mathbf{X}$, then* $\Big(\mathbf{P}$ *is adjacent to* $\mathbf{Q} \Big) \iff \Big(\phi(\mathbf{P})$ *is adjacent to* $\phi(\mathbf{Q}) \Big)$.

(b) $\Big(\mathbf{X}$ *is connected* $\Big) \iff \Big(\mathbf{Y}$ *is connected* $\Big)$.

(c) $\Big(\mathbf{X}$ *is simply connected* $\Big) \iff \Big(\mathbf{Y}$ *is simply connected* $\Big)$.

(d) $\Big(\mathbf{X}$ *is N-connected* $\Big) \iff \Big(\mathbf{Y}$ *is N-connected* $\Big)$.

(e) *If $\mathbf{P} \subset \mathbf{X}$, then* $\Big(\mathbf{P}$ *is connected* $\Big) \iff \Big(\phi(\mathbf{P})$ *is connected* $\Big)$.

Proof: Exercise 36 □

If μ is a utility measure on $\mathbf{X}$, and $\phi : \mathbf{X} \longrightarrow \mathbf{Y}$ is a homeomorphism, then we define a new utility measure $\nu := \phi(\mu)$ on $\mathbf{Y}$ by the equation:

$$\nu[\mathbf{Q}] = \mu\left[\phi^{-1}(\mathbf{Q})\right], \qquad \text{for any } \mathbf{Q} \subset \mathbf{Y}.$$

[Recall: if $\mathbf{Q} \subset \mathbf{Y}$, then $\phi^{-1}(\mathbf{Q}) = \{\mathbf{x} \in \mathbf{X} \, ; \, \phi(\mathbf{x}) \in \mathbf{Q}\}$, hence $\phi^{-1}(\mathbf{Q}) \subset \mathbf{X}$.]

Lemma 6.17 *Let $\phi : \mathbf{X} \longrightarrow \mathbf{Y}$ be a homeomorphism, and let μ be a utility measure on $\mathbf{X}$. Let $\nu = \phi(\mu)$. Then:*

(a) *ν is also a utility measure.*

(b) $\Big($ *μ is nonatomic* $\Big) \iff \Big($ *ν is nonatomic* $\Big)$.

Proof: Exercise 37 □

Homeomorphisms can transform 'good' partitions of $\mathbf{X}$ into 'good' partitions of $\mathbf{Y}$ as follows:

Proposition 6.18 *Let $\mathbf{P}_1, \ldots, \mathbf{P}_N \subset \mathbf{X}$ be some subsets of $\mathbf{X}$. Let $\phi : \mathbf{X} \longrightarrow \mathbf{Y}$ be a homeomorphism, and for all $n \in [1..N]$, define $\mathbf{Q}_n := \phi(\mathbf{P}_n)$. Let $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ and let $\mathcal{Q} = \{\mathbf{Q}_1, \ldots, \mathbf{Q}_N\}$. Then:*

(a) $\Big($ *$\mathcal{P}$ is a partition of $\mathbf{X}$* $\Big) \iff \Big($ *$\mathcal{Q}$ is a partition of $\mathbf{Y}$* $\Big)$.

(b) *For all $n \in [1..N]$,* $\Big($ *$\mathbf{P}_n$ is connected* $\Big) \iff \Big($ *$\mathbf{Q}_n$ is connected* $\Big)$.

(c) *Let $\mu_1, \ldots, \mu_N$ be utility measures on $\mathbf{X}$. Let $\nu_1 = \phi(\mu_1), \ldots, \nu_N = \phi(\mu_N)$. Then*

$$\Big(\mathcal{P} \text{ is a proportional partition of } \mathbf{X}, \text{ relative to } \mu_1, \ldots, \mu_N \Big)$$

$$\iff \Big(\mathcal{Q} \text{ is a proportional partition of } \mathbf{Y}, \text{ relative to } \nu_1, \ldots, \nu_N \Big).$$

Proof: Exercise 38 □

6.3 Envy-freedom

Prerequisites: §6.1 **Recommended:** §6.2.2

If all the players just want to get their 'fair share', then a proportional partition is all we seek, and the Banach-Knaster (§6.2.2) or Dubins-Spanier (§6.2.4) procedure will do the job. However, sometimes the players are jealous or hostile of one another, and each may demand not only that she gets her fair share, but that no other person *more* than his fair share (as *she* sees it).

For example, suppose the three kingdoms *Wei*, *Wu*, and *Shu* are are quibbling over a disputed territory. Each state wants to get at least one third of the territory, but also each wants to make sure that no other state gets *more* territory, because then the other state would have a military advantage in future engagements. For example, even if the partition $\mathcal{P}$ gives Wei 40% of the territory, Wei would find $\mathcal{P}$ unacceptable if (in Wei's perception), $\mathcal{P}$ gives Shu

50% and gives Wu only 10%. The reason is not that Wei cares for the plight of Wu, but rather, that Wei fears the territorial advantage of Shu.

A partition $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ is **envy-free** if, for all n and m, $\mu_n[\mathbf{P}_n] \geq \mu_n[\mathbf{P}_m]$. In other words, each participant believes that he received *at least* as much as any other single participant did. A partition procedure or is **envy free** it always yields an envy-free partition.

Example 6.19: 'I cut, you choose' is envy-free

If Veronique and Waldemar use 'I cut, you choose' (Procedure 6.2), then the outcome will be envy free. To see this, recall that Veronique divides the cake into two portions $\mathbf{P}$ and $\mathbf{Q}$ such that $\mu_v[\mathbf{P}] = \frac{1}{2} = \mu_v[\mathbf{Q}]$. Then Waldemar chooses the portion he thinks is larger —say $\mathbf{Q}$. Thus, Waldemar doesn't envy Veronique, because $\mu_w[\mathbf{Q}] \geq \mu_w[\mathbf{P}]$. Conversely, Veronique doesn't envy Waldemar because $\mu_v[\mathbf{P}] = \mu_v[\mathbf{Q}]$.

Example 6.20: Banach-Knaster is *not* envy free

Suppose $\mathbf{X} = [0, 1]$, and Veronique, Waldemar, and Xavier are dividing $\mathbf{X}$ using 'Last Diminisher' (Game 6.7). We'll give the three players utility measures so that the outcome of 'Last Diminisher' *cannot* be envy free. Suppose the players perceive the cake as in Figure 6.7. Thus:

- Veronique values all parts of $\mathbf{X}$ equally. In other words, for any interval $[a, b] \subset [0, 1]$, $\mu_v[a, b] = (b - a)$.
- Waldemar and Xavier think the left-hand third of the cake is worthless. Xavier also thinks the middle third is worthless. Both Waldemar and Xavier think that the right end of the cake is the most valuable.

Thus, Veronique will choose $r_0 = \frac{1}{3}$, because $\mu_v\left[0, \frac{1}{3}\right] = \frac{1}{3}$. The other two players think this portion is worthless, so they are happy to give it to her untouched. Hence, Veronique receives portion $\mathbf{P}_v = \left[0, \frac{1}{3}\right]$, and exits the game feeling that she has a fair share.

Next, Waldemar will choose $r_1 = \frac{5}{6}$, because he believes $\mu_w\left[\frac{1}{3}, \frac{5}{6}\right] = \frac{1}{2}$ —ie. one half the remaining value of the cake (minus Veronique's 'worthless' portion). Since Xavier also believes $\mu_x\left[\frac{1}{3}, \frac{5}{6}\right] = \frac{1}{2}$, Xavier is happy to let Waldemar have this portion. So Waldemar exits with $\mathbf{P}_w = \left[\frac{1}{3}, \frac{5}{6}\right]$, and Xavier is left with $\mathbf{P}_x = \left[\frac{5}{6}, 1\right]$.

All three players believe they got a fair portion Indeed, Waldemar and Xavier both think they got more than a fair portion, receiving a payoff of $\frac{1}{2}$ each. However, Veronique believes that Waldemar's portion is bigger than her own, because $\mu_v[\mathbf{P}_v] = \frac{1}{3}$, but $\mu_v[\mathbf{P}_w] = \frac{1}{2}$. Hence she envies Waldemar.

'Envy freedom' is only a problem when there are *three or more* people, with *different* utility measures:

Proposition 6.21 *Let $\mathcal{P}$ be a partition.*

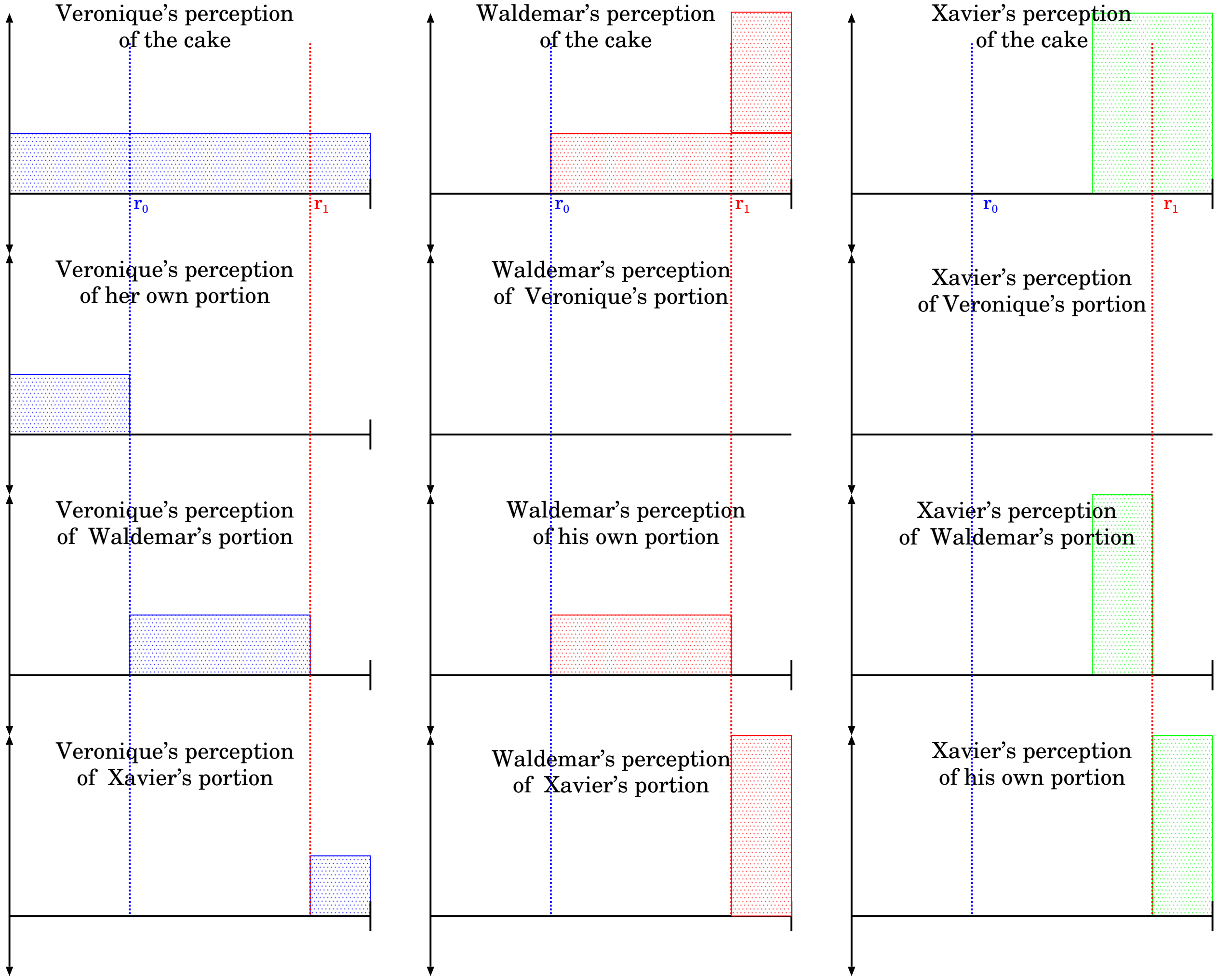

Figure 6.7: The Banach-Knaster procedure is not envy free (Example 6.20).

(a) $\Big($ *$\mathcal{P}$ is envy-free* $\Big) \Longrightarrow \Big($ *$\mathcal{P}$ is proportional* $\Big)$.

(b) *Suppose there are only two people (ie. $\mathcal{V} = \{v, w\}$).*

$$\Big(\mathcal{P} \text{ is envy-free} \Big) \iff \Big(\mathcal{P} \text{ is proportional} \Big).$$

(c) *Suppose all people have the same utility measures (ie. $\mu_1 = \mu_2 = \cdots = \mu_N$). Then*

$$\Big(\mathcal{P} \text{ is envy-free} \Big) \iff \Big(\mathcal{P} \text{ is proportional} \Big).$$

Proof: <u>**Exercise 39**</u> $\square$

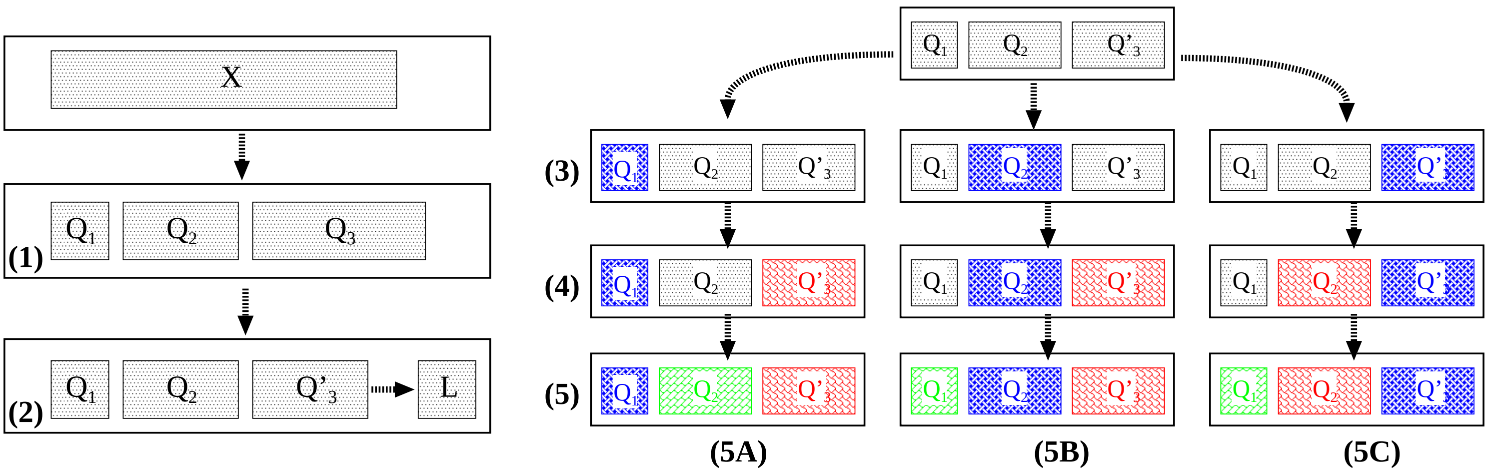

Figure 6.8: The Selfridge-Conway Trimming Procedure (Procedure 6.22)

.

Envy-free cake division becomes a nontrivial problem when there are three or more players. A number of envy-free three-person partition games have been devised; we will discuss one of the earliest and simplest, which was discovered independently by John L. Selfridge and John Horton Conway in the 1960s (but not published by either).

Procedure 6.22: Selfridge-Conway 'Trimming' procedure

Suppose that $\mathcal{V} = \{u, v, w\}$ (say, Ursula, Veronique, and Waldemar). Let these players have utility measures μ_u, μ_v, and μ_w. Let $\mathbf{X}$ be the 'cake'. We refer to Figure 6.22.

(1) Let $\mathcal{Q} = \{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3\}$ be a partition of $\mathbf{X}$ into three portions which Ursula deems of equal size; ie. $\mu_u[\mathbf{Q}_1] = \mu_u[\mathbf{Q}_3] = \mu_u[\mathbf{Q}_3] = \frac{1}{3}$.

(2) Assume without loss of generality (by reordering the portions if necessary) that Veronique ranks these portions in ascending order: $\mu_v[\mathbf{Q}_1] \leq \mu_v[\mathbf{Q}_2] \leq \mu_v[\mathbf{Q}_3]$.

Let $\mathbf{Q}'_3 \subseteq \mathbf{Q}_3$ be a subportion, such that Veronique thinks that portions $\mathbf{Q}_2$ and $\mathbf{Q}'_3$ are 'tied for largest'; ie. $\mu_v[\mathbf{Q}_1] \leq \mu_v[\mathbf{Q}_2] = \mu_v[\mathbf{Q}'_3]$.

Let $\mathbf{L} := \mathbf{Q}_3 \setminus \mathbf{Q}'_3$ (the 'leftover' piece).

(3) Give Waldemar whichever of $\{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}'_3\}$ he thinks is largest. Call this piece $\mathbf{P}'_3$. Observe that, no matter which piece Waldemar takes, at least one of the pieces $\{\mathbf{Q}_2, \mathbf{Q}'_3\}$ must remain.

(4) If only *one* of the two pieces $\{\mathbf{Q}_2, \mathbf{Q}'_3\}$ remains, then give it to Veronique.

If *both* $\mathbf{Q}_2$ and $\mathbf{Q}'_3$ remain, then give Veronique $\mathbf{Q}'_3$ (the one she trimmed).

(Veronique thinks both $\mathbf{Q}_2$ and $\mathbf{Q}'_3$ are equally large, and both are at least as big as $\mathbf{Q}_1$, so she will be happy either way).

(5) Give Ursula the remaining piece, which must be either $\mathbf{Q}_1$ or $\mathbf{Q}_2$, because:

(5A) ...if Waldemar took $\mathbf{Q}_1$, then Veronique took $\mathbf{Q}'_3$, so $\mathbf{Q}_2$ remains.
(5B) ...if Waldemar took $\mathbf{Q}_2$, then Veronique took $\mathbf{Q}'_3$, so $\mathbf{Q}_1$ remains.
(5C) ...if Waldemar took $\mathbf{Q}'_3$, then Veronique took $\mathbf{Q}_2$, so $\mathbf{Q}_1$ remains.

Thus, Ursula always gets an 'untrimmed' piece, which she thinks has size exactly $\frac{1}{3}$.

It remains to dispose of the leftover $\mathbf{L}$. Suppose that the person who got the trimmed piece $\mathbf{Q}'_3$ has surname Short (ie. either Veronique Short or Waldemar Short); and that the person (other than Ursula) who got an *untrimmed* piece has surname Taylor (ie. either Waldemar Taylor or Veronique Taylor). Observe that both Short and Taylor think they got the largest piece of $\{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}'_3\}$ (or at least, one of the two largest pieces), while Ursula thinks she got exactly $\frac{1}{3}$ of the original cake.

Also, observe that, in partitioning $\mathbf{L}$, Ursula has an **irrevocable advantage** over Short; even if Short gets *all* of $\mathbf{L}$, Ursula will still think that Short got no more than $\frac{1}{3}$ of the cake, because $\mu_u[\mathbf{Q}'_3 \sqcup \mathbf{L}] \;=\; \mu_u[\mathbf{Q}_3] = \frac{1}{3}$.

(6) Taylor divides $\mathbf{L}$ into three pieces $\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3$, with $\mu_t[\mathbf{L}_1] \;=\; \mu_t[\mathbf{L}_2] \;=\; \mu_t[\mathbf{L}_3] \;=\; \frac{1}{3}\mu[\mathbf{L}]$.

(7) Short chooses whichever of the three pieces he/she thinks is largest.

(8) Ursula chooses whichever of the two remaining pieces she thinks is largest.

(9) Taylor gets the remaining piece.

At this point:

- Short thinks he/she got the largest of $\{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}'_3\}$ and also the largest of $\{\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3\}$. Hence Short envies no one.
- Ursula thinks both she and Taylor got exactly $\frac{1}{3}$ of the original cake. She also thinks she got a choice of $\{\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3\}$ which is at *least* as large as Taylor's. Hence she does not envy Taylor. Also, Ursula will not envy Short because of her 'irrevocable advantage'.
- Taylor thinks he/she got the largest of $\{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}'_3\}$ and also exactly $\frac{1}{3}$ of $\mathbf{L}$. Hence Taylor envies no one.

Thus, each player envies no one, so the result is an envy-free partition. ______

A number of other envy-free cake division games exist for three players:

- Stromquist's procedure with four moving knives [49].
- Levmore and Cooke's procedure with two orthogonal moving knives [36].
- Webb's extension [51] to three players of Austin's two-person moving knife procedure [5].
- The Brams-Taylor-Zwicker 'rotating pie plate' procedure [46].

A summary of these and other procedures can be found in in [45], or in Chapter 6 of [43].

The envy-free partition problem for more than three players remained unsolved for a long time. The various three-person procedures are based on clever and elegant ideas, but it was not clear how any of them could be generalized. Recently Brams, Taylor, and Zwicker [46] found a four-person envy-free 'moving knife' procedure. Finally, Brams and Taylor [42] invented a general envy-free procedure for N players, which exploits the concept of 'irrevocable advantage' developed in the Selfridge-Conway procedure. The Brams-Taylor procedure is too complex to describe here; we refer the reader to [42] or Chapter 7.4 of [43].

6.4 Pareto Optimality

6.4.1 Introduction

Prerequisites: §6.1 **Recommended:** §6.2.1

Let $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ and $\mathcal{Q} = \{\mathbf{Q}_1, \ldots, \mathbf{Q}_N\}$ be two partitions. We say that $\mathcal{P}$ is **Pareto-preferred** to $\mathcal{Q}$ if:

- For all $n \in [1..N]$, $\mu_n[\mathbf{Q}_n] \leq \mu[\mathbf{P}_n]$.

  (ie. every player gets *at least* as much in $\mathcal{P}$ as he does in $\mathcal{Q}$).

- For some $n \in [1..N]$, $\mu_n[\mathbf{Q}_n] < \mu[\mathbf{P}_n]$

  (ie. at least one player feels that he got strictly *more* in $\mathcal{P}$.)

The partition $\mathcal{P}$ is better (or at least, no worse) for every person. Clearly, given a choice, we should choose partition $\mathcal{P}$ over partition $\mathcal{Q}$.

We say that $\mathcal{P}$ is **Pareto-optimal**[5] if there does not exist any other partition $\mathcal{Q}$ which is Pareto-preferred to $\mathcal{P}$. A partition procedure is **Pareto-optimal** if it always yields a Pareto-optimal outcome. Clearly, this is is desirable. After all, if the procedure produced a partition that was *not* Pareto-optimal, then by definition, we could suggest *another* partition which was at *least* as good for everyone, and strictly *better* for someone.

Example 6.23: 'I cut, you choose' is *not* Pareto Optimal

Suppose the left half of the cake is vanilla, and the right half is walnut cream. Veronique only likes vanilla, and Waldemar only likes walnut cream. Clearly, a Pareto-optimal, proportional partition exists: cut the cake in half, and let Veronique take the vanilla portion and Waldemar take the walnut portion. Both players receive a payoff of 1 (they get everything they value).

Unfortunately, this is *not* the partition which will be generated by 'I cut, you choose' (Game 6.3). Not knowing Waldemar's preferences (or not trusting him to abide by them), Veronique must use her maximin strategy (see Example 6.4), which cuts the cake into two portions,

[5] Sometimes this is called *Pareto efficient*, or even just *efficient*.

each having *half* the vanilla. Waldemar will then choose the portion which has more walnut cream. Waldemar will end up with a payoff of at least $\frac{1}{2}$ (and possibly even a payoff of 1, if one of the portions happens to have all the walnut cream). However, Veronique has *ensured* that she only gets a payoff of $\frac{1}{2}$. This is not Pareto optimal, because, as we've seen, *both* players *could* have gotten a payoff of 1, with the right partition.

Similarly, the Banach-Knaster and Dubins-Spanier procedures are not Pareto-optimal. However, this doesn't mean that proportionality and Pareto-optimality are mutually exclusive.

Lemma 6.24 *Let $\mathcal{Q}$ be a proportional partition. If $\mathcal{P}$ is Pareto-preferred to $\mathcal{Q}$, then $\mathcal{P}$ is also proportional.*

Proof: Exercise 40 □

Thus, given any proportional partition $\mathcal{Q}$ (eg. the outcome of the Banach-Knaster procedure), Lemma 6.24 says we can always find a proportional, Pareto-optimal partition $\mathcal{P}$ which is Pareto-preferred to $\mathcal{Q}$. We can achieve this through 'trade', as we next discuss.

6.4.2 Mutually Beneficial Trade

Prerequisites: §6.4.1

Let $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2\}$ be a two-person partition. A **mutually beneficial trade** is a pair $\mathcal{T} = (\mathbf{T}_1, \mathbf{T}_2)$, where $\mathbf{T}_1 \subset \mathbf{P}_1$ and $\mathbf{T}_2 \subset \mathbf{P}_2$, such that

$$\mu_1[\mathbf{T}_2] \quad \geq \quad \mu_1[\mathbf{T}_1], \qquad \text{and} \qquad \mu_2[\mathbf{T}_1] \quad \geq \quad \mu_2[\mathbf{T}_2],$$

and at least one of these two inequalities is strict. Thus, if player v_1 gives $\mathbf{T}_1$ to v_2 in return for $\mathbf{T}_2$, then at least one of them (and possibly both) will benefit, and neither will suffer.

We say that the partition $\mathcal{Q} = \{\mathbf{Q}_1, \mathbf{Q}_2\}$ is **obtained** from $\mathcal{P}$ via the trade $\mathcal{T}$ if

$$\mathbf{Q}_1 \quad = \quad (\mathbf{P}_1 \setminus \mathbf{T}_1) \sqcup \mathbf{T}_2 \qquad \text{and} \qquad \mathbf{Q}_2 \quad = \quad (\mathbf{P}_2 \setminus \mathbf{T}_2) \sqcup \mathbf{T}_1. \qquad \text{(see Figure 6.9A)}$$

The interpretation is: v_1 gives $\mathbf{T}_1$ to v_2 in return for $\mathbf{T}_2$, while v_2 gives $\mathbf{T}_2$ to v_1 in return for $\mathbf{T}_1$. The **utility** of the trade $\mathcal{T}$ for v_1 is defined:

$$\mu_1(\mathcal{T}) \quad := \quad \mu_1[\mathbf{T}_2] - \mu_1[\mathbf{T}_1].$$

This is the amount v_1 gains by the trade. Likewise, the utility of the trade $\mathcal{T}$ for v_2 is defined:

$$\mu_2(\mathcal{T}) \quad := \quad \mu_2[\mathbf{T}_1] - \mu_2[\mathbf{T}_2].$$

The relation between trade and Pareto-optimality is the following:

Proposition 6.25 *Let $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2\}$ be a two-person partition.*

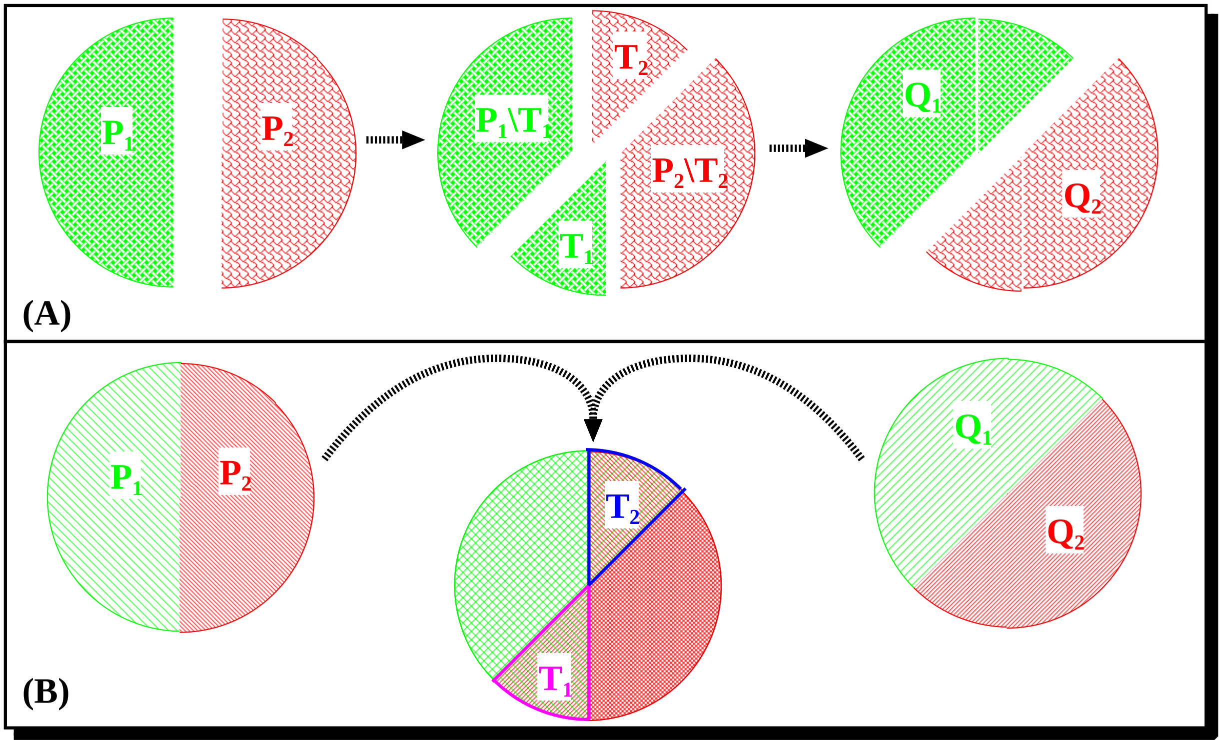

Figure 6.9: **(A)** Starting with partition $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2\}$, player v_1 gives $\mathbf{T}_1$ to v_2 in return for $\mathbf{T}_2$, to obtain partition $\mathcal{Q} = \{\mathbf{Q}_1, \mathbf{Q}_2\}$. **(B)** Given any partitions $\mathcal{P}$ and $\mathcal{Q}$, we can obtain $\mathcal{Q}$ from $\mathcal{P}$ through some kind of trade.

(a) *If $\mathcal{Q} = \{\mathbf{Q}_1, \mathbf{Q}_2\}$ is another partition which is Pareto-preferred to $\mathcal{P}$, then $\mathcal{Q}$ can be obtained from $\mathcal{P}$ by a mutually beneficial trade.*

(b) *Thus, $\mathcal{P}$ is Pareto-optimal iff no mutually beneficial trades can be made from $\mathcal{P}$.*

Proof: **(a)** Suppose $\mathcal{Q}$ is Pareto-preferred to $\mathcal{P}$. Let $\mathbf{T}_1 = \mathbf{Q}_2 \cap \mathbf{P}_1$ and let $\mathbf{T}_2 = \mathbf{Q}_1 \cap \mathbf{P}_2$ (see Figure 6.9B). . It is left as **Exercise 41** to check that:

- $\mathcal{T}$ is a mutually beneficial trade.
- $\mathcal{Q}$ is obtained from $\mathcal{P}$ via the trade $\mathcal{T} = (\mathbf{T}_1, \mathbf{T}_2)$.

(b) follows immediately from **(a)**. $\square$

Given any proportional partition $\mathcal{Q}$, we can always find a proportional, Pareto-optimal partition $\mathcal{P}$ which is Pareto-preferred to $\mathcal{Q}$ through some kind of trade[6]. If the players are able to communicate and trade, then they can achieve a Pareto-optimal partition by 'trading' bits

[6] Economists will recognize this as a version of *Coase's Theorem*, which states that an economy with 'well-defined property rights', 'perfect information', and 'costless transactions' will always converge to a Pareto-optimal state if the participants trade rationally. In the context of fair division theory, the existence of a specific

of their $\mathcal{Q}$-portions with each other. Each player will trade bits he considers 'low value' for bits he considers high value (ie. in Example 6.23, Veronique would give walnut cream to Waldemar, in return for vanilla). Trading can only *increase* the utilities of all the traders, so the post-trade partition is still proportional, and is also Pareto-preferred to the pre-trade partition. We let this process continue until we've reached a partition $\mathcal{P}$ where further no further trades are possible (eg. Veronique runs out of walnut cream, or Waldemar runs out of vanilla). At this point, we have reached a Pareto optimal partition.

The concept of 'mutually beneficial trade' and its relation to Pareto-optimality can be generalized to three or more players. For instance, Barbanel [6] has studied the possibility of *cyclic trades* amongst N players (eg. v_1 gives something to v_2, who gives something to v_3, who gives something to v_4, who gives something to v_1; all four end up better off).

Unfortunately, the 'trading' procedure may destroy the *envy-freedom* of a three-player partition (see §6.3). If there are only two players, then any proportional partition is automatically envy-free (Theorem 6.21), so the 'trading procedure' preserves envy-freedom. However, if we have three players, and two of them trade to their mutual advantage, then the nontrader may end up envying one or both of the two traders.

6.4.3 Utility Ratio Threshold Partitions

Prerequisites: §6.4.2; Elementary integral calculus[7].

To get a nontrivial example of a Pareto-optimal partition for two players, suppose $\mathbf{X} = [0,1]$, and suppose there are **utility functions** $U_1, U_2 : [0,1] \longrightarrow [0,\infty)$ which define the player's utility measures as follows: for any subinterval $[a,b] \subset \mathbf{X}$

$$\mu_1[a,b] \quad = \quad \int_a^b U_1(x)\ dx, \qquad \text{and} \qquad \mu_2[a,b] \quad = \quad \int_a^b U_2(x)\ dx.$$

We assume that $\displaystyle\int_0^1 U_1(x)\ dx \quad = \quad 1 \quad = \quad \int_0^1 U_2(x)\ dx.$

The **utility ratio** is the function $R(x) = U_1(x)/U_2(x)$. Intuitively, if $R(x) > 1$, then player v_1 wants the point x more than player v_2 does, while if $R(x) < 1$, then v_2 wants x more than v_1 does. The **highest bidder** partition is the partition $\mathcal{P}^{(1)} = \{\mathbf{P}_1, \mathbf{P}_2\}$ defined

$$\mathbf{P}_1 \quad \subseteq \quad \{x \in \mathbf{X}\ ;\ R(x) \geq 1\} \quad \text{and} \quad \mathbf{P}_2 \quad \subseteq \quad \{x \in \mathbf{X}\ ;\ R(x) < 1\}.$$

Thus, each player gets those parts of the cake he wants more than the other player. This seems like it should be a fair partition, but it often is not, as the next example shows.

---

partition $\mathcal{P}$ constitutes 'well-defined property rights'. The fact that all players know their own utility measures (Axiom **(Ψ1)** on page 81) constitutes 'perfect information'. We have also tacitly assumed from the beginning that all 'transactions' are costless (ie. the cake is not damaged by repeatedly slicing and recombining pieces, the players never get tired or hungry, etc.)

[7] Antidifferentiation techniques are not required, but a basic understanding of the concept of integration is necessary.

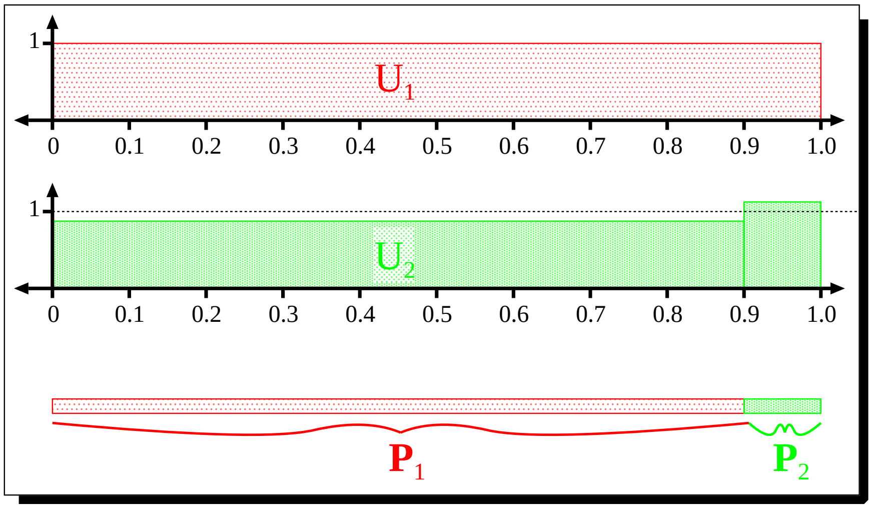

Figure 6.10: Example 6.26: The Highest Bidder Partition favours v_1 over v_2.

Example 6.26: Suppose $U_1(x) = 1$, for all $x \in \mathbf{X}$, while $U_2(x) = \begin{cases} \frac{9}{10} & \text{if} \quad 0 \leq x \leq \frac{9}{10}; \\ \frac{19}{10} & \text{if} \quad \frac{9}{10} < x \leq 1. \end{cases}$

(See Figure 6.10). It can be checked that $\int_0^1 U_1(x)\, dx = 1 = \int_0^1 U_2(x)\, dx$. However,

$$\begin{aligned} \mathbf{P}_1 &= \left[0, \frac{9}{10}\right] && \text{so that} \quad \mu_1[\mathbf{P}_1] = \int_0^{9/10} 1\, dx = 0.9 \\ \text{while} \quad \mathbf{P}_2 &= \left(\frac{9}{10}, 1\right] && \text{so that} \quad \mu_2[\mathbf{P}_2] = \int_{9/10}^1 \frac{19}{10}\, dx = \frac{1}{10} \cdot \frac{19}{10} = 0.19. \end{aligned}$$

Thus, the highest bidder partition is *not* proportional. Player v_1 does very well, but v_2 does very badly, because of the slight asymmetry in v_2's preferences.

The problem here is that we have partitioned the cake using a 'threshold' value of 1 for the utility ratio. This seems like a good idea, but in Example 6.26 it yields an 'unfair' partition. To fix this problem, we need to use a different threshold. Given any $\theta \in [0, \infty)$, a **θ-utility ratio threshold (θ-URT)** partition is a partition $\mathcal{P}^{(\theta)} = \{\mathbf{P}_1, \mathbf{P}_2\}$ such that

$$\mathbf{P}_1 \subseteq \{x \in \mathbf{X}\,;\, R(x) \geq \theta\} \quad \text{and} \quad \mathbf{P}_2 \subseteq \{x \in \mathbf{X}\,;\, R(x) \leq \theta\} \qquad \text{(see Fig.6.11)}$$

Thus, the *highest bidder* partition is a URT partition with $\theta = 1$.

Remark: Note that the allocation of the set $\Delta = \{x \in \mathbf{X}\,;\, R(x) = \theta\}$ is ambiguous here. If the set Δ has zero measure (which is likely), then it doesn't really matter how Δ is split between $\mathbf{P}_1$ and $\mathbf{P}_2$, and there is effectively a unique θ-URT partition. However, if Δ has

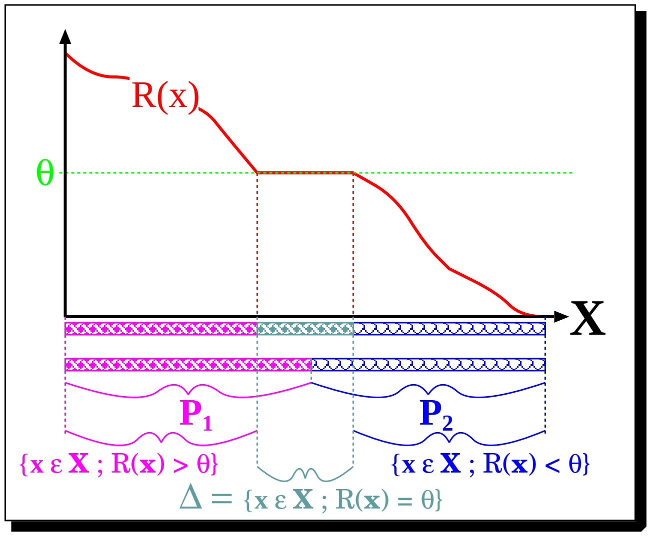

Figure 6.11: A utility ratio threshold partition.

nonzero measure, then there are many ways Δ can be divided, yielding many different θ-URT partitions.

Proposition 6.27 *For any value of θ, any URT partition $\mathcal{P}^{(\theta)}$ is Pareto-optimal.*

Proof: Suppose $\mathcal{P}^{(\theta)}$ were not Pareto-optimal, and suppose $\mathcal{Q}$ was Pareto-prefered to $\mathcal{P}^{(\theta)}$. Then Proposition 6.25(a) says we can obtain $\mathcal{Q}$ from $\mathcal{P}^{(\theta)}$ through some mutually advantageous trade $\mathcal{T} = (\mathbf{T}_1, \mathbf{T}_2)$, where $\mathbf{T}_k \subset \mathbf{P}_k$.

Let $\Delta = \{x \in \mathbf{X} \; ; \; R(x) = \theta\}$.

Claim 1: *We can assume without loss of generality that $\mathbf{T}_1 \cap \Delta = \emptyset$.*

Proof: Idea: *Trading bits of the set Δ benefits no one. Thus, $\mathcal{T}$ can be replaced with modified trade $\mathcal{T}' = (\mathbf{T}'_1, \mathbf{T}'_2)$, which has the same utility to both players, but where $\mathbf{T}'_1 \cap \Delta = \emptyset$.*

Claim 1.1: *If $\mathbf{S} \subset \Delta$ is any subset, then $\mu_1[\mathbf{S}] = \theta \cdot \mu_2[\mathbf{S}]$.*

Proof: $$\mu_1[\mathbf{S}] = \int_{\mathbf{S}} U_1(s)\, ds = \int_{\mathbf{S}} R(s) \cdot U_2(s)\, ds \underset{(*)}{=} \int_{\mathbf{S}} \theta \cdot U_2(s)\, ds = \theta \cdot \int_{\mathbf{S}} U_2(s)\, ds = \theta \cdot \mu_2[\mathbf{S}].$$

Here, $(*)$ is because $R(s) = \theta$ for all $s \in \mathbf{S} \subset \Delta$. ☐ `[Claim 1.1]`

Let $\mathbf{S}_1 := \mathbf{T}_1 \cap \Delta$, and let $s_1 := \mu_2[\mathbf{S}_1]$; then Claim 1.1 says that $\mu_1[\mathbf{S}_1] = \theta \cdot s_1$.

Let $\mathbf{S}_2 := \mathbf{T}_2 \cap \Delta$, and let $s_2 := \mu_2[\mathbf{S}_2]$; then Claim 1.1 says that $\mu_1[\mathbf{S}_2] = \theta \cdot s_2$.

Assume without loss of generality that $s_1 \leq s_2$ (otherwise switch the two players to make this the case). Let $\mathbf{S}'_2 \subseteq \mathbf{S}_2$ be a subset such that $\mu_2[\mathbf{S}'_2] = s_1$; such an $\mathbf{S}'_2$ exists because $s_1 \leq s_2$, and because μ_2 is nonatomic. Claim 1.1 says that $\mu_1[\mathbf{S}'_2] = \theta \cdot s_1$.

Let $\mathbf{T}'_1 = \mathbf{T}_1 \setminus \mathbf{S}_1$ and let $\mathbf{T}'_2 = \mathbf{T}_2 \setminus \mathbf{S}'_2$. Let $\mathcal{T}' = (\mathbf{T}'_1, \mathbf{T}'_2)$.

Claim 1.2: *$\mu_1(\mathcal{T}') = \mu_1(\mathcal{T})$ and $\mu_2(\mathcal{T}') = \mu_2(\mathcal{T})$.*

Proof: First, note that

$$\begin{aligned} \mu_1[\mathbf{T}'_1] &= \mu_1[\mathbf{T}_1] - \mu_1[\mathbf{S}_1] &=& \mu_1[\mathbf{T}_1] - \theta \cdot s_1, \\ \mu_2[\mathbf{T}'_1] &= \mu_2[\mathbf{T}_1] - \mu_2[\mathbf{S}_1] &=& \mu_2[\mathbf{T}_1] - s_1, \\ \mu_1[\mathbf{T}'_2] &= \mu_1[\mathbf{T}_2] - \mu_1[\mathbf{S}'_2] &=& \mu_1[\mathbf{T}_2] - \theta \cdot s_1, \\ \text{and} \quad \mu_2[\mathbf{T}'_2] &= \mu_2[\mathbf{T}_2] - \mu_2[\mathbf{S}'_2] &=& \mu_2[\mathbf{T}_2] - s_1. \end{aligned}$$

$$\begin{aligned} \text{Thus,} \quad \mu_1(\mathcal{T}') &= \mu_1[\mathbf{T}'_2] - \mu_1[\mathbf{T}'_1] &=& \Big(\mu_1[\mathbf{T}_2] - \theta s_1\Big) - \Big(\mu_1[\mathbf{T}_1] - \theta s_1\Big) \\ &= \mu_1[\mathbf{T}_2] - \mu_1[\mathbf{T}_1] &=& \mu_1(\mathcal{T}). \\ \text{Likewise,} \quad \mu_2(\mathcal{T}') &= \mu_2[\mathbf{T}'_1] - \mu_2[\mathbf{T}'_2] &=& \Big(\mu_2[\mathbf{T}_1] - s_1\Big) - \Big(\mu_2[\mathbf{T}_2] - s_1\Big) \\ &= \mu_2[\mathbf{T}_1] - \mu_2[\mathbf{T}_2] &=& \mu_2(\mathcal{T}). \end{aligned}$$

.... □ `[Claim 1.2]`

Thus, we can replace trade $\mathcal{T}$ with a modified trade $\mathcal{T}'$, which has exactly the same value for both players. In the modified trade $\mathcal{T}'$, notice that $\mathbf{T}_1 \cap \Delta = \emptyset$. ..... □ `[Claim 1]`

Claim 2: *Suppose $(\mathbf{T}_1, \mathbf{T}_2)$ is a trade, and $\mathbf{T}_1 \cap \Delta = \emptyset$. Then $(\mathbf{T}_1, \mathbf{T}_2)$ <u>cannot</u> be a mutually advantageous trade. That is: Either $\mu_1[\mathbf{T}_2] < \mu_1[\mathbf{T}_1]$, or $\mu_2[\mathbf{T}_1] < \mu_2[\mathbf{T}_2]$.*

Proof: Suppose

$$\mu_1[\mathbf{T}_1] \quad \leq \quad \mu_1[\mathbf{T}_2]; \tag{6.3}$$

I'll show that $\mu_2[\mathbf{T}_1] < \mu_2[\mathbf{T}_2]$. To see this, note that:

$$\mu_1[\mathbf{T}_2] \;=\; \int_{\mathbf{T}_2} U_1(t)\, dt \;\underset{(\dagger)}{=}\; \int_{\mathbf{T}_2} R(t) \cdot U_2(t)\, dt \;\underset{(*)}{\leq}\; \int_{\mathbf{T}_2} \theta \cdot U_2(t)\, dt \;=\; \theta\, \mu_2[\mathbf{T}_2], \tag{6.4}$$

here, $(\dagger)$ is because $R(t) = U_1(t)/U_2(t)$, and $(*)$ is because $R(t) \leq \theta$ for all $t \in \mathbf{T}_2$. Likewise,

$$\mu_2[\mathbf{T}_1] \;=\; \int_{\mathbf{T}_1} U_2(t)\, dt \;\underset{(\ddagger)}{=}\; \int_{\mathbf{T}_1} \frac{U_1(t)}{R(t)} \cdot dt \;\underset{(\star)}{<}\; \int_{\mathbf{T}_1} \frac{1}{\theta} \cdot U_1(t)\, dt \;=\; \frac{1}{\theta}\, \mu_1[\mathbf{T}_1], \tag{6.5}$$

where $(\ddagger)$ is because is because $R(t) = U_1(t)/U_2(t)$, and $(\star)$ is because $R(t) > \theta$ for all $t \in \mathbf{T}_1$ (because $\mathbf{T}_1 \cap \Delta = \emptyset$). Thus,

$$\mu_2[\mathbf{T}_1] \;\underset{(6.5)}{<}\; \frac{1}{\theta}\, \mu_1[\mathbf{T}_1] \;\underset{(6.3)}{\leq}\; \frac{1}{\theta}\, \mu_1[\mathbf{T}_2] \;\underset{(6.4)}{\leq}\; \frac{\theta}{\theta}\, \mu_2[\mathbf{T}_2] \;=\; \mu_2[\mathbf{T}_2],$$

where **(6.5)** is by eqn.(6.5); **(6.3)** is by hypothesis (6.3); and **(6.4)** is by eqn.(6.4). Hence, $\mu_2[\mathbf{T}_1] < \mu_2[\mathbf{T}_2]$, so the trade is *not* beneficial for player v_2.

Conversely, if we suppose $\mu_2[\mathbf{T}_1] \geq \mu_2[\mathbf{T}_2]$, then similar reasoning shows that $\mu_1[\mathbf{T}_2] < \mu_1[\mathbf{T}_1]$. .... □ `[Claim 2]`

Thus, a mutually beneficial trade is impossible, so Proposition 6.25(B) says that $\mathcal{P}^{(\theta)}$ must be Pareto-optimal. □

In §6.5 a suitable choice of θ will yield a partition which is both Pareto-optimal and envy-free.

Remark: Throughout this section, we assumed $\mathbf{X} = [0,1]$ only for concreteness and simplicity. Actually the definition of a URT partition and the proof of Theorem 6.27 will work if $\mathbf{X}$ is any 'reasonable' subset of $\mathbb{R}^N$, or indeed, if $\mathbf{X}$ is any measure space, and $U_1, U_2 \in \mathbf{L}^1(\mathbf{X})$.

Exercise 42 Generalize the definition of *utility ratio partition* and the proof of Theorem 6.27 to the case when $\mathbf{X} \subset \mathbb{R}^N$.

Exercise 43 Generalize the definition of *utility ratio partition* and the proof of Theorem 6.27 to the case when $\mathbf{X}$ is an abstract measure space.

Further reading: Much work on Pareto-optimality in fair division has been done by Ethan Akin [1] and Julius Barbanel [6, 7].

6.4.4 Bentham Optimality & 'Highest Bidder'

Prerequisites: §6.4.1 **Recommended:** §6.4.3

Let $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ be a partition. The **total utility** of $\mathcal{P}$ is defined:

$$U(\mathcal{P}) \quad = \quad \sum_{n=1}^{N} \mu_n[\mathbf{P}_n].$$

Let $\mathcal{Q} = \{\mathbf{Q}_1, \ldots, \mathbf{Q}_N\}$ be another partition. We say that $\mathcal{P}$ is **Bentham-prefered** to $\mathcal{Q}$ if $U(\mathcal{P}) \ \geq \ U(\mathcal{Q})$. We say that $\mathcal{P}$ is **Bentham-optimal** if there does not exist any other partition $\mathcal{Q}$ which is Bentham-prefered to $\mathcal{P}$. In other words, $\mathcal{P}$ has the maximum total utility of any partition.

Lemma 6.28 *Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions.*

(a) $\Big($ *$\mathcal{P}$ is Pareto-prefered to $\mathcal{Q}$* $\Big) \Longrightarrow \Big($ *$\mathcal{P}$ is Bentham-prefered to $\mathcal{Q}$* $\Big)$.

(b) $\Big($ *$\mathcal{P}$ is Bentham-optimal* $\Big) \Longrightarrow \Big($ *$\mathcal{P}$ is Pareto-optimal* $\Big)$.

(c) *The converses of* **(a)** *and* **(b)** *are false.*

Proof: **Exercise 44** $\Box$

Intuitively, we can achieve Bentham optimality by giving each player the parts of $\mathbf{X}$ which he values more than any other player. Thus, every bit of cake goes to the person who values it most, so the maximum amount of total utility has been 'extracted' from the cake.

For example, let $\mathbf{X} = [0,1]$. Suppose $\mathcal{V} = \{v_1, \ldots, v_N\}$, and suppose there are **utility functions** $U_1, U_2, \ldots, U_n : [0,1] \longrightarrow [0,\infty)$ which define the players' utility measures as follows: for any subinterval $[a,b] \subset \mathbf{X}$

$$\mu_n[a,b] \quad = \quad \int_a^b U_n(x)\, dx, \qquad \text{for all } n \in [1..N]. \tag{6.6}$$

We assume that $\int_{\mathbf{X}} U_n(x)\ dx = 1$, for all $n \in [1..N]$. The **Highest Bidder** partition $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_n\}$ is defined by

$$\mathbf{P}_n = \{x \in \mathbf{X}\ ;\ \text{for all } m \in [1..N], \text{ either } U_n(x) > U_m(x) \text{ or } U_n(x) = U_m(x) \text{ and } n < m\}. \tag{6.7}$$

Remarks: (a) Notice that we break 'bidding ties' in an arbitrary way by awarding the tied portion to the player with the lower number.

(b) Observe that, if $N = 2$, this agrees with the 'highest bidder' partition defined in §6.4.3.

(c) As in §6.4.3, there is nothing special about $\mathbf{X} = [0, 1]$; we could perform a similar construction if $\mathbf{X}$ was any 'reasonable' subset of $\mathbb{R}^N$, or indeed, if $\mathbf{X}$ was any measure space.

Proposition 6.29 *If the utility measures of $v_1, \ldots, v_n$ are defined by utility functions as in eqn.(6.6), then the Highest Bidder partition of eqn.(6.7) is Bentham-optimal.*

Proof: **Exercise 45** □

The problem with the Highest Bidder partition is that it may be far from proportional, as we saw in Example 6.26. Indeed, with three or more players, the Highest Bidder partition may award some players an *empty* portion (**Exercise 46**). To remedy this, we can introduce 'side payments', whereby the losing players are compensated by the winners using some commodity *beyond* the cake (eg. money), until all players feel they have received an equal proportion of the *value* (even if they haven't recieved an equal portion of cake).

6.5 Equitable Partitions & 'Adjusted Winner'

Prerequisites: §6.1, §6.2.1, §6.4.3. **Recommended:** §6.3

In §6.2.4, we discussed how the Dubins-Spanier procedure is 'fairer' than the Banach-Knaster procedure because Dubins-Spanier is *symmetric*, meaning that it doesn't favour the 'first' player over the 'second' player or vice versa. However, Dubins-Spanier may still inadvertentely discriminate against a certain player because of the structure of her utility measure (regardless of whether she is 'first', 'second', or whatever). We seek a procedure which treats all players equally, regardless of the nature of their preferences.

A partition $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ is **equitable** if $\mu_1[\mathbf{P}_1] = \mu_2[\mathbf{P}_2] = \cdots = \mu_N[\mathbf{P}_N]$. In other words, each player's assessment of his own portion is the same as every *other* player's assesment of *her own* portion; no one has somehow been 'favoured' by the partition. A partition procedure (or game) is **equitable** if it always produces an equitable partition.

Example 6.30: 'Moving Knife' is not equitable

Recall the Dubins-Spanier 'Moving Knife' Game (Game 6.12). Suppose v_1 is the first person to say 'cut'; then v_1 gets a portion $\mathbf{P}_1$ such that $\mu_1[\mathbf{P}_1] = \frac{1}{N}$.

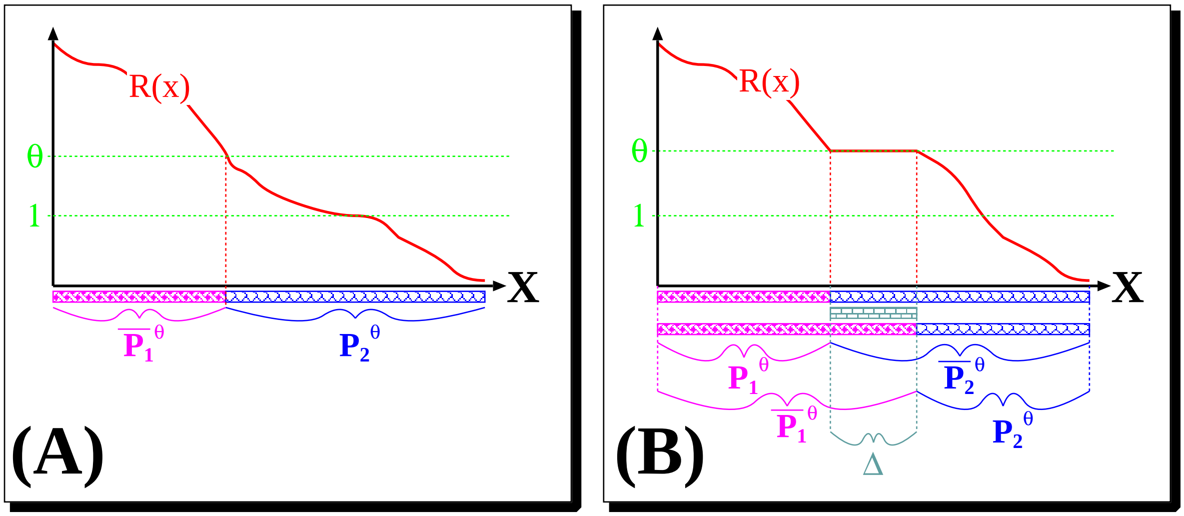

Figure 6.12: The Adjusted Winner Procedure

But if $\mathbf{X}_1 = \mathbf{X} \setminus \mathbf{P}_1$, then everyone *else* thinks $\mathbf{X}_1$ is worth *more* than $\frac{N-1}{N}$ of the total value of $\mathbf{X}$. Thus, if v_2 is the *second* person to say 'cut', then v_2 gets a portion $\mathbf{P}_2$ so that

$$\mu_2[\mathbf{P}_2] \quad = \quad \left(\frac{1}{N-1}\right) \mu_2[\mathbf{X}_1] \quad > \quad \left(\frac{1}{N-1}\right) \cdot \left(\frac{N-1}{N}\right) \quad = \quad \frac{1}{N} \quad = \quad \mu_1[\mathbf{P}_1].$$

Thus, v_2 is favoured over v_1.

Loosely speaking, the person who says 'cut' first (ie. v_1) is the 'least greedy' (at least, measuring things from the left end of the cake), whereas people who wait longer to say 'cut' are 'more greedy'. Thus, the 'Moving-Knife' procedure favours 'greedy' people.

Clearly, equitability implies symmetry, but Example 6.30 shows that symmetry does not imply equitability. Equitable partition procedures have two advantages:

1. No player is favoured or harmed by the intrinsic structure of her preferences (eg. 'greedy' players are not favoured over 'nongreedy' players).

2. Equitability combined with Pareto-optimality yields envy-freedom, by the next lemma.

Lemma 6.31 *Suppose $\mathcal{V} = \{v, w\}$. If $\mathcal{P}$ is an equitable, Pareto-optimal partition, then $\mathcal{P}$ is also envy-free (and thus, proportional).*

Proof: Exercise 47 □

The Brams-Taylor Adjusted Winner partition: Recall the definition of a *utility ratio threshold* (URT) partition $\mathcal{P}^{(\theta)}$ from §6.4.3. We saw that the 'Highest Bidder' partition $\mathcal{P}^{(1)}$ seemed like a good idea, but often produced highly unbalanced outcomes (Example 6.26). However, 'Highest Bidder' is a good starting point from which to build a fair partition.

Proposition 6.32 *There exists a $\theta_0 \in [0, \infty)$ and a θ_0-utility ratio threshold partition $\mathcal{Q}$ which is equitable and Pareto-optimal, and thus, envy-free.*

Proof: Let $R(x) = U_1(x)/U_2(x)$ for all $x \in \mathbf{X}$. For any $\theta \geq 1$, define the θ-URT partition $\mathcal{P}^{(\theta)} = \{\overline{\mathbf{P}}_1^{(\theta)}, \mathbf{P}_2^{(\theta)}\}$ by

$$\overline{\mathbf{P}}_1^{(\theta)} \quad = \quad \{x \in \mathbf{X}\ ;\ R(x) \geq \theta\} \quad \text{and} \quad \mathbf{P}_2^{(\theta)} \quad = \quad \{x \in \mathbf{X}\ ;\ R(x) < \theta\}. \qquad \text{(Fig.6.12A)}$$

Thus, if $\theta = 1$, then $\mathcal{P}^{(1)} = \{\overline{\mathbf{P}}_1^{(1)}, \mathbf{P}_2^{(1)}\}$ is the Highest Bidder partition (see §6.4.3).

Proposition 6.27 says $\mathcal{P}^{(1)}$ is Pareto-optimal. If $\mu_1[\overline{\mathbf{P}}_1^{(1)}] = \mu_2[\mathbf{P}_2^{(1)}]$ then $\mathcal{P}^{(1)}$ is also equitable, so set $\mathcal{Q} := \mathcal{P}^{(1)}$ and we're done.

Otherwise, assume without loss of generality that $\mu_1[\overline{\mathbf{P}}_1^{(1)}] > \mu_2[\mathbf{P}_2^{(1)}]$ like in Example 6.26 (if not, then switch the players). As we increase θ, observe that $\mu_1[\overline{\mathbf{P}}_1^{(\theta)}]$ decreases, while $\mu_2[\mathbf{P}_2^{(\theta)}]$ increases. In the limit as θ goes to ∞, we have $\overline{\mathbf{P}}_1^{(\infty)} = \emptyset$ and $\mathbf{P}_2^{(\infty)} = \mathbf{X}$ (ie. player v_1 gets nothing, and v_2 gets everything).

Let $\Theta \in [1, \infty)$ be the largest value of θ such that $\mu_1[\overline{\mathbf{P}}_1^{(\theta)}] \geq \mu_2[\mathbf{P}_2^{(\theta)}]$.

Case 1: If $\mu_1[\overline{\mathbf{P}}_1^{(\Theta)}] = \mu_2[\mathbf{P}_2^{(\Theta)}]$, then $\mathcal{P}^{(\Theta)}$ is equitable, and Proposition 6.27 already says $\mathcal{P}^{(\Theta)}$ is Pareto-optimal, so set $\mathcal{Q} := \mathcal{P}^{(\Theta)}$, and we're done.

Case 2: If $\mu_1[\overline{\mathbf{P}}_1^{(\Theta)}] > \mu_2[\mathbf{P}_2^{(\Theta)}]$, then define a new Θ-URT partition $\widetilde{\mathcal{P}}^{(\Theta)} = \{\mathbf{P}_1^{(\Theta)}, \overline{\mathbf{P}}_2^{(\Theta)}\}$ by

$$\mathbf{P}_1^{(\Theta)} \quad = \quad \{x \in \mathbf{X}\ ;\ R(x) > \Theta\} \quad \text{and} \quad \overline{\mathbf{P}}_2^{(\Theta)} \quad = \quad \{x \in \mathbf{X}\ ;\ R(x) \leq \Theta\}. \qquad \text{(Fig.6.12B)}$$

Note: $\mathbf{P}_1^{(\Theta)} \subset \overline{\mathbf{P}}_1^{(\Theta)}$ and $\overline{\mathbf{P}}_2^{(\Theta)} \supset \mathbf{P}_2^{(\Theta)}$. Thus, $\mu_1[\mathbf{P}_1^{(\Theta)}] \leq \mu_1[\overline{\mathbf{P}}_1^{(\Theta)}]$ and $\mu_2[\overline{\mathbf{P}}_2^{(\Theta)}] \geq \mu_2[\mathbf{P}_2^{(\Theta)}]$.

Claim 1: $\mu_1[\mathbf{P}_1^{(\Theta)}] \leq \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}]$.

Proof: <u>**Exercise 48**</u> Hint: First, note that, for any $\theta > \Theta$, $\overline{\mathbf{P}}_1^{(\theta)} \subset \mathbf{P}_1^{(\Theta)}$ and $\mathbf{P}_2^{(\theta)} \supset \overline{\mathbf{P}}_2^{(\Theta)}$. Thus, $\mu_1[\overline{\mathbf{P}}_1^{(\theta)}] \leq \mu_1[\mathbf{P}_1^{(\Theta)}]$ and $\mu_2[\mathbf{P}_2^{(\theta)}] \geq \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}]$.

Next, prove that $\lim_{\theta \searrow \Theta} \mu_1[\overline{\mathbf{P}}_1^{(\theta)}] = \mu_1[\mathbf{P}_1^{(\Theta)}]$ and that $\lim_{\theta \searrow \Theta} \mu_2[\mathbf{P}_2^{(\theta)}] = \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}]$.

Now prove the claim by using the fact that $\Theta \in [0, \infty)$ is the *largest* value of θ such that $\mu_1[\overline{\mathbf{P}}_1^{(\theta)}] \geq \mu_2[\mathbf{P}_2^{(\theta)}]$. .... □ [Claim 1]

Now there are two subcases

CASE 2.1: If $\mu_1[\mathbf{P}_1^{(\Theta)}] = \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}]$, then $\widetilde{\mathcal{P}}^{(\Theta)}$ is equitable, and Proposition 6.27 already says $\widetilde{\mathcal{P}}^{(\Theta)}$ is Pareto-optimal, so set $\mathcal{Q} := \widetilde{\mathcal{P}}^{(\Theta)}$, and we're done.

CASE 2.2: Suppose $\mu_1[\mathbf{P}_1^{(\Theta)}] < \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}]$. Then loosely speaking, $\mathbf{P}_1^{(\Theta)}$ is too small, while $\overline{\mathbf{P}}_1^{(\Theta)}$ is too big. Likewise, $\overline{\mathbf{P}}_2^{(\Theta)}$ is too *big*, while $\mathbf{P}_2^{(\Theta)}$ is too *small*. Let

$$\Delta \quad := \quad \{x \in \mathbf{X} \; ; \; R(x) = \Theta\} \quad = \quad \overline{\mathbf{P}}_1^{(\Theta)} \setminus \mathbf{P}_1^{(\Theta)} \quad = \quad \overline{\mathbf{P}}_2^{(\Theta)} \setminus \mathbf{P}_2^{(\Theta)} \qquad \text{(Fig.6.12B)}$$

Claim 2: *There exists a family of subsets* $\{\Delta_r \subseteq \Delta \; ; \; r \in [0,1]\}$ *such that:*

(a) $\Delta_0 = \emptyset$ *and* $\Delta_1 = \Delta$.

(b) *If* $s < r$, *then* $\Delta_s \subset \Delta_r$.

(c) *Define* $f_1(r) := \mu_1[\Delta_r]$ *and* $f_2(r) := \mu_2[\Delta_r]$ *for all* $r \in [0,1]$. *Then* $f_1 : [0,1] \longrightarrow \mathbb{R}$ *and* $f_2 : [0,1] \longrightarrow \mathbb{R}$ *are continuous nondecreasing functions.*

Proof: **Exercise 49** .... □ [Claim 2]

Now, for each $r \in [0,1]$, define $\mathbf{Q}_1^{(r)} := \mathbf{P}_1^{(\Theta)} \sqcup \Delta_r$ and $\mathbf{Q}_2^{(r)} := \overline{\mathbf{P}}_2^{(\Theta)} \setminus \Delta_r$. Let $F(r) := \mu_1[\mathbf{Q}_1^{(r)}] - \mu_2[\mathbf{Q}_2^{(r)}]$.

Claim 3: $F(0) < 0 < F(1)$.

Proof: Observe that $\mathbf{Q}_1^{(0)} = \mathbf{P}_1^{(\Theta)}$ and $\mathbf{Q}_2^{(0)} = \overline{\mathbf{P}}_2^{(\Theta)}$. Hence $F(0) = \mu_1[\mathbf{P}_1^{(\Theta)}] - \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}] < 0$ by hypothesis.

Likewise, $\mathbf{Q}_1^{(1)} = \overline{\mathbf{P}}_1^{(\Theta)}$ and $\mathbf{Q}_2^{(1)} = \mathbf{P}_2^{(\Theta)}$. Hence $F(1) = \mu_1[\overline{\mathbf{P}}_1^{(\Theta)}] - \mu_2[\mathbf{P}_2^{(\Theta)}] > 0$ by hypothesis. .... □ [Claim 3]

Claim 4: $F : [0,1] \longrightarrow \mathbb{R}$ *is continuous*

Proof: Observe that $\mu_1[\mathbf{Q}_1^{(r)}] = \mu_1[\mathbf{P}_1^{(\Theta)}] + f_1(r)$ and $\mu_2[\mathbf{Q}_2^{(r)}] = \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}] - f_2(r)$.

$$\begin{aligned}\text{Thus,} \quad F(r) \quad &= \quad \mu_1[\mathbf{Q}_1^{(r)}] - \mu_2[\mathbf{Q}_1^{(r)}] \quad = \quad \Big(\mu_1[\mathbf{P}_1^{(\Theta)}] \; + \; f_1(r)\Big) - \Big(\mu_2[\overline{\mathbf{P}}_2^{(\Theta)}] \; - \; f_2(r)\Big) \\ &= \quad \Big(\mu_1[\mathbf{P}_1^{(\Theta)}] - \mu_2[\overline{\mathbf{P}}_2^{(\Theta)}]\Big) \; + \; f_1(r) + f_2(r),\end{aligned}$$

and f_1 and f_2 are continuous by Claim 2(c). .... □ [Claim 4]

Thus, Claims 3 and 4 and the Intermediate Value Theorem together imply that there is some $r \in [0,1]$ such that $F(0) = 0$, which means that $\mu_1[\mathbf{Q}_1^{(r)}] = \mu_2[\mathbf{Q}_2^{(r)}]$ Thus, $\mathcal{Q} := \{\mathbf{Q}_1^{(r)}, \; \mathbf{Q}_2^{(r)}\}$ is equitable, and Proposition 6.27 already says $\mathcal{Q}$ is Pareto-optimal, so we're done. ______□

The Brams-Taylor **Adjusted Winner** partition (AWP) is the URT $\mathcal{Q}$ of Proposition 6.32. There is no practical procedure to *exactly* compute the AWP in real situations, because doing so would require complete information about the functions U_1 and U_2, which is potentially an infinite amount of information. In practical applications, we assume that we can divide $\mathbf{X}$ into some fine partition $\mathcal{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\}$ so that U_1 and U_2 are *constant* on each set $\mathbf{R}_m$. We can then ask the players to express their preferences by 'bidding' on each set $\mathbf{R}_m$. We do this as follows:

1. Each player is given some finite collection of 'points' (say 1000).

2. The players can then 'spend' his points to place 'bids' on each of the sets $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M$, with the understanding that the highest bidder will be (initially) awarded each set. ('Fractional' bids are allowed).

3. Each player thus has an incentive to bid more points on those subsets he values most, and not to squander points on subsets he values less. (In other words, his minimax strategy is to bid 'honestly'.) Thus, we expect that the distribution of v_n's bidding points will be a good approximation of his utility function U_n.

4. We then compute the Highest Bidder partition $\mathcal{P}^{(1)}$. If $\mathcal{P}^{(1)}$ equitable, then we're done. If $\mathcal{P}^{(1)}$ is not equitable, we slowly slide the threshold θ up or down (as appropriate), until we reach a partition $\mathcal{P}^{(\theta)}$ which is equitable. Then we stop.

The outcome is that each player gets those parts of the cake he desires 'most'. The balance point between the desires of player v_1 and the desires of v_2 is chosen to make the partition equitable.

Brams and Taylor propose this procedure as a way to divide a collection of goods (eg. in a divorce settlement). We can imagine that each of the subsets $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M$ represents some physical item; the 'bidding' procedure is how the players express their preferences for different items. Brams and Taylor point out one advantage of the AWP: with possibly a single exception, every one of the subsets $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M$ will either fall entirely into $\mathbf{P}_1$ or entirely into $\mathbf{P}_2$. Thus, if the items are difficult or impossible to 'share', then we are reduced to the problem of 'sharing' at most *one* item. (If the players cannot come to an agreement on how to share this item, then the item can be sold and the profits split).

Brams and Taylor caution that the Adjusted Winner procedure is manipulable (see §6.6.4); one player can seek to obtain an advantage by lying about his preferences (ie. bidding dishonestly). A subtle fallacy is to think: 'since the outcome is equitable by definition, any dishonesty on the part of one player will help (or hurt) *both* players equally'. The problem here is that the outcome is only equitable according to the players' *stated* preferences, not their *true* preferences. *Manipulation* occurs exactly when one player lies about his preferences. He thereby engineers an outcome which *appears* equitable when in fact it is not.

Further reading: Brams and Taylor introduce the Adjusted Winner procedure in Chapter 4 of [43]. They also propose a second procedure for producing equitable, envy-free two-player partitions, called *Proportional Allocation* (PA). PA does *not* yield Pareto-optimal partitions, but Brams and Taylor claim PA is probably less manipulable than AWP (although PA is still manipulable).

In Chapter 5 of [43], Brams and Taylor sketch the application of AWP to various disputes (some real, some hypotetical). In a more recent book [44], they have fleshed out the practical applications of AWP in much greater detail, promoting it as a broadly applicable conflict-resolution procedure.

6.6 Other issues

In this section we'll briefly look at four other issues: *entitlements*, *indivisible value*, *chores*, and *manipulation*.

6.6.1 Entitlements

So far we've only considered partition problems where all participants get an 'equal' share. However, in some partition problems, there may be good reasons (either moral, legal, or political) for giving some parties a *larger* share than others. For example:

- In settling an inheritance, the spouse of the deceased may be entitled to $\frac{1}{2}$ of the estate, and each of the three children entitled to $\frac{1}{6}$th of the estate (as each sees it).

- In settling an international territorial dispute, the peace treaty might stipulate that each country gets a portion of territory proportional to its population (or military strength, or ancient historical claim, or whatever). If country A has twice as many people as country B, then country A is entitled to twice as large/valuable a portion (where different countries may value different things).

- In a coalition government, the political parties making up the ruling coalition must divide important cabinet positions amongst themselves. In theory, each political party is entitled to a portion of 'government power' proportional to its popular support (or proportional to the number of seats it obtained in the legislature, which is not the same thing). Different parties will have different estimates of the 'power' of various government ministries (one party may think that the Defense Minister is the more powerful than the Finance Minister; another may think the opposite).

- In a multinational body (eg. the UN, NATO, WTO, etc.), the member states send delegates to a governing Council. The problem is how to allocate power within the Council.

  One solution is that each state should receive numerical representation within the Council proportional to its population (or wealth, or military strength, or whatever). Thus, if state A has three times the population of state B, then state A should have three times as many votes in the Council. However, in §3.4, we saw the 'voting power' of a country is *not* simply the number of votes it has; some countries may have one or more votes, but actually have *no* power, while other countries get too much. So a better solution is that each state should get *voting power* on the Council proportional to its population (or wealth, strength, etc.). However, there are several different 'voting power indices', which give different measures of voting power, and it's not clear which index is correct. Different states may measure their power using different indices, and thus, have different opinions about what 'fair' representation means.

The states must also divide important government positions amongst themselves. For example, who chairs the Council? Who gets to be on important subcommittees? The problem is similar to that of a coalition government, only now with entire states instead of political parties.

If N parties are trying to partition a set $\mathbf{X}$, then an **entitlement vector** is a vector $\mathbf{E} = (e_1, \ldots, e_N)$, where $e_n \geq 0$ for all $n \in [1..N]$, such that $e_1 + \cdots + e_N = 1$. Here, e_n represents the portion that v_n is 'entitled' to receive. The **equidistributed entitlement** $\mathbf{E}_0 = (\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N})$ represents the case where all parties are entitled to an equal share.

A partition $\mathcal{P} = \{\mathbf{P}_1, \ldots \mathbf{P}_N\}$ is called **E-proportional** if, for each $n \in [1..N]$, $\mu[\mathbf{P}_n] \geq e_n$. Thus, each player v_n thinks he got *at least* his 'entitled share' of e_n. Notice that a normal 'proportional' partition is just an $\mathbf{E}_0$-proportional partition, where $\mathbf{E}_0 = (\frac{1}{N}, \ldots, \frac{1}{N})$.

A partition $\mathcal{P} = \{\mathbf{P}_1, \ldots \mathbf{P}_N\}$ is called **E-envy-free** if, for each $n \in [1..N]$, and each $m \neq n$,

$$\frac{\mu_n[\mathbf{P}_n]}{e_n} \quad \geq \quad \frac{\mu_n[\mathbf{P}_m]}{e_m}$$

This means: each player v_n thinks he got at least his entitled share e_n. Furthermore, if he thinks any other player v_m was 'overpaid' more than her entitled share, then he thinks that he himself got overpaid *even more* by comparison. Hence, he will not envy her[8].

Suppose that $\mathbf{E}$ is a **rational** vector, meaning that $e_1, \ldots, e_N$ are rational numbers. Then all of the N-person proportional partition procedures of §6.2 generalize to yield $\mathbf{E}$-proportional partitions. Also, the Brams-Taylor N-person envy-free procedures mentioned at the end of §6.3 generalizes to yield $\mathbf{E}$-envy-free partitions. The key in both cases is to increase the size of N, as follows:

1. Suppose $\mathcal{V} = \{v_1, \ldots, v_N\}$ are N players, with rational entitlement vector $\mathbf{E} = (e_1, \ldots, e_N)$.

   Let D be the greatest common denominator of $e_1, \ldots, e_n$. Thus, we can write:

   $$e_1 \;=\; \frac{c_1}{D}; \quad e_2 \;=\; \frac{c_2}{D}; \quad \ldots \quad e_N \;=\; \frac{c_N}{D}.$$

   for some $c_1, \ldots, c_N \in \mathbb{N}$ such that $c_1 + \cdots + c_N = D$.

[8]Of course, this is assuming that all players feel that the original entitlement vector $\mathbf{E}$ is fair to begin with. If $e_m = 2e_n$, then v_n may 'envy' v_m because she has twice as large an entitlement as he does. But this is beyond the scope of our discussion.

2. Now, construct a new partition problem involving a set $\mathcal{W}$ of D players:

$$\mathcal{W} \quad = \quad \{w_{1,1}, \ldots, w_{1,c_1},\ w_{2,1}, \ldots, w_{2,c_2},\ \ldots, w_{2,N}, \ldots, w_{N,c_N}\}$$

   Imagine that $\{w_{1,1}, \ldots, w_{1,c_1}\}$ are c_1 distinct 'clones' of v_1 (so they all have utility measure μ_1). Likewise, $\{w_{2,1}, \ldots, w_{2,c_2}\}$ are c_2 distinct 'clones' of v_2 (with utility measure μ_2), etc.

3. Apply the partition procedure of your choice to yield a proportional/envy-free partition

$$\mathcal{Q} \quad = \quad \{\mathbf{Q}_{1,1}, \ldots, \mathbf{Q}_{1,c_1},\ \mathbf{Q}_{2,1}, \ldots, \mathbf{Q}_{2,c_2},\ \ldots, \mathbf{Q}_{2,N}, \ldots, \mathbf{Q}_{N,c_N}\}$$

   amongst the players of $\mathcal{W}$.

4. Now define $\mathbf{P}_1 := \mathbf{Q}_{1,1} \sqcup \cdots \sqcup \mathbf{Q}_{1,c_1}$, $\quad \mathbf{P}_2 := \mathbf{Q}_{2,1} \sqcup \cdots \sqcup \mathbf{Q}_{2,c_2}$, etc. Then $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ is a partition of $\mathbf{X}$ amongst $\{v_1, \ldots, v_N\}$.

Observe (**Exercise 50**) that:

$$\Big(\ \mathcal{Q} \text{ is proportional amongst } \mathcal{W}\ \Big) \implies \Big(\ \mathcal{P} \text{ is } \mathbf{E}\text{-proportional amongst } \mathcal{V}\ \Big).$$
$$\Big(\ \mathcal{Q} \text{ is envy-free amongst } \mathcal{W}\ \Big) \implies \Big(\ \mathcal{P} \text{ is } \mathbf{E}\text{-envy-free amongst } \mathcal{V}\ \Big).$$

6.6.2 Indivisible Value

Most of the procedures we've considered assume that the value of the cake is, in principle, infinitely divisible. In other words, they assume that the utility measures of the player contain no *atoms*, or at least, relatively few atoms. However, in *real* fair division problems, there are many large components of indivisible value. For example:

- In an inheritance or divorce settlement, there may be large, indivisible items (a house, a car, a piano) which cannot realistically be shared amongst the disputants. If a mutually agreeable division cannot be found, a standard solution is to liquidate these assets and distribute the proceeds. This may not be satisfactory to anyone, however, because the house (for example) may have a sentimental value which far exceeds its market value. Liquidating the house and dividing the cash is a very suboptimal solution.

- In a territorial dispute, there may be particular sites (eg. mines, cities, religious shrines), which are highly valued by one or more parties, and which cannot be shared between states.

- In dividing government posts amongst the various members of a coalition government, clearly the government posts themselves are indivisible entities. Two political parties cannot 'share' the position of President or of Finance Minister.

In the extreme case, the utility measure is **entirely atomic**, which means $\mathbf{X} = \{x_1, \ldots, x_m\} \sqcup \mathbf{Y}$, where $\mu\{x_1\} + \ldots + \mu\{x_m\} = 1$ and $\mu[\mathbf{Y}] = 0$. In this case, *none* of the methods in §6.2 -§6.5 are applicable.

One elegant procedure which can acommodate entirely atomic measures is Bronislaw Knaster's method of *Sealed Bids* [40]. We do not have time to discuss this procedure; we refer the reader to Section 3.2 of Brams and Taylor [43], or Section 14.9 of Luce and Raiffa [30]. Suffice it to say that Knaster's procedure guarantees a proportional partition of $\{x_1, \ldots, x_m\}$ amongst N players, and has the further advantage that the players need not even agree on the total value of the goods (ie. Veronique may think the whole cake is bigger than Waldemar thinks it is). Knaster resolves these difficulties by introducing an infinitely divisible *numéraire* commodity (ie. money) which the players can exchange to even out the inevitable unfairness of partitioning indivisible commodities. The disadvantage of Knaster's procedure is that it requires each player to enter the division problem with a pre-existing *bankroll* (ie. a stash of cash) which can be used to 'pay off' other players. Thus, Knaster's procedure is inapplicable if the value of the cake exceeds the amount of money which one or more parties can realistically bring to the table.

Other fair division procedures have been proposed for indivisible commodities. For example, William F. Lucas [43, §3.3] has proposed a variant of the Dubins-Spanier procedure, but it is not guaranteed to work in all situations; Lucas must assume a property of 'linearity', which in practice means that there are many very small atoms and few or no large ones. Also, Brams and Taylor suggest that their Adjusted Winner procedure (§6.5) is good for indivisible commodities, because the players will be forced to divide at most *one* of the atoms $\{x_1, \ldots, x_m\}$. In practical situations, this may require liquidating *one* asset, which is certainly better than liquidating *all* of them.

6.6.3 Chores

Instead of partitioning a cake (ie. a 'good' thing), suppose the players must partition a set of 'chores' (ie. a 'bad' thing). Now each player does not want to recieve as *much* as possible, but instead, wants to get as *little* as possible.

For example, a partition $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ of *chores* is **proportional** if $\mu_n[\mathbf{P}_n] \leq \frac{1}{N}$ for all $n \in [1..N]$. The partition $\mathcal{P}$ is **envy-free** if, for all $n, m \in [1..N]$, $\mu_n[\mathbf{P}_n] \leq \mu_n[\mathbf{P}_m]$. Finally, $\mathcal{P}$ is **equitable** if $\mu_1[\mathbf{P}_1] = \mu_2[\mathbf{P}_2] = \cdots = \mu_N[\mathbf{P}_N]$. Pareto-preference and Pareto-optimality are defined in the obvious way.

Most of the 'cake-cutting' procedures in §6.2 -§6.5 generalize to chore-division; we must simply modify them so that each player will systematically choose the 'smallest' portion they can (rather than the 'largest'). Here's one example:

Procedure 6.33: I cut, you choose, for chores

Let $\mathbf{X} = [0, 1]$ be the unit interval (representing a one-dimensional set of chores). Let μ_1 and μ_2 be utility measures on $\mathbf{X}$. Assume μ_1 is at most $\frac{1}{2}$ atomic.

(1) Let $r \in [0,1]$ be such that $\mu_1[0,r) = \frac{1}{2} = \mu_1[r,1]$ (ie. player v_1 divides the chores into two pieces which she perceives have equal size; this is possible because μ_1 is at most $\frac{1}{2}$ atomic)

(2a) If $\mu_2[0,r) \leq \mu_2[r,1]$, then define $\mathbf{P}_2 = [0,r)$ and $\mathbf{P}_1 = [r,1]$. (If v_2 thinks that $[0,r)$ is *smaller*, he takes this piece, and v_1 takes the other one).

(2b) Otherwise, if $\mu_2[0,r) > \mu_2[r,1]$, then define $\mathbf{P}_1 = [0,r)$ and $\mathbf{P}_2 = [r,1]$. (If v_2 thinks that $[r,1]$ is *smaller*, he takes this piece, and v_1 takes the other one).

Now let $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2\}$. (**Exercise 51** Show that that $\mathcal{P}$ is proportional allocation of chores.) ∎

We refer the reader to Sections 3.2.2 and 7.6.3 of [43] for further discussion of chore division.

6.6.4 Nonmanipulability

Recall that a partitioning game Γ *yields* a partition procedure Π if each player of Γ has a unique maximin strategy, and if Γ will produce the same partition as Π when all players play their maximin strategies. A partition game is designed so that your maximin strategy is an 'honest' expression of your preferences. Thus, if you are rational and risk-averse, your 'best' strategy is simply to be honest about what you want. If everyone chooses their 'best' strategy in this way, then everyone will be 'honest', and the outcome will be a 'fair' division.

However, your maximin strategy is only your 'best' strategy when you know *nothing* about the preferences and strategies of the other players. Your maximin strategy simply optimizes your 'worst-case scenario', based on total ignorance of what everyone else is doing. If you *know* (or at least, suspect) what the other players will do, you can exploit this information by picking a strategy which is *not* your maximin, but which will yield a superior outcome if the other people play like you expect them to. This is called **manipulating** the partition game.

Example 6.34: The Divider's Advantage in 'I cut, you choose'

Suppose the cake is half walnut cream, half vanilla, and Veronique knows that Waldemar only likes walnut cream. She likes walnut cream and vanilla equally. Clearly, the 'fairest' partition is into two pieces **A** and **B** such that:

A contains *all* the vanilla (50% of the whole cake).

B contains *all* the walnut cream (the other 50% of the whole cake).

Waldemar will take **B**, and Veronique can take **A**. Both will be happy. Waldemar got everything he wanted, and Veronique herself is indifferent between the two pieces.

However, Veronique can also *exploit* Waldemar as by cutting the cake into two pieces as follows:

A contains *all* the vanilla and 48% of the walnut cream (a total of 74% of the whole cake).

B contains the remaining 52% of the walnut cream (only 26% of the whole cake).

Because **B** has (slightly) more walnut cream, Veronique can expect Waldemar to choose **B**, leaving her with the disproportionately large piece **A** of the cake.

Veronique's opportunity to manipulate 'I cut, you choose' is sometimes called the *Divider's Advantage.*

A partition game is **nonmanipulable** if the maximin strategy for each player is also a *dominant* strategy for that player. In other words, no matter what the other players plan to do, it is *always* rational for you to play your maximin (ie. 'honest') strategy.

At present, no nonmanipulable partition games are known (indeed, there may be an 'impossibility theorem' lurking here). However, Brams and Taylor suggest an interesting strategy [43, §4.5]. Although we cannot find a *single* game where a player's 'honest' maximin strategy is dominant, perhaps we can find a *sequence* of games $\Gamma_1, \Gamma_2, \ldots, \Gamma_M$ (all using the same strategy sets for each player) such that, for each player $v \in \mathcal{V}$,

- v has the same maximin strategy in all of $\Gamma_1, \Gamma_2, \ldots, \Gamma_N$.
- For any N-tuple of strategies $\mathbf{s} = (s_1, \ldots, s_N)$ where v does *not* use his maximin strategy, there is at least one game Γ_m where $\mathbf{s}$ yields an outcome for v which is *strictly worse* than his maximin outcome on Γ_m.

The sequence $\Gamma_1, \ldots, \Gamma_M$ now defines a 'supergame' with the following rules:

1. Everyone plays Γ_1. If everyone is satisfied with the outcome, then the game ends here.
2. If even *one* player is dissatisfied, and feels that Γ_1 has been 'manipulated' by someone else, then the dissatisfied player can unilaterally nullify the results of Γ_1.

   The players move on and play Γ_2, with the proviso that *everyone must use the same strategies they used in* Γ_1*.*
3. If someone is dissatisfied with the results of Γ_2, then the players move onto Γ_3, and so forth.

Thus, if player v suspects player w of a manipulating the outcome of Γ_1 by *not* using his maximin strategy, then v can force w to play the same 'dishonest' strategy in Γ_2, Γ_3, etc. In at least *one* of these games, w's strategy (if it really wasn't his maximin strategy) will produce a strictly inferior outcome for w. The *threat* of this possibility makes it rational for w to be honest in Γ_1.

Further reading: The most extensive resource on fair division theory is by Brams and Taylor [43]; we strongly recommend it to the interested reader.

Conclusion

The theory of social choice is the fascinating intersection of game theory, mathematical economics, and political science, and is an active area of research by scholars in all three areas. This brief introduction discussed some of the major problems in the field, but there are other issues we didn't mention. For example:

Representation & Delegation: In *indirect* democracies, voters ('constituents') elect *representatives* (or *delegates*), and these representatives then vote to decide laws. Even if we assume that the representatives always faithfully represent the wishes of their constituents (a big 'if'), this process introduces many opportunities for unfairness and manipulation, by multiplying the distortions introduced by voting procedures.

For example, suppose that there are four political parties: *New Demagogue*, *Literal*, *Coercitive*, and *Qubekistan Liberation Front* (QLF). If the QLF party controls 40% of the seats in the Parliament, then the QLF can decide any plurality vote with 3 or more alternatives (assuming all QLF representatives vote as a block, and assuming that the opposition vote is split between the other alternatives). Suppose, furthermore, that, in each constituency where QLF was elected, it got only 40% of the popular vote (again, because the opposition votes were split between the three competing parties). Suppose that, in non-QLF constituencies, the QLF got 0% of the vote (because it is a radical ethnic-nationalist party which is only popular in one region). Then really, the QLF party represents only $40\% \times 40\% = 16\%$ of the national vote; nevertheless, this 16% can effectively control the national parliament.

Similar pathologies arise even if we use other 'better' voting methods, like Borda Count. It is a difficult problem to find a 'fair' way to aggregate constituent votes in an indirect democracy.

Ballot Formulation: We have assumed that the voters are presented with a 'ballot' of 'alternatives' amongst which to choose, but we have not examined where this ballot comes from. Sometimes the ballot arises naturally from the problem context (eg. in deciding what movie to see at the cinema, the ballot is simply the set of movies which are currently playing), but in many cases, the ballot must be formulated by *someone*. For example:

- In an election, *someone* must nominate the candidates who will appear on the ballot.
- In a legislature, *someone* must write the bills and amendments which will be debated.

Clearly, this *someone*, who decides the ballot of alternatives, has immense power to influence the outcome of the voting process. For example, one criticism of modern democracy is that there is no 'real choice' amongst the candidates, because all of them are ultimately selected from the same clique of partisan apparatchiks by the same 'establishment' of political cronies, corporate campaign donors, and other vested interests.

Further reading: Saari [33] is a good introduction to voting theory using methods of convex geometry and linear algebra. Kim and Roush [20] develop voting theory using combinatorial methods and boolean matrices. Riker [32] is a good overview of the subject, with emphasis on the ramifications in political philosophy. Another approachable introduction is Taylor [50]. Luce and Raiffa [30] is the classic reference on game theory, and covers bargaining (Chapt. 6), voting theory (Chapt. 14), and fair division (§14.9). Other classic references on voting theory are Sen [37], Fishburn [13] and Arrow [4]. A very readable introduction to the basics of mathematical political science is [50].

Bibliography

[1] Ethan Akin. Vilfredo Pareto cuts the cake. *Journal of Mathematical Economics*, 24:23–44, 1995.

[2] Alan D. Taylor and William S. Zwicker. A characterization of weighted voting. *Proceedings of the AMS*, 115:1089–1094, 1992.

[3] Alan D. Taylor and William S. Zwicker. Weighted voting, multicameral representation, and power. *Games and Economic Behaviour*, 5:170–181, 1993.

[4] Kenneth J. Arrow. *Individual Values and Social Choice*. John Wiley & Sons, New York, 2nd edition, 1963.

[5] A.K. Austin. Sharing a cake. *Mathematical Gazette*, 66(437):212–215, 1982.

[6] Julius Barbanel. Partition ratios, Pareto optimal cake division, and related notions. *Journal of Mathematical Economics*, 32:401–428, 1999.

[7] Julius Barbanel. On the structure of Pareto optimal cake partitions. *Journal of Mathematical Economics*, 33:401–424, 2000.

[8] Anatole Beck. Constructing a fair border. *American Mathematical Monthly*, 94(2):157–162, February 1987.

[9] Duncan Black. *Theory of Committees and Elections*. Cambridge UP, Cambridge, UK, 1958.

[10] Jean-Charles Borda. Memoire sur les elections au Scrutin. *Histoire de l'Academie Royale des Sciences*, 1781.

[11] R.B. Braithwaite. *Theory of games as a tool for the moral philosopher*. Cambridge University Press, Cambridge, UK, 1955.

[12] A.M. Fink. A note on the fair division problem. *Mathematics Magazine*, 37(5):341–342, November 1964.

[13] Peter C. Fishburn. *The Theory of Social Choice*. Princeton Univ. Press, Princeton, NJ, 1973.

[14] Peter Gärdenfors. A concise proof of a theorem on manipulation of socal choice functions. *Public Choice*, 32:137–140, Winter 1977.

[15] John Geanokoplos. Three brief proofs of Arrow's Impossibility Theorem. Technical report, Cowles Foundation for Research in Economics; Yale University, June 2001. `http://cowles.econ.yale.edu`.

[16] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41:587–601, July 1973.

[17] Clifford Hildreth. Alternative conditions for social orderings. *Econometrica*, 21:81–94, 1953.

[18] Theodore P. Hill. Determining a fair border. *American Mathematical Monthly*, 90(7):438–442, August 1983.

[19] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 2nd edition, 1947.

[20] Ki Hang Kim and Fred W. Roush. *Introduction to Mathematical Consensus Theory.* Marcel Dekker, New York, 1980.

[21] Marc D. Kilgour. A formal analysis of the amending formula of Canada's Constitution Act. *Canadian Journal of Political Science*, 16:771–777, 1983.

[22] Bronislaw Knaster. Sur le problem du partage de H. Steinhaus. *Annales de la Societé Polonaise de Mathematique*, 19:228–230, 1946.

[23] Harold W. Kuhn. On games of fair division. In Martin Shubik, editor, *Essays in Mathematical Economics in Honour of Oskar Morgenstern*, pages 29–37. Princeton UP, Princeton, NJ.

[24] L.A. Goodman and Harry Markowitz. Social welfare functions based on individual rankings. *American Journal of Sociology*, 58:257–262, 1952.

[25] Jean-Antoine-Nicolas Caritat (le Marquis de Condorcet). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, 1785.

[26] Lester E. Dubins and Edwin H. Spanier. How to cut a cake fairly. *American Mathematical Monthly*, 68:1–17, 1961.

[27] Lloyd Shapley and Martin Shubik. A method for evaluating the distribution of power in a committee system. *American Political Science Review*, 48:787–792, 1954.

[28] Kenneth May. A set of independent, necessary, and sufficient conditions for simple majority decision. *Econometrica*, 20:680–684, 1952.

[29] John Nash. The bargaining problem. *Econometrica*, 18:155–162, 1950.

[30] R. Duncan Luce and Howard Raiffa. *Games and Decisions.* Dover, New York, 1957,1980.

[31] Howard Raiffa. Arbitration schemes for generalized two-person games. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the Theory of Games* **II**, volume 28 of *Annals of Mathematical Studies*, pages 361–387. Princeton University Press, Princeton, NJ, 1953.

[32] William H. Riker. *Liberalism against populism.* Waveland Press, Prospect Heights, Illinois, 1982.

[33] Donald G. Saari. *Basic Geometry of Voting.* Springer-Verlag, New York, 1995.

[34] Donald G. Saari. Are individual rights possible? *Mathematics Magazine*, 70(2):83–92, April 1997.

[35] Mark Satterthwaite. Strategy proofness and Arrow's conditions. *Journal of Economic Theory*, 10:187–217, October 1975.

[36] Saul X. Levmore and Elizabeth Early Cook. *Super Strategies for Puzzles and Games.* Doubleday, Garden City, NY, 1981.

[37] Amartya K. Sen. *Collective Choice and Social Welfare.* Holden-Day, San Francisco, CA, 1970.

[38] Amartya K. Sen. The impossibility of a Paretian liberal. *Journal of Political Economy*, 78:152–157, 1970.

[39] S.J. Brams and Peter C. Fishburn. Approval voting. *American Political Science Review*, 72:831–847, 1978.

[40] Hugo Steinhaus. The problem of fair division. *Econometrica*, 16:101–104, 1948.

[41] Hugo Steinhaus. Sur la division pragmatique. *Econometrica* supplement, 17:315–319, 1948.

[42] Stephen J. Brams and Alan D. Taylor. An envy-free cake division protocol. *American Mathematical Monthly*, 102(1):9–18, January 1995.

[43] Stephen J. Brams and Alan D. Taylor. *Fair Division: from cake-cutting to dispute resolution.* Cambridge UP, Cambridge, 1996.

[44] Stephen J. Brams and Alan D. Taylor. *The Win-Win Solution: Guaranteeing Fair Shares to Everybody.* Norton, 2000.

[45] Alan D. Taylor Stephen J. Brams and William S. Zwicker. Old and new moving-knife schemes. *Mathematical Intelligencer.*

[46] Alan D. Taylor Stephen J. Brams and William S. Zwicker. A moving knife solution to the four-person envy-free cake-division problem. *Proceedings of the AMS*, 125(2):547–554, February 1997.

[47] Philip Strafin. *Topics in the Theory of Voting.* Birkhauser, Boston, 1980.

[48] Philip Strafin. *Game Theory and Strategy.* Mathematical Association of America, 1993.

[49] Walter Stromquist. How to cut a cake fairly. *American Mathematical Monthly*, 87(8):640–644, October 1980.

[50] Alan D. Taylor. *Mathematics and Politics: Strategy, Voting, Power and Proof.* Springer-Verlag, New York, 1995.

[51] W.A. Webb. An algorithm for a stronger fair division problem.

[52] Douglas Woodall. A note on the fair division problem. *Journal of Combinatorial Theory* **A**, 42:300–301, July 1986.