Draft of

MODELS IN POLITICAL ECONOMY

Collective choice, voting, elections, bargaining, and rebellion

Version 2024.4.26

Latest version available at
https://www.economics.utoronto.ca/osborne/mpe

Comments are welcome. If you notice errors or have suggestions for improvements, including the addition of new material, please send them to martin.j.osborne@gmail.com.

Martin J. Osborne
University of Toronto

Version: 2024.4.26

Contents

Preface vii

Part I Collective choice 1

1 Collective choice with known preferences 3
- 1.1 Collective choice rules 5
- 1.2 Anonymity and neutrality 11
- 1.3 Two alternatives: majority rule 13
- 1.4 Three or more alternatives: Condorcet winners 19
- 1.5 Two domains with Condorcet winners 29
- 1.6 Condorcet winners for two-dimensional sets of alternatives 39
- 1.7 Preference aggregation: Arrow's theorem 49
- 1.8 Preference intensities and interpersonal comparisons 57
- Notes 67
- Solutions to exercises 69

2 Collective choice with privately-known preferences 77
- 2.1 Strategy-proofness of strict Condorcet winner 79
- 2.2 Strategy-proofness for domain of all preference profiles 86
- 2.3 General mechanisms 90
- Notes 93
- Solutions to exercises 94

Part II Voting 97

3 Voting with two alternatives 99
- 3.1 Voting as a strategic game 101
- 3.2 Costly voting with uncertainty about cost 104
- 3.3 Turnout and population size with uncertain preferences 118
- 3.4 Preferences with regret 121
- Notes 124
- Solutions to exercises 124

4 Voting with many alternatives: plurality rule 129
4.1 Plurality rule voting games 130
4.2 Spatial model with concave payoff functions 137
4.3 Strategic and sincere voting: divided majority 139
Notes 148
Solutions to exercises 149

5 Sequential pairwise voting 155
5.1 Binary agendas 157
5.2 Successive agendas 168
5.3 Amendment agendas 170
5.4 Single-peaked preferences and convex agendas 178
Notes 182
Solutions to exercises 183

6 Ethical voting and expressive voting 187
6.1 Ethical voting 188
6.2 Expressive voting 194
Notes 195

7 Voting with shared values and asymmetric information 197
7.1 Strategic abstention 198
7.2 Unanimity rule 220
Notes 226
Solutions to exercises 227

Part III Electoral competition 231

8 Electoral competition: two office-motivated candidates 233
8.1 General model 237
8.2 Spatial model 243
8.3 Candidates uncertain of citizens' preferences 248
8.4 Costly voting 256
8.5 Citizens with preferences over candidates 265
8.6 Electing a legislature 272
8.7 Interpreting Nash equilibrium 278
Notes 279
Solutions to exercises 280

9 Electoral competition: two policy-motivated candidates 287
9.1 Basic model 289
9.2 Uncertain median 299
9.3 Candidates privately informed about policies 305
9.4 Repeated elections 314
Notes 318
Solutions to exercises 318

10 Electoral competition: endogenous candidates 327
10.1 Simultaneous entry with sincere voting 329
10.2 Simultaneous entry with strategic voting 331
10.3 Sequential entry 340
10.4 Citizen-candidates 346
Notes 359
Solutions to exercises 360

11 Distributive politics 369
11.1 Two-candidate competition with exogenous incomes under uncertainty 371
11.2 Voting over transfer systems when income is endogenous 378
11.3 Voting over linear transfer systems 386
11.4 Coalitional bargaining over redistribution 394
Notes 404
Solutions to exercises 405

12 Money in electoral competition 407
12.1 Mobilizing citizens to vote 409
12.2 Informing citizens of candidates' qualities 421
Notes 430
Solutions to exercises 430

13 Two-period electoral competition with imperfect information 433
13.1 Selecting and controlling incumbents with unknown preferences 435
13.2 Inducing an incumbent to exert effort 441
13.3 Candidates who pander to voters 443
Notes 456

Part IV Exercising power 457

14 Bargaining 459
- 14.1 Bargaining game with voting 461
- 14.2 Recurrent distributive bargaining game with voting 486
- Notes 490
- Solutions to exercises 490

15 Rulers threatened by rebellion 493
- 15.1 Threat of revolt 494
- 15.2 Coordinating rebellion 501
- Notes 517
- Solutions to exercises 517

Part V Appendix 521

16 Preferences, profiles, and games 523
- 16.1 Preferences and payoffs 523
- 16.2 Sets of individuals and profiles 528
- 16.3 Brouwer's fixed-point theorem 529
- 16.4 Strategic games 529
- 16.5 Bayesian games 533
- 16.6 Extensive games 535
- 16.7 Coalitional games 546
- 16.8 Optimization 550

References 553

Preface

In many societies, a leader or group of leaders formulates rules that regulate the interactions among individuals and possesses means by which to (imperfectly) enforce these rules. The rules restrict, tax, and subsidize the members' actions. For example, theft may be punishable, economic activity taxed, and members of the society with skills not highly valued by others subsidized. The degree to which the members of a society may influence the selection of the group of leaders and the rules that they enact, and the manner in which they may do so, varies widely. In some societies, elections determine the composition of an assembly that plays a role in determining the rules enacted. However, the nature of the role this assembly plays and the sensitivity of its composition to the preferences of the members of the society varies considerably across societies. These topics are subjects of the field of political economy.

Many other topics fall under the heading of political economy. This book covers only a small part of the field; its title should be read as "*some* models in political economy". In terms of methodology, it is restricted to formal models. In terms of subjects, it concentrates on collective choice, voting, and electoral competition, with brief forays into bargaining and regime change.

I present a small number of models in detail, and make no attempt to survey related work. I personally appreciate a result only when I understand its proof, and for this reason I include full proofs of almost all the results I state.

The models I discuss are the ones that I find appealing—they elegantly express original ideas and help me organize my thoughts about aspects of the world. One way in which they do that is bringing to the fore the common threads in disparate situations. For example, the model of collective decision-making in Chapter 1 highlights the elements common to the problems of the residents of a country choosing a national health policy and a group of friends choosing a restaurant for dinner. In linking situations in this way, the model make me think that I better understand some aspect of the world, although the precise nature of that improvement is often hard to pin down .

In models as in many other spheres, tastes differ. I prefer relatively general models over ones that assume specific functional forms and am wary of refinements of the basic solution concepts of game theory, Nash equilibrium and subgame perfect equilibrium. I find models that assume specific functional forms unsatisfying because they leave open the possibility that their properties depend on the forms, and models that consider only a subset of equilibria unsatisfying because they they leave open the possibility that other equilibria do not have the same properties. But in both cases, the dividing line is fuzzy. After all, every model is an example of a more general model. In some of the models I discuss, the decision-makers' payoff functions are linear, a specific functional form, and the players are assumed not to use weakly dominated actions, a refinement of the standard notions of equilibrium.

One large class of models that I omit is worth mentioning explicitly. Many of the models I discuss involve candidates competing in elections. In all of these models I assume that each candidate aims to maximize her probability of winning. Many models that have been studied assume instead that each candidate aims to maximize the expected number of votes she receives or her expected plurality. These criteria, while often simplifying the analysis, are generally inconsistent with the maximization of the probability of winning: a candidate's expected vote share or expected plurality may be higher in one situation than in another even though her probability of winning is lower.

Format and conventions

The formal content is contained in definitions and propositions, which are self-contained. The text is intended to make the formal content digestible, but the definitions and propositions are intended to be entirely self-contained: if you read nothing else you will not miss only discussion and motivation, not any formal content.

In the electronic version of the book, every term in the boxes containing the formal content that has a technical meaning, other than basic mathematical terms, is hyperlinked to its definition. If you click on the hyperlink you are taken to the definition; your pdf viewer probably allows you to return to where you were by pressing the Alt and left arrow keys. My intention is that these hyperlinks allow you to read any definition or result independently of the other material.

With a few exceptions, the names I attach to concepts, models, and results relate to their content rather than the people who originated them, even when the names of the originators are commonly used by researchers in the area. Names that relate to the originators are convenient shorthands for the cognoscenti, but are unhelpful for the uninitiated or for those of us who are memory-challenged. Further, many models have mixed and unclear parentage, and in those cases assigning one name implies a misplaced certainty regarding their origin.

Acknowledgments

For comments and discussions I am grateful to Avidit Acharya, Jean-Pierre Benoît, Daniel Bernhardt, Craig Brett, Peter Buisseret, Francesco De Sinopoli, John Duggan, Jon Eguia, Jean Guillaume Forand, Yoram Halevy, Helios Herrera, Sean Horan, Haruo Imai, Navin Kartik, Hervé Moulin, Dominik Peters, Carolyn Pitchik, Jeffrey S. Rosenthal, Ariel Rubinstein, Remzi Sanver, Alejandro Saporiti, Colin Stewart, Matthew Turner, Richard van Weelden, John Weymark, Ruolong Xiao, and Huseyin Yildirim.

I am particularly grateful to Sean Horan, Carolyn Pitchik, and Ariel Rubinstein, each of whom made detailed comments on several chapters of a draft of the book that led to big improvements.

Jon Eguia read the entire book and provided countless comments. He pointed out errors, made suggestions for cutting material and for adding it, and showed me how to improve the exposition. His advice led me make many major improvements. I am enormously grateful for the many hours he spent with my drafts.

My discussions with Ariel Rubinstein over the many years of our collaboration have been enormously influential in my thinking about economic modeling and the presentation of economic theory.

I am grateful to Hsin-Po Wang (StackExchange user `Symbol1`) for technical help with the ternary plots in Chapters 4 and 14.

During the long gestation period for this book, I had the pleasure of working on it at various locations. I am grateful for the hospitality I was generously afforded during my visits to the Kyoto Institute of Economic Research, the Department of Economics at the National University of Singapore, the School of Economics at the University of New South Wales, the College of Administrative Sciences and Economics at Koç University, and the Centro de Economía Aplicada at the University of Chile.

I cite the sources of the models that I discuss in the "Notes" section at the end of each chapter. In addition, I found Moulin (1988), Austen-Smith and Banks (1999, 2005), and Mueller (2003), as well as unpublished notes by John Duggan and by Stephen Coate, particularly helpful in understanding and appreciating work in the field.

MODELS IN POLITICAL ECONOMY

I Collective choice

1 Collective choice with known preferences 3
2 Collective choice with privately-known preferences 77

1 Collective choice with known preferences

1.1 Collective choice rules 5
1.2 Anonymity and neutrality 11
1.3 Two alternatives: majority rule 13
1.4 Three or more alternatives: Condorcet winners 19
1.5 Two domains with Condorcet winners 29
1.6 Condorcet winners for two-dimensional sets of alternatives 39
1.7 Preference aggregation: Arrow's theorem 49
1.8 Preference intensities and interpersonal comparisons 57

The residents of a country have to choose a national health policy. The members of an organization have to choose a board of directors. A group of friends has to choose a restaurant for dinner. These problems all involve collective choice. In each case, a group of people whose members may disagree about the desirability of the available alternatives has to select a common action.

The study of collective choice lies at the heart of political economy. One line of inquiry analyzes the properties of specific mechanisms for choosing an action. This chapter discusses a more ambitious avenue that involves formulating a list of properties that appear to be desirable and determining the mechanisms that satisfy these properties.

What do we mean by a mechanism? In this chapter I assume that the individuals' preferences are known, and a mechanism takes these preferences and the set of available alternatives as inputs and generates a subset of the available alternatives as output. (Ideally, this subset consists of a single alternative.) In the next chapter, I consider models in which the individuals' preferences are not known.

What can we know about the individuals' preferences? For each pair of alternatives, we may know the one that each individual prefers. We may in addition have information about the intensity of each individual's preference for one alternative rather than another, and about the well-being each individual derives from each alternative and how these well-beings differ among the individuals.

Synopsis

The models in Sections 1.1 through 1.7 include information only about the individuals' rankings of alternatives. A collective choice problem consists of a set of individuals, a set of alternatives, and a specification of the individuals' preferences over the alternatives. Proposition 1.1, known as May's theorem, shows that for a problem with two alternatives, the mechanism that selects the alternative favored by a majority of individuals is the only one for which the outcome does not depend on the names of the individuals (anonymity) or the names of the alternatives (neutrality) and responds sensibly when the individuals' preferences change (positive responsiveness).

For collective choice problems with three or more alternatives, the existence of mechanisms for selecting alternatives with desirable properties depends on the nature of the individuals' preferences. Suppose that for a given problem there exists an alternative x such that, for every other alternative y, a majority of individuals prefer x to y. Such an alternative is called the strict Condorcet winner of the problem. Proposition 1.2 shows that if every member of a set of problems has a strict Condorcet winner then the mechanism that selects that alternative for each problem in the set is the only one that satisfies properties similar to, though less obviously compelling than, the properties in May's theorem.

Under what conditions does a collective choice problem have a strict Condorcet winner? One such condition is that the alternatives may be ordered so that every individual's preferences are single-peaked: as we move through the ordering, every individual initially becomes better off, and then, after we pass her favorite alternative, worse off. Proposition 1.4 shows that for a problem that satisfies this condition, the median of the individuals' favorite alternatives is the strict Condorcet winner if the number of individuals is odd. Another condition under which a collective choice problem has a strict Condorcet winner is that the individuals' preferences satisfy the single-crossing condition: the individuals may be ordered so that for every individual i and any alternatives x and y, if i likes x at least as much as y then either (*a*) all individuals who precede i in the ordering or (*b*) all individuals who follow i in the ordering prefer x to y. Proposition 1.5 shows that for a problem that satisfies this condition, if each median individual with respect to the ordering of individuals has a unique favorite alternative then each such alternative is a Condorcet winner of the problem. Further, if the number of individuals is odd then for any alternatives x and y, the (unique) median individual prefers x to y if and only if a majority of individuals do so, and hence in particular the favorite alternative of the median individual is the strict Condorcet winner.

For sets of problems that are more than slightly larger than the set of all problems with a strict Condorcet winner, the conclusion is negative. Proposition 1.3 shows that for any set of problems that includes all problems with a strict Condorcet winner plus all problems that would have a strict Condorcet winner if the preferences of a single individual were changed in a certain way, no mechanism for selecting an alternative satisfies the properties in Proposition 1.2.

Section 1.7 presents a different approach. Instead of considering mechanisms for selecting alternatives in collective choice problems, it studies the problem of aggregating the individuals' preferences. The objective is to find a single (societal) preference relation that reasonably reflects the individuals' preferences. One motivation for finding such a preference relation is that we do not know the collective choice problem the society will face, and we want to be prepared for whatever problem arises. Proposition 1.9 (Arrow's impossibility theorem) shows that no mechanism for constructing a societal preference relation from the individuals' preference relations satisfies three appealing properties.

The model in Section 1.8 includes information about each individual's welfare for each alternative, not merely her preferences. A social welfare ordering ranks welfare profiles. Three examples are the utilitarian ordering, which ranks profiles according to their sum, the Nash ordering, which ranks positive profiles according to their product, and the leximin ordering, which ranks profiles according to the welfare of the worst-off individual. All three of these orderings are anonymous and indicate an increase in social welfare when all individuals' welfares increase. Propositions 1.11, 1.12, and 1.13 show the implications of adding one more requirement. Proposition 1.11 shows that the leximin ordering results when the additional requirement imposes a particular type of equity. Proposition 1.12 shows that the utilitarian ordering results when the additional requirement is that the welfare index is invariant to transformations of the individuals' welfares that preserve the rankings of welfare differences but not necessarily those of welfare levels. Proposition 1.13 shows that the Nash ordering results when the additional requirement is that the welfare index is invariant to transformations of the individuals' welfares that preserve the rankings of welfare ratios.

1.1 Collective choice rules

A society consists of a set of individuals and a set of alternatives.

Definition 1.1: Society

A *society* $\langle N, X\rangle$ consists of a set N (of *individuals*) and a set X (of *alternatives*). The society $\langle N, X\rangle$ is *finite* if N and X are finite.

A collective choice problem consists of a society and, for each individual, a preference relation over the set of alternatives. The preference relation $\succsim_i$ of individual i models her preferences: for all alternatives x and y, we interpret $x \succsim_i y$ to mean that i likes x at least as much as y. For any preference relation $\succsim_i$ we define the binary relation $\succ_i$ by

$$x \succ_i y \quad \Leftrightarrow \quad x \succsim_i y \text{ and not } y \succsim_i x \tag{1.1}$$

and interpret $x \succ_i y$ to mean that i prefers x to y, and we define the binary relation $\sim_i$ by

$$x \sim_i y \quad \Leftrightarrow \quad x \succsim_i y \text{ and } y \succsim_i x \tag{1.2}$$

and interpret $x \sim_i y$ to mean that i is indifferent between x and y (that is, she likes them equally well). (See Section 16.1 for more discussion of preference relations.) I refer to an assignment of preference relations to the individuals as a preference profile.

Definition 1.2: Preference profile

A *preference profile* for a society $\langle N, X\rangle$ is a function that associates with each individual (member of N) a preference relation on X. A preference profile is *strict* if every individual's preference relation is strict (i.e. no individual is indifferent between any two alternatives).

I denote by $(\succsim_i)_{i\in N}$ the preference profile in which the preference relation of each individual $i \in N$ is $\succsim_i$. I denote this profile also simply by $\succsim$; the absence of a subscript indicates that the symbol denotes a profile rather than a preference relation.

Definition 1.3: Collective choice problem

A *collective choice problem* $\langle N, X, \succsim\rangle$ consists of a society $\langle N, X\rangle$ for which both N and X have at least two members and a preference profile $\succsim$ for $\langle N, X\rangle$. The problem is *finite* if the society is finite.

Example 1.1: Collective choice problem

An example of a collective choice problem is $\langle\{1,2,3\},\{a,b,c\},\succsim\rangle$ where

$$a \succ_1 c \succ_1 b$$
$$b \succ_2 a \sim_2 c$$
$$a \sim_3 b \sim_3 c$$

(with $\succ_i$ and $\sim_i$ derived from $\succsim_i$ in (1.1) and (1.2)). Here is an evocative representation of this problem.

1	2	3
a	b	$a\,b\,c$
c	$a\,c$	
b		

Each column shows the preference relation of the individual whose name heads the column. In each column, the alternatives are listed in order of preference, with the best at the top. Multiple alternatives in a cell indicate indifferences. For example, the middle column indicates that individual 2 likes b best and is indifferent between a and c.

A target of the analysis in this chapter is to specify, for each collective choice problem, alternatives that are reasonable compromises given the individuals' (possibly divergent) preferences. A function that specifies a set of alternatives for each collective choice problem is called a collective choice rule. I formulate properties for collective choice rules that seem to be desirable and look for rules that satisfy these properties.

To require that the properties hold for all collective choice problems is demanding. In some environments, some preference profiles are not plausible, and we may be content for the properties to be satisfied for only a limited set of profiles. For example, if we are studying the choice of a political position from the set {*left*, *center*, *right*}, we might assume that *center* is not the worst alternative for any individual: an individual whose favorite position is *left* prefers *center* to *right*, and an individual whose favorite position is *right* prefers *center* to *left*. So in this environment it may be enough that a collective choice rule specifies outcomes for preference profiles in which *center* is not the worst alternative for any individual. To accommodate such cases, I allow a collective choice rule to apply to only a subset of the set of all collective choice problems. Following the conventional terminology, I call such a set of collective choice problems a domain.

A domain that fits the example in the previous paragraph is the set of collective choice problems $\langle N, \{left, center, right\}, \succsim\rangle$ for which the preference relation $\succsim_i$ of each individual $i \in N$ satisfies *left* $\succsim_i$ *center* $\succsim_i$ *right*, *center* $\succsim_i$ *left* $\succsim_i$ *right*, *center* $\succsim_i$ *right* $\succsim_i$ *left*, or *right* $\succsim_i$ *center* $\succsim_i$ *left*. Another domain consists of all problems $\langle N, X, \succsim\rangle$ for a given set N where X has three members. In this case the preference relation $\succsim_i$ of each individual $i \in N$ is one of the thirteen possible preference relations over a three-member set (the one for which all three alternatives are indifferent, the six for which exactly two alternatives are indifferent, and the six for which no two alternatives are indifferent). A domain that is conveniently assumed in some models is the set of collective choice problems for a given set of individuals in which every individual's preference relation is strict.

A collective choice rule intended to recommend the alternative to be chosen is most useful if it specifies, for each collective choice problem, a single alternative. However, this requirement conflicts with the requirement that the alternatives be treated symmetrically. Suppose, for example, that the number of individuals is even, the set of available alternatives is $\{a, b\}$, and half of the individuals prefer a to b while the other half prefer b to a. Then if we treat the alternatives symmetrically we have to declare a tie between a and b. For this reason I define a collective choice rule for a domain to specify a *set* of alternatives for each collective choice problem in the domain.

Definition 1.4: Collective choice rule

For any set D of collective choice problems, a *collective choice rule for D* is a function that associates with every collective choice problem $\langle N, X, \succsim\rangle$ in D a nonempty subset of X (the alternatives selected by the rule).

Perhaps the most well-known collective choice rule is plurality rule, which selects the alternative (or alternatives, in the case of a tie) that is ranked first by the largest number of individuals. To define this rule precisely, I first define an individual's set of favorite alternatives: the alternatives she likes at least as much as every other alternative.

Definition 1.5: Favorite alternatives

For any set X (of alternatives) and any preference relation $\succsim_i$ on X, the set of *favorite alternatives in X for $\succsim_i$* is

$$\{x \in X : x \succsim_i y \text{ for all } y \in X\}.$$

If, for example, $X = \{a, b, c\}$ and $a \sim_i b \succ_i c$, then the set of favorite alternatives in X for $\succsim_i$ is $\{a, b\}$. Note that if X has infinitely many members then the set of favorite alternatives for a preference relation on X may be empty. For example, if

$X = [0,1)$ and the preference relation $\succsim_i$ is defined by $x \succsim_i y$ if and only if $x \geq y$ then the set of favorite alternatives in X for $\succsim_i$ is empty.

Plurality rule assigns to a collective choice problem the alternatives that are favorites of the largest number of individuals.[1]

Definition 1.6: Plurality rule

Let D be a set of collective choice problems $\langle N, X, \succsim \rangle$ for which N is finite and for each $i \in N$ the set X_i^* of favorite alternatives in X for $\succsim_i$ is nonempty. *Plurality rule* is the collective choice rule for D that assigns to each collective choice problem $\langle N, X, \succsim \rangle \in D$ the set

$$\{x \in X : |\{i \in N : x \in X_i^*\}| \geq |\{i \in N : y \in X_i^*\}| \text{ for all } y \in X\}.$$

For the collective choice problem in Example 1.1, plurality rule selects $\{a, b\}$, because a and b are both favorite alternatives of two individuals and c is a favorite alternative of only one individual.

The alternatives selected by the plurality rule collective choice rule depend only on the individuals' favorite alternatives. A rule that takes into account the individuals' preferences among the alternatives they rank below their favorite alternatives was proposed by Jean-Charles de Borda (1733–1799). This rule is defined only for collective choice problems in which all preference relations are strict (no individual is indifferent between any two alternatives). It assigns to each alternative x in each individual's preferences a number of points equal to the number of alternatives the individual ranks lower than x. Then it chooses the alternative (or alternatives, in the case of a tie) for which the sum of the number of points over all individuals is largest.

Definition 1.7: Borda rule

Let D be a set of collective choice problems $\langle N, X, \succsim \rangle$ for which N is finite and the preference relation $\succsim_i$ of each individual $i \in N$ is strict. The *Borda rule* is the collective choice rule for D that assigns to each collective choice problem $\langle N, X, \succsim \rangle \in D$ the set of alternatives $x \in X$ that maximize

$$\sum_{i \in N} p_i(x),$$

where $p_i(x) = |\{z \in X : x \succ_i z\}|$ for each $i \in N$ and $x \in X$, the number of

[1] The name plurality rule is used also in a different context, for a voting mechanism in which each individual selects (votes for) one alternative (not necessarily her favorite), and the alternative selected by the most individuals wins. I analyze this voting mechanism in Chapters 3 and 4.

alternatives that $\succsim_i$ ranks below x.

Example 1.2: Borda rule

Consider the collective choice problem $\langle\{1,2,3\},\{a,b,c,d\},\succsim\rangle$ in which the individuals' preferences are given in the following table.

1	2	3
a	a	b
b	b	c
c	c	d
d	d	a

We have $p_1(a) = p_2(a) = 3$ and $p_3(a) = 0$, $p_1(b) = p_2(b) = 2$ and $p_3(b) = 3$, and $p_1(c) = p_2(c) = 1$ and $p_3(c) = 2$, so the Borda rule selects $\{b\}$. By contrast, plurality rule selects $\{a\}$. The Borda rule takes into account that even though b is ranked first by only one individual, it is ranked second by the other two, whereas a is ranked last by the third individual.

The model of a collective choice problem includes only information about the individuals' preference rankings. One interpretation of the Borda rule is that it imbues these rankings with interpersonally-comparable cardinal significance: it treats each rung in each individual's ranking as equivalent to each rung in every other individual's ranking. If, for example, an alternative goes up one rung in one individual's preferences and down one rung in another's, the total number of points it receives remains the same. In some circumstances, using the individuals' preference rankings in this way seems inappropriate. For example, if, in the situation that the collective choice problem in Example 1.2 models, there is a meaningful scale on which individuals 1 and 2 regard b as much worse than a whereas individual 3 regards c, d, and a as only slightly worse than b, then $\{a\}$ may be a more reasonable choice for the group than $\{b\}$.

For a collective choice problem with k alternatives, the Borda rule assigns $k-1$ points to the top alternative in an individual's preferences, $k-2$ points to the next alternative, and so on. A *scoring rule* is a generalization of the Borda rule in which for some numbers $r^1 \geq r^2 \geq \cdots \geq r^k$, r^1 points are assigned to the top alternative in an individual's preference, r^2 points are assigned to the next alternative, as so forth. Like the Borda rule, such a rule may be interpreted as imbuing the individuals' preference rankings with interpersonally-comparable cardinal significance. It differs from the Borda rule in the weights it assigns to the rungs in the individuals' rankings. Note that a scoring rule for which r^1 is a positive number and $r^j = 0$ for $j = 2, \ldots, k$ is plurality rule.

Exercise 1.1: Scoring rules

Consider the collective choice problem $\langle N, \{a,b,c\}, \succsim\rangle$ for ten individuals in which one individual prefers a to b to c, four prefer b to a to c, three prefer c to a to b, and two prefer c to b to a. For each number $p \in (0,1)$ find the alternatives selected by the scoring rule that assigns 1 point to the top alternative in each individual's preferences, p points to the middle alternative, and 0 points to the bottom alternative.

1.2 Anonymity and neutrality

Plurality rule and the Borda rule both treat the individuals symmetrically—every individual's preferences have the same influence on the outcome. They also treat the alternatives symmetrically—no alternative has any special significance. I now define these properties precisely.

First define a permutation on a finite set to be a one-to-one function from the set to itself, and a permutation of a profile to be a one-to-one reassignment of the elements in the profile.

Definition 1.8: Permutation

For any finite set Y, a *permutation on* Y is a one-to-one function from Y to Y. For any finite set N and profile $(x_i)_{i\in N}$, the profile $(y_i)_{i\in N}$ is a *permutation of* $(x_i)_{i\in N}$ if for some permutation π on N we have $y_i = x_{\pi(i)}$ for all $i \in N$.

One permutation on the set $N = \{1,2,3\}$, for example, is the function π for which $\pi(1) = 2$, $\pi(2) = 1$, and $\pi(3) = 3$, and the corresponding permutation of the profile (x_1, x_2, x_3) is the profile (x_2, x_1, x_3). There are five other permutations on the set $\{1,2,3\}$, including the one that maps each member of the set to itself.

Now let N be a set of individuals, $\succsim$ a preference profile for N, and $\succsim'$ a permutation of $\succsim$. For example, for the collective choice problem $\langle\{1,2,3\}, \{a,b,c\}, \succsim\rangle$ in Example 1.1 and the permutation π on $\{1,2,3\}$ for which $\pi(1) = 2$, $\pi(2) = 1$, and $\pi(3) = 3$, the preference profile $\succsim'$ is

1	2	3
b	a	abc
ac	c	
	b	

A collective choice rule is anonymous if it assigns the same set of alternatives to the collective choice problems $\langle N, X, \succsim\rangle$ and $\langle N, X, \succsim'\rangle$ whenever $\succsim'$ is a permutation of $\succsim$. That is, the set of alternatives assigned by an anonymous collective choice rule depends only on the collection of preference relations, not on which individual has which preference relation.

Definition 1.9: Anonymous collective choice rule

Let D be a set of finite collective choice problems. A collective choice rule F for D is *anonymous* if for every collective choice problem $\langle N, X, \succsim\rangle \in D$ and every permutation $\succsim'$ of $\succsim$ for which $\langle N, X, \succsim'\rangle \in D$ we have $F(N, X, \succsim') = F(N, X, \succsim)$.

Plurality rule and the Borda rule are both anonymous. Dictatorship is decidedly not.

Example 1.3: Dictatorship

Let N be a set (of individuals). For any individual $i \in N$, *dictatorship by individual i* is the collective choice rule (for any set of collective choice problems) that selects for any collective choice problem $\langle N, X, \succsim\rangle$ the favorite alternatives in X for $\succsim_i$.

Although dictatorship does not treat the individuals equally, it treats the alternatives equally, as do plurality rule and the Borda rule. To be precise, let $\langle N, X, \succsim\rangle$ be a collective choice problem, let σ be a permutation on X, and consider the collective choice problem $\langle N, X, (\succsim_i^\sigma)_{i\in N}\rangle$ in which $\succsim_i^\sigma$ is the preference relation defined by

$$x \succsim_i y \text{ if and only if } \sigma(x) \succsim_i^\sigma \sigma(y).$$

That is, i's preference between x and y according to $\succsim_i$ is her preference between $\sigma(x)$ and $\sigma(y)$ according to $\succsim_i^\sigma$.

Consider again, for example, the collective choice problem $\langle\{1,2,3\}, \{a, b, c\}, \succsim\rangle$ in Example 1.1. Define the permutation σ on X by $\sigma(a) = c$, $\sigma(b) = a$, and $\sigma(c) = b$. Then $(\succsim_i^\sigma)_{i\in N}$ is given by

1	2	3
c	a	abc
b	bc	
a		

A collective choice rule is neutral if for any permutation on the set of alternatives, the alternatives the rule selects for the permuted problem are the permutations of the alternatives it selects for the original problem. If, for example, $\{a\}$

is selected for the problem in Example 1.1, then $\sigma(a)$, namely c, is selected for the problem $\langle\{1,2,3\},\{a,b,c\},(\succsim_i^\sigma)_{i\in N}\rangle$ just defined.

Definition 1.10: Neutral collective choice rule

Let D be a set of finite collective choice problems. A collective choice rule F for D is *neutral* if for every collective choice problem $\langle N,X,\succsim\rangle \in D$ and every permutation σ on X for which $\langle N,X,\succsim^\sigma\rangle \in D$, we have

$$F(N,X,\succsim^\sigma) = \{x \in X : x = \sigma(z) \text{ for some } z \in F(N,X,\succsim)\},$$

where for each individual $i \in N$ the preference relation $\succsim_i^\sigma$ is defined by

$$x \succsim_i y \text{ if and only if } \sigma(x) \succsim_i^\sigma \sigma(y).$$

The set of alternatives selected by a neutral collective choice rule depends only on the individuals' preferences over the alternatives, not on the names of the alternatives. If for one collective choice problem a neutral rule selects alternative a but not alternative b, for example, then for the problem in which the individuals' rankings of a and b are reversed, the rule selects b but not a. In particular, a neutral collective choice rule gives no significance to any given alternative, like the status quo if one exists. The idea is that the priority of any given alternative is reflected in the individuals' preferences; no alternative has any special status independent of its rankings by the individuals.

Many collective choice rules, including ones that do not seem sensible, are anonymous and neutral. For example, the rule that selects the alternative ranked *lowest* by the largest number of individuals is anonymous and neutral. I now discuss additional properties that appear to be desirable. I first analyze the case of two alternatives, which turns out to differ significantly from that of three or more alternatives.

1.3 Two alternatives: majority rule

For collective choice problems in which the set of alternatives contains two members, many collective choice rules, including both plurality rule and the Borda rule, are equivalent to majority rule. That is, they select the alternative favored by more than 50% of the individuals, or, if both alternatives are favored by exactly 50% of the individuals, they select both alternatives, in a tie.

Definition 1.11: Majority rule

Majority rule is the collective choice rule for any set of finite collective choice problems with two alternatives that selects both alternatives (i.e. declares a tie) when each alternative is a favorite alternative of the same number of individuals, and otherwise selects the alternative that is a favorite alternative of a strict majority of individuals. An alternative with the latter property is a *strict majority winner*.

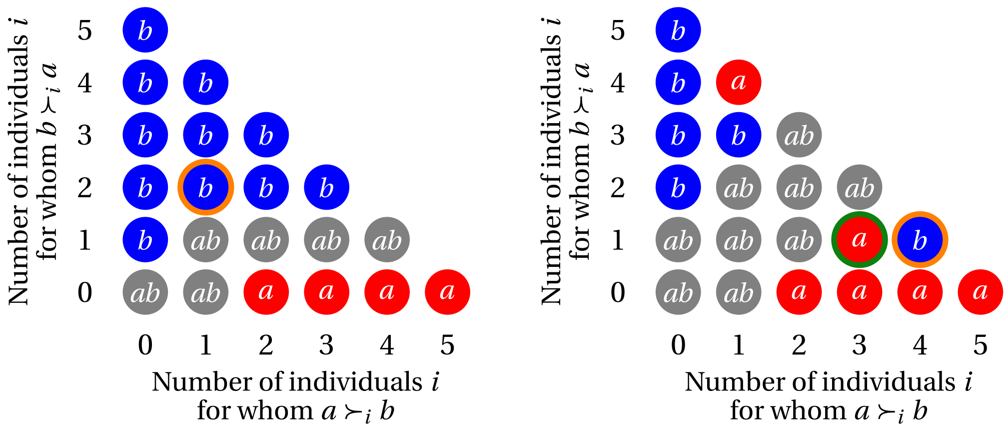
Note that if majority rule selects a single alternative, that alternative is not necessarily preferred to the other alternative by a majority of individuals. For example, if one individual prefers a to b and all the remaining individuals are indifferent between a and b then a is the alternative selected by majority rule.

One anonymous and neutral rule that is not equivalent to majority rule is minority rule, which selects the alternative favored by *fewer than* 50% of the individuals. What property distinguishes majority rule from minority rule?

We can represent an anonymous collective choice rule for the set of problems with two alternatives and no restrictions on the individuals' preferences in a diagram like those in Figure 1.1. These diagrams represent rules for problems with five individuals in which the alternatives are a and b. In each diagram, each disk represents a set of collective choice problems. (Ignore for the moment the letters inside the disks.) The disk at the point (x, y) represents the problems in which x individuals prefer a to b, y prefer b to a, and the remainder are indifferent between a and b. For example, the disk surrounded by an orange circle in Figure 1.1a represents problems in which one individual prefers a to b, two prefer b to a, and two are indifferent between a and b.

In each diagram, the letters inside the disks define an anonymous collective choice rule: each disk indicates the outcome selected by the rule for the problems that the disk represents. For example, in Figure 1.1a the label ab on the disk at position $(1,0)$ indicates that the rule the figure depicts assigns a tie to every problem in which one individual prefers a to b and the remainder are indifferent between a and b. For easy identification, disks labeled a are red, those labeled b are blue, and those labeled ab are gray.

The rule shown in Figure 1.1a is not neutral. Denote by $G(x, y)$ the set of alternatives assigned by the rule when x individuals prefer a to b, y prefer b to a, and the remainder are indifferent between a and b. For the rule to be neutral, we need $G(x, y) = \{a\}$ if and only if $G(y, x) = \{b\}$, and $G(x, y) = \{a, b\}$ if and only if $G(y, x) = \{a, b\}$. In particular, we need $G(x, x) = \{a, b\}$ for all values of x. Thus a rule is neutral if and only if the pattern of outcomes in a diagram like those in Figure 1.1 is symmetric about the main diagonal, so that in particular every



(a) A non-neutral rule. (b) A neutral rule.

Figure 1.1 Two anonymous collective choice rules for collective choice problems with five individuals and two alternatives, a and b. (The rules indicated are not intended to be sensible.)

outcome on the diagonal is $\{a, b\}$. The rule in Figure 1.1a is not neutral because, for example, it assigns b to problems in which two individuals prefer a to b, two prefer b to a, and the remaining individual is indifferent between a and b.

The rule given in Figure 1.1b is neutral, but it has an unattractive feature: as the disk circled in green indicates, when three individuals prefer a to b, one prefers b to a, and the remaining individual is indifferent between a and b, the outcome is a, but, as the disk circled in orange indicates, if the individual who is indifferent switches to preferring a then the outcome switches to b. That is, when *more* individuals favor a, the outcome moves *away* from a. (Symmetrically, the outcome is b when three individuals prefer b to a, one prefers a to b, and the remaining individual is indifferent between a and b, and is a when the individual who is indifferent switches to preferring b.)

To eliminate collective choice rules that behave in this way, we can require that if an alternative moves up in some individuals' preferences then the outcome specified by the collective choice rule moves in the direction of that alternative. To define this requirement precisely, I first define a preference profile $\succsim'$ to be an improvement of a profile $\succsim$ for an alternative x relative to an alternative y if the profiles differ only in the preferences of the individuals in some set J and for every individual $j \in J$, either x is ranked equal to or below y by $\succsim_j$ and above y by $\succsim'_j$ or x is ranked below y by $\succsim_j$ and equal to or above y by $\succsim'_j$, while all other alternatives are ranked the same by $\succsim_j$ and $\succsim'_j$. I give a definition of the property for collective choice problems with any number of alternatives, because I use it later for problems with more than two alternatives.

1	2	3
a	b	abc
c	ac	
b		

Original problem

1	2	3
a	b	abc
bc	ac	

Improvement of b relative to c

1	2	3
a	b	b
c	ac	ac
b		

Improvement of b relative to c

Figure 1.2 The collective choice problem in Example 1.1, at the left, and two problems in which the preference profiles are improvements for b relative to c. For the problem on the right, the preference profile is also an improvement for b relative to a.

Definition 1.12: Improvement of preference profile for alternative

Let $\langle N, X, \succsim \rangle$ be a finite collective choice problem, let $\succsim'$ be another preference profile for $\langle N, X \rangle$, and let $J \subseteq N$ be the set of individuals j for whom $\succsim_j$ differs from $\succsim'_j$. The profile $\succsim'$ is an *improvement of $\succsim$ for $x \in X$ relative to $y \in X$* if for each $j \in J$ we have

$$\text{either } y \succsim_j x \text{ and } x \succ'_j y, \text{ or } y \succ_j x \text{ and } x \succsim'_j y$$

and

$$w \succsim'_j z \text{ if and only if } w \succsim_j z \text{ for all } w \in X \setminus \{x\} \text{ and all } z \in X \setminus \{x\}.$$

Consider the collective choice problem in Example 1.1, which is shown at the left in Figure 1.2. The preference profiles for the other two problems in the figure are improvements for b relative to c. In the problem in the middle, b moves up in individual 1's preferences, to become indifferent with c. (If b moves further up, to lie between a and c, to be indifferent with a, or to be above a, the resulting problem is also an improvement for b relative to c.) In the problem on the right, b moves up in individual 3's preferences to become preferred to a and c rather than indifferent with them. (This profile is also an improvement for b relative to a.)

For a collective choice problem $\langle N, X, \succsim \rangle$ with two alternatives, a and b, the preference profile $\succsim'$ is an improvement of $\succsim$ for a relative to b if for every individual j in some set J we have (*i*) $b \succ_j a$ and $a \succ'_j b$, (*ii*) $a \sim_j b$ and $a \succ'_j b$, or (*iii*) $b \succ_j a$ and $a \sim'_j b$, and all other individuals' preferences remain the same. Thus the number of individuals for whom $a \succ_i b$ either remains the same or increases and the number for whom $b \succ_i a$ either remains the same or decreases. In terms of Figures 1.1a and 1.1b, for any problem at position (x, y) and any position (x', y') to the east, southeast, or south of (x, y), some problem at (x', y') is an improvement for a relative to b.

I now define the property of positive responsiveness, which rules out the unattractive feature of the rule shown in Figure 1.1b. For now I define the property only for problems with two alternatives. In this case, a collective choice rule is positively responsive if, when an alternative x goes up in some individuals' rankings, the rule selects $\{x\}$ if it did so originally or if originally it selected both alternatives, in a tie. (In Definition 1.15 I extend the definition to problems with three or more alternatives.)

Definition 1.13: Positively responsive collective choice rule for two alternatives

Let $\langle N, X\rangle$ be a finite society for which X consists of two alternatives, and let D be a set of finite collective choice problems $\langle N, X, \succsim\rangle$. A collective choice rule F for D is *positively responsive* if for every preference profile $\succsim$ for $\langle N, X\rangle$ for which $\langle N, X, \succsim\rangle \in D$ and every improvement $\succsim'$ of $\succsim$ for alternative x relative to the other alternative for which $\langle N, X, \succsim'\rangle \in D$ we have

$$x \in F(N, X, \succsim) \quad \Rightarrow \quad F(N, X, \succsim') = \{x\}.$$

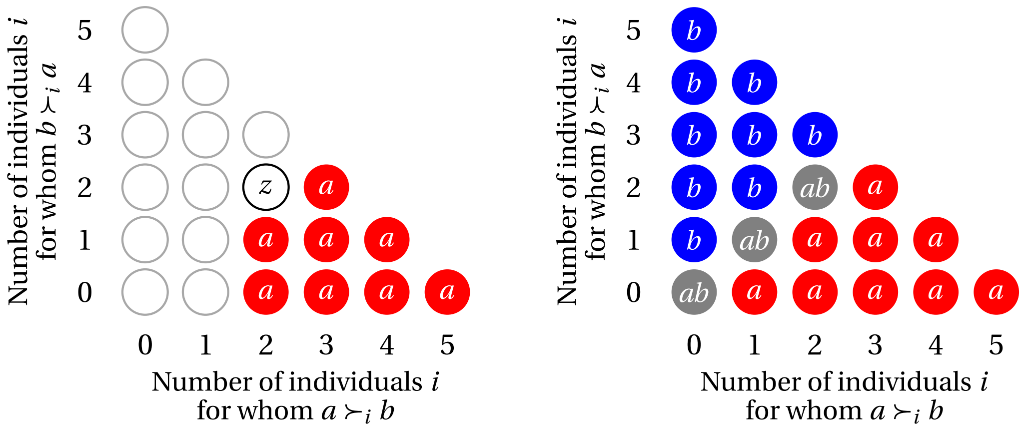
In terms of the diagrams, a rule is positively responsive if for every position assigned to $\{a\}$ or $\{a, b\}$, like the one labeled z in Figure 1.3a, the positions to the east, southeast, and south are assigned to $\{a\}$, and for every problem assigned to $\{b\}$ or $\{a, b\}$, the problems to the west, northwest, and north are assigned to $\{b\}$.

Majority rule, shown in Figure 1.3b, is positively responsive: if a goes up in some individuals' preferences, then either the number of individuals for whom a is a favorite increases or the number of individuals for whom b is a favorite decreases, so that if the rule originally selected a then it still does so, and if it originally selected $\{a, b\}$ then it switches to selecting a.

In fact, majority rule is the only anonymous, neutral, and positively responsive collective choice rule for collective choice problems with two alternatives: the requirements of neutrality (Figure 1.1b) and positive responsiveness (Figure 1.3a) generate Figure 1.3b. This result is known as May's theorem, after its originator, Kenneth O. May (1915–1977).

Proposition 1.1: May's theorem

Let $\langle N, X\rangle$ be a society for which N is finite and X has two members, and let D be (*a*) the set of all collective choice problems $\langle N, X, \succsim\rangle$, (*b*) the set of all such problems in which each individual's preference relation is strict, or (*c*) the set of all such problems with a strict majority winner. A collective



(a) If z is a or ab then for a positively responsive collective choice rule, the outcomes indicated are a.

(b) Majority rule, the only anonymous, neutral, and positively responsive collective choice rule.

Figure 1.3 Anonymous collective choice rules for collective choice problems with five individuals and two alternatives, a and b. (The diagram on the left only partially specifies a rule.) See the text discussing Figure 1.1 for an explanation of the way in which the diagrams represent collective choice rules.

choice rule for D is anonymous, neutral, and positively responsive if and only if it is majority rule.

Proof

I have argued that majority rule is anonymous, neutral, and positively responsive (for any domain).

Now consider a collective choice rule that is anonymous, neutral, and positively responsive on D. I argue that it is majority rule. Given that the rule is anonymous, it may be represented in a diagram like Figure 1.1. (For the domains in cases (b) and (c), the relevant diagrams contain only a subset of the positions in the triangle.) By neutrality, every position on the main diagonal (if any exist for the domain D) is labeled ab. Suppose that some position (x, y) with $x < y$ (so that (x, y) is above the main diagonal) is labeled a. Then by neutrality, (y, x) is labeled b. But each problem associated with (x, y) is an improvement for b relative to a for a problem associated with (y, x), so by positive responsiveness (x, y) is labeled b, a contradiction. Similarly, every position (x, y) with $x > y$ is labeled a. Thus the collective choice rule is majority rule (Figure 1.3b).

Exercise 1.2: Collective choice rules consistent with two of the three conditions

As I note, the rule shown in Figure 1.1b is anonymous and neutral but not positively responsive. Give examples of collective choice rules for problems with two alternatives that are (*a*) anonymous and positively responsive but not neutral and (*b*) neutral and positively responsive but not anonymous.

The condition of positive responsiveness is fairly strong. It requires that any increase in the number of individuals favoring an outcome breaks a tie. Suppose we relax the condition to nonnegative responsiveness: if the number of individuals favoring alternative x increases then either the outcome remains x or changes from a tie to x (as for positive responsiveness) *or remains a tie*. More precisely, the collective choice rule F is *nonnegatively responsive* if for any improvement $\succsim'$ of $\succsim$ for a relative to b we have

$$F(N, X, \succsim) = \{a\} \Rightarrow F(N, X, \succsim') = \{a\} \quad \text{and} \quad F(N, X, \succsim) = \{a, b\} \Rightarrow a \in F(N, X, \succsim')$$

and symmetrically for an improvement for b relative to a. Rules other than majority rule are consistent with anonymity, neutrality, and nonnegative responsiveness.

Exercise 1.3: Nonnegatively responsive collective choice rules

In a diagram like Figure 1.1, characterize the collective choice rules for sets of alternatives with two members that are anonymous, neutral, and nonnegatively responsive (but not necessarily positively responsive).

1.4 Three or more alternatives: Condorcet winners

For a collective choice problem with two alternatives, an alternative is selected by majority rule if it beats or ties the other alternative. For collective choice problems with three or more alternatives, a natural extension of majority rule selects the alternatives that beat or tie every other alternative in pairwise contests. These alternatives are called Condorcet winners, after Marie Jean Antoine Nicolas de Caritat, Marquis of Condorcet (1743–1794).

Definition 1.14: Condorcet winner

Let $\langle N, X, \succsim \rangle$ be a collective choice problem for which N is finite and let $x \in X$ and $y \in X$ be alternatives. Then x *beats* y if the number of individuals $i \in N$ for whom $x \succ_i y$ exceeds the number for whom $y \succ_i x$, and *ties with* y if these two numbers are equal. The alternative x is

- a *Condorcet winner* of $\langle N, X, \succsim \rangle$ if it beats or ties every other alternative
- a *strict Condorcet winner* of $\langle N, X, \succsim \rangle$ if it beats every other alternative.

A problem can have at most one strict Condorcet winner; if it has no strict Condorcet winner then it may have more than one Condorcet winner.

Exercise 1.4: Collective choice problem with no strict Condorcet winner but unique Condorcet winner

Given an example of a collective choice problem with no strict Condorcet winner but a unique Condorcet winner.

Exercise 1.5: Alternative that ties with Condorcet winner

Is an alternative that ties with a Condorcet winner necessarily a Condorcet winner?

For a collective choice problem with two alternatives, the set of Condorcet winners is exactly the set of alternatives selected by majority rule, and if one alternative is preferred to the other by a strict majority of individuals, then that alternative is the strict Condorcet winner.

The collective choice problem in Example 1.1, in which there are three alternatives, has two Condorcet winners, a and b, and no strict Condorcet winner. The next example shows that the strict Condorcet winner may differ from the alternatives selected by plurality rule and the Borda rule.

Example 1.4: Condorcet winner, plurality winner, and Borda winner

For the following collective choice problem, c is the strict Condorcet winner: four individuals prefer it to a, four individuals prefer it to b, and four individuals prefer it to d. By contrast, a is selected by plurality rule and b is selected by the Borda rule.

1	2	3	4	5	6	7
a	a	a	c	c	b	d
c	b	b	b	b	d	c
b	c	d	d	d	c	b
d	d	c	a	a	a	a

Some collective choice problems with three or more alternatives have no Condorcet winner.

Example 1.5: Condorcet cycle

For the following collective choice problem, known as a Condorcet cycle, two individuals prefer a to b, but only one prefers a to c, so that a is not a Condorcet winner; b and c are not Condorcet winners for similar reasons.

1	2	3
a	c	b
b	a	c
c	b	a

Among all logically possible collective choice problems with three alternatives in which all individuals' preferences are strict, the percentage with a Condorcet winner declines as the number of individuals increases; for any number of individuals it is at least 91%. For problems with more alternatives, this bound on the percentage is smaller. For example, it is about 58% for problems with eight alternatives (Gehrlein and Fishburn 1976). In some environments, some preference profiles may be unreasonable. In Section 1.5 I discuss two domains for which every collective choice problem has a Condorcet winner.

For the domain of collective choice problems that have Condorcet winners, we can generalize May's theorem. To do so, we need to first generalize the property of positive responsiveness to many alternatives. One generalization requires that if an alternative x selected for a collective choice problem improves relative to some other alternative y, then (i) x is still selected for the new problem, (ii) y is not selected for the new problem, and (iii) an alternative is selected for the new problem only if it was selected for the original problem.

Definition 1.15: Positively responsive collective choice rule

For any set D of collective choice problems, a collective choice rule F for D is *positively responsive* if, for every collective choice problem $\langle N, X, \succsim\rangle \in D$, every alternative $x \in F(N, X, \succsim)$, every alternative $y \in X$, and every improvement $\succsim'$ of $\succsim$ for x relative to y for which $\langle N, X, \succsim' \rangle \in D$,

1. $x \in F(N, X, \succsim')$
2. $y \notin F(N, X, \succsim')$
3. $F(N, X, \succsim') \subseteq F(N, X, \succsim)$.

Note that if a collective choice rule selects $\{x\}$ for some collective choice problem and an alternative y improves relative to x, then the property of positive responsiveness does not have any implications for the alternatives selected by the rule for the new problem.

For the domain of collective choice problems with a strict Condorcet winner, the collective choice rule that assigns the strict Condorcet winner to each collective choice problem is positively responsive: if a strict Condorcet winner improves in some individuals' preferences then it remains a strict Condorcet winner.

Plurality rule satisfies condition 1 of positive responsiveness: if x is one of the alternatives the rule selects (maybe the only one) and goes up in some individuals' preferences, it remains one of the selected alternatives. But plurality rule does not satisfy condition 2: if x and y are both members of the set it selects and y improves relative to x, then x may still be among the selected alternatives. The reason is that the alternatives plurality rule selects are not affected by the relative ranking of alternatives that are not at the top of an individual's preferences. Consider the following example.

1	2	3
a	b	c
b	a	b
c	c	a

1	2	3
a	b	c
b	a	a
c	c	b

In the problem on the left, plurality rule selects $\{a, b, c\}$ (all three alternatives, in a tie). The preference profile for the problem on the right is an improvement for a relative to b (via individual 3), so that condition 2 of positive responsiveness requires that b not be selected for this problem. Since plurality rule continues to select $\{a, b, c\}$, condition 2 is violated.

Exercise 1.6: Plurality rule with runoff

Plurality rule with runoff is the collective choice rule that selects the alternative that is the favorite of the largest number of individuals if this number is more than half the number of individuals, and otherwise first identifies the two alternatives that are the favorites of the largest number of individuals and then selects the one of these two that beats the other in a two-alternative contest (or both, if they tie). The idea behind the rule is that the result of the first round determines which two alternatives have the most support, and the second round allows everyone to express a preference between these alternatives. Use the following example, in which each column is a preference ordering, to show that this rule does not satisfy condition 1 of positive responsiveness.

6 individuals	5 individuals	4 individuals	2 individuals
a	c	b	b
b	a	c	a
c	b	a	c

That plurality rule is not positively responsive when there are three or more alternatives is not surprising. When there are two alternatives, knowing an individual's favorite alternative tells us the individual's complete preference relation. When there are three or more alternatives, that is no longer the case, and the identity of the individuals' favorite alternatives, the only information used by plurality rule, seems inadequate to make a collective decision that fully reflects the individuals' preferences.

The Borda rule is designed to address this problem: it gives weight to individuals' rankings of alternatives below their favorites, and is indeed positively responsive. If an alternative x improves in an individual's preferences relative to y, then the number of points assigned to x goes up and the number of points assigned to every other alternative either goes down or remains the same. So if x was selected for the original problem it is the only alternative selected for the new problem.

However, when there are three or more alternatives, the Borda rule has an undesirable property. Consider the following collective choice problem.

1	2	3	4	5
b	b	b	a	a
a	a	a	c	c
c	c	c	b	b

The Borda rule selects a for this problem. (It gets 7 points; b gets 6 points, and c gets 2 points.) Now suppose that c is no longer available, so that we have the following problem.

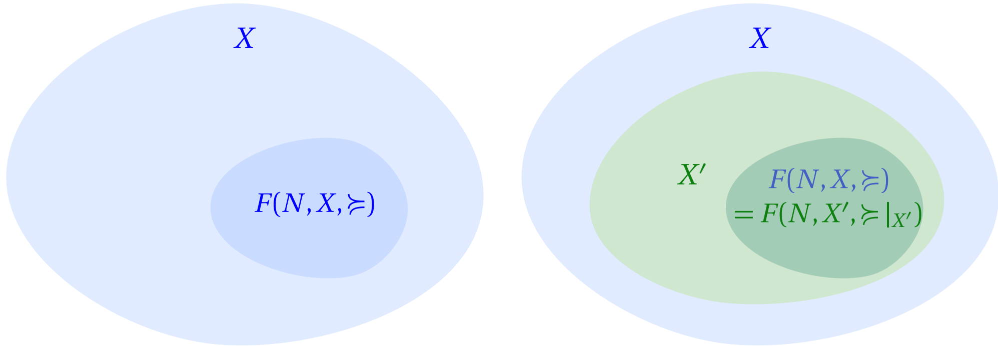


Figure 1.4 An illustration of Nash independence. In the problem on the right, X' is a subset of X that contains $F(N,X,\succsim)$. Nash independence requires that the set $F(N,X',\succsim|_{X'})$ of alternatives selected for $\langle N,X',\succsim|_{X'}\rangle$ is equal to the set $F(N,X,\succsim)$ selected for $\langle N,X,\succsim\rangle$.

1	2	3	4	5
b	b	b	a	a
a	a	a	b	b

Originally, c was not chosen, so one might expect that its unavailability should not affect the alternative selected: a should remain the selected alternative. However, the Borda rule selects b for the new problem. Thus the fact that c, which was not selected initially, becomes unavailable, causes b to be chosen rather than a.

To restrict to collective choice rules that do not behave in this way we can require, in addition to anonymity, neutrality, and positive responsiveness, that removing unchosen alternatives from the set of available alternatives does not affect the set of alternatives selected. More precisely, if, when the set of available alternatives is X, the set of alternatives selected is Y, then when the set of available alternatives is a subset of X that includes Y, the set of alternatives selected remains Y. This condition is illustrated in Figure 1.4. It is a version of one proposed by Nash (1950) and is named after him. In the following definition, $\succsim|_{X'}$ is the restriction of the preference profile $\succsim$ to the set X': for a, $b \in X'$, $a \succsim|_{X'} b$ if and only if $a \succsim b$.

Definition 1.16: Nash independence

For any set D of collective choice problems, the collective choice rule F for D is *Nash independent* if for every collective choice problem $\langle N,X,\succsim\rangle \in D$ and every set $X' \subset X$ for which $\langle N,X',\succsim|_{X'}\rangle \in D$ and $F(N,X,\succsim) \subseteq X'$, we have $F(N,X',\succsim|_{X'}) = F(N,X,\succsim)$.

On the domain of problems that have a strict Condorcet winner, the rule that selects the strict Condorcet winner satisfies this property: if an alternative beats every other in pairwise contests then it continues to do so if some of these alternatives are eliminated.

Thus for the domain of collective choice problems that have a strict Condorcet winner, among the examples of collective choice rules I have described only the one that selects the strict Condorcet winner is anonymous, neutral, positively responsiveness, and Nash independent. In fact, if every individual's preference relation is strict, it is the only rule that satisfies these properties among *all* possible rules.

Definition 1.17: Strict Condorcet domain

For any finite sets N and A, the *strict Condorcet domain for* (N,A), denoted $C(N,A)$, is the set of all collective choice problems $\langle N,X,\succsim\rangle$ for which (*i*) $X \subseteq A$, (*ii*) $\succsim_i$ is strict for all $i \in N$, and (*iii*) $\langle N,X,\succsim\rangle$ has a strict Condorcet winner.

Proposition 1.2: Generalization of May's theorem to many alternatives

For any finite sets N and A, a collective choice rule F for the strict Condorcet domain $C(N,A)$ is anonymous, neutral, positively responsive, and Nash independent if and only if, for every collective choice problem $\langle N,X,\succsim\rangle \in C(N,A)$, $F(N,X,\succsim)$ contains only the strict Condorcet winner of $\langle N,X,\succsim\rangle$.

Proof

I have argued that the collective choice rule for $C(N,A)$ that selects the strict Condorcet winner satisfies the four properties.

Now let F be a collective choice rule for $C(N,A)$ that is anonymous, neutral, positively responsive, and Nash independent. If A contains two alternatives then by May's theorem $F(N,X,\succsim) = \{c\}$, where c is the strict Condorcet winner of $\langle N,X,\succsim\rangle$.

Now suppose that A contains three or more alternatives.

Step 1 *Let $\langle N,X,\succsim\rangle$ be a collective choice problem in $C(N,A)$ and let c be its strict Condorcet winner. If $F(N,X,\succsim)$ contains an alternative different from c then it contains at least two alternatives different from c.*

Proof. Suppose, to the contrary, that $F(N,X,\succsim) = \{x,c\}$ or $F(N,X,\succsim) = \{x\}$ for some $x \in X \setminus \{c\}$. The alternative c is a strict Condorcet winner of $\langle N,\{x,c\},\succsim|_{\{x,c\}}\rangle$, so this problem is in $C(N,A)$, and $F(N,X,\succsim) \subseteq$

$\{x,c\}$, so by Nash independence $F(N,\{x,c\},\succsim|_{\{x,c\}}) = F(N,X,\succsim)$ and hence $F(N,\{x,c\},\succsim|_{\{x,c\}}) = \{x,c\}$ or $F(N,\{x,c\},\succsim|_{\{x,c\}}) = \{x\}$. Now, given that F is anonymous, neutral, and positively responsive, in particular it satisfies these properties for two-alternative problems in $C(N,A)$. Thus by May's theorem $F(N,\{x,c\},\succsim|_{\{x,c\}}) = \{c\}$, a contradiction. ◁

Step 2 *If for some problem* $\langle N,X,\succsim\rangle \in C(N,A)$ *the set* $F(N,X,\succsim)$ *has* k *members with* $k \geq 2$*, there exists a problem* $\langle N,X,\succsim'\rangle \in C(N,A)$ *for which* $F(N,X,\succsim')$ *has fewer than* k *members, including at least one different from its strict Condorcet winner.*

Proof. Let $\langle N,X,\succsim\rangle \in C(N,A)$, let c be its strict Condorcet winner, and let $F(N,X,\succsim) = \{x_1,x_2,\ldots,x_k\}$ with $k \geq 2$. By Step 1, at least two members of $F(N,X,\succsim)$ differ from c. Suppose that $x_1 \neq c$ and $x_2 \neq c$. A strict majority of individuals prefer c to x_1 and a strict majority prefer c to x_2, so at least one individual prefers c to both x_1 and x_2. Suppose that $c \succ_i x_1 \succ_i x_2$. Modify i's preference relation by raising x_2 to come between x_1 and c; keep the preference relation of every other individual the same. Denote the new preference profile by $\succsim'$. The alternative c is the strict Condorcet winner for $\langle N,X,\succsim'\rangle$, so in particular this problem has a strict Condorcet winner and hence is in $C(N,A)$. Thus by positive responsiveness $F(N,X,\succsim')$ is a subset of $\{x_1,x_2,\ldots,x_k\}$ that contains x_2 but not x_1. Hence it contains fewer alternatives than does $F(N,X,\succsim)$, among them an alternative different from c. ◁

An implication of Step 2 is that for some problem $\langle N,X,\succsim\rangle$ the set $F(N,X,\succsim)$ contains a single alternative different from c, contradicting Step 1.

Is any collective choice rule anonymous, neutral, positively responsive, and Nash independent on a domain larger than the strict Condorcet domain? For domains that are more than slightly larger than the strict Condorcet domain, the answer is no. Define the Condorcet-plus domain to consist of all problems with a strict Condorcet winner *plus* all problems for which some improvement for a single individual generates a problem that has a strict Condorcet winner. (This domain, unlike the strict Condorcet domain, contains problems for which the preference relations are not strict.)

Definition 1.18: Condorcet-plus domain

For any finite sets N and A, the *Condorcet-plus domain for* (N,A), denoted $C^{+}(N,A)$, is the set of collective choice problems $\langle N,X,\succcurlyeq\rangle$ such that $X\subseteq A$ and either (*i*) $\langle N,X,\succcurlyeq\rangle$ has a strict Condorcet winner or (*ii*) there is a preference profile $\succcurlyeq'$ for $\langle N,X\rangle$ that differs from $\succcurlyeq$ only in the preference relation of one individual and is an improvement of $\succcurlyeq$ such that the problem $\langle N,X,\succcurlyeq'\rangle$ has a strict Condorcet winner.

If A contains two alternatives then for any finite set N the Condorcet-plus domain for (N,A) contains all collective choice problems with two alternatives: for any two-alternative problem that does not have a strict Condorcet winner, the alternatives tie, and an improvement for any individual breaks this tie.

If A is a finite set containing three or more alternatives then for any finite set N the Condorcet-plus domain for (N,A) contains the Condorcet cycle in Example 1.5: if a is raised above c in individual 2's preferences then it becomes a strict Condorcet winner. But some three-alternative problems are not in $C^{+}(N,A)$. An example is the following problem, which does not have a strict Condorcet winner and for which no improvement for a single individual produces a strict Condorcet winner.

1	2	3	4	5	6
a	a	c	c	b	b
b	b	a	a	c	c
c	c	b	b	a	a

Each alternative beats one of the other alternatives four to two and loses to the remaining alternative two to four. If an alternative improves in any one individual's preferences, it improves its performance against the alternative that previously beat it only to a tie, so the new problem, like the old one, has no strict Condorcet winner.

I now show that if N and A are finite sets each containing at least three elements, then no collective choice rule is anonymous, neutral, positively responsive, and Nash independent on $C^{+}(N,A)$ or any domain that contains $C^{+}(N,A)$.

Proposition 1.3: No rule is anonymous, neutral, positively responsive, and Nash independent on Condorcet-plus domain

Let N and A be finite sets, each containing at least three elements. No collective choice rule for any set of finite collective choice problems that contains the Condorcet-plus domain $C^{+}(N,A)$ is anonymous, neutral, positively responsive, and Nash independent.

Proof

Let F be an anonymous, neutral, positively responsive, and Nash independent collective choice rule for a set of collective choice problems that includes $C^+(N,A)$. Let $N = \{1,2,\ldots,n\}$ and $X = \{a,b,c\}$ and suppose that the preferences of individuals 1, 2, and 3 form the Condorcet cycle in Example 1.5 while the other individuals are indifferent between a, b, and c:

1	2	3	4	$\cdots$	n
a	c	b	abc	$\cdots$	abc
b	a	c			
c	b	a			

This problem, $\langle N,X,\succsim\rangle$, is in $C^+(N,A)$ because raising b above a in individual 1's preferences makes it a strict Condorcet winner.

I argue that the fact that F satisfies neutrality and anonymity implies that $F(N,X\succsim) = \{a,b,c\}$. Consider the permutation σ on X given by $\sigma(a) = c$, $\sigma(b) = a$, and $\sigma(c) = b$. The problem $\langle N,X,\succsim^\sigma\rangle$, where $\succsim^\sigma$ is the preference profile derived from $\succsim$ as given in Definition 1.10, is

1	2	3	4	$\cdots$	n
c	b	a	abc	$\cdots$	abc
a	c	b			
b	a	c			

This problem is in $C^+(N,A)$ because raising a above c in individual 1's preferences makes it a strict Condorcet winner. So by neutrality, (*i*) $F(N,X,\succsim^\sigma)$ consists of the alternatives $\sigma(z)$ for each $z\in F(N,X,\succsim)$. But $\succsim^\sigma$ is the permutation of $\succsim$ obtained by mapping individual 1 into 3, 2 into 1, and 3 into 2. Thus by anonymity (*ii*) $F(N,X,\succsim^\sigma) = F(N,X,\succsim)$. Conditions (*i*) and (*ii*) are satisfied if and only if $F(N,X,\succsim) = \{a,b,c\}$.

Now starting with $\langle N,X,\succsim\rangle$ raise b to be indifferent with a in individual 1's preferences, generating the problem

1	2	3	4	$\cdots$	n
ab	c	b	abc	$\cdots$	abc
c	a	c			
	b	a			

Denote this problem $\langle N,X,\succsim'\rangle$. It is in $C^+(N,A)$ because raising b above a in individual 1's preferences makes b a strict Condorcet winner, so that by positive responsiveness, $b\in F(N,X,\succsim')$ and $a\notin F(N,X,\succsim')$.

If $F(N,X,\succsim') = \{b\}$, remove c from the set of alternatives, to generate the problem

1	2	3	4	$\cdots$	n
ab	a	b	ab	$\cdots$	ab
	b	a			

This problem is in $C^{+}(N,A)$ because raising b above a in individual 1's preferences makes b a strict Condorcet winner, so by Nash independence F assigns $\{b\}$ to it, contradicting May's theorem, which assigns $\{a,b\}$ to it.

If $F(N,X,\succsim') = \{b,c\}$, remove a from the set of alternatives, to generate the problem

1	2	3	4	$\cdots$	n
b	c	b	bc	$\cdots$	bc
c	b	c			

This problem is in $C^{+}(N,A)$ because b is a strict Condorcet winner, so by Nash independence F assigns $\{b,c\}$ to it, contradicting May's theorem, which assigns $\{b\}$ to it.

Exercise 1.7: Nash independent rules consistent with two of the three other conditions

For each pair of the conditions anonymity, neutrality, and positive responsiveness, find a collective choice rule that satisfies that pair of conditions and is Nash independent for the domain of all collective choice problems.

Proposition 1.2 shows that the collective choice rule that selects the strict Condorcet winner is the only rule for the domain of problems that have a strict Condorcet winner that satisfies the properties of anonymity, neutrality, positive responsiveness, and Nash independence. Proposition 1.3 shows that this result cannot be extended much beyond this domain. For any domain that includes problems for which no improvement in any individual's preferences generates a problem with a strict Condorcet winner, no collective choice rule satisfies the four properties.

1.5 Two domains with Condorcet winners

The results in the previous section show that for a domain consisting of collective choice problems with strict Condorcet winners, the collective choice rule that selects the strict Condorcet winner has attractive properties. In this section I discuss two such domains.

1.5.1 Single-peaked preferences

Suppose that for some ordering of the alternatives, individual i has a favorite alternative, say a_i^*, and prefers a_i^* to b to c whenever $c < b < a_i^*$ or $a_i^* < b < c$. We say that her preference relation is single-peaked with respect to the ordering.

Definition 1.19: Single-peaked preference relation

Let X be a set (of alternatives) and let $\trianglerighteq$ be a linear order on X. A preference relation $\succsim_i$ over X is *single-peaked with respect to* $\trianglerighteq$ if it has a single favorite alternative, say a^*, and

$$c \lhd b \lhd a^* \quad \text{or} \quad a^* \lhd b \lhd c \quad \Rightarrow \quad a^* \succ_i b \succ_i c. \tag{1.3}$$

For any strict preference relation, there are linear orders of the alternatives with respect to which the preference relation is single-peaked. (One such order, for example, arranges the alternatives from most preferred to least preferred.) If a preference profile has the property that every individual's preference relation is single-peaked with respect to the *same* linear order of the alternatives, we say that the profile is single-peaked (with respect to the order). Assuming that a preference profile has this property may be reasonable, for example, if the alternatives are the amounts of money society spends on a certain endeavor: some people like low spending better than moderate spending better than high spending, some people have the reverse preference, and some people like moderate spending best, but no one likes both low spending and high spending better than moderate spending.

Definition 1.20: Single-peaked preference profile

Let $\langle N, X\rangle$ be a society and let $\trianglerighteq$ be a linear order on X. A preference profile $\succsim$ for $\langle N, X\rangle$ is *single-peaked with respect to* $\trianglerighteq$ if every preference relation $\succsim_i$ for $i \in N$ is single-peaked with respect to $\trianglerighteq$. A collective choice problem $\langle N, X, \succsim\rangle$ has *single-peaked preferences* if for some linear order $\trianglerighteq$ on X the preference profile $\succsim$ is single-peaked with respect to $\trianglerighteq$.

The name single-peaked comes from the shape of the payoff functions that represent the preferences. An example is shown in Figure 1.5. In this diagram, there is a small colored disk for each alternative and each individual, and the fact that the y-coordinate of the disk for some alternative x of a certain color is larger than the y-coordinate of the disk for another alternative x' of the same

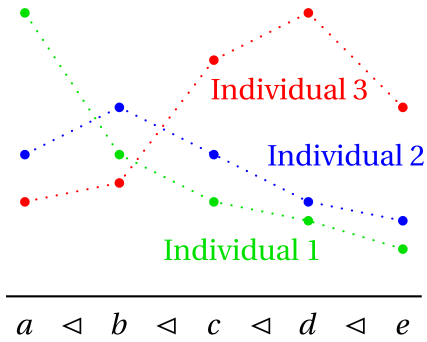


Figure 1.5 A single-peaked preference profile.

color means that the individual associated with the color prefers x to x'. (The magnitude of the difference between the y-coordinates has no significance.) For example, individual 3's preference relation $\succsim_3$ is given by $d \succ_3 c \succ_3 e \succ_3 b \succ_3 a$. Note that each individual's preference relation in a single-peaked preference profile is strict on each side of the individual's favorite alternative, but the individual may be indifferent between alternatives on different sides of her favorite alternative. In the preference profile in Figure 1.5, for example, individual 2 is indifferent between a and c.

The next result involves the median of the individuals' favorite alternatives. An odd number of ordered alternatives has a single median, namely the middle alternative in the order; an even number of ordered alternatives has two medians, namely the two middle alternatives.

Definition 1.21: Median of finite collection

Let X be a set and $\trianglerighteq$ a linear order on X. Suppose that $x_i \in X$ for $i = 1, \dots, n$, where n is a positive integer, with $x_1 \trianglerighteq x_2 \trianglerighteq \cdots \trianglerighteq x_n$. If n is odd then the *median of* $(x_1, \dots, x_n)$ *with respect to* $\trianglerighteq$ is x_k for $k = \frac{1}{2}(n+1)$, and if n is even then the *medians of* $(x_1, \dots, x_n)$ *with respect to* $\trianglerighteq$ are x_l and x_{l+1} for $l = \frac{1}{2}n$.

For example, for the preferences illustrated in Figure 1.5, for which the individuals' favorite alternatives are a, b, and d, we have $a \lhd b \lhd d$, so that b is the median of the favorite alternatives with respect to $\trianglerighteq$.

I now show that for a collective choice problem with single-peaked preferences, an alternative is a median of the individuals' favorite alternatives if and only if it is a Condorcet winner, and if the number of individuals is odd then the unique median favorite alternative is the strict Condorcet winner. The result has the name "median voter theorem" although the median is defined with respect to an ordering of the alternatives, not the individuals (voters).

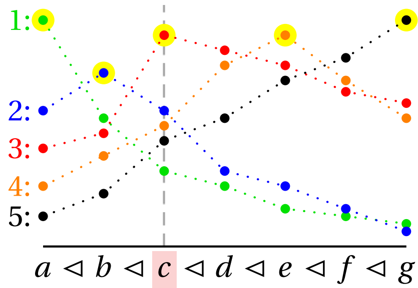


Figure 1.6 An illustration of Proposition 1.4 for a collective choice problem with five individuals. The median of the individuals' favorite positions is c.

Proposition 1.4: Median voter theorem for single-peaked preferences

Consider a collective choice problem in which the number of individuals is finite. If the problem has single-peaked preferences with respect to a linear order $\trianglerighteq$, an alternative is a Condorcet winner if and only if it is a median of the individuals' favorite alternatives with respect to $\trianglerighteq$. If the number of individuals is odd, the unique median is the strict Condorcet winner.

Proof

Denote by n the number of individuals and let m be a median of the individuals' favorite alternatives with respect to the ordering $\trianglerighteq$. (In Figure 1.6, for example, the individuals' favorite alternatives are a, b, c, e, and g (indicated in yellow), so that $m = c$.)

For every individual i, denote i's favorite alternative by x_i^*. If $x \triangleright m$ then for all individuals i for whom $m \trianglerighteq x_i^*$ we have $m \succ_i x$ by the single-peakedness of preferences. The number of such individuals is $\frac{1}{2}(n+1)$ if n is odd and at least $\frac{1}{2}n$ if n is even, because m is a median of the favorite alternatives. Similarly, if $m \triangleright x$ then for all i with $x_i^* \trianglerighteq m$ we have $m \succ_i x$, and the number of such individuals is $\frac{1}{2}(n+1)$ if n is odd and at least $\frac{1}{2}n$ if n is even. Thus m is a Condorcet winner of the collective choice problem, and is a strict Condorcet winner if n is odd.

If an alternative x is not a median of the individuals' favorite alternatives, it is beaten by any median, and hence is not a Condorcet winner.

Exercise 1.8: Median rule and Borda winner

Give an example to show that for a single-peaked preference profile the median favorite alternative may differ from the alternative selected by the

Borda rule.

In some models, the assumption that the individuals' preferences are single-peaked is unreasonable, but a weaker version of this condition is acceptable. One such version allows each individual's preferences to have a single plateau, rather than a single peak, with strict preferences on each side of the plateau. Precisely, for some linear order $\trianglerighteq$ on the set of alternatives and, for each individual $i \in N$, alternatives $\underline{a}^i$ and $\overline{a}^i$ (not necessarily distinct), we have $a \prec_i b$ if $a \lhd b \unlhd \underline{a}^i$, $a \sim_i b$ if $\underline{a}^i \unlhd a \lhd b \unlhd \overline{a}^i$, and $a \succ_i b$ if $\overline{a}^i \lhd a \lhd b$. Denote the number of individuals by n and let $z_1, z_2, \ldots, z_{2n}$ be an ordering of $\{\underline{a}^1, \underline{a}^2, \ldots, \underline{a}^n, \overline{a}^1, \overline{a}^2, \ldots, \overline{a}^n\}$ such that $z_1 \unlhd z_2 \unlhd \cdots \unlhd z_{2n}$. Fishburn (1973, Theorem 9.3) shows that in this case an alternative x is a Condorcet winner of the collective choice problem $\langle N, X, \succsim \rangle$ if and only if $z_n \unlhd x \unlhd z_{n+1}$. The next exercise asks you to find an example in which each individual's preferences have a single plateau and the median of some collection of the individuals' favorite alternatives does not satisfy this condition, and hence is not a Condorcet winner.

Exercise 1.9: Median of collection of favorite alternatives and Condorcet winner for single-plateau preferences

Give an example of a collective choice problem $\langle N, X, \succsim \rangle$ with an odd number of individuals in which each individual's preferences have a single plateau rather than a single peak (as defined in the previous paragraph) and the median of some collection $(a_i^*)_{i \in N}$ of alternatives for which each a_i^* is a favorite alternative of individual i differs from the unique Condorcet winner.

Now suppose that each individual has a single favorite alternative but may be indifferent among alternatives on the same side of her favorite. Then the (unique) median of the individuals' favorite alternatives may differ from the Condorcet winner, as you are asked to demonstrate in the next exercise.

Exercise 1.10: Median of favorite alternatives and Condorcet winner for variant of single-peaked preferences with indifference

Consider a collective choice problem $\langle N, X, \succsim \rangle$ with an odd number of individuals in which each individual has a single favorite alternative and, for some linear order $\trianglerighteq$ of the alternatives, each individual's preferences satisfy the variant of (1.3) in which the right-hand side is $a^* \succsim_i b \succsim_i c$. Give an example of such a problem that does not have a Condorcet winner and also an example with a unique Condorcet winner that differs from the median of the individuals' favorite alternatives.

The next exercise leads you through an alternative, inductive, proof of the median voter theorem (Proposition 1.4) for a finite collective choice problem. A variant of this proof establishes an elegant generalization of that result, discussed after the exercise.

Exercise 1.11: Another proof of existence of Condorcet winner when preferences are single-peaked

Let $\langle N, X, \succsim \rangle$ be a finite collective choice problem with an odd number of individuals that has single-peaked preferences with respect to a linear order $\trianglerighteq$. Denote the alternatives arranged according to this order by $(x_1, x_2, \ldots, x_k)$ and for any $t \in \{1, \ldots, k\}$ let $X_t = \{x_1, \ldots, x_{t-1}\}$ and $Z_t = \{x_t, \ldots, x_k\}$. (*a*) Show that if for some $t \in \{1, \ldots, k\}$ (*i*) no alternative $x_j \in X_t$ is the favorite alternative in Z_j of a majority of individuals and (*ii*) x_t is the favorite alternative in Z_t of a majority of individuals, then x_t is the strict Condorcet winner of $\langle N, X, \succsim \rangle$. (*b*) Use this result to give an inductive proof of the existence of a strict Condorcet winner of $\langle N, X, \succsim \rangle$ in which at each step the first remaining alternative according to $\trianglerighteq$ is selected if it is the favorite among the remaining alternatives of a majority of individuals and otherwise is removed from the set of alternatives.

A tree is a connected graph with no cycles. Define a preference profile to be single-peaked on a tree if for some tree each individual's preferences are single-peaked on every path through the tree. The proof in Exercise 1.11 may be adapted to show that every finite collective choice problem in which the preference profile is single-peaked on a tree has a strict Condorcet winner. At each step in the inductive argument, one of the terminal nodes of the tree is either selected or removed. An example of a collective choice problem with preferences that are single-peaked on a tree but not on a line is given in Figure 1.7. These preferences are not single-peaked in the sense of Definition 1.20 because the individuals' worst alternatives (d, c, and b) number more than two. But they are single-peaked on all the paths in the tree on the right of Figure 1.7, and a is the strict Condorcet winner of the problem.

1.5.2 *Single-crossing preferences*

The single-crossing property is another condition on preference profiles that is sufficient for the existence of a Condorcet winner.

Suppose that the individuals can be ordered in such a way that for any alter-

1	2	3
a	a	a
b	d	c
c	b	d
d	c	b

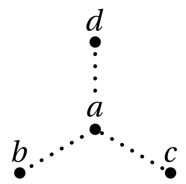


Figure 1.7 A collective choice problem (left) with preferences that are single-peaked on the tree on the right, but are not single-peaked on a line.

natives x and y, if a median individual in the ordering likes x at least as much as y then either all individuals earlier in the ordering or all individuals later in the ordering prefer x to y. Then if a median individual prefers x to y, a majority of individuals prefer x to y, so that in particular if a median individual has a unique favorite alternative then that alternative is a Condorcet winner. Further, if the number of individuals is odd and the (unique) median individual has a unique favorite alternative, then that alternative is the strict Condorcet winner. In addition, under these conditions the preferences of the median individual coincide with those of the majority in the sense that for any alternatives x and y the median individual prefers x to y if and only if a majority of individuals do so. That is, the median individual's preference relation coincides with the majority relation, defined as follows.

Definition 1.22: Majority relation

Let $\langle N, X, \succsim \rangle$ be a finite collective choice problem. The binary relation $\trianglerighteq$ on X defined by

> $x \trianglerighteq y$ if and only if the number of individuals i for whom $x \succ_i y$ is at least the number for whom $y \succ_i x$

is the *majority relation for* $\langle N, X, \succsim \rangle$.

A preference profile has the single-crossing property if the individuals can be ordered in such a way that the preferences of every individual, not only those of the median individual, satisfy the condition in the previous paragraph. That is, there is a linear order $\geq$ of the individuals such that for every individual i and all alternatives x and y, if $x \succsim_i y$ then either (a) $x \succ_j y$ for every individual $j < i$ or (b) $x \succ_j y$ for every individual $j > i$.

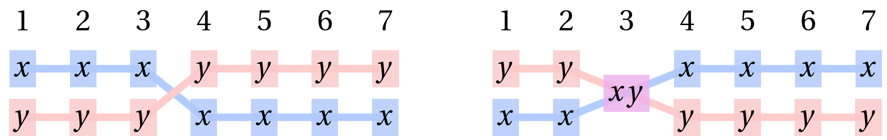


Figure 1.8 An illustration of the individuals' preferences between the alternatives x and y for two collective choice problems with single-crossing preferences and seven individuals. Each column in each diagram indicates the preference between x and y for one individual, with the alternative that the individual prefers at the top. In the diagram on the right, individual 3 is indifferent between x and y.

Definition 1.23: Single-crossing preferences

Let $\langle N, X, \succsim \rangle$ be a collective choice problem for which N is finite and let $\geq$ be a linear order on N. The problem $\langle N, X, \succsim \rangle$ has *single-crossing preferences with respect to* $\geq$ if for every $i \in N$, $x \in X$, and $y \in X \setminus \{x\}$,

$$x \succsim_i y \quad \Rightarrow \quad x \succ_j y \text{ either } (a) \text{ for all } j < i \text{ or } (b) \text{ for all } j > i.$$

The problem has *single-crossing preferences* if this condition is satisfied for some linear order $\geq$ on N.

If a collective choice problem has single-crossing preferences and $x \succsim_i y$ for some individual i then either $x \succ_j y$ for every individual j or there is a unique individual i^* such that $x \succsim_{i^*} y$ and either (*a*) $x \succ_i y$ for all $i < i^*$ and $x \prec_i y$ for all $i > i^*$ or (*b*) $x \prec_i y$ for all $i < i^*$ and $x \succ_i y$ for all $i > i^*$. Figure 1.8 shows two examples and in doing so motivates the term "single-crossing". This implication of Definition 1.23 leads to an equivalent definition of single-crossing that involves a linear order of the alternatives in addition to a linear order of the individuals.

Exercise 1.12: Alternative definition of single-crossing

Show that a collective choice problem $\langle N, X, \succsim \rangle$ has single-crossing preferences if and only if there is a linear order $\geq$ of the individuals and a linear order $\trianglerighteq$ of the alternatives such that whenever $x \succsim_i y$, (*a*) $x \triangleleft y \Rightarrow x \succ_j y$ for all $j < i$ and (*b*) $y \triangleleft x \Rightarrow x \succ_j y$ for all $j > i$.

That individuals' preferences satisfy the single-crossing property is plausible in some economic models in which individuals differ in a characteristic like their earning-power or their degree of risk aversion or altruism, by which they can be ordered. Section 11.3.2 contains an example.

The next exercise concerns the relation between single-peaked and single-crossing preferences: (*a*) single-peaked preferences may not be single-crossing,

(b) if every individual's preferences are strict and every alternative is some individual's favorite, single-crossing preferences are single-peaked, and (c) otherwise they may not be.

Exercise 1.13: Single-crossing and single-peakedness

a. Show that the collective choice problem $\langle\{1,2,3\},\{a,b,c,d\},\succsim\rangle$ for which $a \succ_1 b \succ_1 c \succ_1 d$, $b \succ_2 c \succ_2 d \succ_2 a$, and $c \succ_3 b \succ_3 a \succ_3 d$ has single-peaked but not single-crossing preferences.

b. Suppose that a collective choice problem with a finite number of individuals in which every individual's preference relation is strict has single-crossing preferences with respect to a linear order $\geq$, and every alternative is the favorite of some individual. Show that the problem has single-peaked preferences with respect to the ordering of the alternatives given by the preference relation of the first individual according to $\geq$.

c. Specify a collective choice problem that has single-crossing preferences but does not satisfy the conditions in (b) and does not have single-peaked preferences. (Such a problem exists with three individuals and three alternatives.)

Here is a precise version of the claims at the start of this section.

Proposition 1.5: Median voter theorem for single-crossing preferences

Consider a collective choice problem with a finite number of individuals that has single-crossing preferences with respect to a linear order $\geq$. Suppose that each median individual with respect to $\geq$ has a unique favorite alternative.

a. The favorite alternative of each median individual is a Condorcet winner of the problem, and if the number of individuals is odd, the favorite alternative of the (unique) median individual is the strict Condorcet winner of the problem.

b. If the number of individuals is odd, the preference relation of the (unique) median individual coincides with the majority relation.

Proof

Denote the collective choice problem by $\langle N, X, \succsim\rangle$ and the number of individuals (members of N) by n. Let i be a median individual according to $\geq$ and let x^* be her favorite alternative.

a. For every alternative $x \in X \setminus \{x^*\}$ we have $x^* \succ_i x$, and hence by single-crossing either $x^* \succ_j x$ for all $j < i$ or $x^* \succ_j x$ for all $j > i$. Thus $x^* \succ_j x$ for at least $\frac{1}{2}(n+1)$ individuals j if n is odd and for at least $\frac{1}{2}n$ individuals j if n is even. Hence x^* is a Condorcet winner, and a strict one if n is odd.

b. If i prefers x to y, then by single-crossing either all individuals $j < i$ or all individuals $j > i$ prefer x to y, so that a majority of individuals prefer x to y. If i is indifferent between x and y, then by single-crossing either all individuals $j < i$ prefer x to y and all individuals $j > i$ prefer y to x, or vice versa, so that the number of individuals who prefer x to y is equal to the number who prefer y to x.

For a problem for which the number of individuals is even, note that, unlike the companion result Proposition 1.4 for single-peaked preferences, this result claims only that the favorite alternative of a median individual is a Condorcet winner, not the converse. Here is an example that shows that the converse is false. There are three alternatives, a, b, and c, and four individuals. Two individuals prefer a to c to b and the other two prefer b to c to a. This problem has single-crossing preferences with respect to any ordering of the individuals in which the two individuals who prefer a to b to c are either first and second or third and fourth. The alternative c is a Condorcet winner but for no ordering of the individuals is it the favorite alternative of either median individual.

The proof of part a of the result applies the single-crossing condition only to the case in which i is the median individual and x is her favorite alternative. Thus, as the argument at the start of this section suggests, the assumption in the result that the problem has single-crossing preferences may be replaced with the assumption that for some linear order $\geq$ on N and all $y \in X \setminus \{x^*\}$ we have

$$x^* \succsim_{i^*} y \quad \Rightarrow \quad \text{either } (a)\ x^* \succ_j y \text{ for all } j < i^* \text{ or } (b)\ x^* \succ_j y \text{ for all } j > i^*$$

where i^* is the median individual according to $\geq$ and x^* is her favorite alternative.

The property in part b of the result does not hold for problems with single-peaked preferences: there are such problems in which no individual's preference relation coincides with the majority relation.

Exercise 1.14: Single-peaked preferences and no individual with preferences of majority

Give an example of a collective choice problem with single-peaked preferences in which for no individual i is it the case that for all alternatives x and y, $x \succcurlyeq_i y$ if and only if $x \succcurlyeq_j y$ for a majority of individuals j. (An example exists with three individuals and four alternatives.)

If the assumption in Proposition 1.5 that each median individual has a unique favorite alternative is removed, each favorite alternative of the median individual is a Condorcet winner (but not necessarily a strict one) if the number of individuals is odd but a favorite alternative of a median individual may not be a Condorcet winner if the number of individuals is even, as you are asked to show in the next exercise.

Exercise 1.15: Single-crossing preferences and individual with multiple favorite alternatives

Show that for a collective choice problem with single-crossing preferences and an even number of individuals, a favorite alternative of a median individual with multiple favorite alternatives may not be a Condorcet winner.

Exercise 1.16: Condorcet winner exists but preferences not single-peaked or single-crossing

Give an example of a collective choice problem that has a strict Condorcet winner but for which the preference profile is not single-peaked and does not satisfy the single-crossing property.

Propositions 1.4 and 1.5 and the results in Exercises 1.13 and 1.16 imply that the sets of collective choice problems that have a Condorcet winner, have single-peaked preferences, and have single-crossing preferences are related in the way indicated in Figure 1.9.

1.6 Condorcet winners for two-dimensional sets of alternatives

For a collective choice problem in which the set of alternatives is a subset of a a one-dimensional space, the assumptions of single-peaked or single-crossing preferences, which require the alternatives to be linearly ordered, may be appropriate, and if they are then the collective choice problem has a Condorcet winner by Proposition 1.4 or 1.5. For a problem in which the set of alternatives is a sub-

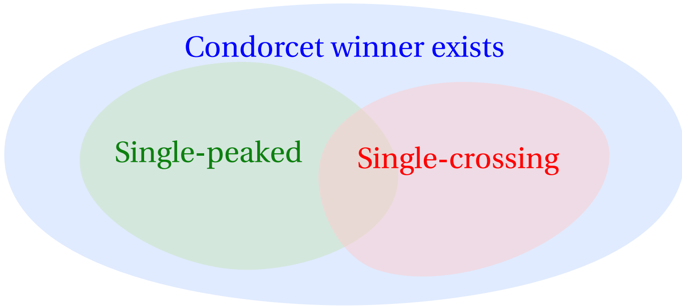


Figure 1.9 The structure of the set of collective choice problems that have a Condorcet winner.

set of a space with two or more dimensions, these assumptions seem less likely to be appropriate. In this section I consider conditions unrelated to the conditions of single-peaked or single-crossing preferences under which a collective choice problem for which the set of alternatives is $\mathbb{R}^2$, the set of pairs of real numbers, has a Condorcet winner. As in the previous section, I assume that the set of individuals is finite.

1.6.1 City block preferences

In a city with streets that form a grid with square blocks, you have a favorite location and you evaluate other locations according to their walking distance from that location. Suppose that streets that run north–south are called avenues and ones that run east–west are called streets. Then if, for example, your favorite location is (5th Avenue, 10th Street), you are indifferent between (3rd Avenue, 9th Street) and (2nd Avenue, 10th Street): both of these locations are three blocks away. We say you have city block preferences. The following definition does not restrict the alternatives to be located on a grid: the set of alternatives is $\mathbb{R}^2$.

Definition 1.24: City block preferences

A preference relation $\succsim_i$ on $\mathbb{R}^2$ reflects *city block preferences* if for some alternative $(x_{i1}^*, x_{i2}^*) \in \mathbb{R}^2$ it is represented by the payoff function u_i defined by

$$u_i(x_1, x_2) = -|x_1 - x_{i1}^*| - |x_2 - x_{i2}^*| \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2.$$

The alternative x_i^* in this definition is the individual's favorite alternative, and each of her indifference sets has the form

$$\{(x_1, x_2) \in \mathbb{R}^2 : |x_1 - x_{i1}^*| + |x_2 - x_{i2}^*| = c\}$$

for some number c, a square with center x_i^* whose diagonals are vertical and

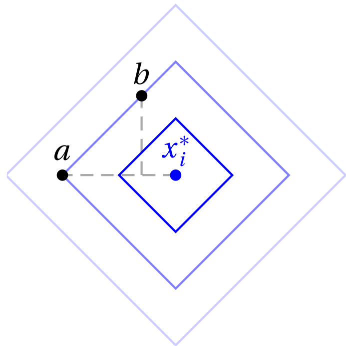


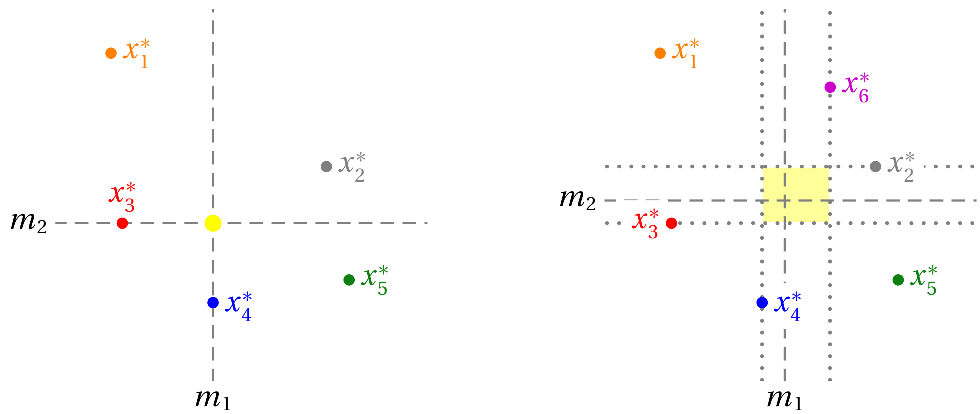
Figure 1.10 Some indifference sets of an individual with city block preferences. The individual's favorite alternative is x_i^*, and darker indifference sets correspond to more preferred alternatives. The city-block distance from x_i^* to a, for example, is the same as the city-block distance from x_i^* to b.

horizontal. Figure 1.10 shows an example. For such a preference relation, two alternatives are indifferent if they are equidistant from x^*, where the distance between two alternatives is the length of the shortest path between them that consists only of horizontal and vertical line segments.

I show that every collective choice problem in which each individual has city block preferences has a Condorcet winner. Each component of this winner is a median of the individuals' favorite values of that component. More precisely, for a problem in which the number of individuals is odd, denote by m_j, for $j = 1$, 2, the (unique) median of $(x_{ij}^*)_{i \in N}$, the collection of the jth components of the individuals' favorite alternatives. The alternative (m_1, m_2) is the strict Condorcet winner of such a problem. (Figure 1.11a shows an example.) For a problem in which the number of individuals is even, every pair (m_1, m_2) for which, for $j = 1$, 2, $m_j \leq x_{ij}^*$ for exactly half of the individuals (and hence $m_j \geq x_{ij}^*$ for the other half) is a Condorcet winner. (Figure 1.11b shows an example.)

Suppose that the number of individuals is odd and (x_1, x_2) is a point for which $|m_1 - x_1| \neq |m_2 - x_2|$. Then the reason that the number of individuals who prefer (m_1, m_2) to (x_1, x_2) exceeds the number with the opposite preference may be seen geometrically. Figure 1.12a shows a case in which (x_1, x_2) lies northeast of (m_1, m_2). An individual whose favorite alternative is in the area shaded green prefers (m_1, m_2) to (x_1, x_2). For every value of (x_1, x_2) with $|m_1 - x_1| \neq |m_2 - x_2|$, this area includes all alternatives (z_1, z_2) with $z_2 \leq m_2$, so that by the definition of m_2 it includes the favorite positions of a majority of individuals. The arguments for the cases in which (x_1, x_2) lies in one of the three other quadrants are symmetric.

If (x_1, x_2) is a point for which $|m_1 - x_1| = |m_2 - x_2|$ then the favorite positions of the individuals who prefer (m_1, m_2) to (x_1, x_2) lie in the area shaded green in Fig-



(a) A problem with five individuals. The strict Condorcet winner is (m_1, m_2), indicated by the yellow disk.

(b) A problem with six individuals. Each alternative in the yellow rectangle (for example (m_1, m_2)) is a Condorcet winner.

Figure 1.11 Collective choice problems in which the set of available alternatives is $\mathbb{R}^2$ and each individual has city block preferences.

ure 1.12b. The argument that the favorite positions of a majority of individuals lie in this area is given in the proof of the result.

Proposition 1.6: Condorcet winner for collective choice problem with two-dimensional set of alternatives and city block preferences

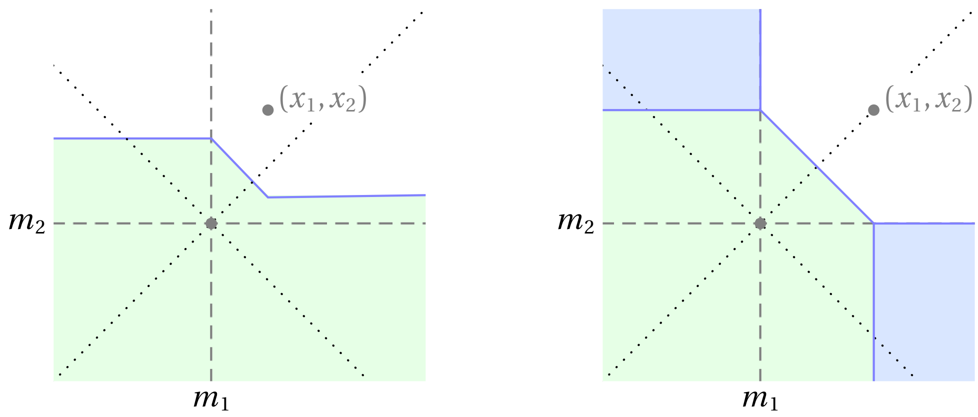
Let $\langle N, X, \succsim \rangle$ be a collective choice problem for which the set N of individuals is finite, the set X of alternatives is $\mathbb{R}^2$, and each individual has city block preferences.

If the number of individuals is odd, for $j \in \{1, 2\}$ let m_j be the median of the jth components of the individuals' favorite alternatives with respect to the linear order $\geq$. Then the alternative (m_1, m_2) is the strict Condorcet winner of $\langle N, X, \succsim \rangle$.

If the number of individuals is even, for $j \in \{1, 2\}$ let $\underline{m}_j$ and $\overline{m}_j$, with $\underline{m}_j \leq \overline{m}_j$, be the medians of the jth components of the individuals' favorite alternatives with respect to $\geq$. Then the Condorcet winners of $\langle N, X, \succsim \rangle$ are the alternatives (m_1, m_2) with $m_j \in [\underline{m}_j, \overline{m}_j]$ for $j = 1, 2$.

Proof

Denote the number of individuals by n. First suppose that n is odd. Let $(x_1, x_2) \in \mathbb{R}^2$ be an alternative different from (m_1, m_2). I argue that a strict



(a) $|m_1 - x_1| \neq |m_2 - x_2|$. (b) $|m_1 - x_1| = |m_2 - x_2|$.

Figure 1.12 Collective choice problems in which the set of available alternatives is $\mathbb{R}^2$ and each individual has city block preferences. In each case, an individual whose favorite alternative is on the blue line or in the area shaded blue is indifferent between (m_1, m_2) and (x_1, x_2), and one whose favorite alternative is in the area shaded green prefers (m_1, m_2) to (x_1, x_2).

majority of the individuals prefer (m_1, m_2) to (x_1, x_2), so that (m_1, m_2) is a strict Condorcet winner. Denote the favorite position of each individual i by x_i^*.

First suppose that $x_2 > m_2 + |x_1 - m_1|$ (as in the example in Figure 1.12a). Then every individual i whose favorite position is at most m_2 (individuals 3, 4, and 5 in Figure 1.11a) prefers (m_1, m_2) to (x_1, x_2):

$$\begin{aligned} |x_1 - x_{i1}^*| + |x_2 - x_{i2}^*| &> |x_1 - x_{i1}^*| + |m_2 - x_{i2}^* + |x_1 - m_1|| \\ &= |x_1 - x_{i1}^*| + |m_1 - x_1| + |m_2 - x_{i2}^*| \text{ (given } m_2 \geq x_{i2}^*) \\ &\geq |m_1 - x_{i1}^*| + |m_2 - x_{i2}^*|. \end{aligned}$$

By the definition of m_2, the number of such individuals is $\frac{1}{2}(n+1)$, a strict majority.

Now suppose that $x_2 = m_2 + |x_1 - m_1|$, with $x_1 > m_1$, as in the example in Figure 1.12b, so that (x_1, x_2) lies on the black line in Figure 1.13 (which excludes the point (m_1, m_2)). Divide the space into four regions, as in the figure; each region consists of an area shaded in a light color plus the boundaries with the corresponding dark color, if any. Denote the numbers of the individuals' favorite alternatives in the regions by n_1, n_2, n_3, and n_4, as shown in the figure. (In the case shown in Figure 1.11a, for example, $n_1 = 2$ and $n_2 = n_3 = n_4 = 1$.) Given the definitions of m_1 and m_2, we have $n_1 + n_2 > n_3 + n_4$ and $n_1 + n_3 > n_2 + n_4$, so that $n_1 > n_4$. By variants

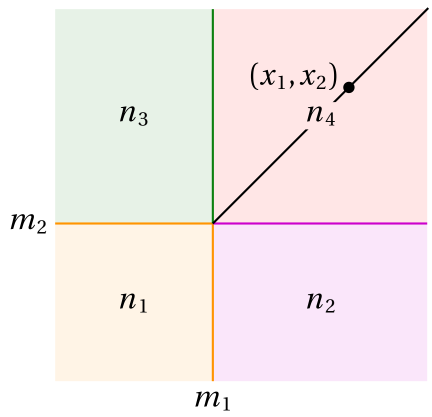


Figure 1.13 The regions in the second part of the proof of Proposition 1.6.

of the argument for the previous case, the n_1 individuals with favorite alternatives in the orange region prefer (m_1, m_2) to (x_1, x_2) and the $n_2 + n_3$ individuals with favorite alternatives in the green and violet regions like (m_1, m_2) at least as much as (x_1, x_2). Thus given $n_1 > n_4$, more individuals prefer (m_1, m_2) to (x_1, x_2) than prefer (x_1, x_2) to (m_1, m_2).

Similar arguments apply to every alternative (x_1, x_2) with $|m_1 - x_1| = |m_2 - x_2|$ and $x_1 < m_1$, so that (m_1, m_2) is a strict Condorcet winner.

If n is even, for any (m_1, m_2) satisfying the condition in the result, similar arguments establish that for any other alternative (x_1, x_2), at least as many individuals prefer (m_1, m_2) to (x_1, x_2) as prefer (x_1, x_2) to (m_1, m_2), so that (m_1, m_2) is a Condorcet winner.

This result does not generalize to problems in higher dimensional spaces: not all problems with city block preferences in spaces of three or more dimensions have Condorcet winners (Wendell and Thorson 1974, Example 3.1).

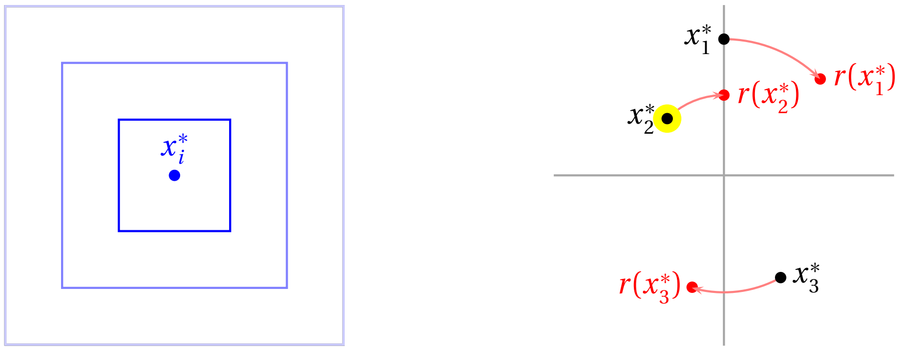
1.6.2 Max preferences

Suppose that each individual focuses exclusively on the dimension for which an alternative differs most from her favorite alternative.

Definition 1.25: Max preferences

A preference relation $\succsim_i$ on $\mathbb{R}^2$ reflects *max preferences* if for some alternative $(x_{i1}^*, x_{i2}^*) \in \mathbb{R}^2$ it is represented by the payoff function u_i defined by

$$u_i(x_1, x_2) = -\max\{|x_1 - x_{i1}^*|, |x_2 - x_{i2}^*|\} \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2.$$



(a) Indifference sets of an individual with max preferences and favorite alternative x_i^*. Darker indifference sets correspond to higher payoffs.

(b) A collective choice problem with three individuals in which each individual has max preferences. The Condorcet winner is x_2^*.

Figure 1.14

The indifference sets for such preferences are squares centered at x_i^* with sides parallel to the axes, as in Figure 1.14a.

Given Proposition 1.6 and the fact that these indifference sets are rotations by 45 degrees of the indifference sets for city block preferences, we can find the Condorcet winners of a collective choice problem in which each individual has max preferences as follows. Rotate the set of the individuals' favorite positions by 45 degrees and for $j \in \{1,2\}$ let m_j be a median of the jth components of the rotated alternatives. A Condorcet winner of the problem is the alternative obtained by applying the inverse of the rotation to (m_1, m_2). (The outcome is independent of the center and direction of the rotation.)

Proposition 1.7: Condorcet winner for collective choice problem with two-dimensional set of alternatives and max preferences

Let $\langle N, X, \succsim \rangle$ be a collective choice problem for which the set N of individuals is finite, the set X of alternatives is $\mathbb{R}^2$, and each individual has max preferences. Denote the favorite alternative of each individual $i \in N$ by x_i^* and let $r : \mathbb{R}^2 \to \mathbb{R}^2$ be a 45° rotation of the plane. An alternative x^* is a Condorcet winner of $\langle N, X, \succsim \rangle$ if and only if for $j \in \{1,2\}$ the number $r_j(x^*)$ is a median of the set of points $r_j(x_i^*)$ for $i \in N$.

Figure 1.14b shows an example with three individuals in which I use a clockwise rotation about $(0,0)$. For an arbitrary problem, a Condorcet winner is not necessarily the favorite alternative of any individual, but in this example the strict Condorcet winner is the favorite alternative of individual 2. Note that the value of its first dimension, x^*_{21}, is the smallest value of the first dimension over all the individuals' favorite alternatives. It is the strict Condorcet winner of the problem because for each of the other individuals, the second dimension, not the first, is salient: $|x^*_{j2} - x^*_{22}| > |x^*_{j1} - x^*_{21}|$ for $j \in \{1,3\}$. To verify directly that it is the strict Condorcet winner, let x be an alternative different from x^*_2. If $x_2 < x^*_{22}$ then individuals 1 and 2 prefer x^*_2 to x; if $x_2 > x^*_{22}$ then individuals 2 and 3 prefer x^*_2 to x; and if $x_2 = x^*_{22}$ then individual 2 prefers x^*_2 to x and individuals 1 and 3 are indifferent between x^*_2 and x.

1.6.3 Euclidean preferences

Finally, suppose that each individual evaluates an alternative x according to the length of the line segment between her favorite alternative and x, the Euclidean distance between these alternatives.

Definition 1.26: Euclidean preferences

A preference relation $\succsim_i$ on $\mathbb{R}^2$ reflects *Euclidean preferences* if for some alternative $x^*_i \in \mathbb{R}^2$ it is represented by the payoff function u_i defined by

$$u_i(x) = -\|x - x^*_i\| \quad \text{for all } x \in \mathbb{R}^2$$

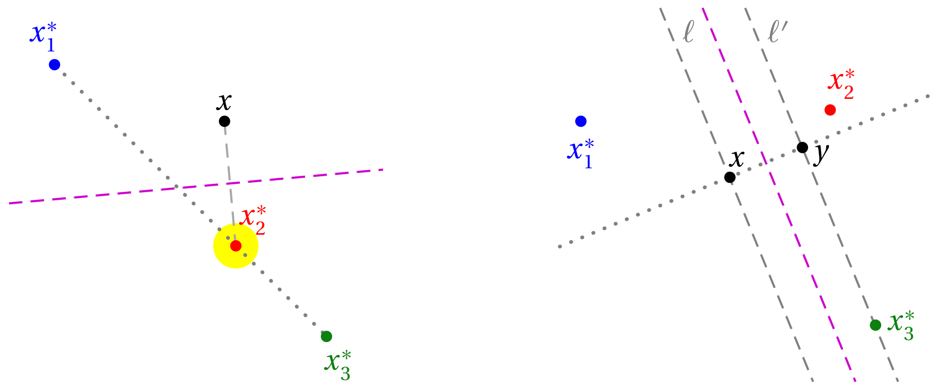
where $\|x - x^*_i\|$ is the Euclidean distance between x and x^*_i (the length of the line segment joining x and x^*_i).

An indifference set of an individual with Euclidean preferences has the form

$$\{x \in \mathbb{R}^2 : \|x - x^*_i\| = c\}$$

for some number c, a circle with center x^*_i.

Some collective choice problems in which every individual has Euclidean preferences have Condorcet winners. Suppose, for example, that there are three individuals and their favorite alternatives are the ones given in Figure 1.15a. The favorite alternative of individual 2, x^*_2, is a Condorcet winner of this collective choice problem. Take any other alternative, x. An individual prefers x^*_2 to x if and only if x^*_2 is closer to her favorite alternative than is x. The alternatives equidistant from x^*_2 and x lie on the perpendicular bisector of the line segment joining x^*_2 and x, indicated by the violet dashed line in Figure 1.15a. All individuals with favorite alternatives on the x^*_2-side of this perpendicular bisector prefer x^*_2 to x, and all individuals with favorite alternatives on the x-side of it prefer x to x^*_2. Thus x^*_2 beats the alternative x in the figure two to one, and you should be able to convince yourself that it beats every other alternative by at least two to one. Thus x^*_2 is a Condorcet winner—in fact, the strict Condorcet winner.



(a) A collective choice problem with a Condorcet winner (x_2^*).

(b) A collective choice problem with no Condorcet winner.

Figure 1.15 Collective choice problems in which the set of available alternatives is $\mathbb{R}^2$, the set of all pairs of real numbers.

The fact that x_3^* is a Condorcet winner in this example depends on the fact that it lies on the line through x_1^* and x_3^*. Consider the collective choice problem in Figure 1.15b, in which x_2^* does not have this property. This problem has no Condorcet winner. For every alternative x, there is a line ℓ through x with the property that the favorite alternatives of more individuals lie on one side of ℓ than on the other side, where an alternative on ℓ is counted as being on both sides. For example, in the figure, the favorite alternatives of individuals 2 and 3 lie on the right of ℓ and the favorite alternative of individual 1 lies on the left of it. Now select the favorite alternative on the side of ℓ where a majority of favorite alternatives lie that is closest to ℓ (x_3^* in the figure) and draw a line ℓ' through it parallel to ℓ. Finally, choose the point y that is on both ℓ' and a line through x perpendicular to ℓ. Then more individuals prefer y to x than x to y, showing that x is not a Condorcet winner.

The next result generalizes these arguments to collective choice problems with any number of individuals.

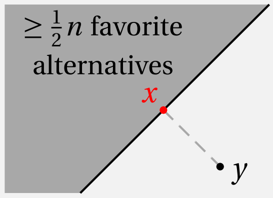
Proposition 1.8: Condorcet winner when set of alternatives is two-dimensional and preferences are Euclidean

Consider a collective choice problem $\langle N, X, \succsim \rangle$ in which the set N of individuals is finite, the set X of alternatives is $\mathbb{R}^2$, and each individual has Euclidean preferences. An alternative x is a Condorcet winner of $\langle N, X, \succsim \rangle$ if and only if the favorite alternatives of at least half of the individuals lie on each side of every line through x (with a point on the line counted as being on both sides).

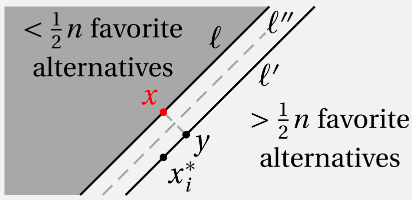
Proof

Denote the number of individuals by n.

I first argue that if x is an alternative with the property that the favorite alternatives of at least half of the individuals lie on each side of every line through x then x is a Condorcet winner. Let y be another alternative. Draw the line through x perpendicular to the line segment joining x and y. By assumption, the favorite alternatives of at least $\frac{1}{2}n$ individuals lie on the side of this perpendicular line opposite to y, so that those individuals prefer x to y. Thus x is a Condorcet winner.

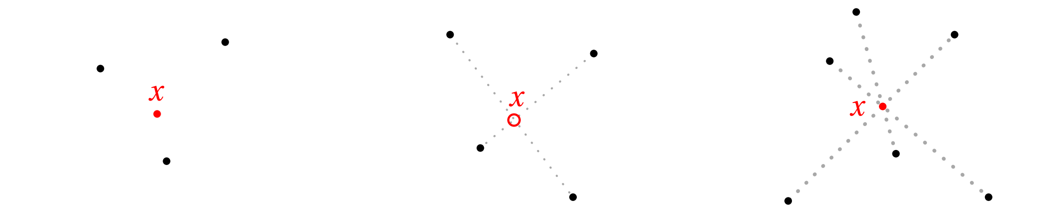


I now argue that if x does not satisfy the property then it is not a Condorcet winner. Given that fewer than $\frac{1}{2}n$ individuals' favorite alternatives lie on one side of some line ℓ through x, more than $\frac{1}{2}n$ individuals' favorite alternatives lie *strictly* on the other side of ℓ. Let x_i^* be a favorite alternative closest to ℓ among those on the other side of ℓ, and draw a line ℓ' parallel to ℓ through x_i^*.



Choose the alternative y at the intersection of ℓ' and the line perpendicular to ℓ (and to ℓ') through x. Let ℓ'' be the line midway between ℓ and ℓ' (the dashed line in the diagram). The alternative y is preferred to x by all individuals with favorite alternatives on the opposite side (relative to x) of ℓ'', who number more than $\frac{1}{2}n$, so x is not a Condorcet winner.

The condition in the result for an alternative to be a Condorcet winner is aptly characterized as requiring the alternative to be a "median in every direction". For every collective choice problem in which the set of alternatives is two-dimensional and there are four individuals, an alternative exists that satisfies the condition. If one of the individual's favorite alternatives lies within the triangle formed by the other individuals' favorite alternatives, as in Figure 1.16a, then this alter-



(a) Four individuals, Condorcet winner x.

(b) Four individuals, Condorcet winner x.

(c) Seven individuals, (strict) Condorcet winner x.

Figure 1.16 Configurations of favorite alternatives satisfying the condition in Proposition 1.8 for x to be a Condorcet winner. Each small disk, black or red, is an individual's favorite alternative. (In the middle case, x, indicated by a circle, is not a favorite alternative of any individual.)

native satisfies the condition and hence is a Condorcet winner. Otherwise, as in Figure 1.16b, the favorite alternatives can be divided into two pairs with the property that the two lines connecting the members of each pair intersect inside the quadrilateral formed by the alternatives. In this case, the alternative at the intersection of the line (and only this alternative) is a Condorcet winner.

For most collective choice problems with a two-dimensional set of alternatives and either three individuals or at least five, no alternative is a median in every direction. Suppose, for example, that the number n of individuals is odd. Then $\frac{1}{2}n$ is not an integer, and the condition requires the favorite alternatives of at least $\frac{1}{2}(n+1)$ individuals to lie on each side of every line through x. Thus x must be the favorite alternative of some individual. Further, as a line through x rotates, whenever it passes through an alternative different from x it must at the same time pass through another such alternative, to keep $\frac{1}{2}(n+1)$ alternatives on each side of it. If $n = 3$ only configurations of the individuals' favorite alternatives that lie on a line satisfy this condition; the middle alternative is the Condorcet winner. An example of a configuration satisfying the condition for $n = 7$ is shown in Figure 1.16c.

1.7 Preference aggregation: Arrow's theorem

The object of study of the previous sections is a collective choice rule, which associates with each collective choice problem an alternative or set of alternatives, based on the individuals' preferences. This section considers the possibility of associating with each collective choice problem a (societal) preference relation over the set of alternatives; that is, the possibility of aggregating the individuals' preferences. Rather than looking for an alternative that best reflects the individuals' possibly diverse preference relations, we look for an entire ranking of the alternatives that best reflects these preference relations. One motivation for this line of inquiry is that the members of the society do not currently know the collective choice problem they will face, and they wish to be prepared for whatever problem arises. Armed with a preference relation for society, for any set of alternatives that arises they can select the alternative that is best according to the preference relation. The main result is due to Kenneth J. Arrow (1921–2017), whose work is the foundation of social choice theory.

I assume that every individual has a strict preference relation over the set of alternatives. Our job is to associate with every profile of such preference relations a single (not necessarily strict) preference relation over the set of alternatives, which I call the social preference relation. That is, we seek a preference aggregation function, defined as follows.

Definition 1.27: Preference aggregation function

Let $\langle N, X\rangle$ be a society and D a set of strict preference profiles for $\langle N, X\rangle$. A *preference aggregation function* for $(\langle N, X\rangle, D)$ is a function that assigns a ("social") preference relation over X to every collective choice problem $\langle N, X, \succsim\rangle$ with $\succsim \in D$.

Example 1.6: Borda aggregation

Assign to each collective choice problem the preference relation $\trianglerighteq$ over the set of alternatives that ranks x at least as highly as y if and only if the number of points $\sum_{i\in N} p_i(x)$ that the Borda rule assigns to x is at least the number it assigns to y. For instance, for the problem in Example 1.2, this preference relation is given by $b \triangleright a \triangleright c \triangleright d$ (b is assigned 7 points, a is assigned 6, c is assigned 4, and d is assigned 1).

Example 1.7: Condorcet aggregation

For each pair (x, y) of alternatives, say that x beats y if a majority of individuals prefer x to y, and x and y tie if the number of individuals who prefer x to y is the same as the number who prefer y to x. Assign to each alternative one point for every alternative it beats and half a point for every alternative with which it ties. Rank the alternatives by the number of points received. For example, for the problem in Example 1.2, the preference relation $\trianglerighteq$ thus defined is given by $a \triangleright b \triangleright c \triangleright d$ (a beats b, c, and d, b beats c and d, c beats d, and d beats no alternative).

The following property of a preference aggregation function plays a central role in the analysis: the ranking of two alternatives according to the social preference relation depends only on the individuals' preferences between these two alternatives. More precisely, if for two preference profiles $\succsim$ and $\succsim'$ and two alternatives x and y, the preference of every individual i between x and y is the same according to $\succsim_i$ as it is according to $\succsim'_i$, then the social preference between x and y generated by $\succsim$ is the same as the social preference between these alternatives generated by $\succsim'$. This property is known as *independence of irrelevant alternatives* (a name that seems calculated to convince the reader that the property is reasonable).

Definition 1.28: Independence of irrelevant alternatives (IIA)

Let $\langle N, X\rangle$ be a society and D a set of strict preference profiles for $\langle N, X\rangle$. A preference aggregation function F for $(\langle N, X\rangle, D)$ is *independent of irrelevant alternatives* (IIA) if for all preference profiles $\succsim \in D$ and $\succsim' \in D$ for $\langle N, X\rangle$ and any alternatives $x \in X$ and $y \in X$ with

$$x \succ_i y \text{ if and only if } x \succ'_i y \quad \text{for every } i \in N$$

we have

$$x \trianglerighteq y \text{ if and only if } x \trianglerighteq' y,$$

where $\trianglerighteq = F(N, X, \succsim)$ and $\trianglerighteq' = F(N, X, \succsim')$.

The Borda preference aggregation function defined in Example 1.6 does not satisfy this property for a domain that includes the following two collective choice problems.

1	2	3
a	a	a
b	b	b
c	c	d
d	d	c

1	2	3
c	a	d
d	b	a
a	c	b
b	d	c

Each individual ranks c relative to d in the same way for both problems, but for the problem on the left c gets more points than d (2 versus 1) whereas for the problem on the right the opposite is true (c gets 4 points and d gets 5), so that according to the Borda rule c is socially preferred to d for the left problem and d is socially preferred to c for the right problem.

The Condorcet aggregation function defined in Example 1.7 also does not satisfy independence of irrelevant alternatives for a domain that includes these two problems. For the problem on the left, c beats only d whereas d does not beat any alternative, so that c is socially preferred to d, but for the problem on the right, c beats only d whereas d beats a and b, so that d is socially preferred to c.

One preference aggregation function that does satisfy independence of irrelevant alternatives assigns to every preference profile the social preference relation in which all alternatives are indifferent. In addition to failing to meaningfully aggregate the individuals' preference relations, this function violates the natural requirement that the social preference relation ranks one alternative above another one whenever all the individuals do so. This requirement is called the Pareto property, after Vilfredo Pareto (1848–1923), who introduced the idea that one social state is better than another if all individuals prefer it.

Definition 1.29: Pareto property of preference aggregation function

Let $\langle N, X\rangle$ be a society and let D be a set of strict preference profiles for $\langle N, X\rangle$. A preference aggregation function F for $(\langle N, X\rangle, D)$ has the *Pareto property* if for every preference profile $\succsim \in D$ and all $x \in X$ and $y \in X$ with $x \succ_i y$ for all $i \in N$ we have $x \rhd y$, where $\trianglerighteq = F(N, X, \succsim)$ (and $\rhd$ is the strict relation associated with $\trianglerighteq$).

A simple argument shows that no preference aggregation function satisfies independence of irrelevant alternatives, the Pareto property, and adaptations of anonymity and neutrality to preference aggregation. Consider a society with two individuals, 1 and 2, and three alternatives, a, b, and c. Here are three possible preference profiles.

1	2
b	c
a	a
c	b

1	2
c	c
b	a
a	b

1	2
b	c
c	a
a	b

Alternatives b and c are symmetric in the first problem, so anonymity and neutrality imply they are socially indifferent. Similarly, a and b are socially indifferent in the second problem. Now, each individual's rankings of b and c in the first and third problems are the same, so by independence of irrelevant alternatives these alternatives are socially indifferent in the third problem, and each individual's rankings of a and b in the second and third problems are the same, so these alternatives are socially indifferent in the third problem. By the transitivity of social preferences, a and c are thus socially indifferent in the third problem. This conclusion conflicts with the implication of the Pareto property that, because both individuals prefer c to a in this problem, c is socially preferred to a.

The next result, Arrow's impossibility theorem, does not impose the assumptions of anonymity and neutrality on a preference aggregation function. Without these assumptions, preference aggregation functions that satisfy independence of irrelevant alternatives and the Pareto property exist: for any individual $i^* \in N$ assign i^*'s preference relation, $\succsim_{i^*}$, to every preference profile. Such a preference aggregation function is, naturally, called dictatorial.

Definition 1.30: Dictatorial preference aggregation function

Let $\langle N, X\rangle$ be a society and let D be a set of strict preference profiles for $\langle N, X\rangle$. A preference aggregation function F for $(\langle N, X\rangle, D)$ is *dictatorial* if for some individual $i^* \in N$, for every preference profile $\succsim \in D$ we have $x \trianglerighteq y$ if and only if $x \succsim_{i^*} y$, where $\trianglerighteq = F(N, X, \succsim)$.

Arrow's theorem shows that a preference aggregation function satisfies independence of irrelevant alternatives and the Pareto property only if it is dictatorial.

Proposition 1.9: Arrow's impossibility theorem

Let $\langle N, X\rangle$ be a finite society for which X contains at least three alternatives and let D be the set of all strict preference profiles for $\langle N, X\rangle$. A preference aggregation function for $(\langle N, X\rangle, D)$ satisfies the Pareto property and independence of irrelevant alternatives if and only if it is dictatorial.

Proof

If a preference aggregation function is dictatorial then it satisfies the Pareto property and independence of irrelevant alternatives (IIA). I now show the converse.

Let F be a preference aggregation function for $(\langle N, X\rangle, D)$ that satisfies the Pareto property and IIA. Let $N = \{1, \ldots, n\}$ and fix $b \in X$.

Step 1 *Let $\succsim$ be a preference profile for $\langle N, X\rangle$ for which b is either at the top or the bottom of each individual's ranking, and let $\trianglerighteq = F(N, X, \succsim)$. Then b is either the unique maximal alternative for $\trianglerighteq$ or the unique minimal alternative for $\trianglerighteq$ (that is, either $b \rhd x$ for all $x \in X \setminus \{b\}$ or $x \rhd b$ for all $x \in X \setminus \{b\}$).*

Proof. Assume to the contrary that for two other alternatives a and c we have $a \trianglerighteq b \trianglerighteq c$. Let $\succsim'$ be a preference profile that is obtained from $\succsim$ by moving c just above a for every individual i for whom $a \succ_i c$ (so that it remains below b for all individuals for whom b is best), as illustrated in the following example.

1	2	3	⋯	$n-1$	n		1	2	3	⋯	$n-1$	n
b	b		⋯	a	b		b	b		⋯	c	b
	a	a	⋯					c	c	⋯	a	
c	c		⋯	c	a	$\rightarrow$	c	a	a	⋯		c
		c	⋯		c					⋯		a
a		b	⋯	b			a		b	⋯	b	
			$\succsim$							$\succsim'$		

Let $\trianglerighteq' = F(N, X, \succsim')$. By the Pareto property, $c \rhd' a$. We have $a \succ_i b$ if and only if $a \succ'_i b$ for every $i \in N$ and $a \trianglerighteq b$, so that $a \trianglerighteq' b$ by IIA. Also, $b \succ_i c$ if and only if $b \succ'_i c$ for every $i \in N$ and $b \trianglerighteq c$, so that $b \trianglerighteq' c$ by IIA. Thus $a \trianglerighteq' c$ by the transitivity of $\trianglerighteq'$, contradicting $c \rhd' a$. ◁

Step 2 *Consider two preference profiles $\succsim$ and $\succsim'$ for $\langle N, X\rangle$ for which b is either at the top or the bottom of each individual's ranking and the set of individuals who rank b at the top is the same. Let $\trianglerighteq = F(N, X, \succsim)$ and $\trianglerighteq' = F(N, X, \succsim')$. Then b is either the unique maximal alternative for both $\trianglerighteq$ and $\trianglerighteq'$ or the unique minimal alternative for both $\trianglerighteq$ and $\trianglerighteq'$.*

Proof. By Step 1, b is either the unique maximal alternative for $\trianglerighteq$ or the unique minimal alternative for $\trianglerighteq$. Suppose the former. The ranking of b relative to any other alternative x is the same in $\succsim_i$ and $\succsim'_i$ for every individual i and hence by IIA b and x are ranked in the same way by both $\trianglerighteq$ and $\trianglerighteq'$. Thus b is the unique maximal alternative for $\trianglerighteq'$. If b is the unique minimal alternative for $\trianglerighteq$ the argument is analogous. ◁

Step 3 *For some individual $i^* \in N$,*

i. for every preference profile $\succsim$ for $\langle N, X\rangle$ for which $1, \ldots, i^* - 1$ rank b at the top and $i^*, \ldots, n$ rank it at the bottom, the preference relation $F(N, X, \succsim)$ ranks b uniquely at the bottom

ii. for every preference profile $\succsim$ for $\langle N, X\rangle$ for which $1, \ldots, i^*$ rank b at the top and $i^* + 1, \ldots, n$ rank it at the bottom, the preference relation $F(N, X, \succsim)$ ranks b uniquely at the top.

Proof. Let $\succsim$ be a preference profile in which b is at the bottom of all individuals' preferences. By the Pareto property, b is the unique minimal alternative for $F(N, X, \succsim)$. Now, for each individual i in turn, starting with individual 1, move b from the bottom to the top of i's preferences. By

Step 1, b is always either the unique maximal or unique minimal alternative for the resulting social preferences. By the Pareto property, it is the unique maximal alternative after it moves to the top of all individuals' preferences. Let i^* be the first individual for whom the change in her preferences moves b from the bottom to the top of the social preferences. By Step 2, the identity of i^* does not depend on the individuals' rankings of the other alternatives. ◁

Step 4 *For any preference profile $\succsim$ for $\langle N, X\rangle$ and all alternatives a and c different from b we have $a \rhd c$ if and only if $a \succ_{i^*} c$, where $\unrhd = F(N, X, \succsim)$ and i^* is the individual identified in Step 3.*

Proof. Assume to the contrary that for some preference profile $\succsim$ for $\langle N, X\rangle$ we have $a \succ_{i^*} c$ and $c \unrhd a$. Let $\succsim'$ be the profile obtained from $\succsim$ by raising b to the top of the preferences of individuals $1, \ldots, i^*-1$, lowering it to the bottom of the preferences of individuals $i^*+1, \ldots, n$, and moving it between a and c for individual i^*, as in the following example.

1	$\cdots$	i^*-1	i^*	i^*+1	$\cdots$	n		1	$\cdots$	i^*-1	i^*	i^*+1	$\cdots$	n
	$\cdots$	c		a	$\cdots$	b		b	$\cdots$	b		a	$\cdots$	
b	$\cdots$		a		$\cdots$				$\cdots$	c	a		$\cdots$	c
a	$\cdots$			c	$\cdots$	c	$\to$	a	$\cdots$		b	c	$\cdots$	a
	$\cdots$	a	c	b	$\cdots$	a			$\cdots$				$\cdots$	
c	$\cdots$	b	b		$\cdots$			c	$\cdots$	a	c	b	$\cdots$	b
			$\succsim$								$\succsim'$			

Let $\unrhd' = F(N, X, \succsim')$. The relative positions of a and c are the same in $\succsim$ and $\succsim'$, so $c \unrhd' a$ by IIA. In $\succsim'$, the individuals' rankings of a relative to b are the same as they are in any profile in which b is ranked at the top by $1, \ldots, i^*-1$ and at the bottom by the remaining individuals, so that by Step 3*i* and IIA we have $a \rhd' b$. Similarly, using Step 3*ii* and IIA, $b \rhd' c$. Thus by transitivity $a \rhd' c$, contradicting $c \unrhd a$. ◁

Step 4 says that i^* is the dictator regarding any two alternatives other than b. It remains to show that i^* is also the dictator regarding the comparison of b with any other alternative.

Step 5 *For any preference profile $\succsim$ for $\langle N, X\rangle$ and every alternative a we have $a \rhd b$ if and only if $a \succ_{i^*} b$, where $\unrhd = F(N, X, \succsim)$ and i^* is the individual identified in Step 3.*

Proof. Let $\succsim$ be a preference profile for which $a \succ_{i^*} b$. Let c be an arbitrary third alternative. Let $\succsim'$ be a preference profile obtained from $\succsim$ by moving c in i^*'s ranking to between b and a (if it is not already there) and raising c to the top of all the other individuals' rankings, as in the following example.

1	$\cdots$	i^*-1	i^*	i^*+1	$\cdots$	n
	$\cdots$	c		a	$\cdots$	b
c	$\cdots$		a		$\cdots$	
a	$\cdots$		b	c	$\cdots$	c
	$\cdots$	b		b	$\cdots$	
b	$\cdots$	a	c		$\cdots$	a
			$\succsim$			

$\rightarrow$

1	$\cdots$	i^*-1	i^*	i^*+1	$\cdots$	n
c	$\cdots$	c		c	$\cdots$	c
	$\cdots$		a	a	$\cdots$	b
a	$\cdots$		c		$\cdots$	
	$\cdots$	b	b	b	$\cdots$	
b	$\cdots$	a			$\cdots$	a
			$\succsim'$			

Let $\trianglerighteq' = F(N, X, \succsim')$. By the Pareto property, $c \rhd' b$. By Step 4, given $a \succ'_{i^*} c$ we have $a \rhd' c$ and hence $a \rhd' b$ by transitivity. Since $a \succ_i b$ if and only if $a \succ'_i b$ for all $i \in N$, we thus have $a \rhd b$ by IIA. $\lhd$

This result has the same flavor as Proposition 1.3: no aggregation method satisfies a list of attractive properties for the domain of all possible preference profiles. The results differ both in the type of aggregation considered—collective choice rule or preference aggregation function—and in the nature of the properties imposed, but the message is similar.

Proposition 1.2 shows that for the domain of collective choice problems that have a strict Condorcet winner, a collective choice rule that satisfies a list of attractive properties *does* exist: the one that assigns to each problem its strict Condorcet winner. An analogous result holds for preference aggregation functions.

For a problem $\langle N, X, \succsim\rangle$ with the property that for every subset X' of X (including X itself) the problem $\langle N, X', \succsim|_{X'}\rangle$ has a strict Condorcet winner, the majority relation is transitive, so that the function that assigns it to each problem in the domain of problems with this property is a preference aggregation function. This preference aggregation function also satisfies the Pareto property and independence of irrelevant alternatives.

Proposition 1.10: Preference aggregation with a strict Condorcet winner

Let $\langle N, X\rangle$ be a finite society for which the number of individuals is odd and let D be the set of strict preference profiles $\succsim$ for $\langle N, X\rangle$ for which for every subset X' of X the collective choice problem $\langle N, X', \succsim|_{X'}\rangle$ has a strict Condorcet winner. For any preference profile $\succsim \in D$, the majority relation for $\langle N, X, \succsim\rangle$ is a preference relation, and the preference aggregation function for $(\langle N, X\rangle, D)$ that for each $\succsim \in D$ assigns to $\langle N, X, \succsim\rangle$ the majority relation for $\langle N, X, \succsim \rangle$ satisfies the Pareto property and independence of irrelevant alternatives.

Proof

By definition the majority relation for $\langle N, X, \succsim \rangle$ is complete. I now argue that it is transitive. Let $x \in X$, $y \in X$, and $z \in X$ and suppose that a majority of individuals prefer x to y and a majority prefer y to z. Let $X' = \{x, y, z\}$. By assumption, the collective choice problem $\langle N, X', \succsim|_{X'} \rangle$ has a strict Condorcet winner. This alternative is not y, because a majority of individuals prefer x to y, and it is not z, because a majority of individuals prefer y to z, so it must be x. Thus a majority of individuals prefer x to z and hence the majority relation for $\langle N, X, \succsim \rangle$ is transitive. We conclude that the majority relation for $\langle N, X, \succsim \rangle$ is a preference relation.

Denote the majority relation by $\trianglerighteq$. If for any alternatives x and y all individuals prefer x to y then $x \vartriangleright y$, so the Pareto property is satisfied. Whether $x \trianglerighteq y$ or $y \trianglerighteq x$ depends only on the individuals' preferences between x and y, so also independence of irrelevant alternatives is satisfied.

If a preference profile on a set X is single-peaked with respect to a linear order $\trianglerighteq$, then the restriction of the profile to a subset of X is single-peaked with respect to the restriction of $\trianglerighteq$ to the subset, so by Proposition 1.4 the domain of single-peaked preference profiles for a society $\langle N, X \rangle$ for which the number of individuals is odd satisfies the property in this result. Hence for this domain the function that assigns to each collective choice problem its majority relation is a preference aggregation function that satisfies the Pareto property and independence of irrelevant alternatives. By Proposition 1.5, the same is true for the domain of single-crossing preference profiles for a society with an odd number of individuals, and moreover for this domain the majority relation is the preference relation of the median individual.

1.8 Preference intensities and interpersonal comparisons

The information in a collective choice problem about the individuals' preferences concerns only the individuals' rankings of the alternatives. For some collective choice problems, this information appears to be an inadequate basis for the selection of a collective action.

Consider, for example, the simplest of all collective choice problems: two individuals have to decide between two alternatives, a and b. One individual prefers a to b and the other prefers b to a. To select one of these alternatives, we need more information about the individuals' evaluations of the alternatives.

Now suppose that two individuals have to decide among three alternatives, a, b, and c. One individual prefers a to b to c and the other prefers c to b to a. The collective choices consistent with anonymity and neutrality are $\{a,c\}$, $\{b\}$, and $\{a,b,c\}$. The last is a non-choice, and to choose between $\{a,c\}$ and $\{b\}$ again we need more more information about the individuals' evaluations of the alternatives.

One piece of information that we can use to choose an alternative in both of these examples is a comparison of the individuals' welfares for the alternatives. Such information not only allows us to select alternatives in problems like these two, but may also overturn our resolutions of other problems. For example, suppose that three individuals have to decide between the alternatives a and b. Two individuals prefer a and the third prefers b. Then a is the strict Condorcet winner, so that it is selected by any anonymous, neutral, positively responsive, and Nash independent collective choice rule (Proposition 1.2), and is ranked first in the social preference relation generated by a preference aggregation function that satisfies the Pareto property and independence of irrelevant alternatives (Proposition 1.10). But if we have information on the intensity of the individuals' preferences, and the first two individuals' preferences for a over b are slight compared with the last individual's preference for b over a, we may decide to select b rather than a, especially if the first two individuals' welfares for b are higher than the last individual's.

One reason that the models in the previous sections do not include interpersonally comparable measures of welfare is the difficulty of quantifying welfare and comparing it across individuals. We can in principle obtain information about individuals' ordinal preferences by observing their choices, but these choices do not directly reveal the individuals' welfares, and especially do not allow us to compare one individual's welfare with another's. However, individuals' welfares can be assessed and compared in other ways. We can ask individuals to report their well-beings on a common scale; we can assess the extent to which each individual experiences certain states, like hunger and ill-health; and we can observe individuals' incomes and wealths.

I assume for the remainder of this section that associated with each alternative is a profile of numbers, interpreted as the individuals' welfares. How should such profiles be ranked? My approach, as in the previous sections, is axiomatic: I state some properties for an ordering of welfare profiles that embody certain principles and investigate the orderings that satisfy these properties and hence are consistent with the principles.

A social welfare ordering is a complete transitive binary relation—that is, a preference relation—over welfare profiles.

Definition 1.31: Social welfare ordering

A *social welfare ordering* over a set X^n of welfare profiles for a set of n individuals is a preference relation over X^n.

For two of the specific social welfare orderings that I discuss, the set X is $\mathbb{R}$, the set of all real numbers, and for the remaining ordering it is $\mathbb{R}_{++}$, the set of positive numbers.

The interpretation of a social welfare ordering $\succsim$ is that for any profiles u and v of the individuals' welfares, u is at least as socially desirable as v if and only if $u \succsim v$. (In this section I use u and v to denote profiles of numbers, rather than functions, as elsewhere.)

One example of a social welfare ordering is the utilitarian ordering, which ranks welfare profiles according to the sum of the welfares.

Definition 1.32: Utilitarian social welfare ordering

Let $N = \{1, \ldots, n\}$ be a set of individuals. A social welfare ordering $\succsim$ over $\mathbb{R}^n$ for N is the *utilitarian ordering* if $u \succsim v$ if and only if $\sum_{i \in N} u_i \geq \sum_{i \in N} v_i$.

This ordering pays no attention to inequality in welfare. For a society of two individuals, for example, it makes the welfare profile $(1,1)$ indifferent to the profiles $(2,0)$ and $(5,-3)$, and ranks the profile $(101,0)$ above the profile $(50,50)$.

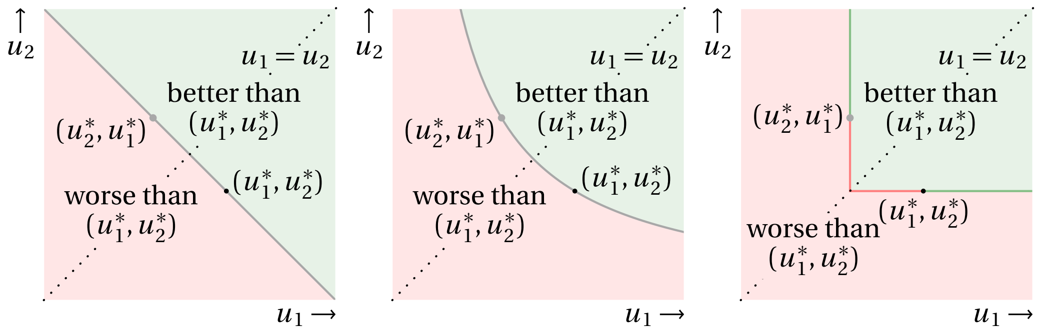
A related social welfare ordering ranks profiles of positive welfares according to the product of the welfares.

Definition 1.33: Nash social welfare ordering

Let $N = \{1, \ldots, n\}$ be a set of individuals. A social welfare ordering $\succsim$ over $\mathbb{R}^n_{++}$ for N is the *Nash ordering* if $u \succsim v$ if and only if $\prod_{i \in N} u_i \geq \prod_{i \in N} v_i$.

This ordering puts some weight on the equality of welfare. For a society of two individuals, for example, it ranks the welfare profile $(3,3)$ above the profile $(8,1)$.

An example of a social welfare ordering that is even more sensitive to inequality is the leximin ordering, which gives priority to the smallest welfare. When comparing the profiles u and v according to this ordering, we arrange their components according to size, smallest to largest, and rank the profiles according to the first components in this order that differ. For example, for $u = (1,4,2,1,5)$ and $v = (3,1,8,2,1)$, the orderings by component size are $(1,1,2,4,5)$ and $(1,1,2,3,8)$, so because these vectors are the same up to their third components and the fourth component of the first vector, 4, exceeds the fourth component of the second vector, 3, u is ranked above v. (The utilitarian and Nash orderings rank



(a) Utilitarian ordering (b) Nash ordering (c) Leximin ordering

Figure 1.17 The pairs (u_1, u_2) ranked better than (u_1^*, u_2^*) (green), worse than it (red), and equal to it (gray) by three welfare orderings. For the leximin ordering, each region includes its boundaries with a dark shade of the color of the region, and the only pair indifferent to (u_1^*, u_2^*) is (u_2^*, u_1^*).

v above u.)

Definition 1.34: Leximin social welfare ordering

Let $N = \{1, \ldots, n\}$ be a set of individuals and for any $u \in \mathbb{R}^n$ and $k \in N$ let $k(u)$ be the kth smallest component of u (so that $u_{k(u)} \leq u_{(k+1)(u)}$ for all $k = 1, \ldots, n-1$). The social welfare ordering $\succsim$ over $\mathbb{R}^n$ for N is the *leximin ordering* if $u \succsim v$ if and only if either u is a permutation of v or there exists $k \in N$ such that $u_{j(u)} = v_{j(v)}$ for $j = 1, \ldots, k-1$ and $u_{k(u)} > v_{k(v)}$.

Note that this ordering is not continuous. For example, for a society of two individuals, $(1,3)$ is better than $(1,2)$, but for all $\varepsilon > 0$, $(1-\varepsilon, 3)$ is worse than $(1,2)$.

For the case of two individuals ($n = 2$), the sets of welfare pairs ranked better than, equal to, and worse than a pair (u_1^*, u_2^*) by the three orderings are shown in Figure 1.17.

I now present axiomatic characterizations of these orderings. The characterizations share two properties, adaptations of the anonymity property for collective choice rules and the Pareto property for a preference aggregation function, both of which are appealing. The anonymity property says that an ordering does not depend on the individuals' names.

Definition 1.35: Anonymous social welfare ordering

A social welfare ordering $\succsim$ over a set X^n of welfare profiles for n individuals is *anonymous* if for every $u \in X^n$ and every permutation v of u we have $u \sim v$.

The Pareto property says that if the welfare of every individual is at least as high in u as it is in v and is higher for at least one individual, then u is ranked above v.

Definition 1.36: Strong Pareto condition for social welfare ordering

A social welfare ordering $\succsim$ over a set X^n of welfare profiles for a set $N = \{1,\ldots,n\}$ of individuals satisfies the *strong Pareto condition* if for all $u \in X^n$ and $v \in X^n$,

$$u_i \geq v_i \text{ for all } i \in N \text{ with at least one inequality} \quad \Rightarrow \quad u \succ v.$$

Each characterization involves one additional property. For the leximin ordering this property says that the welfare profile u is ranked above the profile v if these profiles differ only in the welfares of two individuals, say i and j, i's welfare is higher in u than in v, j's welfare is higher in v than in u, and j's welfare in u is higher than i's welfare in u. Thus moving from v to u makes i better off and j worse off, and both before and after the change j is better off than i. The property is named after its originator, Peter J. Hammond.

Definition 1.37: Hammond-equitable social welfare ordering

A social welfare ordering $\succsim$ over $\mathbb{R}^n$ for a set $N = \{1,\ldots,n\}$ of individuals is *Hammond-equitable* if for all $i \in N$ and $j \in N$ with $i \neq j$ and all $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ with $u_k = v_k$ for all $k \in N \setminus \{i,j\}$,

$$v_j > u_j > u_i > v_i \quad \Rightarrow \quad u \succ v.$$

All three welfare orderings that I have defined are anonymous and satisfy the strong Pareto condition, but only the leximin ordering is Hammond equitable: both the utilitarian and Nash orderings rank the profile $(1,7)$ above the profile $(2,3)$, in violation of Hammond equity. In fact, among all possible orderings, the leximin ordering is the only one that satisfies all three conditions.

Proposition 1.11: Characterization of leximin social welfare ordering

A social welfare ordering over $\mathbb{R}^n$ for a set of n individuals is anonymous and Hammond-equitable and satisfies the strong Pareto condition if and only if it is the leximin ordering.

I present a proof of this result only for a society consisting of two individuals.

Proof for two individuals

The leximin ordering satisfies the three conditions. I now show that if a social welfare ordering for a set of two individuals satisfies the three conditions then it is the leximin ordering.

First suppose that $u_1^* > u_2^*$ and consider the ordering of the pair (u_1^*, u_2^*) of welfares relative to any other pair. Refer to Figure 1.18.

- By the strong Pareto condition, pairs in the green region are ranked above (u_1^*, u_2^*) and ones in the violet region are ranked below (u_1^*, u_2^*).
- By Hammond equity, pairs in the yellow region are ranked above (u_1^*, u_2^*) and ones in the red region are ranked below (u_1^*, u_2^*).
- Given these rankings, anonymity implies that pairs in the blue region are ranked above (u_1^*, u_2^*) and ones in the brown region are ranked below (u_1^*, u_2^*).
- The pairs that remain are (u_2^*, u_1^*) and the ones on the dashed black line.

  The pair (u_2^*, u_1^*) is equivalent to (u_1^*, u_2^*) by anonymity.

  Finally, for any pair (u_1, u_2) on the dashed black line there is a pair (u_1', u_2') in the yellow region with $u_1 > u_1'$ and $u_2 > u_2'$, so that by the strong Pareto condition (u_1, u_2) is ranked above (u_1', u_2'). Given that every pair in the yellow region is ranked above (u_1^*, u_2^*), the transitivity of a social welfare ordering implies that (u_1, u_2) is ranked above (u_1^*, u_2^*).

A symmetric argument applies to pairs (u_1^*, u_2^*) with $u_1^* < u_2^*$. The only comparisons that remain are between pairs (u_1^*, u_2^*) and (u_1, u_2) with $u_1^* = u_2^*$ and $u_1 = u_2$. The strong Pareto condition implies that (u_1^*, u_2^*) is ranked above (u_1, u_2) if $u_1^* > u_1$ and below it if $u_1^* < u_1$.

We conclude that the rankings of all pairs of welfares are the ones given by the leximin ordering, shown in Figure 1.17c.

I now present a property that, in addition to anonymity and the Pareto condition, characterizes the utilitarian social welfare ordering. The character of this property is completely different from that of Hammond-equity. It says that if one welfare profile is ranked above another, then certain transformations of the first profile are ranked above the same transformations of the second profile. Specifically, if u is ranked above v, then for any profile $(a_i)_{i \in N}$ of numbers, the profile u' defined by $u_i' = u_i + a_i$ for all $i \in N$ is ranked above the profile v' defined by

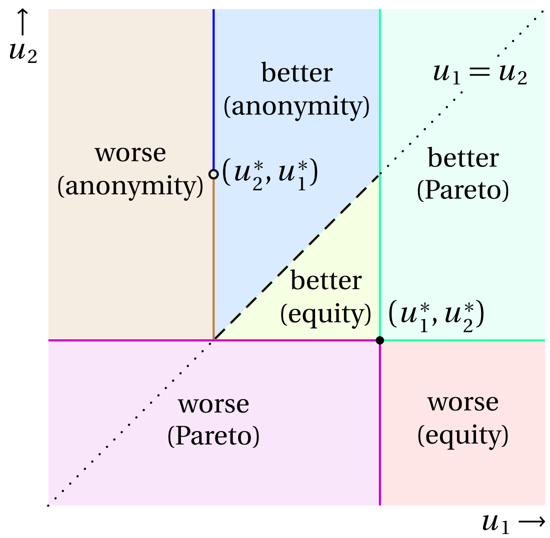


Figure 1.18 The proof of Proposition 1.11 when there are two individuals. (Each region includes any of its boundaries with a dark shade of the color of the region.)

$v_i' = v_i + a_i$ for all $i \in N$.

Definition 1.38: Invariance of social welfare ordering with respect to additive transformations of welfares

A social welfare ordering $\succsim$ over $\mathbb{R}^n$ for a set of n individuals is *invariant with respect to additive transformations of welfares* if whenever $u \succsim v$ we have $u + a \succsim v + a$ for any $a \in \mathbb{R}^n$.

This condition may make sense if we cannot observe the individuals' welfare levels but we can observe whether the difference in one individual's welfare between two welfare profiles is bigger or smaller than the difference in any other individual's welfare between two other welfare profiles. The reason is that the transformations of welfare in the definition preserve comparisons of differences in welfare: $u_i - v_i \geq w_j - y_j$ if and only if $u_i + a_i - (v_i + a_i) \geq w_j + a_j - (y_j + a_j)$.

Proposition 1.12: Characterization of utilitarian welfare ordering

A social welfare ordering over $\mathbb{R}^n$ for n individuals is anonymous and invariant with respect to additive transformations of welfares and satisfies the strong Pareto condition if and only if it is the utilitarian ordering.

For a society consisting of two individuals, this result may be given a simple proof. Let v be a welfare pair for which $v_1 = v_2$, let w be a pair for which $w_1 + w_2 = v_1 + v_2$, and let $w' = (w_2, w_1)$, as in Figure 1.19. By anonymity, $w \sim w'$. Now let

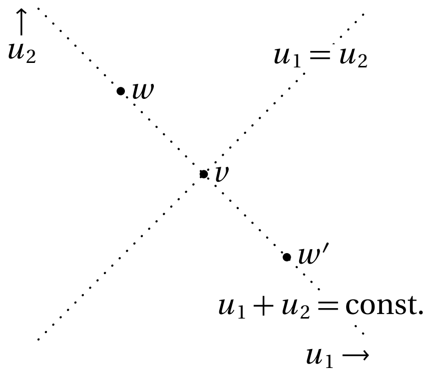


Figure 1.19 An illustration for the proof of Proposition 1.12 for a society consisting of two individuals.

$a_1 = v_1 - w_1$ and $a_2 = v_2 - w_2$. Then $v_i = w_i + a_i$ and $w'_i = v_i + a_i$ for $i = 1, 2$. Thus if $w \succ v$ then $v \succ w'$ by invariance with respect to additive transformations of welfares, and hence $w \succ w'$, contradicting $w \sim w'$. Similarly, if $w \prec v$ then $v \prec w'$, contradicting $w \sim w'$. Thus $w \sim v$. Finally, the strong Pareto condition implies that $(v_1 + \alpha, v_2 + \alpha) \succ (v_1, v_2)$ for any $\alpha > 0$, so that $v' \succ v$ if and only if $v'_1 + v'_2 > v_1 + v_2$.

I now present a proof of the result for an arbitrary number of individuals. The main part of the argument shows that if a social welfare ordering $\succcurlyeq$ satisfies the conditions in the result then for any two welfare profiles u and v for which $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i$ we have $u \sim v$. To reach this conclusion, u and v are repeatedly transformed. Let $u^0 = u$ and $v^0 = v$. First, the components of u^0 and v^0 are put in order, from smallest to largest, to generate $\hat{u}^0$ and $\hat{v}^0$. By the anonymity condition, $\hat{u}^0 \sim u^0$ and $\hat{v}^0 \sim v^0$. Then for each component i, the smaller of $\hat{u}_i^0$ and $\hat{v}_i^0$ is subtracted from both $\hat{u}_i^0$ and $\hat{v}_i^0$, to generate u^1 and v^1. By the invariance condition, $u^1 \sim \hat{u}^0$ and $v^1 \sim \hat{v}^0$, so that $u^1 \sim u^0$ and $v^1 \sim v^0$. For each value of i, either $u_i^1 = 0$ or $v_i^1 = 0$, and at least one component of u^1 and one component of v^1 is zero (given that it is not the case that every component of u exceeds the corresponding component of v, or vice versa). The components of u^1 and v^1 are then put in order to generate $\hat{u}^1$ and $\hat{v}^1$, so that the first components of $\hat{u}^1$ and $\hat{v}^1$ are 0, and the process is repeated. At step n, all components of the resulting profiles u^n and v^n are 0, so that $u \sim u^n \sim v^n \sim v$.

Proof of Proposition 1.12

The utilitarian ordering satisfies the conditions in the result.

To show that it is the only social welfare ordering that does so, let $\succcurlyeq$ be a social welfare ordering that satisfies the conditions in the result and let

$N = \{1, \ldots, n\}$. Let $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ with $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i$. I argue that $u \sim v$.

To do so, I apply a sequence of transformations to u and v and argue that the invariance condition implies that the welfare profiles generated at each step are socially indifferent to the profiles from the previous step. Denote the welfare profiles generated by the transformations at each step $t \geq 1$ by u^t and v^t, and let $u^0 = u$ and $v^0 = v$. For $t \geq 1$, u^t and v^t are generated from u^{t-1} and v^{t-1} as follows.

1. Let $\hat{u}^{t-1}$ be a permutation u^{t-1} of with $\hat{u}_1^{t-1} \leq \hat{u}_2^{t-1} \leq \cdots \leq \hat{u}_n^{t-1}$ and let $\hat{v}^{t-1}$ be a permutation v^{t-1} of with $v_1^{t-1} \leq v_2^{t-1} \leq \cdots \leq v_n^{t-1}$.
2. For each $i \in N$, let $u_i^t = \hat{u}_i^{t-1} - \min\{\hat{u}_i^{t-1}, \hat{v}_i^{t-1}\}$ and $v_i^t = \hat{v}_i^{t-1} - \min\{\hat{u}_i^{t-1}, \hat{v}_i^{t-1}\}$.

Step 1 *For all $t \geq 1$ we have $u^t \sim u^{t-1}$ and $v^t \sim v^{t-1}$.*

Proof. By the anonymity of $\succsim$, $\hat{u}^{t-1} \sim u^{t-1}$ and $\hat{v}^{t-1} \sim v^{t-1}$, and by its invariance with respect to additive transformations of welfares, $u^t \sim \hat{u}^{t-1}$ and $v^t \sim \hat{v}^{t-1}$. ◁

Step 2 *For $t = 1, \ldots, n$ we have $\hat{u}_i^t = 0$ and $\hat{v}_i^t = 0$ for $i = 1, \ldots, t$.*

Proof. At each step $t \geq 1$, for each $i \in N$ we have

$$\begin{aligned} u_i^t = 0 \text{ and } v_i^t \geq 0 \quad & \text{if } \hat{u}_i^{t-1} \leq \hat{v}_i^{t-1} \\ u_i^t \geq 0 \text{ and } v_i^t = 0 \quad & \text{if } \hat{u}_i^{t-1} \geq \hat{v}_i^{t-1}. \end{aligned}$$

Thus in particular $u_i^t \geq 0$, $v_i^t \geq 0$, and either $u_i^t = 0$ or $v_i^t = 0$ (or both). Further, $\sum_{i=1}^n u_i^t = \sum_{i=1}^n v_i^t$ if $\sum_{i=1}^n u_i^{t-1} = \sum_{i=1}^n v_i^{t-1}$, and thus since $\sum_{i=1}^n u_i^0 = \sum_{i=1}^n v_i^0$ we have $\sum_{i=1}^n u_i^t = \sum_{i=1}^n v_i^t$ for $t = 1, \ldots, n$. Hence $u_i^t = 0$ for at least one value of i and $v_i^t = 0$ for at least one value of i.

In particular, $\hat{u}_1^1 = 0$ and $\hat{v}_1^1 = 0$ and hence $\hat{u}_1^t = 0$ and $\hat{v}_1^t = 0$ for all $t \geq 1$.

I now argue by induction. Let $1 \leq t \leq n-1$ and suppose that $\hat{u}_i^t = 0$ and $\hat{v}_i^t = 0$ for $i = 1, \ldots, t$, so that $\hat{u}_i^{t+1} = 0$ and $\hat{v}_i^{t+1} = 0$ for $i = 1, \ldots, t$. Then $\sum_{i=t+1}^n \hat{u}_i^t = \sum_{i=t+1}^n \hat{v}_i^t$, so that $u_i^{t+1} = 0$ for at least one value of $i \geq t+1$ and $v_i^{t+1} = 0$ for at least one value of $i \geq t+1$. Thus $\hat{u}_{t+1}^{t+1} = 0$ and $\hat{v}_{t+1}^{t+1} = 0$ and hence $\hat{u}_i^{t+1} = 0$ and $\hat{v}_i^{t+1} = 0$ for $i = 1, \ldots, t+1$. ◁

Step 3 $u \sim v$.

Proof. By Step 2, $u_i^n = v_i^n = 0$ for $i = 1, \ldots, n$, so that $u^n \sim v^n$ and hence by Step 1, $u^t \sim v^t$ for $t = 0, \ldots, n$. Thus in particular $u \sim v$. ◁

The strong Pareto condition implies that if $\sum_{i=1}^n u_i > \sum_{i=1}^n v_i$ then $u \succ v$, so that $\succsim$ is the utilitarian ordering.

If you find the invariance condition appealing, this result may make the utilitarian social welfare ordering more appealing (or less unappealing). I am not in this camp. We can plausibly assess at least imperfectly whether one individual is better off than another, while assessing how one individual's gain in welfare compares with another's seems an order of magnitude more difficult. So assuming that welfares cannot be compared but differences in welfare can seems backwards.

The Nash social welfare ordering is characterized by a different invariance property, in conjunction with anonymity and the Pareto condition.

Definition 1.39: Invariance of social welfare ordering with respect to multiplicative transformations of welfares

A social welfare ordering $\succsim$ social welfare ordering over $\mathbb{R}^n_{++}$ for a set $N = \{1, \ldots, n\}$ of individuals is *invariant to multiplicative transformations of welfares* if whenever $u \succsim v$ we have $(b_1 u_1, b_2, u_2, \ldots, b_n u_n) \succsim (b_1 v_1, b_2 v_2, \ldots, b_n v_n)$ for any $b \in \mathbb{R}^n$ with $b_i > 0$ for all $i \in N$.

This condition may make sense if we cannot observe the individuals' welfare levels but we can observe whether the ratio u_i / v_i of one individual's welfare between two welfare profiles is bigger or smaller than the ratio w_j / y_j of another individual's welfare between two other welfare profiles, because the transformations of welfare in the definition preserve these ratios. The next result is a corollary of Proposition 1.12.

Proposition 1.13: Characterization of Nash welfare ordering

A social welfare ordering on $\mathbb{R}^n_{++}$ is anonymous and invariant with respect to multiplicative transformations of welfares and satisfies the strong Pareto condition if and only if it is the Nash ordering.

Proof

The Nash ordering satisfies the conditions in the result.

Now let $\succsim$ be a social welfare ordering on $\mathbb{R}^n_{++}$ that satisfies the conditions in the result. Define the social welfare ordering $\succsim^*$ on $\mathbb{R}^n$ by

$$u \succsim^* v \text{ if and only if } (e^{u_1}, e^{u_2}, \ldots, e^{u_n}) \succsim (e^{v_1}, e^{v_2}, \ldots, e^{v_n}).$$

Then $\succsim^*$ is invariant with respect to additive transformations of welfares by the following argument. For any $a \in \mathbb{R}^n$, let $b_i = e^{a_i}$ for all $i \in N$, where N is the set of individuals. Suppose that $u \succsim^* v$, so that $(e^{u_1}, e^{u_2}, \ldots, e^{u_n}) \succsim (e^{v_1}, e^{v_2}, \ldots, e^{v_n})$. Then $(b_1 e^{u_1}, b_2 e^{u_2}, \ldots, b_n e^{u_n}) \succsim (b_1 e^{v_1}, b_2 e^{v_2}, \ldots, b_n e^{v_n})$ by the assumption that $\succsim$ is invariant with respect to multiplicative transformations of welfares, and hence $u + a \succsim^* v + a$.

Now, the fact that $\succsim$ is anonymous and satisfies the strong Pareto condition means that $\succsim^*$ satisfies these conditions, so by Proposition 1.12, $\succsim^*$ is the utilitarian ordering. That is, $u \succsim^* v$ if and only if $\sum_{i \in N} u_i \geq \sum_{i \in N} v_i$. Thus $u \succsim v$ if and only if $\sum_{i \in N} \ln u_i \geq \sum_{i \in N} \ln v_i$ or equivalently $\prod_{i \in N} u_i \geq \prod_{i \in N} v_i$, so that $\succsim$ is the Nash ordering.

In the models of collective choice in the remainder of the book, the outcome is determined either by voting or by the balance of power, as determined by the availability of actions to individuals and groups that can affect other individuals and groups. For the most part, the models assume that all individuals are selfish—no individual's welfare is directly affected by the other individuals' well-beings—and so considerations of relative welfare, which dominate the analysis of this section, are absent.

Notes

Proposition 1.1 is due to May (1952). The notion of a Condorcet winner is due to Condorcet (1785). Section 1.4 is based on Horan et al. (2019); Proposition 1.2 is Theorem 1 in the paper and Proposition 1.3 is a weak version of Theorem 2. (The notion I call positive responsiveness is called full positive responsiveness in the paper.) The proof I give for Proposition 1.2 is a simplification due to Ariel Rubinstein of the argument in Horan et al. (2019) and the proof for Proposition 1.3 is due to him. (The stronger Theorem 2 in Horan et al. 2019 requires a different proof.) The results in this section are related to those of Dasgupta and Maskin (2008). Their model has a continuum of individuals and they require a collective choice rule to assign a single alternative to almost every preference profile. They identify a set of conditions that are satisfied by the collective choice rule that assigns to each collective choice problem its set of Condorcet winners and show, roughly, that no other collective choice rule satisfies these conditions on a larger domain of problems. The conditions include anonymity, neutrality, and a relative of Nash independence that treats each alternative, rather than each set of alternatives, as a unit.

Proposition 1.4 is due to Black (1958) (see the theorems on pp. 16 and 18). The generalization to trees discussed after Exercise 1.11 is due to Demange (1982). The use of single-crossing preferences in the context of collective choice has its origin in Roberts (1977); Proposition 1.5 is due to Rothstein (1990) (who uses the term "order-restricted preferences") and Gans and Smart (1996). My presentation of the material in Section 1.5.2 benefitted from discussions with Navin Kartik.

Proposition 1.6 is due to Rae and Taylor (1971, 77) and Wendell and Thorson (1974, Theorem 3.2). Wendell and Thorson (1974, Example 3.1) show that the result does not generalize to three dimensions (contrary to the claim on p. 78 of Rae and Taylor 1971). McKelvey and Wendell (1976) and Humphreys and Laver (2010) further investigate Condorcet winners when alternatives are multidimensional. Proposition 1.8 is due to Davis et al. (1972); see also Enelow and Hinich (1984, Section 3.6).

Proposition 1.9 is due to Arrow (1951). The proof I give is due to Geanakoplos (2005), and is taken almost verbatim from Osborne and Rubinstein (2023, Proposition 20.1). Other parts of Section 1.7 are based on Chapter 20 of that book; some are quoted verbatim. I am grateful to Ariel Rubinstein for permitting me to include the material here. The preference aggregation function in Example 1.7 (Condorcet aggregation) has a long history, dating back at least to Ramon Llull (c. 1232–1315/16) (see Hägele and Pukelsheim 2001). It is sometimes called the Copeland method, after Arthur H. Copeland (1898–1970), who (re)proposed it in 1951 (see, for example, Goodman 1954, 42; apparently no copy of the mimeographed notes he cites survive).

The study of collective choice based on interpersonal comparisons of welfare was initiated by Sen (1970). The Nash social welfare ordering is named for its relation with the bargaining solution of Nash (1950). Proposition 1.11 is due to Hammond (1976, Theorem 7.2). The proof for two individuals that I present is taken from Blackorby et al. (1984, Theorem 6.1) and Bossert and Weymark (2004, Theorem 12.2). Proposition 1.12 is due to d'Aspremont and Gevers (1977, Theorem 3). The proof for two individuals is taken from Blackorby et al. (1984, 351–352). My presentation of both results draws upon Bossert and Weymark (2004). Proposition 1.13 is a slight variant of Moulin (1988, Theorem 2.3).

Exercise 1.3 is based on Fishburn (1974, 67) (see also Moulin 1988, Exercise 11.2). The example in Exercise 1.6 is taken from Moulin (1988, 235). The argument in Exercise 1.11 and its extension to trees is taken from Exercise 10.4 in Moulin (1988, 279). The examples in parts *a* and *c* of Exercise 1.13 are taken from Austen-Smith and Banks (1999, Example 4.6). The result in part *b* is Corollary 3

points	$1\times$	$4\times$	$3\times$	$2\times$
1	a	b	c	c
p	b	a	a	b
0	c	c	b	a

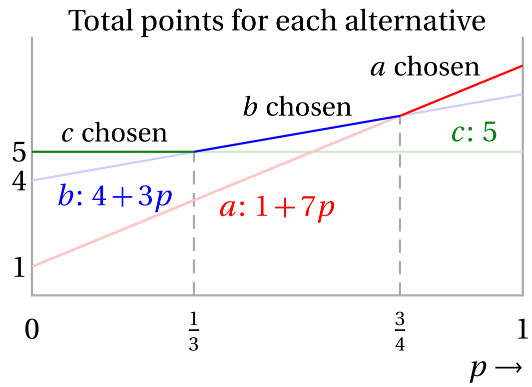


Figure 1.20 The collective choice problem in Exercise 1.1.

of Puppe (2018); the proof that I give is taken from Elkind et al. (2022, Proposition 3.19). The observation in Exercise 1.14 is taken from Rothstein (1990).

Solutions to exercises

Exercise 1.1

Refer to Figure 1.20. The set of alternatives selected by the rule is

$$\begin{cases} \{c\} & \text{if } 0 < p < \frac{1}{3} \\ \{b,c\} & \text{if } p = \frac{1}{3} \\ \{b\} & \text{if } \frac{1}{3} < p < \frac{3}{4} \\ \{a,b\} & \text{if } p = \frac{3}{4} \\ \{a\} & \text{if } \frac{3}{4} < p < 1. \end{cases}$$

Exercise 1.2

(*a*) For an anonymous and positively responsive collective choice rule, the pattern of outcomes in a diagram like those in Figure 1.1 has to satisfy the condition in Figure 1.3a and the symmetric condition for b. An example is shown in Figure 1.21. The rule shown in not neutral because the pattern of outcomes is not symmetric about the main diagonal. Another example is the rule that selects $\{a\}$ if more than $\frac{2}{3}$ of the individuals who are not indifferent between a and b prefer a to b, selects $\{b\}$ if fewer do so, and selects $\{a,b\}$ if exactly $\frac{2}{3}$ do so.

(*b*) A collective choice rule that is neutral and positively responsive but not anonymous is dictatorship by any individual.

Exercise 1.3

For an anonymous and neutral collective choice rule, the pattern of outcomes

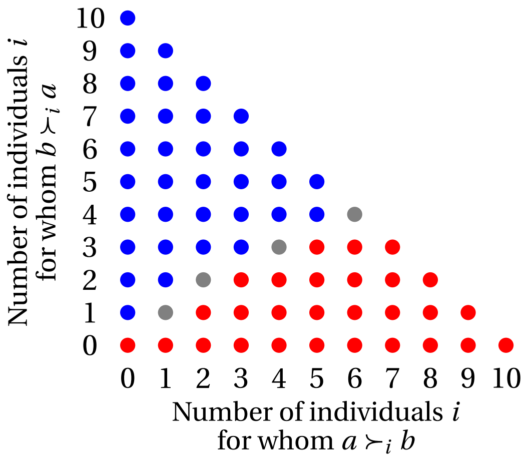


Figure 1.21 A collective choice rule that is anonymous and positively responsive but not neutral. See the discussion of Figure 1.1 for an explanation of the way in which the diagram represents a collective choice rule.

in a diagram like those in Figure 1.1 has to be symmetric about the main diagonal.

If a nonnegatively responsive collective choice rule selects $\{a\}$ for some problem then it selects $\{a\}$ for every problem in the region to the east, southeast, and south; if it selects $\{b\}$ for some problem then it selects $\{b\}$ for every problem in the region to the west, northwest, and north; and if it selects $\{a,b\}$ for some problem then it selects either $\{a\}$ or $\{a,b\}$ for every problem in the region to the east, southeast, and south and either $\{b\}$ or $\{a,b\}$ for every problem to the west, northwest, and north (cf. Figure 1.3a).

A pattern that satisfies these conditions is given in Figure 1.22. Another example is the rule that selects $\{a\}$ if more than $\frac{2}{3}$ of the individuals who are not indifferent between a and b prefer a to b, selects $\{b\}$ if more than $\frac{2}{3}$ of the individuals who are not indifferent between a and b prefer b to a, and otherwise selects $\{a,b\}$.

Exercise 1.4

In the following problem, a is the unique Condorcet winner, but is not a strict Condorcet winner. It beats b, but ties with c (which loses to b).

1	2	3
a	c	b
b	a	ac
c	b	

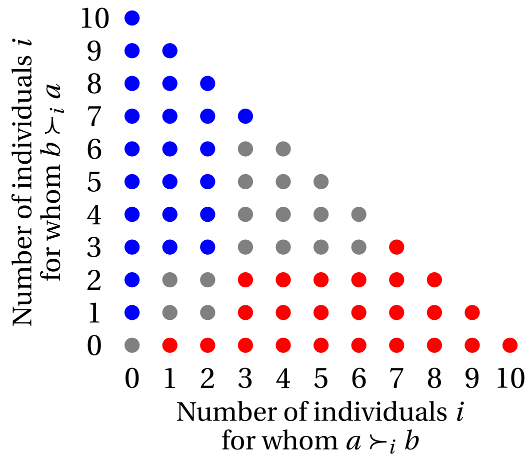


Figure 1.22 A collective choice rule that is anonymous and nonnegatively responsive but not positively responsive. See the discussion of Figure 1.1 for an explanation of the way in which the diagram represents a collective choice rule.

Exercise 1.5

The following collective choice problem shows that the answer is negative: a is a Condorcet winner and b ties with it, but is not a Condorcet winner.

1	2	3	4
a	a	b	c
c	c	a	b
b	b	c	a

Exercise 1.6

On the first round a and b are selected, and on the second round a wins (11 to 6). Now raise a above b in the preferences of the last two individuals, so that their preferences become $a \succ b \succ c$. Then on the first round a and c are selected and on the second round c wins. Thus after a's ranking improves, it is no longer selected.

Exercise 1.7

Neutrality, positive responsiveness, and Nash independence: Dictatorship by individual i: for any individual i, select i's favorite alternatives.

Anonymity, positive responsiveness, and Nash independence: For an arbitrary alternative, select that alternative for every collective choice problem.

Anonymity, neutrality, and Nash independence: Select the set of all alternatives.

Exercise 1.8

Suppose there are five alternatives, a, b, c, d, and e and five individuals. Two individuals have the preference ordering $b \succ_i a \succ_i c \succ_i d \succ_i e$, one has the ordering $c \succ_i b \succ_i a \succ_i d \succ_i e$, and two have the ordering $d \succ_i c \succ_i b \succ_i e \succ_i a$. This preference profile is single-peaked relative to the ordering $\lhd$ with $a \lhd b \lhd c \lhd d \lhd e$. The median favorite alternative is c and the Borda winner is b. (This alternative gets 15 points; c gets 14 points, d gets 11 points, a gets 8 points, and e gets 1 point.)

Exercise 1.9

Consider the collective choice problem in which the set of individuals is $\{1,2,3\}$, the set of alternatives is $\{a,b,c\}$, and the individuals' preferences are given by $a \succ_1 b \succ_1 c$, $a \sim_2 b \succ_2 c$, and $a \sim_3 b \sim_3 c$. This preference profile satisfies the single-plateau condition for the linear order $\unrhd$ for which $a \lhd b \lhd c$. Then $a_1^* = a$, $a_2^* = b$, and $a_3^* = c$ are favorite alternatives for the individuals. The median of these alternatives with respect to $\unrhd$ is b, but the only Condorcet winner of the collective choice problem is a.

Exercise 1.10

Consider the collective choice problem in which the set of individuals is $\{1,2,3,4,5\}$, the set of alternatives is $\{a,b,c,d,e\}$, and the preference profile is given as follows.

1	2	3	4	5
a	b	c	d	e
$bcde$	a	b	c	d
	cde	a	b	c
		de	a	b
			e	a

Each individual has a single favorite alternative and the profile satisfies the variant of (1.3) for the linear order $\unrhd$ for which $a \lhd b \lhd c$. Alternatives a and b are beaten by c, d and e are beaten by a, and c is beaten by d, so the problem has no Condorcet winner.

Now suppose that with the same set of individuals the set of alternatives is $\{a,b,c\}$ and the preference profile is given as follows.

1	2	3	4	5
a	a	b	c	c
bc	bc	a	ab	ab
		c		

Each individual has a single favorite alternative and the profile satisfies the variant of (1.3) for the linear order $\unrhd$ for which $a \lhd b \lhd c$. The median of the individuals' favorite alternatives with respect to $\unrhd$ is b, but the only

Condorcet winner of the collective choice problem is a.

Exercise 1.11

a. By (*ii*), x_t beats every other alternative in Z_t. Suppose, contrary to the claim, that some $x_j \in X_t$ beats x_t. That is, a majority of individuals prefer x_j to x_t. Then by the single-peakedness of preferences, the same individuals prefer x_{t-1} to x_t and hence also prefer x_{t-1} to every other alternative in Z_t, so that x_{t-1} is the favorite alternative in Z_{t-1} of a majority of individuals, contrary to (*i*) for $j = t-1$. Thus no $x_j \in X_t$ beats x_t, and hence x_t is the strict Condorcet winner of $\langle N, X, \succsim \rangle$.

b. At step t, if x_t is the favorite alternative in Z_t of a majority of individuals, select it and terminate. Otherwise proceed to step $t+1$. This procedure terminates at latest at step k. If it terminates at step t then it selects x_t, which by part *a* is the strict Condorcet winner of $\langle N, X, \succsim \rangle$.

Exercise 1.12

If the condition in the exercise is satisfied then the condition in Definition 1.23 is satisfied because for any alternatives x and y either $x \lhd y$ or $y \lhd x$.

Now suppose that the condition in Definition 1.23 is satisfied. I construct a linear order on X such that the condition in the exercise is satisfied.

Define the binary relation $\trianglerighteq$ on X as follows. Let x and y be alternatives. If $x \succ_j y$ for all $j \in N$, then $x \lhd y$. Otherwise, if $x \succsim_i y$ for some individual i, let i^* be the unique individual implied by Definition 1.23 for whom $x \succsim_{i^*} y$ and either (*a*) $x \succ_j y$ for all $j < i^*$ and $x \prec_j y$ for all $j > i^*$ or (*b*) $x \prec_j y$ for all $j < i^*$ and $x \succ_j y$ for all $j > i^*$. Define $x \lhd y$ in case (*a*) and $x \rhd y$ in case (*b*). If $x \prec_j y$ for all $j \in N$, then $x \rhd y$.

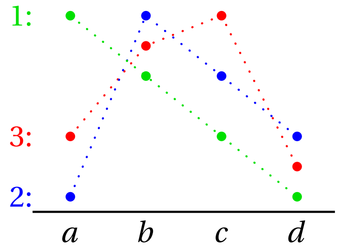
Given that the collective choice problem satisfies the condition in Definition 1.23, the binary relation $\trianglerighteq$ is complete. It is transitive by the following argument. If $x \lhd y$ and $y \lhd z$ then by the transitivity of each individual's preference relation either $x \succ_j z$ for every individual j or there exists an individual j^* such that $x \succ_j z$ for all $j < j^*$ and $x \prec_j z$ for all $j > j^*$. Thus $x \lhd z$, so that $\trianglerighteq$ is transitive. Finally, $x \lhd y$ and $y \lhd x$ are possible only if $x = y$. Thus $\trianglerighteq$ is a linear order.

If $x \lhd y$ then either $x \succ_j y$ for all $j \in N$ or for some individual i^* we have $x \succ_i y$ for all $i < i^*$ and $x \prec_i y$ for all $i > i^*$, so that (*a*) in the condition in the exercise is satisfied; if $y \lhd x$ then (*b*) is satisfied.

Exercise 1.13

a. The preference profile is given as follows.

1	2	3
a	b	c
b	c	b
c	d	a
d	a	d



The profile is single-peaked for the orderings $a \lhd b \lhd c \lhd d$ and $d \lhd c \lhd b \lhd a$, and only for these orderings.

It is not single-crossing because the condition in Definition 1.23 is not satisfied for any ordering of the individuals:

- $a \succ_1 b$ but $b \succ_2 a$ and $b \succ_3 a$, so the condition is not satisfied by any ordering in which 1 is in the middle
- $d \succ_2 a$ but $a \succ_1 d$ and $a \succ_3 d$, so the condition is not satisfied by any ordering in which 2 is in the middle
- $c \succ_3 b$ but $b \succ_1 c$ and $b \succ_2 c$, so the condition is not satisfied by any ordering in which 3 is in the middle.

b. Denote by $\succsim_1$ the preference ordering of the first individual according to $\geq$. Suppose, contrary to the result, that the preferences of some individual i are not single-peaked with respect $\succsim_1$. That is, for some alternatives x, y, and z with $x \prec_1 y \prec_1 z$ we have $x \succ_i y$ and $z \succ_i y$. Let j be an individual with favorite alternative y, so that $y \succ_j x$ and $y \succ_j z$. Given that $x \prec_1 y$ and $x \succ_i y$, by single-crossing every individual after i prefers x to y. Thus j comes before i. But $y \succ_j z$ and $y \prec_1 z$, so everyone after j prefers y to z, contradicting the fact that i comes after j and prefers z to y.

c. Consider the collective choice problem given as follows.

1	2	3
a	c	c
b	a	b
c	b	a

This problem has single-crossing preferences with respect to the ordering $\geq$ of the individuals for which $1 < 2 < 3$.

The preference profile is not single-peaked with respect to any ordering of the alternatives because for single-peakedness the middle alternative cannot be ranked lowest for any individual, and every alternative is ranked lowest by one individual.

Exercise 1.14

Consider the problem with three individuals and four alternatives in which the individuals' preferences are $a \succ_1 b \succ_1 c \succ_1 d$, $b \succ_2 c \succ_2 d \succ_2 a$, and $c \succ_3 b \succ_3 a \succ_3 d$. This problem has single-peaked preferences with respect to the ordering $a \lhd b \lhd c$, but no individual's preference relation coincides with that of the majority, for which b beats c beats a beats d.

Exercise 1.15

Consider the problem with four individuals and four alternatives in which the individuals' preferences are $a \succ_1 b \succ_1 c \succ_1 d$, $a \sim_2 b \succ_2 c \succ_2 d$, $c \succ_3 b \succ_3 d \succ_3 a$, and $d \succ_4 c \succ_4 b \succ_4 a$. This problem has single-crossing preferences with respect to the ordering 1, 2, 3, 4 of individuals, and a, a favorite alternative of individual 2, a median individual, is not a Condorcet winner (it loses to b).

Exercise 1.16

Take the preference profile in the solution of Exercise 1.13*a*, which is single-peaked but not single-crossing, and raise d to between b and c in individual 1's preferences. For the resulting profile, b is a strict Condorcet winner. However, the resulting profile is not single-peaked and does not have the single-crossing property.

It is not single-peaked because the alternatives the individuals rank lowest, c, a, and d, are distinct and thus cannot all be the smallest or largest alternative according to the ordering of alternatives, as single-peakedness requires.

It is not single-crossing because the condition in Definition 1.23 is not satisfied for any ordering of the individuals:

- $a \succ_1 c$ but $c \succ_2 a$ and $c \succ_3 a$, so the condition is not satisfied by any ordering in which 1 is in the middle
- $d \succ_2 a$ but $a \succ_1 d$ and $a \succ_3 d$, so the condition is not satisfied by any ordering in which 2 is in the middle
- $c \succ_3 b$ but $b \succ_1 c$ and $b \succ_2 c$, so the condition is not satisfied by any ordering in which 3 is in the middle.

2 Collective choice with privately-known preferences

2.1 Strategy-proofness of strict Condorcet winner 79
2.2 Strategy-proofness for domain of all preference profiles 86
2.3 General mechanisms 90

A mechanism for selecting an alternative in a collective choice problem should sensibly use information about the individuals' preferences. The models in the previous chapter assume that these preferences are known to the designer of the mechanism. The models in the remainder of the book assume that each individual's preferences are known only to the individual.

One way for a mechanism designer to proceed if she does not know the individuals' preferences is to ask each individual to report a preference relation from a set of the designer's choosing and to select an alternative based on these reports. Suppose that when the profile of reported preference relations is $\succsim$, the mechanism designer selects the alternative $g(\succsim)$. The individuals are assumed to know the function g, and can be modeled as players in a strategic game. In this game, the set of actions of each individual is the set of permitted preference relations, say P, and the action profile $\succsim$ (the profile of reported preference relations) results in the alternative $g(\succsim)$ being selected, so that each individual i prefers to report an action profile $\succsim^1$ to an action profile $\succsim^2$ if and only if she prefers $g(\succsim^1)$ to $g(\succsim^2)$. Assume for expositional convenience that the solution concept we apply to the game assigns a unique action profile to every true preference profile of the individuals. The task of the mechanism designer is then to specify the set P of permitted reports and the function g in such a way that, for any given true preference profile $\succsim^*$ for the individuals, the alternative $g(\succsim)$, where $\succsim$ is the preference profile given by the solution concept of the game when the true preference profile is $\succsim^*$, varies reasonably with $\succsim^*$.

Synopsis

Section 2.1 studies environments in which for some set P of preference relations, every collective choice problem in which each individual's preference relation is in P has a strict Condorcet winner. For such an environment, consider the mechanism in which each individual is restricted to report a preference relation in P and the function g selects the strict Condorcet winner of the collective choice problem in which the preference profile is the one reported. Proposition 2.1 shows that for this mechanism, whenever the true preference relation of each individual is in P, no individual can do better than report her true preference relation, regardless of the preference relations reported by the other individuals. In the argot of the field, that is, the mechanism is strategy-proof. If each individual does in fact report her true preference relation, then the mechanism selects the strict Condorcet winner for the true preference profile, which Proposition 1.2 suggests is a reasonable choice.

Suppose, for example, that the number of individuals is odd and for some linear order on the set of alternatives, P is the set of all preference relations that are single-peaked with respect to the order. Then by Proposition 1.4, every collective choice problem in which each individual's preference relation is in P has a strict Condorcet winner. Thus by Proposition 2.1, if every individual's preference relation is in P then for the mechanism that asks each individual to report a preference relation in P and selects the strict Condorcet winner for the reported preference profile, no individual can do better than report her true preference relation. In assessing the significance of this result, a reasonable question to ask is: how could a mechanism designer know that every individual's preference relation is single-peaked?

Suppose that the mechanism designer does not want to pre-judge the character of the individuals' preferences, and thus restricts herself to mechanisms in which the set P of permitted reports is the set of all possible preference relations. Call a collective choice rule *acceptable* if (*i*) no individual is a dictator—for no individual i is the outcome i's favorite alternative regardless of the other individuals' preferences—and (*ii*) whenever every individual's favorite alternative is the same, the rule selects that alternative. The main result in Section 2.2, Proposition 2.3 (known as the Gibbard-Satterthwaite theorem) is a strong contrast to Proposition 2.1. It shows that if there are at least three alternatives and every individual is free to report any preference relation, then no acceptable collective choice rule that selects a single alternative for every profile of reported preference relations is strategy-proof.

This result does not by itself imply that a mechanism designer who has no information about the individuals' preferences can implement only collective choice rules that are not acceptable. Suppose that you want to implement a collective choice rule that assigns a single alternative $f(N,X,\succsim)$ to each collective choice problem $\langle N,X,\succsim\rangle$. Doing so is possible if, when you announce that you will select the alternative given by the collective choice rule, every individual reports her true preference relation. Proposition 2.3 rules out this possibility for any acceptable collective choice rule. Another route you could consider is less direct: you announce that for some mapping g from preference profiles to alternatives that differs from the collective choice rule, you will select the alternative $g(\succsim')$ if the reported preference profile is $\succsim'$, and you design g in such a way that each individual i reports a preference relation $\succsim'_i$, which may differ from her true preference relation $\succsim_i$, such that $g(\succsim') = f(N, X, \succsim)$. That is, you compensate for the distortion in the reported preferences by distorting the mapping from reported preferences to alternatives. The argument in Section 2.3 shows that under certain conditions you cannot gain by such a tactic. Specifically, suppose there is a mechanism that induces the individuals to submit reports that generate the alternative specified by some (single-valued) collective choice rule, where "induces" means that each individual's report is optimal for her regardless of the other individuals' reports. Then Proposition 2.4 (known as a "revelation principle") shows that the collective choice rule is strategy-proof. But then Proposition 2.3 implies that the collective choice rule is not acceptable: it either is dictatorial or does not respect unanimity.

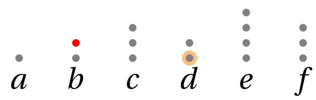
2.1 Strategy-proofness of strict Condorcet winner

Proposition 1.2 in the previous chapter shows that for collective choice problems that have a strict Condorcet winner, the collective choice rule that selects that alternative has singular appeal. Can this rule be implemented if we do not know the individuals' preferences? Consider the mechanism in which each individual reports a preference relation and the alternative chosen is the strict Condorcet winner for the profile of reported preference relations. Can we expect self-interested individuals to report their true preference relations?

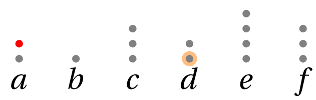
The answer is a qualified yes. Let P be a set of preference relations such that every collective choice problem for which every individual's preference relation is in P has a strict Condorcet winner. (By Proposition 1.4, if the number of individuals is odd, the set of single-peaked preference relations is one such set.) I argue that if each individual is restricted to report a member of P and the alternative selected for any profile of reported preference relations is the strict Condorcet winner for that profile, no individual can do better than report her true preference relation, regardless of the preference relations reported by the other individuals. We say that the collective choice rule that selects the strict Condorcet winner is strategy-proof over P.

If there are two alternatives, the number of individuals is odd, and P is the set of strict preference relations, this result is clear: an individual's saying that she favors b when in fact she favors a either does not affect the strict Condorcet winner or changes it from $\{a\}$ to $\{b\}$, which is worse for her.

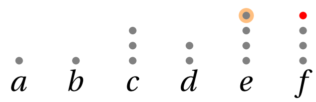
When there are three or more alternatives, the idea behind the result may be conveyed by an example. Suppose that the number of individuals is odd, the alternatives are a, b, c, d, e, and f, and every individual's preferences are single-peaked relative to that ordering. By Proposition 1.4, the strict Condorcet winner of the reported preference profile is the median of the favorite alternatives for the reported profile, so we can think of everyone simply reporting an alternative, rather than her entire preference relation, and the mechanism selecting the median of these reports. Suppose that you favor b and that, given the alternatives reported by everyone else, if you report b then the median reported alternative is d.



What can you do? The crucial point is that you can do nothing to bring the median reported alternative closer to b. If you switch to reporting a or c, the median does not change: the number of reported alternatives less than d remains the same.



If you switch to d the median also does not change. If you switch to e or f, the median might change, but if it does then it changes to e, which is worse for you than c.



Thus by changing your report, you may be able to move the selected alternative away from your favorite, but you can do nothing to move it closer. So whatever the other individuals report, you can do no better than report your favorite alternative.

I now show a general result; subsequently I return to problems with single-peaked preferences. Given a set P of preference relations, the set of collective choice problems in which every individual's preference relation is in P is called the domain generated by P.

Definition 2.1: Domain generated by set of preference relations

Let $\langle N, X\rangle$ be a finite society. For any set P of preference relations over X, the *domain* $\mathscr{D}(N, X, P)$ is the set of all collective choice problems $\langle N, X, \succsim\rangle$ where $\succsim_i \in P$ for each $i \in N$.

The subsequent results concern collective choice rules that assign a single alternative to each collective choice problem. I call such a rule a collective choice *function*.

Definition 2.2: Collective choice function

For any set D of collective choice problems, a *collective choice function for D* is a function that associates with every collective choice problem $\langle N, X, \succsim\rangle$ in D a single member of X (the alternative selected by the rule).

A collective choice function is strategy-proof over a set P of preference relations if, for every collective choice problem in which every individual's preference relation is in P, no individual can induce a better outcome according to her true preference relation by reporting a preference relation in P different from her true preference relation, regardless of the preference relations (in P) submitted by the other individuals.

Definition 2.3: Strategy-proof collective choice function

Let $\langle N, X\rangle$ be a finite society, let P be a set of preference relations over X, and let f be a collective choice function for the domain $\mathscr{D}(N, X, P)$. Then f is *strategy-proof over P* if for every $\langle N, X, \succsim\rangle \in \mathscr{D}(N, X, P)$ and every individual $i \in N$,

$$f(N, X, \succsim) \succsim_i f(N, X, (\succsim'_i, \succsim_{-i})) \quad \text{for every } \succsim'_i \in P,$$

where $(\succsim'_i, \succsim_{-i})$ is the preference profile that differs from $\succsim$ only in that i's preference relation is $\succsim'_i$ rather than $\succsim_i$.

This concept is closely related to that of weak domination for an action in a strategic game. Consider the strategic game in which the players are the individuals, each player's set of actions is a set P of preference relations, and each player i prefers the action profile $\succsim$ to the action profile $\succsim'$ if and only if $f(N, X, \succsim) \succ_i f(N, X, \succsim')$. Then the collective choice rule F is strategy-proof over P if and only if, for each individual i and each preference relation $\succsim'_i \in P$ different from i's true preference relation $\succsim_i$, i's action $\succsim_i$ either weakly dominates $\succsim'_i$ or is equivalent to $\succsim_i'$ in the sense that i is indifferent between the two actions regardless of the other individuals' actions. To see that an individual's true preference relation may be equivalent to another preference relation in this game, suppose that $X = \{a, b, c\}$ and P contains exactly two preference relations: a preferred to b preferred to c, and a preferred to c preferred to b. Then for every preference profile the strict Condorcet winner is a, and in particular the winner is the same whether an individual reports her true preference relation or the other possible preference relation.

Let P be a set of preference relations and suppose that every collective choice problem in the domain $\mathscr{D}(N, X, P)$ has a strict Condorcet winner. Consider the collective choice function that selects the strict Condorcet winner for the submitted preference profile, let that alternative be a if every individual submits her true preference relation, and let b be an alternative that individual i prefers to a. Can i cause b to become the selected alternative by submitting a preference relation that ranks b higher than it is in her true preferences? No: the fact that a is the strict Condorcet winner for the true preference profile, in which i ranks b above a, means that it beats all other alternatives, including b; i's submitting a preference relation in which b is ranked even higher than it is in her true preferences does not change that fact. Individual i may be able to change the outcome to an alternative, say c, that she ranks *below* a, by submitting a preference relation in which c is ranked above rather than below a, but that makes her worse off, not better off. She cannot change the outcome to one that she prefers to a. Thus the collective function that selects the strict Condorcet winner is strategy-proof.

Proposition 2.1: Collective choice function for strict Condorcet domain is strategy-proof

Let $\langle N, X\rangle$ be a finite society and let P be a set of preference relations over X for which every collective choice problem in the domain $\mathscr{D}(N, X, P)$ has a strict Condorcet winner. The collective choice function for $\mathscr{D}(N, X, P)$ that assigns to each collective choice problem in $\mathscr{D}(N, X, P)$ its strict Condorcet winner is strategy-proof over P.

Proof

Let $\langle N, X, \succsim\rangle \in \mathscr{D}(N, X, P)$ be a collective choice problem and let a be its strict Condorcet winner. Suppose that for $\succsim_i' \in P$, the strict Condorcet winner of $\langle N, X, \succsim'\rangle$, where $\succsim' = (\succsim_i', \succsim_{-i})$, is $b \neq a$. Then the number of individuals j for whom $a \succ_j b$ exceeds the number for whom $b \succ_j a$, and the number for whom $b \succ_j' a$ exceeds the number for whom $a \succ_j' b$. The preference profiles $\succsim$ and $\succsim'$ differ only in i's preference relation, so $a \succ_i b$ (and $b \succ'_i a$), establishing the result.

The next exercise shows that this result cannot be extended to domains with Condorcet winners that are not strict.

Exercise 2.1: Non-strict Condorcet winners and strategy-proofness

Suppose that $N = \{1,2,3\}$, $X = \{a,b,c\}$, and $c \succ_1 b \succ_1 a$, $a \succ_2 c \succ_2 b$, and $a \sim_3 b \succ_3 c$. Consider the collective choice rule that assigns to every collective choice problem its set of (not necessarily strict) Condorcet winners. Suppose that individual 1 prefers one set of alternatives to another if and only if she prefers the alternative she likes best in the first set to the one she likes best in the second set. Show that by reporting a preference relation different from her true relation, individual 1 can induce an outcome that she prefers. Construct an example to show that the same is true if she (pessimistically) evaluates a set of alternatives according to the worst alternative for her in the set.

For a set of collective choice problems each of which has a strict Condorcet winner, Proposition 2.1 suggests a way of implementing the rule that selects the strict Condorcet winner even if we do not know the individuals' preferences: ask each individual to submit a preference relation from an appropriate set and select the strict Condorcet winner of the submitted relations. The result establishes that each individual can do no better than submit her true preference relation, regardless of the preference relations submitted by the other individuals.

But if we do not know the individuals' preferences, how can we know whether the collective choice problem they face has a strict Condorcet winner? And how can we select an appropriate set of preference relations from which the individuals are allowed to choose? If the number of individuals is odd, two sufficient conditions for a collective choice problem to have a strict Condorcet winner are that the problem has single-peaked preferences (Proposition 1.4) and that it has single-crossing preferences (Proposition 1.5). If the set of alternatives is naturally one-dimensional (e.g. the amount of money to spend on a public good) we may have reason to believe that one or other of these properties is satisfied. Nevertheless, it is hard to see how we can be sure that is the case, with the consequence that if we ask each individual to submit a preference relation, we need to handle submitted relations that are not single-peaked or do not belong to the collection of relations with the single-crossing property that we have in mind.

One option in a one-dimensional environment is to prohibit non-compliant submissions by restricting each individual to report a single alternative and selecting the median of the submitted alternatives. In a single-peaked or single-crossing domain, the strict Condorcet winner is the median of the individuals' favorite alternatives, so if every individual reports her favorite alternative, this mechanism yields the strict Condorcet winner. I formulate the mechanism as a strategic game.

Definition 2.4: Median-based collective choice game

A *median-based collective choice game* $\langle N, X, \trianglerighteq, (\succsim_i)_{i \in N} \rangle$, where N is a finite set (of individuals) with an odd number of members, X is a set (of alternatives), $\trianglerighteq$ is a linear order on X, and, for each $i \in N$, $\succsim_i$ is a preference relation on X, is a strategic game with the following components.

Players
The set N.

Actions
The set of actions of each player is X.

Preferences
The outcome of an action profile a is the median $m(a)$ of the individuals' actions with respect to $\trianglerighteq$, so each player i prefers the action profile a to the action profile b if and only if $m(a) \succ_i m(b)$.

An implication of Proposition 2.1 is that in such a game in which each individual's preference relation is single-peaked with respect to the order, no individual can do better than choose her favorite alternative, regardless of the alternatives chosen by the other individuals. In fact, each individual's action of choosing her favorite alternative weakly dominates all her other actions.

Proposition 2.2: Collective choice game with single-peaked preferences

For a median-based collective choice game $\langle N, X, \trianglerighteq, (\succsim_i)_{i \in N} \rangle$ in which the set X of alternatives is finite and the preference relation $\succsim_i$ of each individual i is single-peaked with respect to $\trianglerighteq$, each individual's action of choosing her favorite alternative weakly dominates all her other actions.

Proof

Let P be the set of single-peaked preference relations over X. A collective choice problem $\langle N, X, \succsim \rangle$ is in the domain generated by P if N is finite and $\succsim_i \in P$ for each $i \in N$. By Proposition 1.4 the median of the individuals' favorite alternatives is the strict Condorcet winner of any such problem.

Thus by Proposition 2.1 the collective choice rule that assigns the median of the individuals' favorite alternatives to each collective choice problem in the domain generated by P is strategy-proof. That is, for each individual i and each alternative b different from her favorite alternative, a^*, the action a^* either weakly dominates b or is equivalent to b in the collective choice game. But no action is equivalent to a^*: for any action $b \neq a^*$, if the other individuals' actions are equally split between a^* and b, then for i's action a^* the median action is a^* and for her action b the median action is b.

Exercise 2.2: Collective choice game with preferences not single-peaked

Given an example of a median-based collective choice game for which the preference profile is not single-peaked and the action profile in which each individual chooses her favorite alternative is not a Nash equilibrium.

Exercise 2.3: Game in which outcome is smallest chosen alternative

Consider a variant of a median-based collective choice game in which the outcome is the smallest chosen alternative (according to the ordering of alternatives) rather than the median. Show that if each individual's preference relation is single-peaked with respect to the ordering of alternatives, each individual's action of choosing her favorite alternative weakly dominates all her other actions, as it does in Proposition 2.2.

The models I have presented so far assume that all individuals participate in the mechanisms. If participation is costly, they may not. To communicate her preferences, an individual may have to attend a meeting or file a report. These activities take time and effort, and some individuals may decide that the expected return does not justify the cost. If so, which information is lost, and how is the chosen alternative affected? The next two exercises ask you to analyze examples.

Exercise 2.4: Collective choice game with costly reporting

Consider a variant of a median-based collective choice game in which reporting an alternative is optional, and an individual who does so incurs a cost. For simplicity, assume that each individual is restricted to either report her favorite alternative or not submit a report. Assume also that alternatives are real numbers, the number of individuals is $2k+1$ for some positive integer k, and each integer from $-k$ to k is the favorite alternative of exactly one individual. Assume further that the outcome is the mean of the two medians of the submitted reports if the number of individuals who submit reports is positive and even, and the outcome is 0 if no individual submits a report. Finally, assume that the payoff of each individual i is $-|x - x_i^*| - c$ if she submits a report and $-|x - x_i^*|$ if she does not, where x_i^* is her favorite alternative and c, the cost of reporting, is a positive number. Show that for some positive number h, the game has a Nash equilibrium in which an individual submits a report if and only if her favorite alternative is at most $-h$ or at least h.

Exercise 2.5: Collective choice game with shareable reporting costs

Consider a variant of the game in the previous exercise in which each individual may report any alternative (she is not restricted to reporting her favorite alternative) and the cost she incurs is decreasing in the number of other individuals who report the same alternative (but is always positive). Show that in a Nash equilibrium (*i*) no more than four distinct alternatives are reported, (*ii*) if one alternative is reported then it is reported by exactly one individual, is the individual's favorite alternative, and differs from 0, and (*iii*) if three or four alternatives are reported then exactly one individual reports each middle alternative.

2.2 Strategy-proofness for domain of all preference profiles

Is any collective choice function strategy-proof for the domain of all preference profiles? If there are two alternatives, the number of individuals is odd, and the individuals' preferences are strict, then the alternative favored by a majority is a strict Condorcet winner, so that the collective choice rule that assigns this alternative is strategy-proof over the set of all strict preference relations by Proposition 2.1. Other rules are strategy-proof over this set also. For example, denote the alternatives a and b and let k be a nonnegative integer at most equal to the number of individuals. Then the rule that assigns a to a problem if and only if at least k individuals favor a is strategy-proof.

If there are three or more alternatives, a dictatorship is strategy-proof over the set of all preference relations: if, for some individual i, the rule selects i's favorite alternative then no individual can induce an outcome she prefers by reporting any preference relation different from her own. I now show (Proposition 2.3) that if there are three or more alternatives, then among collective choice functions that respect the individuals' unanimous agreement regarding the best alternative, dictatorship is the only one that is strategy-proof over the set of all strict preference profiles.

Definition 2.5: Unanimous collective choice function

Let D be a set of collective choice problems for which every individual's preference relation is strict. A collective choice function f for D is *unanimous* if for any collective choice problem $\langle N, X, \succsim\rangle \in D$,

$$x \text{ is the favorite alternative in } X \text{ for } \succsim_i \text{ for every } i \in N \;\Rightarrow\; f(N, X, \succsim) = x.$$

The result is named for its originators, Allan Gibbard and Mark A. Satterthwaite. The proof I present uses Arrow's impossibility theorem (Proposition 1.9). For any collective choice function f, it defines a preference aggregation function G and shows that if f is unanimous and strategy-proof then G satisfies the Pareto property and independence of irrelevant alternatives. Thus by Arrow's theorem G is dictatorial, which implies that f is also dictatorial.

Proposition 2.3: Gibbard-Satterthwaite theorem

Let $\langle N, X\rangle$ be a finite society for which X contains at least three alternatives and let P be the set of all strict preference relations over X. Let f be a collective choice function for the domain $\mathscr{D}(N, X, P)$ generated by P. If f is unanimous and strategy-proof over P then it is a dictatorship: for some individual $i^* \in N$, $f(N, X, \succsim)$ is the favorite alternative of individual i^* for every profile $\succsim \in P$.

Proof

Let f be a unanimous and strategy-proof collective choice function for the domain generated by P. Throughout the argument, every preference relation is strict and every preference profile consists of strict preference relations.

Step 1 *Let $f(N, X, \succsim) = x$ and let $\succsim'$ be a preference profile that differs from $\succsim$ only in the preference relation of individual j and, for some alternative $y \neq x$, (i) $a \succ'_j b$ if and only if $a \succ_j b$ for all alternatives a and b different from y and (ii) $y \prec_j z$ and $y \succ'_j z$ for some alternative z. (That is, $\succsim'_i$ differs from $\succsim_i$ in that y is raised relative to at least one other alternative while the ordering of all other alternatives is maintained.) Then $f(N, X, \succsim') \in \{x, y\}$.*

Proof. Suppose to the contrary that $f(N, X, \succsim') = w$ for some $w \notin \{x, y\}$. If

$w \succ_j x$ then f is not strategy-proof because

$$f(N,X,(\succsim'_j,\succsim_{-j})) = f(X,N,\succsim') = w \succ_j f(N,X,\succsim)$$

(if j's preference relation is $\succsim_j$ she is better off reporting $\succsim'_j$ than reporting $\succsim_j$ when every other individual i reports $\succsim_i$). If $x \succ_j w$ then $x \succ'_j w$ (because x and w differ from y) so that f is also not strategy-proof because

$$f(N,X,(\succsim_j,\succsim'_{-j})) = f(X,N,\succsim) \succ'_j w = f(N,X,\succsim')$$

(if j's preference relation is $\succsim'_j$ she is better off reporting $\succsim_j$ than reporting $\succsim'_j$ when every other individual i reports $\succsim'_i$). ◁

Step 2 *For any alternatives x and y and any preference profile $\succsim$ in which x and y are the top two alternatives for every individual, $f(N,X,\succsim) \in \{x,y\}$.*

Proof. Assume to the contrary that there is a preference profile $\succsim$ for which x and y are the top two alternatives for every individual but $f(N,X,\succsim) \notin \{x,y\}$. Let $\succsim$ be such a profile with the maximal number of individuals who prefer x to y, say k, among such profiles. Then $k < |N|$ because by unanimity we have $f(N,X,\succsim) = x$ if x is the favorite alternative of every individual. Let $f(N,X,\succsim) = z$, let j be an individual for whom $y \succ_j x$, and let $\succsim'_j$ be a preference relation for which x is at the top and y is ranked second. Then the number of individuals who prefer x to y according to $(\succsim'_j,\succsim_{-j})$ is $k+1$, so that $f(N,X,(\succsim'_j,\succsim_{-j})) \in \{x,y\}$, which, given $x \succ_j z$ and $y \succ_j z$, contradicts the strategy-proofness of f. ◁

Step 3 *Let $\succsim$ and $\succsim'$ be preference profiles for which x and y are the top two alternatives for every individual and for every $i \in N$ we have $x \succ_i y$ if and only if $x \succ'_i y$. Then $f(N,X,\succsim) = f(N,X,\succsim')$.*

Proof. By Step 2, $f(N,X,\succsim) \in \{x,y\}$. Without loss of generality assume that $f(N,X,\succsim) = x$. We can transform $\succsim$ into $\succsim'$ by a sequence of moves, at each of which we raise one alternative, other than x or y, in one individual's preferences, keeping x and y as the top two alternatives for all individuals. By Step 1 the alternative given by f after each move is either the raised alternative, which is not x or y, or the alternative given by f before the move. By Step 2 the alternative given by f after each move is either x or y. Thus the alternative given by f after every move, $f(N,X,\succsim')$, is x. ◁

Step 4 *For any alternatives a and b and any preference profile $\succsim$, for each individual i define $\succsim_i^{ab}$ to be the preference relation obtained from $\succsim_i$ by moving a and b to the top, keeping them in the same order as they are in $\succsim_i$ and not changing the order of the other alternatives. For any preference profile $\succsim$, define the binary relation $\trianglerighteq$ on X by $x \trianglerighteq y$ if and only if $f(N, X, \succsim^{xy}) = x$, for all $x \in X$ and $y \in X$. The binary relation $\trianglerighteq$ is a strict preference relation.*

Proof. From Step 2, for all alternatives x and y we have $f(N, X, \succsim^{xy}) \in \{x, y\}$, so that either $x \trianglerighteq y$ or $y \trianglerighteq x$, and hence $\trianglerighteq$ is complete.

To verify that $\trianglerighteq$ is transitive, assume to the contrary that there exist alternatives a, b, and c for which $a \trianglerighteq b \trianglerighteq c \trianglerighteq a$. Consider the profile $\succsim''$ obtained from $\succsim$ by moving a, b, and c to the top, preserving their order, in every individual's preference relation. By an argument analogous to the proof of Step 2, $f(N, X, \succsim'') \in \{a, b, c\}$. Without loss of generality, let $f(N, X, \succsim'') = a$. Now let $\gtrsim$ be the preference profile obtained from $\succsim''$ by moving b to the third position in all preferences. The conclusions of the following two arguments are contradictory. (i) By Step 2, $f(N, X, \gtrsim) \in \{a, c\}$ because the top two alternatives in every preference relation $\gtrsim_i$ are a and c. The preference profile $\succsim''$ may be obtained from $\gtrsim$ by a sequence of changes in each of which b is raised in one individual's preferences. Thus by Step 1, $f(N, X, \succsim'') \in \{b, x\}$, where $x = f(N, X, \gtrsim)$. Given that $f(N, X, \succsim'') = a$, we have $x = a$, so that $f(N, X, \gtrsim) = a$. (ii) By definition, given $c \trianglerighteq a$ we have $f(N, X, \succsim^{ac}) = c$. For each individual the relative order of a and c in the profiles $\gtrsim$ and $\succsim$ is the same, so that Step 3 applied to $\succsim^{ac}$ and $\gtrsim$ imply that $f(N, X, \gtrsim) = c$.

Finally, given $\succsim^{xy} = \succsim^{yx}$ for all alternatives x and y, if $x \trianglerighteq y$ then it is not the case that $y \trianglerighteq x$, so that the ordering is strict. ◁

Step 5 *Let G be the preference aggregation function for $(\langle N, X\rangle, P)$ that maps a preference profile $\succsim$ into a preference relation $\trianglerighteq$ as described in Step 4. This preference aggregation function is dictatorial.*

Proof. I argue that G satisfies the conditions of Proposition 1.9 (Arrow's impossibility theorem).

Suppose that $x \succ_i y$ for all $i \in N$. Then x is the favorite alternative of every individual in the preference profile $\succsim^{xy}$ defined in Step 4, so that because F is unanimous, $f(N, X, \succsim^{xy}) = x$, and hence $x \trianglerighteq y$. Thus G satisfies the Pareto property.

Let x and y be two alternatives and let $\succsim$ and $\succsim'$ be two preference profiles for which for every $i \in N$ we have $x \succ_i y$ if and only if $x \succ'_i y$. Then by Step 3 we have $f(N,X,\succsim^{xy}) = f(N,X,\succsim'^{xy})$, so that $x \trianglerighteq y$ if and only if $x \trianglerighteq' y$ where $\trianglerighteq = G(N,X,\succsim)$ and $\trianglerighteq' = G(N,X,\succsim')$. Thus G satisfies independence of irrelevant alternatives.

The conclusion that G is dictatorial follows from Proposition 1.9. ◁

Step 6 *There exists an individual i^* such that $f(N,X,\succsim)$ is the favorite alternative of i^* for all $\succsim \in P$.*

Proof. By Step 5 there is an individual i^* such that $G(N,X,\succsim) = \succsim_{i^*}$ for all $\succsim \in P$. Let $\succsim \in P$, let $f(N,X,\succsim) = x$, and let y be another alternative. By Step 2 we have $f(N,X,\succsim^{xy}) \in \{x,y\}$. The profile $\succsim$ may be obtained from $\succsim^{xy}$ by a sequence of steps in each of which one alternative other than x and y is raised in one individual's preferences. By Step 1, after each step, the alternative selected by f is either the raised alternative or the alternative selected previously. Given that $f(N,X,\succsim) = x$, we conclude that $f(N,X,\succsim^{xy}) = x$. Thus by the definition of G, $x \succ_{i^*} y$. ◁

2.3 General mechanisms

So far I have discussed only mechanisms in which each individual reports a preference relation and the alternative selected is the one given by a collective choice rule for the reported preference profile. More generally, a mechanism designer can ask each individual to select a report from a set of the designer's choosing and base her choice of an alternative in an arbitrary fashion on the profile of submitted reports.

Definition 2.6: Mechanism

Let $\langle N,X\rangle$ be a finite society. A *mechanism* $\langle (S_i)_{i\in N}, g\rangle$ *for* $\langle N,X\rangle$ consists of a set S_i (of *reports*) for each individual $i \in N$ and a function $g : \times_{i\in N} S_i \to X$ (the *outcome function*).

For the mechanisms considered in the previous sections, S_i is a set of preference profiles for $\langle N,X\rangle$ and $g(\succsim)$ is the alternative selected by a collective choice rule for the collective choice problem $\langle N,X,\succsim\rangle$. The question addressed is: for which collective choice rules does every individual optimally report her true preference relation, regardless of the preference relations reported by the other individuals? For the case in which there are at least three alternatives and S_i is the set of all strict preference profiles, the answer given by the Gibbard-Satterthwaite theorem is: among unanimous rules, only dictatorships. Can we do better with a general mechanism?

To be more precise about the meaning of doing better, define a mechanism to implement a collective choice rule in quasi-dominant strategies if, for every preference profile $\succsim$, there is a report profile $\sigma(\succsim)$ such that the outcome $g(\sigma(\succsim))$ of the mechanism is the alternative selected by the collective choice rule for $\succsim$ and no individual can induce an outcome that she prefers by choosing a different report, regardless of the other individuals' reports.

Definition 2.7: Mechanism implementing collective choice function in quasi-dominant strategies

Let $\langle N, X\rangle$ be a finite society, let D be the set of all collective choice problems $\langle N, X, \succsim\rangle$, and let f be a collective choice function for D. The mechanism $\langle (S_i)_{i\in N}, g\rangle$ for $\langle N, X\rangle$ *implements f in quasi-dominant strategies* if for every individual $i \in N$ and every preference relation $\succsim_i$ on X for individual i there exists a report $\sigma_i(\succsim_i) \in S_i$ such that

$$g(\sigma(\succsim)) = f(N, X, \succsim), \tag{2.1}$$

where $\sigma(\succsim) = (\sigma_i(\succsim_i))_{i\in N}$, and

$$g(\sigma_i(\succsim_i), s_{-i}) \succsim_i g(s'_i, s_{-i}) \text{ for all } s'_i \in S_i, \text{ all } s_{-i} \in S_{-i}, \text{ and all } i \in N. \tag{2.2}$$

This concept is sometimes called implementation in dominant strategies. However, the report $\sigma_i(\succsim_i)$ in the definition does not necessarily weakly dominate individual i's other possible reports in the sense of Definition 16.18 because it may not satisfy the second condition in this definition. For this reason I attach the prefix quasi.

The question now is: are there collective choice functions that are not strategy-proof but can be implemented in quasi-dominant strategies? The answer is negative. Here's the argument, illustrated for the case of two individuals in Figure 2.1. If a collective choice function can be implemented in quasi-dominant strategies, then for some collection of sets of permitted reports and some outcome function g, for every preference profile $\succsim$ there is a permitted report profile $\sigma(\succsim)$ for which (*i*) the outcome $g(\sigma(\succsim))$ is the alternative $f(N, X, \succsim)$ and (*ii*) no individual has a different permitted report that induces an outcome she prefers for any list of the other individuals' reports. Now consider the mechanism in which the set of permitted reports of each individual is the set of preference relations and the outcome of the report profile $\succsim$ is $g(\sigma(\succsim))$. We can think of this mechanism as one

$$
\begin{array}{r|c}
 & \sigma_2(\succcurlyeq_2) \\ \hline
\sigma_1(\succcurlyeq_1) & g(\sigma(\succcurlyeq_1,\succcurlyeq_2)) \\
 & \underline{\curlyvee} \\
\sigma_1(\succcurlyeq_1') & g(\sigma(\succcurlyeq_1',\succcurlyeq_2))
\end{array}
\qquad\qquad
\begin{array}{r|c}
 & \succcurlyeq_2 \\ \hline
\succcurlyeq_1 & g(\sigma(\succcurlyeq_1,\succcurlyeq_2)) = f(N,X,(\succcurlyeq_1,\succcurlyeq_2)) \\
 & \underline{\curlyvee} \\
\succcurlyeq_1' & g(\sigma(\succcurlyeq_1',\succcurlyeq_2)) = f(N,X,(\succcurlyeq_1',\succcurlyeq_2))
\end{array}
$$

Mechanism implements collective choice function in quasi-dominant strategies $\Rightarrow$ Collective choice function is strategy-proof

Figure 2.1 An illustration for the case of two individuals of the structure of the argument that if a mechanism implements a collective choice rule in quasi-dominant strategies then the collective choice rule is strategy-proof.

in which each individual i reports $\succcurlyeq_i$ and the mechanism operator then reports $\sigma_i(\succcurlyeq_i)$ to the first mechanism on her behalf. For any given preference profile $\succcurlyeq$, the outcomes of the two mechanisms are the same: $g(\sigma(\succcurlyeq))$. Further, under the first mechanism, no individual i is better off changing her report from $\sigma_i(\succcurlyeq_i)$ to the report $\sigma_i(\succcurlyeq_i')$ she would submit if her preference relation were $\succcurlyeq_i'$, for any $\succcurlyeq_i'$, regardless of the other individuals' reports. Thus in particular i is not better off changing her report from $\sigma_i(\succcurlyeq_i)$ to $\sigma_i(\succcurlyeq_i')$ for the reports the other individuals would choose for any given specification of their preference relations. Thus i is not better off changing her report from $\succcurlyeq_i$ to $\succcurlyeq_i'$ under the second mechanism, regardless of the preference relations reported by the other individuals. That is, the collective choice rule is strategy-proof.

Proposition 2.4: Revelation principle for implementation in quasi-dominant strategies

Let $\langle N, X\rangle$ be a finite society, let D be the set of all collective choice problems $\langle N, X, \succcurlyeq\rangle$, and let f be a collective choice function for D. If some mechanism for $\langle N, X\rangle$ implements f in quasi-dominant strategies then f is strategy-proof over the set of all preference relations over X.

Proof

Suppose that the mechanism $\langle (S_i)_{i\in N}, g\rangle$ for $\langle N, X\rangle$ implements f in quasi-dominant strategies. For each preference profile $\succcurlyeq$ for $\langle N, X\rangle$ and each individual $i \in N$, let $\sigma_i(\succcurlyeq_i)$ be the member of S_i given in Definition 2.7. Then for any preference profile $\succcurlyeq$ and any preference relation $\succcurlyeq_i'$ different from $\succcurlyeq_i$, substitute $s_i' = \sigma_i(\succcurlyeq_i')$ and $s_{-i} = (\sigma_j(\succcurlyeq_j))_{j\in N\setminus\{i\}}$ into (2.2) to obtain

$$g(\sigma(\succcurlyeq)) \succcurlyeq_i g(\sigma(\succcurlyeq_i', \succcurlyeq_{-i})) \text{ for all } \succcurlyeq_i' \text{ and all } i \in N.$$

Using (2.1), this condition is equivalent to

$$f(N,X,\succsim) \succsim_i f(N,X,(\succsim_i', \succsim_{-i})) \text{ for all } \succsim_i' \text{ and all } i \in N,$$

so that f is strategy-proof.

An implication of this result and Proposition 2.3 (the Gibbard-Satterthwaite theorem) is that if a unanimous single-valued collective choice rule for the domain of all strict preference profiles can be implemented in quasi-dominant strategies then it is dictatorial.

Corollary 2.1: Unanimous collective choice rule that can be implemented in dominant strategies is dictatorial

For a finite society in which there are at least three alternatives, any collective choice function for the set of all strict preference profiles that is unanimous and can be implemented in quasi-dominant strategies is dictatorial: for some individual i^* it selects the favorite alternative of i^* for every collective choice problem.

Notes

Proposition 2.1 is due to Black (1948b, 32); it is a special case of Moulin (1980, Proposition 1) (see also Moulin 1988, Lemma 10.3). Proposition 2.3 is due to Gibbard (1973) and Satterthwaite (1975). My presentation of this result and its proof are taken from Osborne and Rubinstein (2023); I am grateful to Ariel Rubinstein for allowing me to use this material. Proposition 2.4 was first established by Gibbard (1973); several versions have been demonstrated in various models subsequently. For discussions of the implementation of a collective choice rule via solution concepts other than equilibrium in dominant strategies, see Osborne and Rubinstein (1994, Sections 10.4 and 10.5) and Austen-Smith and Banks (2005, Section 3.3).

The result in Exercise 2.3 is generalized by Moulin (1980), who characterizes the collective choice rules that are strategy-proof in a single-peaked domain. Saporiti (2009) provides an analogous characterization for a single-crossing domain. Osborne et al. (2000) study a model that generalizes the example in Exercise 2.4 and Osborne and Tourky (2008) study a model that generalizes the example in Exercise 2.5.

Solutions to exercises

Exercise 2.1

If individual 1 reports her true preference relation, then there is a single Condorcet winner, a. (Alternatives a and b tie, a beats c, and c beats b.) If she instead reports the ordering $b \succ_1 c \succ_1 a$ then the Condorcet winners are a and b. (Alternatives a and b tie, and both a and b beat c.) So by reporting a preference relation different from her true preference relation she can induce the outcome $\{a, b\}$, in which the best outcome for her is b.

Now suppose that $N = \{1,2,3\}$, $b \succ_1 a \succ_1 c$, $a \succ_2 b \succ_2 c$, and $c \succ_3 a \sim_3 b$. Then the set of Condorcet winners is $\{a, b\}$. If individual 1 switches to reporting $b \succ_1 c \succ_1 a$, then the set of Condorcet winners becomes $\{b\}$.

Exercise 2.2

Suppose the game has three individuals and three alternatives, a, b, and c, with the ordering $a \lhd b \lhd c$. Individual 1 prefers a to b to c, individual 2 prefers b to a to c, and individual 3 prefers c to a to b. If every individual chooses her favorite alternative, the outcome is b. If individual 3 deviates and chooses a, the outcome changes to a, which she prefers to b.

Exercise 2.3

Suppose that an individual i chooses her favorite alternative, say a_i^*. Denote the smallest alternative chosen by all individuals by $\underline{a}$. If $a_i^* = \underline{a}$ then the outcome is a_i^*, and i can do no better by choosing another alternative. If $a_i^* > \underline{a}$ then i can affect the outcome only by choosing an alternative smaller than $\underline{a}$, which is worse for her than $\underline{a}$, given that her preferences are single-peaked. Thus the first condition in Definition 16.18 is satisfied. Now let a' be an alternative different from a_i^*. If all the individuals other than i choose alternatives larger than a', the outcome is better for i is she chooses a_i^* than if she chooses a', so that the second condition in Definition 16.18 is satisfied.

Exercise 2.4

Let h be a positive integer and consider the action profile in which an individual i submits a report if and only if $|x_i^*| \geq h$. The outcome is 0 (the mean of the two submitted reports that are smallest in absolute value). First suppose that $x_i^* \geq h$. Then i's payoff is $-x_i^* - c$. If she deviates to not submit a report, the outcome changes to $-h$, so that her payoff becomes $-x_i^* - h$. Thus her submission of a report is optimal for her if and only if $h \geq c$. If $x_i^* \leq -h$, a symmetric argument yields the same conclusion. Now suppose that $0 \leq x_i^* < h$. Then i's payoff is $-x_i^*$. If she deviates to submit a report, the outcome changes to her favorite position, x_i^*, and her payoff becomes $-c$. Thus her non-submission of a report is optimal for her if and only if $x_i^* \leq c$. If

$-h < x_i^* \leq 0$, a symmetric argument yields the same conclusion.

In summary, the action profile is a Nash equilibrium if and only if $h \geq c$ and $x_i^* \leq c$ whenever $x_i^* < h$. If c is an integer, these conditions are satisfied if and only if $h = c$ or $h = c + 1$, and if c is not an integer, the conditions are satisfied if and only if h is the smallest integer that is at least c.

Exercise 2.5

Two factors limit the number of distinct alternatives reported. First, by switching from reporting an alternative that is being reported by k individuals to reporting a different one that is being reported by at least k individuals, an individual reduces her reporting cost. Second, some switches in the alternative an individual reports do not affect the outcome—for example, that is the case if both alternatives are less than an alternative that is in turn less than the outcome. These two factors run through the following arguments.

i. Denote the outcome of the equilibrium x^*. First suppose that three or more distinct alternatives greater than x^* are reported. Denote the two largest alternatives reported by x and y, and suppose that the number of individuals who report x is at least the number who report y. Then if an individual who is reporting y switches to reporting x, she reduces her reporting cost and does not affect the outcome, and hence is better off. Thus in any equilibrium at most two distinct alternatives greater than the outcome are reported. A symmetric argument shows that also at most two distinct alternatives less than the outcome are reported.

We conclude that at most five distinct alternatives are reported, and if five are reported then the outcome is the middle reported alternative. In this last case, let x and y be the two smallest reported alternatives, with the number of individuals reporting x at least the number reporting y. Then, as in the previous paragraph, an individual reporting y who switches to report x reduces her reporting cost and does not change the outcome, and hence is better off. Thus at most four distinct alternatives are reported.

ii. If one alternative is reported, it is reported by only one individual, because if two or more individuals report it, any one of them can switch to not reporting without changing the outcome, and thereby save the cost of reporting. The alternative is the individual's favorite because if it is not, she can switch to reporting her favorite, which changes the outcome to that alternative. Her favorite alternative must differ from 0, because if it is 0 she is better off switching to not reporting.

iii. Suppose that three distinct alternatives are reported, $x < y < z$. Suppose that the outcome is greater than y. If two or more individuals report y then an individual who switches from reporting x to reporting y, or vice versa, does not affect the outcome. Thus if two or more individuals report y then if the number who report x is at least the number who report y, an individual who reports y can benefit by switching to report x, and if the number who report y is at least the number who report x, an individual who reports x can benefit by switching to report y. So only one individual reports y. A symmetric argument leads to the same conclusion if the outcome is less than y. If the outcome is equal to y, the number of individuals who report x and the number who report z are at least the number who report y, otherwise an individual reporting x or z is better off switching to reporting y. But then if two or more individuals report y, either the outcome does not change if one of them switches to x or the outcome does not change if one of them switches to z. In both cases the individual who switches is better off, so only one individual reports y.

Now suppose that four distinct alternatives are reported, $w < x < y < z$. If the outcome is at most x an individual benefits from switching from y to z or vice versa, and if the outcome is at least y an individual benefits from switching from w to x or vice versa. So the outcome is the mean of x and y. Then as for the case of three reported alternatives, the number reporting w is at least the number reporting x, so that if two or more individuals report x one of them is better off switching to report w. Hence only one individual reports x. Similarly, only one individual reports y.

II Voting

3 Voting with two alternatives 99
4 Voting with many alternatives: plurality rule 129
5 Sequential pairwise voting 155
6 Ethical voting and expressive voting 187
7 Voting with shared values and asymmetric information 197

3 Voting with two alternatives

3.1 Voting as a strategic game 101
3.2 Costly voting with uncertainty about cost 104
3.3 Turnout and population size with uncertain preferences 118
3.4 Preferences with regret 121

A group of individuals selects one of two alternatives by voting, which may entail a cost. How do the individuals' decisions to vote depend on their preferences and voting costs? How does the fraction of individuals who vote depend on the population size?

Synopsis

Section 3.1 models voting as a strategic game. In the game, an individual's voting for her favorite alternative weakly dominates her voting for the other alternative and, if her cost of voting is zero, also weakly dominates abstention (Proposition 3.1). In particular, if every individual's cost of voting is zero, an action profile is a Nash equilibrium in which no individual's action is weakly dominated if and only if each individual who is not indifferent between the alternatives votes for her favorite alternative (Corollary 3.1). If the individuals' voting costs are positive, most voting games in which the costs are not so high that no individual optimally votes have no Nash equilibrium, rendering the model of little use for understanding voting behavior when voting is costly.

A variant of the model that does have equilibria assumes that each individual knows her own voting cost but is uncertain of the other individuals' costs. Sections 3.2 and 3.3 explore a model in which each individual believes that every other individual's voting cost is drawn independently from a given distribution and each individual's preferences are represented by the expected value of a Bernoulli payoff function. As in the model in Section 3.1, each individual is assumed to know the other individuals' preferences between the alternatives. In every equilibrium of the game I formulate, there is a number c_i^* for each individual i such that i votes for her favorite alternative if her cost is less than c_i^* and abstains if it is greater than c_i^* (Lemma 3.1). Proposition 3.2 characterizes these threshold costs for equilibria in which every individual who favors the same alternative has the same threshold cost (symmetric equilibria) and Proposition 3.3 shows that such an equilibrium exists.

Intuition suggests that if many individuals vote then the probability of an individual's vote affecting the outcome—the probability that among the other individuals the vote is tied or nearly tied—is small, so that only individuals with small voting costs vote. If all voting costs are drawn from distributions with positive lower limits, an implication is that equilibria in which a large number of individuals vote are impossible. Proposition 3.4 formalizes this idea. It gives conditions under which as the number of individuals increases without bound the expected number of individuals who vote remains finite. Section 3.3 explores the idea further in a model in which each individual is uncertain of the other individuals' preferences between the alternatives.

Section 3.4 considers an alternative to the assumption that each individual's preferences are represented by the expected value of a payoff function. Instead, each individual considers, for each of her actions and each list of the other individuals' actions, how much better off she would have been had she chosen a different action. She chooses the action for which the largest value of this gain over all lists of the other individuals' actions is smallest. That is, she chooses the action for which her maximal possible regret is smallest. If she votes, then the outcomes that generate the most regret are that her favored alternative loses or wins by two votes or more, in which case the outcome would have been the same had she not voted (and thereby saved the cost of doing so). If she abstains then the outcomes that generate the most regret are that her favored alternative ties or loses by one vote, so that had she voted she would have increased the probability of her favored alternative's winning by one-half. Thus if her cost of voting is not too high she optimally votes, regardless of the number of individuals, so that this model can generate high turnout even in large populations.

When an individual chooses to abstain in any of these models, she does so because the cost of voting exceeds the benefit from doing so. Another rationalization for abstention is that individuals feel they are insufficiently informed to make a choice, and prefer to delegate the decision to those who are informed. I explore this approach in Chapter 7.

3.1 Voting as a strategic game

I formulate voting as a strategic game. The model includes the option for each individual not to vote and the possibility that voting is costly. The players are individuals, each of whom can vote for one of two alternatives, a and b, or abstain. If one alternative, say z, receives more votes than the other, it wins, and the payoff of each individual i is $u_i(z) - c_i$ if she votes and $u_i(z)$ if she does not, where u_i is a real-valued function on $\{a, b\}$ and c_i is a nonnegative number. If the alternatives receive the same number of votes, the payoff of each individual i is $\frac{1}{2}(u_i(a) + u_i(b)) - c_i$ if she votes and $\frac{1}{2}(u_i(a) + u_i(b))$ if she does not. One rationale for this specification of an individual's payoff in the case of a tie is that in this event each alternative is selected with probability $\frac{1}{2}$ and u_i is a Bernoulli payoff function whose expected value represents the player's preferences over lotteries over outcomes.

Definition 3.1: Two-alternative voting game

The *two-alternative voting game* $\langle N, \{a, b\}, (u_i)_{i \in N}, (c_i)_{i \in N} \rangle$, where N is a finite set (of individuals) with at least two members, a and b are alternatives, each u_i is a real-valued function on $\{a, b\}$, and each c_i is a nonnegative number, is the following strategic game.

Players

The set N.

Actions

For each player i, the set of actions is $\{\textit{vote for } a, \textit{vote for } b, \textit{abstain}\}$.

Payoffs

For any action profile x, denote by $W(x) \subseteq \{a, b\}$ the set of alternatives that receive the most votes: $W(x) = \{z\}$ if more individuals vote for z than for the other alternative and $W(x) = \{a, b\}$ if the same number of individuals vote for each alternative. The payoff of each player i for x is

$$\begin{cases} \sum_{w \in W(x)} u_i(w)/|W(x)| & \text{if } x_i = \textit{abstain} \\ \sum_{w \in W(x)} u_i(w)/|W(x)| - c_i & \text{if } x_i \in \{\textit{vote for } a, \textit{vote for } b\}. \end{cases}$$

3.1.1 Costless voting

A two-alternative voting game with three or more individuals in which every individual's voting cost is zero has many Nash equilibria. For example, any action profile in which every individual chooses (votes for) the same alternative is an equilibrium, because no change in any individual's action affects the outcome. (Remember that in a Nash equilibrium, no change in any player's action makes her better off, but changes may not make her worse off, either.) The next exercise asks you to describe all Nash equilibria in the case that (for simplicity) no individual is indifferent between the alternatives.

Exercise 3.1: Nash equilibria of two-alternative voting game with zero costs

Find all the Nash equilibria of a two-alternative voting game $\langle N, \{a, b\}, (u_i)_{i \in N}, (c_i)_{i \in N} \rangle$ in which $c_i = 0$ and $u_i(a) \neq u_i(b)$ for each $i \in N$.

Among the Nash equilibria are ones in which at least one individual votes for the alternative she likes least. Such equilibria seem implausible, because an individual who votes for such an alternative can gain no possible advantage by doing so and, intuitively, risks influencing the outcome in favor of that alternative. The notion of Nash equilibrium assumes that no individual wavers from her equilibrium action, so that an equilibrium is not affected by such risk. The idea is captured, instead, by the notion of a weakly dominated action: an action a_i for which there is another action b_i that yields i at least as high a payoff as does a_i for *all* actions of the other players and a higher payoff than does a_i for *some* actions of the other players. An individual's voting for her favorite alternative weakly dominates her voting for the other alternative: no matter how the other individuals vote, an individual is not worse off voting for her favorite alternative than voting for the other alternative, and for some configurations of the other individuals' votes she is better off. If an individual's voting cost is zero, her voting for her favorite alternative also weakly dominates abstention.

Proposition 3.1: Weak domination in two-alternative voting game

Let $\langle N, \{a, b\}, (u_i)_{i \in N}, (c_i)_{i \in N} \rangle$ be a two-alternative voting game and let i be an individual for whom $u_i(a) \neq u_i(b)$. Individual i's action of voting for her favorite alternative weakly dominates her action of voting for the other alternative and, if $c_i = 0$, weakly dominates *abstain*.

This result is closely related to Proposition 2.1, but the following proof employs an argument independent of that result.

Proof

Table 3.1 shows an individual's payoffs as a function of her action and the winning margin in favor of a among the other individuals' votes. For an

	winning margin for a among other individuals				
	≥ 2	1	0	-1	≤ -2
vote a	$u_i(a) - c_i$	$u_i(a) - c_i$	$u_i(a) - c_i$	$\frac{1}{2}(u_i(a) + u_i(b)) - c_i$	$u_i(b) - c_i$
vote b	$u_i(a) - c_i$	$\frac{1}{2}(u_i(a) + u_i(b)) - c_i$	$u_i(b) - c_i$	$u_i(b) - c_i$	$u_i(b) - c_i$
abstain	$u_i(a)$	$u_i(a)$	$\frac{1}{2}(u_i(a) + u_i(b))$	$u_i(b)$	$u_i(b)$

Table 3.1 The payoffs of individual i in a two-alternative voting game as a function of her action and the winning margin for a among the other individuals.

individual who prefers a to b, for each column, the entry in the top cell is at least the entry in the middle cell, and for the second and fourth columns, the entry in the top cell is larger than the entry in the middle cell. So for such an individual, voting for a weakly dominates voting for b. If $c_i = 0$ then for each column, the entry in the top cell is at least the entry in the bottom cell, and for the third and fourth columns, the entry in the top cell is larger than the entry in the bottom cell, so that voting for a weakly dominates abstention. Symmetric arguments apply to an individual who prefers b to a.

An immediate corollary of this result is that in any Nash equilibrium of a two-alternative voting game with costless voting in which no individual's action is weakly dominated, every individual votes for her favorite alternative.

Corollary 3.1: Nash equilibrium in weakly undominated actions in two-alternative voting game with zero costs

Let $\langle N, \{a, b\}, (u_i)_{i \in N}, (c_i)_{i \in N} \rangle$ be a two-alternative voting game in which $c_i = 0$ for all $i \in N$. An action profile is a Nash equilibrium in which no individual's action is weakly dominated if and only if every individual who is not indifferent between the alternatives votes for her favorite alternative.

Because of this result, most models in which individuals vote between two alternatives and voting is costless assume that no individual uses a weakly dominated action.

3.1.2 Costly voting

When voting costs are positive, the nature of the equilibria is completely different: most two-alternative voting games in which every individual's voting cost is positive have no Nash equilibria unless the voting costs are sufficiently high, when they have equilibria in which no one votes or a single individual does so.

When voting is costly, an individual optimally votes only if doing so makes a difference to the outcome—if her vote is *pivotal*. If the numbers of votes for the two alternatives differ, then no vote on the losing side is pivotal, so in any equilibrium with votes for both alternatives, the alternatives tie. Consider such an equilibrium. If individual i votes for a, then by switching to abstention she changes the outcome from a tie to b. Thus her voting for a is optimal only if $\frac{1}{2}(u_i(a)+u_i(b))-c_i \geq u_i(b)$, or equivalently $\frac{1}{2}(u_i(a)-u_i(b)) \geq c_i$. Similarly, her voting for b is optimal only if $\frac{1}{2}(u_i(b)-u_i(a)) \geq c_i$, and her abstaining is optimal only if $\frac{1}{2}(u_i(a)+u_i(b)) \geq \max\{u_i(a), u_i(b)\} - c_i$.

Thus if $c_i \neq \frac{1}{2}|u_i(a)-u_i(b)|$ for all $i \in N$ then in any equilibrium in which both alternatives receive votes, an individual i votes for a if and only if $\frac{1}{2}(u_i(a)-u_i(b)) > c_i$ and votes for b if and only if $\frac{1}{2}(u_i(b)-u_i(a)) > c_i$, so that given that the alternatives tie,

$$|\{i \in N : \tfrac{1}{2}(u_i(a)-u_i(b)) > c_i\}| = |\{i \in N : \tfrac{1}{2}(u_i(b)-u_i(a)) > c_i\}|.$$

Most two-alternative voting games do not satisfy this condition, so that they have no Nash equilibria in which both alternatives receive votes.

Games in which the voting costs are high enough have equilibria in which no one votes and may have equilibria in which one individual votes. These equilibria are of little interest.

These arguments lead to the conclusion that the notion of Nash equilibrium for a two-alternative voting game with positive voting costs is not useful for understanding voting behavior. One option is to consider instead the notion of mixed strategy equilibrium. However, interpretations of mixed strategy equilibria do not fit many environments that voting games are intended to model. Instead, I discuss a related approach, in which each individual is uncertain of the other individuals' voting costs.

3.2 Costly voting with uncertainty about cost

Suppose that each individual knows her own voting cost but is uncertain of the the other individuals' voting costs. In this section I specify a model, show that it has an equilibrium, and characterize its equilibria. A focus of the analysis is the fraction of individuals who vote in a large population. An individual in the model is motivated to vote by the possibility that doing so affects the outcome, which happens only if the votes for the alternatives among the other individuals are tied or almost tied. Intuition suggests that if the other individuals' characteristics are uncertain then a tie or near tie is unlikely if a large number of the other individuals vote. As a consequence, voting is optimal for the remaining individual only if her voting cost is small, so that when the number of individuals is large, only

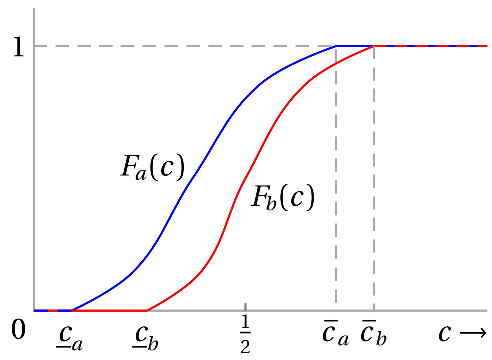


Figure 3.1 An example of distributions of voting costs of the form assumed in a two-alternative voting game with uncertain voting costs.

individuals with small costs vote and thus turnout is low. The analysis leads to Proposition 3.4, which gives conditions under which this intuition is correct.

I retain the assumption in a two-alternative voting game that the number of individuals who favor each alternative is known. For many elections this assumption is not reasonable, but the resulting model captures the fact that each individual is uncertain about the other individuals' voting behavior, and limiting the uncertainty to one source keeps the model relatively simple.

The game-theoretic model of a Bayesian game accommodates these assumptions. It differs from a strategic game in that it includes a specification of the uncertainty that each individual faces (and a payoff function for each individual that applies to the uncertain environment). It models the uncertainty by specifying a set of possible states, the information each individual has about the state, and each individual's belief regarding aspects of the state about which she is not informed.

The uncertainty in the game we wish to analyze concerns the individuals' voting costs, so a state is a profile of such costs. I assume that every individual i knows her own cost, c_i, and believes that the cost of every individual who favors a given alternative is drawn from the same distribution. I assume specifically that the voting cost of every individual who favors $x \in \{a, b\}$ is drawn independently from a nonatomic distribution F_x with support $[\underline{c}_x, \bar{c}_x]$ where $0 < \underline{c}_x < \bar{c}_x$. The assumption that the distributions are nonatomic means that no single cost has a positive probability, so that the probability distribution functions are continuous, like those in Figure 3.1.

Definition 3.2: Two-alternative voting game with uncertain voting costs

A *two-alternative voting game with uncertain voting costs* $\langle(N_a, N_b), \{a, b\}, (F_a, F_b)\rangle$, where N_a and N_b are finite sets, a and b are alternatives, and F_a and F_b are nonatomic probability distribution functions whose supports are intervals of positive numbers, is the following Bayesian game, where $[\underline{c}_x, \bar{c}_x]$ (with $0 < \underline{c}_x < \frac{1}{2} < \bar{c}_x$) denotes the support of F_x for each $x \in \{a, b\}$.

Players

The set $N = N_a \cup N_b$ (N_a consists of individuals who favor a and N_b consists of individuals who favor b).

States

The set of states is the set of profiles $(c_j)_{j\in N}$ of voting costs, with $c_j \in [\underline{c}_a, \bar{c}_a]$ for every player $j \in N_a$ and $c_j \in [\underline{c}_b, \bar{c}_b]$ for every player $j \in N_b$. An individual with a given voting cost is referred to as a *type* of the individual.

Actions

The set of actions of each player is $\{$*vote for a*, *vote for b*, *abstain*$\}$.

Signals

The signal function τ_i of each player i is given by $\tau_i((c_j)_{j\in N}) = c_i$ (every individual knows her own cost, but has no information about any other individual's cost).

Prior beliefs

Every player believes that for each $x \in \{a, b\}$ the voting cost of each player $i \in N_x$ is drawn from F_x, and that each player's cost is drawn independently of every other player's cost.

Payoffs

The Bernoulli payoff function of each player over the set of pairs of action profiles and states is defined as follows.

For each player $i \in N_a$ (who favors a), let $u_i(a) = 1$ and $u_i(b) = 0$, and for each $i \in N_b$ (who favors b), let $u_i(b) = 1$ and $u_i(a) = 0$. For any state $(c_j)_{j\in N}$ and any $x \in \{a, b\}$, if more players choose *vote for x* than *vote for y*, the other alternative, then the Bernoulli payoff of player i is $u_i(x) - c_i$ if she votes and $u_i(x)$ if she abstains; if the number of players who vote for each alternative is the same, then $\frac{1}{2}(u_i(a) + u_i(b))$ replaces $u_i(x)$ in each case.

Comments

- The assumption that each individual's preferences are represented by the expected value of a Bernoulli payoff function is not innocuous. Section 3.4 presents an analysis based on a different assumption about the individuals' preferences.

- The assumption that $u_i(x) = 1$ and $u_i(x) = 0$ for an individual i who favors x means that no individual is indifferent between the alternatives. It implies that a voting cost of 1 is equivalent to the difference between an individual's payoff to her favorite alternative and her payoff to the other alternative. Given that the payoffs are the same in every game, they do not appear as parameters in the specification $\langle(N_a, N_b), \{a, b\}, (F_a, F_b)\rangle$.

- The assumption that $\underline{c}_a$ and $\underline{c}_b$ are less than $\frac{1}{2}$ rules out equilibria in which no type of any individual votes, because if no one votes then an individual who deviates to voting for her favorite alternative changes the outcome from a tie to a win for her favorite alternative, changing her payoff from $\frac{1}{2}$ to $1 - c$, which is a increase if $c < \frac{1}{2}$.

- The assumption that $\bar{c}_a$ and $\bar{c}_b$ are greater than $\frac{1}{2}$ means that with positive probability an individual's cost is high enough that she does not optimally vote even if her vote is pivotal. (Perhaps she faces a medical emergency at the time of the vote.) This assumption rules out equilibria in which every type of every individual votes, because if everyone votes then an individual who deviates from voting for x to abstention either does not change the outcome or changes it from a win for x to a tie, or from a tie to a win for the other alternative, and hence increases her payoff by at least $\frac{1}{2} - (1 - c) = 0 - (\frac{1}{2} - c) = c - \frac{1}{2}$, which is positive if $c > \frac{1}{2}$.

An individual's strategy specifies an action for each of her possible types. That is, for each $x \in \{a, b\}$, a strategy for an individual who favors x is a function from $[\underline{c}_x, \bar{c}_x]$ to $\{\textit{vote for } a, \textit{vote for } b, \textit{abstain}\}$. A strategy profile s^* is a Nash equilibrium if the action $s_i(c)$ of each type c of each individual i is optimal given i's belief about the other individuals' types and the action $s_j^*(c')$ chosen by each type c' of every other individual j, which together generate a probability distribution in i's mind over the combination of actions of the other individuals.

If you are not a frequent user of the notion of a Bayesian game, you may wonder why we require an equilibrium strategy to specify an action for every type of every individual, given that each individual knows her own type. The reason is that in an equilibrium we want the actions that each individual believes each type of every other individual will take to be the ones that the type would in fact take. The way we do that is to specify, and verify the optimality of, each individual's strategy, which includes a specification of the action chosen by each of her possible types. No individual knows the type of any other individual, and the actions an individual's strategy specifies for types different from her actual type function as the other individuals' beliefs about the actions the individual would take were she to have these types.

Given an individual's belief about the other individuals' actions, which action should she choose? Intuition suggests that she should vote for her favorite alternative if her voting cost is low and abstain if her voting cost is high. That is, she should employ a threshold strategy.

Definition 3.3: Threshold strategy in two-alternative voting game with uncertain voting costs

Let $\langle (N_a, N_b), \{a, b\}, (F_a, F_b) \rangle$ be a two-alternative voting game with uncertain voting costs and denote the support of F_x by $[\underline{c}_x, \overline{c}_x]$ for each $x \in \{a, b\}$. For each $x \in \{a, b\}$ and each player $i \in N_x$ (who favors x), a strategy $s_i : [\underline{c}_x, \overline{c}_x] \to \{\textit{vote for a}, \textit{vote for b}, \textit{abstain}\}$ of player i is a *threshold strategy* if for some number $c_i^* \in [\underline{c}_x, \overline{c}_x]$ (the *threshold* for individual i) we have

$$s_i(c_i) = \begin{cases} \textit{vote for } x & \text{if } \underline{c}_x \le c_i < c_i^* \\ \textit{abstain} & \text{if } c_i^* < c_i \le \overline{c}_x \end{cases}$$

and $s_i(c_i^*) \in \{\textit{vote for } x, \textit{abstain}\}$.

The intuition that each individual's best response to any strategies of the other individuals is a threshold strategy is confirmed by the following analysis.

Consider an individual who favors a. For each of her actions, the outcome generated by each combination of actions of the other individuals, and hence her payoff, depends only on the winning margin for a among the other individuals' actions. Her payoffs when her voting cost is c are given in Table 3.2. For $c > 0$, these payoffs have several relevant features.

- Abstention strictly dominates voting for b.
- If $c > \frac{1}{2}$ then abstention strictly dominates voting for either alternative.
- If $c < \frac{1}{2}$ then voting for a is better than abstaining if and only if the winning margin for a among the other individuals is 0 or -1 (the highlighted cells in Table 3.2), and the gain in payoff to switching from abstention to voting for a for these two winning margins is the same, equal to $\frac{1}{2} - c$.

Specifically, for an individual with voting cost $c > 0$ who favors a, a vote for a is a

	winning margin for a among other individuals				
	≥ 2	1	0	-1	≤ -2
vote for a	$1-c$	$1-c$	$1-c$	$\frac{1}{2}-c$	$-c$
vote for b	$1-c$	$\frac{1}{2}-c$	$-c$	$-c$	$-c$
abstain	1	1	$\frac{1}{2}$	0	0

Table 3.2 The payoffs of an individual of type c who favors a in a two-alternative voting game with uncertain voting costs as a function of her action and the winning margin for a among the other individuals. For the highlighted cells, the individual's voting for a and abstaining generate different outcomes.

best response to the other individuals' strategies if and only if

$$\tfrac{1}{2}\left[\Pr(\text{tie among others' votes}) + \Pr(a \text{ loses by 1 among others' votes})\right] \geq c$$

and abstention is a best response if and only if this inequality is reversed ($\leq$); voting for b is never a best response. In particular, if voting for a is a best response for an individual when her cost is c, it is a best response also when her cost is less than c, and if abstention is a best response when her cost is c, it is a best response also when her cost exceeds c. The same considerations apply to an individual who favors b, so every best response of any individual is a threshold strategy.

Lemma 3.1: Equilibrium of two-alternative voting game with uncertain voting costs

In every Nash equilibrium of a two-alternative voting game with uncertain voting costs, each individual's strategy is a threshold strategy.

I consider only Nash equilibria that are symmetric in the sense that every individual who favors a uses the same strategy and every individual who favors b uses the same strategy.

Definition 3.4: Symmetric equilibrium of two-alternative voting game with uncertain voting costs

A Nash equilibrium s^* of a two-alternative voting game with uncertain voting costs $\langle (N_a, N_b), \{a, b\}, (F_a, F_b)\rangle$ is *symmetric* if there are strategies s_a and s_b such that $s_i^* = s_a$ for every individual $i \in N_a$ and $s_i^* = s_b$ for every individual $i \in N_b$.

How do the thresholds c_a and c_b in a symmetric equilibrium depend on the individuals' characteristics? Given the other individuals' strategies, an individual who favors x is indifferent between voting for x and abstaining when her cost is c_x. Her expected payoff in each case depends on the other individuals' strategies and her belief about their types. I now explore these payoffs in detail.

Consider a profile of threshold strategies in which some types of each individual vote and some abstain; let i be an individual who favors a. What probability does i assign to the event that a and b are tied among the other individuals' votes? Only individuals who favor a vote for a, and only those who favor b vote for b, so the maximum number k such that a and b both get k votes among the other individuals is $\min\{n_a - 1, n_b\}$, where $n_a = |N_a|$ and $n_b = |N_b|$. The probability i assigns to the event that an individual who favors a votes is the probability that the individual's cost is at most c_a, which is $F_a(c_a)$, and the probability she assigns to the event that an individual who favors b votes is $F_b(c_b)$. Thus the probability i assigns to a tie between a and b among the other individuals' votes is the sum from $k = 0$ to $k = \min\{n_a, -1, n_b\}$ of the probability the voting cost of k of the other $n_a - 1$ individuals who favor a is at most c_a and the voting cost of k of the individuals who favor b is at most c_b.

To write a compact expression for this probability, for any positive integer n, integer l with $0 \le l \le n$, and number $p \in [0,1]$), denote by $B(n,l,p)$ the probability of exactly l successes in n independent trials when the probability of success on each trial is p:

$$B(n,l,p) = \binom{n}{l} p^l (1-p)^{n-l}.$$

Then the probability that i assigns to a tie among the other individuals' votes when each of the other individuals who favors a votes with probability p_a and each of the individuals who favors b votes with probability p_b is

$$P_a^0(p_a, p_b, n_a, n_b) = \sum_{k=0}^{\min\{n_a-1, n_b\}} B(n_a - 1, k, p_a) B(n_b, k, p_b).$$

Similarly, the probability i assigns to a's losing by one vote among the other individuals is

$$P_a^1(p_a, p_b, n_a, n_b) = \sum_{k=0}^{\min\{n_a-1, n_b-1\}} B(n_a - 1, k, p_a) B(n_b, k+1, p_b).$$

Thus i's expected gain from voting is

$$G_a(p_a, p_b, n_a, n_b) = \tfrac{1}{2}\left[P_a^0(p_a, p_b, n_a, n_b) + P_a^1(p_a, p_b, n_a, n_b)\right]. \qquad (3.1)$$

She optimally votes if this gain is greater than her cost, abstains if it is less than her cost, and is indifferent between voting and abstaining if it is equal to her cost.

For the threshold strategies we are considering, an individual who favors a votes with probability $F_a(c_a)$, the probability that her voting cost is less than c_a, and an individual who favors b votes with probability $F_b(c_b)$, so the condition in terms of c_a and c_b for a symmetric equilibrium in which some types of each individual who favors a vote and some abstain is $c_a^* = G_a(F_a(c_a^*), F_b(c_b^*), n_a, n_b)$, the condition for an equilibrium in which all types abstain is $c_a^* \geq G_a(F_a(c_a^*), F_b(c_b^*), n_a, n_b)$, and the condition for an equilibrium in which all types vote is $c_a^* \leq G_a(F_a(c_a^*), F_b(c_b^*), n_a, n_b)$.

For an individual who favors b, the expressions differ only in that the roles of a and b are interchanged, so a pair of threshold strategies with thresholds c_a^* and c_b^* is a symmetric equilibrium if and only if

$$
\begin{aligned}
c_a^* \begin{Bmatrix} \geq \\ = \\ \leq \end{Bmatrix} G_a(F_a(c_a^*), F_b(c_b^*), n_a, n_b) &\begin{cases} \text{if } F_a(c_a^*) = 0 \\ \text{if } 0 < F_a(c_a^*) < 1 \\ \text{if } F_a(c_a^*) = 1 \end{cases} \\
c_b^* \begin{Bmatrix} \geq \\ = \\ \leq \end{Bmatrix} G_b(F_a(c_a^*), F_b(c_b^*), n_a, n_b) &\begin{cases} \text{if } F_b(c_b^*) = 0 \\ \text{if } 0 < F_b(c_b^*) < 1 \\ \text{if } F_b(c_b^*) = 1. \end{cases}
\end{aligned}
\qquad (3.2)
$$

By Lemma 3.1, each individual's strategy in a Nash equilibrium is a threshold strategy, so this argument establishes the following result.

Proposition 3.2: Symmetric equilibrium of two-alternative voting game with uncertain voting costs

A strategy profile s^* of a two-alternative voting game with uncertain voting costs $\langle (N_a, N_b), \{a, b\}, (F_a, F_b) \rangle$ is a symmetric Nash equilibrium if and only if for numbers c_a^* and c_b^* that satisfy (3.2) and threshold strategies s_a and s_b with thresholds c_a^* and c_b^* respectively we have $s_i^* = s_a$ for every individual $i \in N_a$ and $s_i^* = s_b$ for every individual $i \in N_b$.

Exercise 3.2: Two-alternative voting game with uncertain voting costs

a. Find a Nash equilibrium of a two-alternative voting game with uncertain voting costs $\langle (N_a, N_b), \{a, b\}, (F_a, F_b) \rangle$ in which $|N_a| = |N_b| = 2$ and F_a and F_b are uniform on $[0, 1]$ (so that $F_a(x) = F_b(x) = x$ for $x \in [0, 1]$).

b. If F_a is uniform on $[0, 1]$, for which distributions F_b, if any, does the game have a Nash equilibrium in which no type of any individual in N_b votes?

Informal analysis for $n_a = n_b$ and $F_a = F_b$

To get an idea of the nature of a symmetric equilibrium and the way it varies with the number of individuals, suppose first that the number of individuals who favor a is the same as the number who favor b and the probability distribution of every individual's voting cost is the same: $n_a = |N_a| = n_b = |N_b| = m$ $(= \frac{1}{2}|N|)$ and $F_a = F_b = F$, with $\underline{c}_a = \underline{c}_b = \underline{c}$ and $\bar{c}_a = \bar{c}_b = \bar{c}$.

Under these assumptions, it is reasonable to look for an equilibrium in which all individuals have the same cost thresholds for voting, and hence the same probabilities of voting. If they all vote with probability p, then the expected gain of any individual from voting is $G_a(p, p, m, m) = G_b(p, p, m, m)$. Denote the common function by G, and for any value of m define the function g_m by $g_m(p) = G(p, p, m, m)$ for all p. That is, $g_m(p)$ is each individual's expected gain from voting when m individuals favor each alternative and every other individual votes with probability p. By (3.2), the condition for an equilibrium in which $p_a^* = p_b^* = p^*$ with $0 < p^* < 1$ is $F^{-1}(p^*) = g_m(p^*)$.

Consider the function g_m. Let i be an individual who favors a. If the probability that each of the other individuals votes is zero, a vote by i certainly changes the outcome, from a 0–0 tie to a win for a, so that i's gain from voting is $\frac{1}{2}$. If the probability that each of the other individuals votes is positive but close to zero, the probability of a tie between the alternatives among the other individuals' votes is relatively high, and hence i's voting is likely, but not certain, to change the outcome; thus the benefit to her of voting is less than $\frac{1}{2}$, but not much less. As the probability that each of the other individuals votes increases, the probability of a tie among those individuals' votes decreases, reducing i's benefit of voting. When the probability that each of the other individuals votes is $\frac{1}{2}$, the probability of a tie among their votes reaches a minimum, and hence i's gain from voting is at a minimum. As the probability that each of the other individuals votes increases above $\frac{1}{2}$, the probability of a tie increases, raising i's benefit of voting; when the probability reaches 1, her benefit from voting is again $\frac{1}{2}$. For any value of p, an increase in the number m of individuals who favor each alternative reduces the probability of a tie among the other individuals' votes, and hence reduces i's gain from voting.

Figure 3.2 shows examples of the function g_m for a few values of m, as well as an example of F^{-1}. For any given value of m, the equilibrium values of p are those for which $F^{-1}(p) = g_m(p)$.

This analysis suggests the following results.

- If F is continuous, an equilibrium exists: given $\underline{c} \leq \frac{1}{2}$, the graph of F^{-1} crosses each colored line in Figure 3.2 at least once.

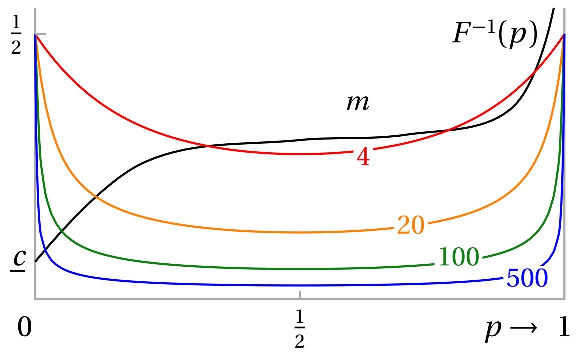


Figure 3.2 The colored curves are graphs of the function g_m defined by $g_m(p) = G(p, p, m, m)$, an individual's expected gain from voting when every other individual votes for her favorite alternative with probability p and abstains with probability $1 - p$ and m individuals favor each alternative, for various values of m.

- Multiple equilibria may exist: for example, for $m = 4$ in Figure 3.2 the game has three equilibria.
- For every value of p with $0 < p < 1$ the value of $g_m(p)$ decreases to zero as m increases without bound, so that if m is large enough the game has only one equilibrium.
- As m increases without bound the equilibrium probability that an individual votes goes to zero.

The general case

Existence of an equilibrium The first observation for the special case, that the continuity of the probability distribution function of the voting cost ensures that an equilibrium exists, applies also to the general model.

Proposition 3.3: Nash equilibrium of two-alternative voting game with uncertain voting costs

Every two-alternative voting game with uncertain voting costs has a symmetric Nash equilibrium and every such equilibrium is a threshold strategy profile.

Proof

Denote the game $\langle (N_a, N_b), \{a, b\}, (F_a, F_b)\rangle$ and define the function $\widehat{G}_a : [0,1]\times[0,1] \to [0,1]\times[0,1]$ by

$$\widehat{G}_a(c_a, c_b) = G_a(F_a(\min\{\max\{c_a, \underline{c}_a\}, \bar{c}_a\}), F_b(\min\{\max\{c_b, \underline{c}_b\}, \bar{c}_b\}), n_a, n_b)$$

and the function $\widehat{G}_b$ analogously. Given that F_a and F_b are continuous and G_a and G_b are continuous in p_a and p_b, $\widehat{G}_a$ and $\widehat{G}_b$ are continuous, so that by Brouwer's fixed point theorem there is a pair $(\hat{c}_a, \hat{c}_b) \in [0,1]\times[0,1]$ such that

$$\begin{aligned} \hat{c}_a &= \widehat{G}_a(\hat{c}_a, \hat{c}_b) \\ \hat{c}_b &= \widehat{G}_b(\hat{c}_a, \hat{c}_b). \end{aligned} \tag{3.3}$$

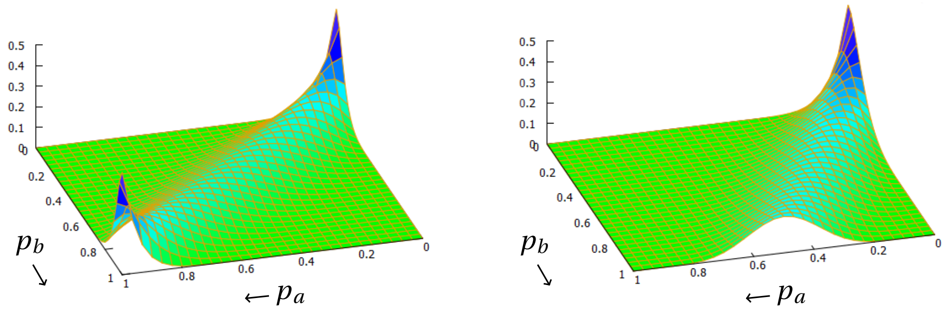
Let $c_a^* = \min\{\max\{\hat{c}_a, \underline{c}_a\}, \bar{c}_a\}$ and $c_b^* = \min\{\max\{\hat{c}_b, \underline{c}_b\}, \bar{c}_b\}$. I argue that (c_a^*, c_b^*) satisfies (3.2), so that by Proposition 3.2 the threshold strategy profile in which the threshold of every individual who favors a is c_a^* and the threshold of every individual who favors b is c_b^* is a Nash equilibrium.

To see that (c_a^*, c_b^*) satisfies (3.2), first suppose that $F_a(c_a^*) = 0$. Then $\hat{c}_a \le \underline{c}_a = c_a^*$ from the definition of c_a^* in terms of $\hat{c}_a$, so that the first condition in (3.3) implies that $c_a^* \ge G_a(F_a(c_a^*), F_b(c_b^*), n_a, n_b)$, using the definition of $\widehat{G}_a$. Now suppose that $0 < F_a(c_a^*) < 1$. Then $\hat{c}_a = c_a^*$, so that the first condition in (3.3) implies that $c_a^* = G_a(F_a(c_a^*), F_b(c_b^*), n_a, n_b)$. The other cases follow similarly.

Every equilibrium is a threshold strategy profile by Lemma 3.1.

Properties of equilibrium with large number of individuals How does equilibrium turnout vary as the number of individuals increases? My informal analysis suggests that if $n_a = n_b$ and $F_a = F_b$ then when the number of individuals is large the game has only one symmetric equilibrium in which $p_a^* = p_b^*$, and in this equilibrium the common probability of voting goes to 0 as the number of individuals increases.

To study the equilibria more generally, first consider how an individual's expected gain from voting varies with the probabilities with which the other individuals vote. Figure 3.3 shows two examples of this gain for an individual who favors a, as a function of the probability p_a of voting for each of the other $n_a - 1$ individuals who favor a and the probability p_b of voting for each of the n_b individuals who favor b. Figure 3.3a shows an example in which n_a and n_b are equal and Figure 3.3b shows one in which they differ. (The restriction of the surface in Figure 3.3a to the line $p_a = p_b$ is the orange curve in Figure 3.2.) Notice that when



(a) $n_a = n_b = 20$.

(b) $n_a = 20$, $n_b = 10$.

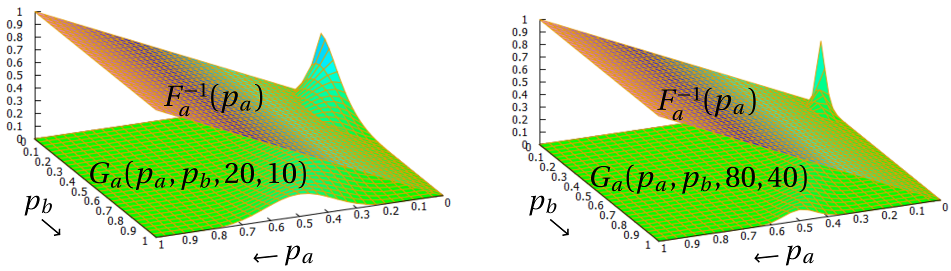
Figure 3.3 The expected gain $G_a(p_a, p_b, n_a, n_b)$ from voting for an individual who favors a as a function of the probability p_a of voting for each of the other $n_a - 1$ individuals who favor a and the probability p_b of voting for each of the n_b individuals who favor b.

n_a and n_b are equal, the expected gain from voting is close to $\frac{1}{2}$ when (p_a, p_b) is close to either $(0,0)$ (no one else votes) or $(1,1)$ (everyone else votes), and when n_a and n_b differ, it is close to $\frac{1}{2}$ only when (p_a, p_b) is close to $(0,0)$.

As n_a and n_b increase, the expected gain decreases at all points except $(0,0)$ and, if $n_a = n_b$, $(1,1)$, where it remains $\frac{1}{2}$. Suppose that the numbers of individuals who favor each alternative increase proportionately. That is, consider a sequence of games in which $n_x = rk_x$ for each $x \in \{a, b\}$, where k_x is a given positive integer and $r = 1, 2, \ldots$. For $(k_a, k_b) = (20, 10)$, Figure 3.4 shows two examples of G_a: the left panel is for $r = 1$ and the right panel is for $r = 4$.

Figure 3.4 shows also the function F_a^{-1} in the case that F_a is uniform on $[0,1]$. The first equilibrium condition in (3.2) is $F_a^{-1}(p_a) = G_a(p_a, p_b, n_a, n_b)$ for $0 < p_a < 1$, which means that an equilibrium lies on the intersection of the two surfaces in Figure 3.4. The second equilibrium condition is the analogue for F_b^{-1} and G_b, a function that has the same general form as G_a when n_a and n_b are large. An equilibrium pair (p_a, p_b) lies at the intersection of (*i*) the intersection of the two surfaces in Figure 3.4 and (*ii*) the intersection of the analogous surfaces for F_b^{-1} and G_b. The figures suggest that as r, and hence the number of individuals who favor each alternative, increases, the possible equilibrium values of p_a and p_b converge to 0.

I now present a result that shows that this property of an equilibrium holds generally: if $k_a \neq k_b$ then in the equilibria of the game in which rk_a individuals favor a and rk_b favor b, as r increases without bound the probability that any individual votes goes to zero. Further, the expected number of individuals who vote for each alternative remains bounded as the number of individuals increases without bound. The reason is that if the expected number of individuals who vote increases without bound, the probability of any individual's vote affecting the outcome goes to zero, so that no individual optimally votes. But if no



(a) $n_a = 20$, $n_b = 10$. (b) $n_a = 80$, $n_b = 40$.

Figure 3.4 The expected gain $G_a(p_a, p_b, n_a, n_b)$ from voting for an individual who favors a and the function $F_a^{-1}(p_a)$ in the case that F_a is uniform on $[0, 1]$.

individual votes, any individual with voting cost less than $\frac{1}{2}$ optimally votes.

Proposition 3.4: Voting probability converges to zero as population increases

Fix positive integers k_a and k_b with $k_a \neq k_b$ and for any positive integer r let $N_a(r)$ be a set with rk_a members and $N_b(r)$ a set with rk_b members. Let $\Gamma(r) = \langle (N_a(r), N_b(r)), \{a, b\}, (F_a, F_b) \rangle$ be a two-alternative voting game with uncertain voting costs and for each $x \in \{a, b\}$ denote the support of F_x by $[\underline{c}_x, \bar{c}_x]$. Let $(c_a^*(r), c_b^*(r))$ be the thresholds in a symmetric Nash equilibrium threshold strategy profile of $\Gamma(r)$.

a. The limits of $c_a^*(r)$ and $c_b^*(r)$ as r increases without bound are $\underline{c}_a$ and $\underline{c}_b$, so that the limiting probability that any individual votes is zero.

b. The limits of the expected numbers of individuals favoring each alternative who vote, $\lim_{r\to\infty} rk_a F_a(c_a^*(r))$ and $\lim_{r\to\infty} rk_b F_b(c_b^*(r))$, are finite.

The result follows from the properties of the binomial distribution. I do not establish these properties, but instead refer to a known result.

Proof

a. The conditions for an equilibrium in threshold strategies are given by (3.2) with $n_a = rk_a$ and $n_b = rk_b$. For all values of (p_a, p_b) except those close to $(0, 0)$, the values of $G_a(p_a, p_b, rk_a, rk_b)$ and $G_b(p_a, p_b, rk_a, rk_b)$ converge to zero as r increases without bound. Precisely, for every $\varepsilon > 0$,

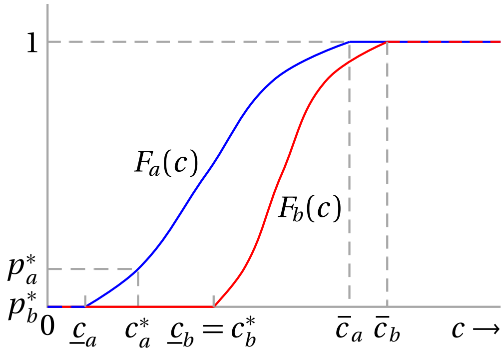


Figure 3.5 The thresholds and probabilities of voting in an equilibrium of a two-alternative voting game with uncertain voting costs in which individuals who favor b do not vote.

for all (p_a, p_b) with $p_a \geq \varepsilon$ and $p_b \geq \varepsilon$,

$$\lim_{r\to\infty} \max_{p_a, p_b}\{G_x(p_a, p_b, rk_a, rk_b) : p_a \geq \varepsilon \text{ or } p_b \geq \varepsilon\} = 0$$

for $x = a, b$. This result, suggested by Figure 3.3, follows from the properties of the binomial distribution, and specifically from Rosenthal (2020, Corollary 10).

Combined with the equilibrium conditions (3.2), the result implies that the equilibrium probability that any individual votes converges to zero, so that each threshold $c_x^*(r)$ converges to $\underline{c}_x$.

b. If the expected number of individuals voting among those who favor a given alternative x increases without bound as r increases, then for r sufficiently large the expected gain $G_x(p_a^*(r), p_b^*(r), rk_a, rk_b)$ is less than $\underline{c}_x$, so that $p_x^*(r) = 0$. But $(p_a, p_b) = (0, 0)$ is not an equilibrium because then a vote by any individual changes the outcome from a tie to a win for the individual's favored alternative, so that every individual whose voting cost is less than $\frac{1}{2}$ optimally votes.

In an equilibrium, one of the alternatives may receive no votes: none of the individuals who favor that alternative may vote (see Figure 3.5). In such an equilibrium, the number of votes for the other alternative is zero with positive probability, so that the alternative that certainly receives no votes ties with positive probability. The alternative that receives votes may win with high probability, even if the number of individuals who favor it is much smaller than the number who favor the other alternative, as the following exercise shows.

Exercise 3.3: Equilibrium in which low-cost minority is likely to win

Let $\langle (N_a, N_b), \{a, b\}, (F_a, F_b) \rangle$ be a two-alternative voting game with uncertain voting costs. Suppose that $n_a = |N_a| = 2$ and let $w \in (\frac{1}{2}, 1)$. Find distributions of voting costs F_a and F_b such that the game has a symmetric Nash equilibrium in which individuals who favor b do not vote and a wins with probability w regardless of the number of individuals who favor b.

Exercise 3.4: Voluntary and mandatory voting

Compare the outcome when voting is mandatory with the symmetric equilibria of a two-alternative voting game with uncertain voting costs for the parameters in Exercise 3.2a and Exercise 3.3.

Note that the analysis of this section is limited to symmetric equilibria. The game may in addition have equilibria in which the individuals who favor a given alternative use different strategies.

Note also, more fundamentally, that the notion of Nash equilibrium may be inappropriate as a solution concept for a model of an election. The interpretation of the notion of Nash equilibrium is most appealing for situations in which individuals repeatedly and anonymously interact. In such situations, it may be reasonable to assume that each individual's long experience playing the game allows her to form accurate beliefs about the actions the other individuals will take. Most elections do not fit into that category: they are unique events, so that individuals have scant basis to form accurate beliefs about each other's strategies. The problem is particularly significant for a voting game with imperfect information, where the notion of Nash equilibrium requires that each individual form accurate beliefs about the action taken by every type (voting cost) of every other individual. In many elections, the source of the information an individual could use to form such beliefs is unclear.

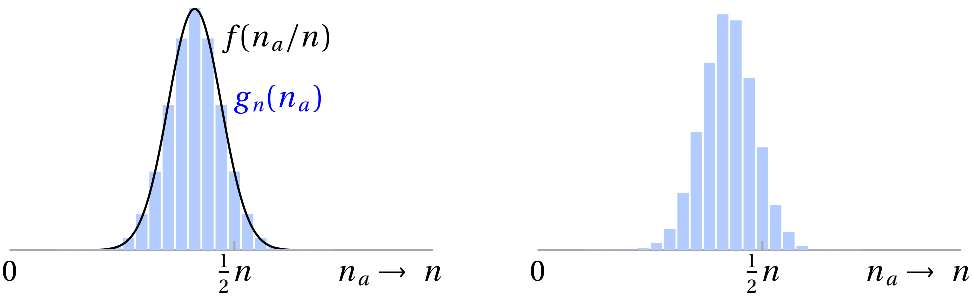
3.3 Turnout and population size with uncertain preferences

Is the turnout in elections consistent with a model in which individuals decide to vote by comparing the costs and benefits? Theory alone cannot answer this question, but can provide a framework for thinking about it. I present a brief analysis of the rate of change of the probability that an individual's vote affects the outcome of an a election as the population increases.

An individual motivated to vote by the chance that her vote will affect the electoral outcome must form a belief about that probability. The basis of her belief is plausibly information about the voting intentions of the other individuals. Sources of such information include polls, reports in media, the individual's own observations, and other individuals. A model of equilibrium can impose discipline on these beliefs by requiring that they be correct. For the purposes of this analysis, the model of the previous section, a two-alternative voting game with uncertain voting costs, seems inadequate, because it assumes that the number of individuals who favor each alternative is known; the only uncertainty in the model concerns the other individuals' voting costs. An essential feature of an individual's estimate of the probability that her vote will affect the outcome of an election appears to be uncertainty about the other individuals' preferences.

One option is to assume that each individual believes that the option favored by every other individual is drawn independently from a known distribution. But this assumption implies that in a large population the distribution of preferences, if not the preference of any one individual, is known almost with certainty. For example, if each individual believes that every other individual independently favors a with probability p and b with probability $1-p$, then if the number of individuals is large, she knows that the fraction who favor a is close to p and the fraction who favor b is close to $1-p$. In a large population, if $p > \frac{1}{2}$ then the probability that individuals who favor a are in a majority is close to 1 and if $p < \frac{1}{2}$ then the probability that they are in a minority is close to 1. Thus the model is hardly more appealing for a large population than one in which the fractions of individuals who favor the alternatives are known.

An alternative assumption, which is consistent with uncertainty about the majority preference surviving in a large population, is that each individual believes that the probability p with which every other individual independently favors a is itself uncertain. One specific assumption is that every individual has the same prior belief about the distribution of p, and her only private information concerns her own preferences. That is, every individual views her own preference for a or b, as well as every other individual's preference, as being the outcome of a random draw: with probability p she has been assigned a preference for a, and with probability $1-p$ a preference for b. She treats her realized preference as evidence regarding the distribution of p. If she favors a, she concludes that the this distribution is skewed towards large values, and if she favors b, she concludes that it is skewed towards small values. If, for example, the mean of the prior distribution of p is $\frac{1}{2}$, then the posterior probability density that an individual who favors a assigns to $p = \frac{1}{3}$ is (using Bayes' rule) $\frac{2}{3}$ of the prior density, whereas the density that an individual who favors b assigns to this value of p is $\frac{4}{3}$ of the prior density. The assumption that individuals view their own preferences as having been generated by a random process and that they make inferences from their own preferences about this process seems odd, and its implication re-



(a) A common prior over the number of individuals who favor a, fitted to a continuous distribution.

(b) Posterior belief over the number of individuals who favor a for an individual who favors a.

Figure 3.6 Beliefs about the number of individuals who favor a.

garding the individuals' posterior beliefs seems implausible. For these reasons, I analyze a different model.

Assume that each individual starts with a prior directly over the number of individuals who favor each alternative, rather than deriving this prior from a model of the probabilistic determination of each individual's preferences. Assume that every individual has the same prior. We want to study how the individuals' beliefs change as the total number n of individuals increases, so assume specifically that each individual has in mind a continuous probability density f of the fraction q of individuals in the population who favor a, and f is independent of n. One possible source of this belief is a poll, which might suggest, for example, that 48% of the population favors a, with a margin of error of 2%. For any given population size n, each individual derives her prior belief about the distribution of the number of individuals who favor a by approximating f by a discrete distribution g_n over the numbers 1 through n, as illustrated in Figure 3.6a. She knows her own political preference, and derives her posterior belief about the distribution of the number of individuals who favor a in the standard way from g_n, using Bayes' rule. (If, for example, she favors a, then the information she uses to update her prior is that at least one individual favors a.) Under this assumption, when n is large, the posterior distribution for an individual who favors a differs only slightly from the prior distribution (and thus also only slightly from the posterior distribution for an individual who favors b). (Figure 3.6b shows the posterior for an individual who favors a given the prior in Figure 3.6a.)

If the number n of voters is odd, the vote of an individual who favors a makes a difference to the outcome if the number of other individuals who vote in favor of a is $\frac{1}{2}(n-1)$; if n is even, it makes a difference if this number is $\frac{1}{2}(n-2)$. When n is large, the individual's posterior belief assigns a probability of approximately $f(\frac{1}{2})/n$ to each of these events. Thus the individual believes that the probability that her vote will make a difference to the outcome is proportional to $1/n$. When the population of voters doubles, for example, the probability that the individual's vote makes a difference to the outcome halves.

To get an idea of the magnitude of the probability, suppose that, based on a poll that reports that 48% of the individuals who intend to vote support a, with margin of error 2%, f is a (truncated) normal distribution with mean 0.48 and standard deviation $2/1.96 \approx 1.02$. (The truncation makes little difference, because the density of the distribution outside $[0, 1]$ is close to zero.) Then $f(\frac{1}{2}) \approx 0.39$. Thus the probability of an individual's vote making a difference to the outcome of an election in a population of n voters is approximately $0.39/n$. For example, in an electoral district with 10,000 voters, the probability is approximately 0.000039. Thus to make voting worthwhile for an individual in such a district, the benefit to her from the election of her favorite candidate rather than the other candidate has to be about 25,000 (= 1/0.000039) times her voting cost.

Throughout this analysis I have assumed that every individual's voting cost is nonnegative. For an individual who has to travel far or to wait a long time in line to vote, the cost may be significant. But for an individual who can vote by mail or online, or has only to walk a short distance to her polling station, the cost may be trivial, to the extent that she treats it as zero when making her decision. Further, an individual may derive satisfaction from endorsing a candidate she likes, or may feel good about carrying out a task that she believes is her duty, so that her cost is effectively negative. In a two-alternative voting game with uncertain voting costs, such an individual optimally votes regardless of her beliefs about the other individuals' behavior—and hence also regardless of the expected closeness of the election.

In the models I have discussed so far, every individual is self-interested; she considers only the change in her personal welfare that her vote might cause. Suppose instead that individuals consider the benefit to society. The size of this benefit is plausibly proportional to the number n of individuals in the society, so that even if the probability of an individual's vote affecting the outcome of the election is proportional to $1/n$, the expected benefit of voting may be large. Evren (2012) studies a model of this type. Chapter 6 presents a different model of voting in which individuals are public-spirited. Under some conditions, the equilibria of this model entail positive turnout even in an arbitrarily large population.

3.4 Preferences with regret

A key assumption of the model of a two-alternative voting game with uncertain voting costs is that each individual's preferences over uncertain electoral outcomes are represented by the expected value of a payoff function over the possible deterministic outcomes. This function assigns a number to each outcome, independently of the lotteries among which the individual is choosing. But an individual's evaluation of an outcome may depend not only on the outcome itself but also on the outcomes that would have occurred had she taken a different action. For example, if an individual who favors a abstains, she may experience regret if the outcome is a tie between a and b, because in that case her voting would have benefitted her. If the outcome is that a wins, she experiences no such regret. If the outcome is that a loses by a large margin, she may also experience no regret, but if it loses by a small margin, she may regret that she did not vote and did not do more to persuade her a-favoring friends to vote. Regret may be experienced not only by an individual who abstains, but also by one who votes: if an individual who favors a votes and the outcome is that a loses or wins by two votes or more, she may regret that she needlessly incurred the cost of voting.

An individual's knowing that she will experience regret for certain outcomes plausibly influences the action she chooses. For example, an individual may be inclined to vote if she knows she will deeply regret abstaining if the vote turns out to be a tie but will only mildly regret voting if the outcome is far from a tie.

The assumption that each individual maximizes the expected value of a payoff function embodies another premise: each individual has a precise belief about the probabilities of the various possible outcomes. In some elections, individuals have few sources of information about this probability distribution. Consider, for example, the election of a legislator in a district in which the candidates have not previously competed against each other, and few opinion polls exist. When the basis for forming beliefs about the probabilities is unclear, individuals may use an alternative calculus. One option is for an individual to choose an action that minimizes the most she will regret from taking the action.

Consider an individual who favors a and has a cost c of voting that is less than $\frac{1}{2}$. Suppose that she abstains. If the vote among the other individuals is a tie, she obtains the payoff $\frac{1}{2}$, but could have obtained the higher payoff of $1 - c$ by voting for a, so her regret from abstaining is $1 - c - \frac{1}{2} = \frac{1}{2} - c$. If among the other individuals the winning margin for a is 1, then she obtains the payoff 1; no other action yields a higher payoff, so her regret from abstaining in this case is 0. Table 3.3 gives the regret for each action and each possible winning margin for a among the other individuals (given $c < \frac{1}{2}$). The entry in the cell in row r and column w of the table is the difference between the largest entry in column w of Table 3.2 (which gives the individual's payoffs) and the entry in row r and column w of that table.

We see that the most the individual regrets from voting for a is c, which happens when it turns out that her vote makes no difference. The most she regrets

	winning margin for a among other individuals				
	≥ 2	1	0	-1	≤ -2
vote for a	c	c	0	0	c
vote for b	c	$\frac{1}{2}+c$	1	$\frac{1}{2}$	c
abstain	0	0	$\frac{1}{2}-c$	$\frac{1}{2}-c$	0

Table 3.3 The regret for an individual who favors a and has voting cost $c < \frac{1}{2}$ as a function of her action and the winning margin for a among the other individuals. For each action, the highest regrets are highlighted.

from voting for b is 1, which happens when the other votes are tied, so that her vote makes a difference in the wrong direction, and the most she regrets from abstaining is $\frac{1}{2}-c$, which happens when her vote would have made a difference to the outcome in her favor. So given $c < \frac{1}{2}$, the action that minimizes her maximal regret is voting for a if $c < \frac{1}{2}-c$, or $c < \frac{1}{4}$, and abstaining if $c > \frac{1}{4}$. A similar analysis for $c > \frac{1}{2}$ shows that in this case abstention always minimizes her maximal regret.

The logic can be stated succinctly as follows. If the individual abstains, the outcome for which she incurs the most regret for not having voted is a tie; if she votes for a, the outcome for which she incurs the most regret for having voted is a win for a by two votes or more. If $c < \frac{1}{4}$ then the amount of her maximal regret if she votes for a is less than the amount of her maximal regret if she abstains, so she votes for a. That is, if an individual acts to minimize her maximal regret, she votes (for her favorite alternative) if her voting cost is less than one-quarter of her payoff from her favored alternative, regardless of the number of individuals. Hence turnout is independent of the size of the population.

Exercise 3.5: Minmax regret individual with three alternatives

Suppose that there are three alternatives, a, b, and z, rather than two. An individual's payoffs to the alternatives are $u(a) = 1$, $u(b) = k$, and $u(z) = 0$, with $0 < k < 1$. If $k > \frac{1}{2}$ and $c < \frac{1}{2}k$, the maximum regret for each of the individual's actions is achieved for the following events. Vote for a: among the other individuals, z wins by 1 vote over b and a gets fewer votes than b. Vote for b, vote for z, or abstain: among the other individuals, a and z are tied and b gets at least two fewer votes than a and z. Which action minimizes the individual's maximum regret? The same action minimizes the individual's regret for all other parameter values. Do these conclusions make sense?

Notes

The methods of noncooperative game theory were first applied to the study of voting by Robin Farquharson in his doctoral thesis of 1958, published as Farquharson (1969). Section 3.1.2 draws on Palfrey and Rosenthal (1983, Proposition 1). Section 3.2 is based on Ledyard (1981) and Palfrey and Rosenthal (1985), and draws also on Krasa and Polborn (2009) and Taylor and Yildirim (2010). Proposition 3.3 is due to Ledyard (1981, Proposition 2) and Proposition 3.4 is due to Palfrey and Rosenthal (1985, Theorem 2). The first two models of the relation between turnout and population size discussed in Section 3.3 are analyzed by Chamberlain and Rothschild (1981) and Myatt (2015). Section 3.4 (including Exercise 3.5) is based on Ferejohn and Fiorina (1974). General models of decision-making under uncertainty that incorporate regret are explored by Loomes and Sugden (1982) and Bell (1982); Bleichrodt and Wakker (2015) discuss these theories.

Campbell (1999) (see also Taylor and Yildirim 2010, Proposition 2) explores the effect of differences in the distributions of voting costs on the winning alternative, an issue touched upon in Exercise 3.3.

Börgers (2004) and Krasa and Polborn (2009) explore the difference between voluntary and mandatory voting considered in Exercise 3.4.

Solutions to exercises

Exercise 3.1

An action profile is a Nash equilibrium if and only if it satisfies one of the following conditions, where the winning margin is the difference between the number of votes for the winner and the number of votes for the other alternative.

1. The winning margin is at least three votes.
2. The winning margin is two votes and every individual who votes for the winning alternative prefers that alternative to the other alternative.
3. The winning margin is one vote and every individual who either votes for the winning alternative or abstains prefers the winning alternative to the other alternative.
4. Each alternative receives the same number of votes, all individuals vote, and every individual votes for her favorite alternative.

(Case 4 is possible only if the number of individuals is even and each alternative is the favorite of exactly half of the individuals.)

Exercise 3.2

a. We have

$$\begin{aligned}P_a^0(F_a(c_a),F_b(c_b),2,2) &= B(1,0,F_a(c_a))B(2,0,F_b(c_b))+B(1,1,F_a(c_a))B(2,1,F_b(c_b))\\ &= (1-F_a(c_a))\times(1-F_b(c_b))^2+F_a(c_a)\times 2F_b(c_b)(1-F_b(c_b))\\ &= (1-c_a)(1-c_b)^2+2c_ac_b(1-c_b)\end{aligned}$$

and

$$\begin{aligned}P_a^1(F_a(c_a),F_b(c_b),2,2) &= B(1,0,F_a(c_a))B(2,1,F_b(c_b))+B(1,1,F_a(c_a))B(2,2,F_b(c_b))\\ &= (1-F_a(c_a))\times 2F_b(c_b)(1-F_b(c_b))+F_a(c_a)\times(F_b(c_b))^2\\ &= 2(1-c_a)c_b(1-c_b)+c_ac_b^2.\end{aligned}$$

After some algebra we get $G_a(F(c_a),F(c_b),2,2)=\frac{1}{2}(1-c_a+2c_ac_b-c_b^2)$, and similarly $G_b(F(c_a),F(c_b),2,2)=\frac{1}{2}(1-c_b+2c_ac_b-c_a^2)$. Thus condition (3.2) for a symmetric equilibrium in which some types of each individual vote and some abstain and the thresholds are c_a and c_b is

$$\begin{aligned}c_a &= \tfrac{1}{2}(1-c_a+2c_ac_b-c_b^2)\\ c_b &= \tfrac{1}{2}(1-c_b+2c_ac_b-c_a^2)\end{aligned}$$

or

$$\begin{aligned}3c_a &= 1+2c_ac_b-c_b^2\\ 3c_b &= 1+2c_ac_b-c_a^2.\end{aligned}$$

Subtracting the second equation from the first we get $3(c_a-c_b)=(c_a-c_b)(c_a+c_b)$, so that if $c_a\neq c_b$ then $c_a+c_b=3$, which is not possible. Thus $c_a=c_b$. Denote the common value c. Then the condition for an equilibrium is $1-3c+c^2=0$ or $c=\frac{1}{2}(3-\sqrt{5})\approx 0.382$. (The other root of the equation is greater than 1.)

Thus the game has a symmetric Nash equilibrium in which each individual votes if her cost is less than $\frac{1}{2}(3-\sqrt{5})$ and abstains if her cost is greater than $\frac{1}{2}(3-\sqrt{5})$.

b. If no individual who favors b votes, then for an individual who favors a we have $P_a^0(F_a(c_a),F_b(c_b),2,2)=1-F_a(c_a)$ (the alternatives are tied if and only if the other individual who favors a does not vote) and $P_a^1(F_a(c_a),F_b(c_b),2,2)=0$ (the winning margin for b is never positive). Thus $G_a(F_a(c_a),F_b(c_b),2,2)=\frac{1}{2}(1-c_a)$, and hence each individual who favors a is indifferent between voting and abstaining if $c_a=\frac{1}{3}$.

Now consider an individual who favors b. If each individual who favors a votes if and only if their cost is less than c_a and the other individual who favors b does not vote, the probability assigned by an individual who favors b to a tie between a and b among the other voters is $(1-c_a)^2$ (neither of the individuals who favor a vote) and the probability that the winning margin for a is 1 is $2c_a(1-c_a)$ (exactly one of the individuals who favors a votes). Thus the expected gain from voting for an individual who favors b is $\frac{1}{2}[(1-c_a)^2+2c_a(1-c_a)] = \frac{1}{2}(1-c_a^2) = \frac{4}{9}$. Hence any distribution F_b for which the lower limit of the support is at least $\frac{4}{9}$ is consistent with an equilibrium in which no type of individual who favors b votes. In such an equilibrium, a wins with probability $\frac{5}{9}$ (the probability that neither of the individuals who favor a votes) and a and b tie with probability $\frac{4}{9}$.

Exercise 3.3

Denote by p_a the probability with which each individual who favors a votes. Then

$$P_a^0(p_a, 0, 2, n_b) = 1 - p_a$$
$$P_a^1(p_a, 0, 2, n_b) = 0,$$

so that $G_a(p_a, 0, 2, n_b) = \frac{1}{2}(1-p_a)$, and

$$P_b^0(p_a, 0, 2, n_b) = (1-p_a)^2$$
$$P_b^1(p_a, 0, 2, n_b) = 2p_a(1-p_a),$$

so that $G_b(p_a, 0, 2, n_b) = \frac{1}{2}(1-p_a^2)$. Now, if neither of the individuals who favor a votes then the vote is a tie, so that a wins with probability $\frac{1}{2}$; otherwise a wins. So the probability that a wins is $\frac{1}{2}(1-p_a)^2 + 1 - (1-p_a)^2 = \frac{1}{2}(1+2p_a - p_a^2)$. Let p_a^* be the (unique) number such that this probability is w. Then $p_a^* \in (0,1)$. Construct F_a with the property that $F_a(\frac{1}{2}(1-p_a^*)) = p_a^*$ and F_b with the property that $F_b(\frac{1}{2}(1-(p_a^*)^2)) = 0$, as in Figure 3.7. Then the game has a threshold equilibrium in which the thresholds are $c_a^* = \frac{1}{2}(1-p_a^*)$ and $c_b^* = \frac{1}{2}(1-(p_a^*)^2)$ and a wins with probability w, regardless of the number n_b of individuals who favor b.

Exercise 3.4

For the game in Exercise 3.2a, the outcome of the symmetric Nash equilibrium is that each alternative is selected with probability $\frac{1}{2}$. If voting is mandatory, the outcome is the same, but every individual incurs the voting cost with certainty. Thus voluntary voting is better than mandatory voting.

In the symmetric equilibrium of the game in Exercise 3.3, a wins unless both individuals who favor a abstain, which occurs with probability $(1-p_a^*)^2$, in which case both a and b win with probability $\frac{1}{2}$. Thus the probability that b wins is $\frac{1}{2}(1-p_a^*)^2$. Hence each of the n_b individuals who favor b obtains the payoff 1 with probability $\frac{1}{2}(1-p_a^*)^2$ and the payoff 0 otherwise.

If voting is mandatory, b wins with certainty and each individual i who favors b obtains the payoff $1-c_i$. If $\bar{c}_b < 1$, then all of these payoffs are positive, so

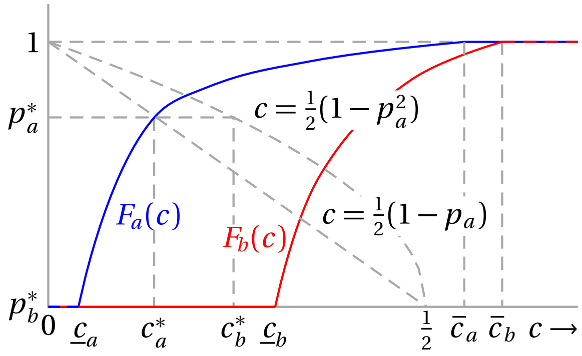


Figure 3.7 The thresholds and probabilities of voting in an equilibrium of a two-alternative voting game with uncertain voting costs in which individuals who favor b do not vote (Exercise 3.3).

for n_b large enough the sum of the individuals' payoffs under mandatory voting exceeds the sum of the payoffs under voluntary voting (regardless of p_a^*). Thus the utilitarian welfare criterion leads to the conclusion that mandatory voting is better than voluntary voting. However, the outcome under mandatory voting does not dominate the outcome under voluntary voting: the individuals who favor a are worse off under mandatory voting. If $\bar{c}_b > 1$ then some individuals who favor b are worse off under mandatory voting, so the comparison between voluntary and mandatory voting according to the utilitarian welfare criterion depends on the forms of the distributions F_a and F_b of voting costs.

Exercise 3.5

Given that the maximum regret for each of the individual's actions is achieved for the events given in the question, the individual's maximal regrets are

$$\begin{aligned}
\text{vote for } a: &\text{ payoff } -c\text{; switch to } b \Rightarrow \text{payoff } \tfrac{1}{2}k - c, \text{ so regret } \tfrac{1}{2}k \\
\text{vote for } b: &\text{ payoff } \tfrac{1}{2} - c\text{; switch to } a \Rightarrow \text{payoff } 1 - c, \text{ so regret } \tfrac{1}{2} \\
\text{vote for } z: &\text{ payoff } -c\text{; switch to } a \Rightarrow \text{payoff } 1 - c, \text{ so regret } 1 \\
\text{abstain}: &\text{ payoff } \tfrac{1}{2}\text{; switch to } a \Rightarrow \text{payoff } 1 - c, \text{ so regret } \tfrac{1}{2} - c.
\end{aligned}$$

Given $k < 1$, the action that minimizes the individual's maximal regret is to vote for a, her favorite alternative.

If an individual has no information about the probabilities of the voting behavior of the other individuals, and wants to choose an action that minimizes her regret, voting for her favorite alternative makes sense. One possibility is that the votes among the other individuals make voting for her second choice, b, optimal: if among the other individuals b and z are tied, and a is two or more votes behind, then voting for a yields a payoff of $\frac{1}{2}k$ whereas voting for b yields a payoff of k. However, voting for b could lead to a bigger regret: if a and z are tied and b is trailing for two votes or more, voting for b yields a payoff of $\frac{1}{2}$ whereas voting for a yields a payoff of 1. Voting for a is a safer option for an individual who wants to minimize her regret, because it is guaranteed to generate a regret of at most $\frac{1}{2}k$.

Plausibly an individual who chooses the action that minimizes her maximal regret votes for her favorite alternative regardless of the number of alternatives, but I do not know whether that is in fact the case.

4 Voting with many alternatives: plurality rule

4.1 Plurality rule voting games 130
4.2 Spatial model with concave payoff functions 137
4.3 Strategic and sincere voting: divided majority 139

One method a group of individuals can use to select an alternative from a set of many alternatives is plurality rule: each individual votes for a single alternative and the winners are the alternatives that receive the most votes. This chapter analyzes models of this mechanism under the assumption that voting is costless.

Synopsis

The main model, a plurality rule voting game, is an extension to many alternatives of a two-alternative voting game with the restriction that all voting costs are zero. This game has many Nash equilibria. For example, for every alternative, the strategy profile in which every individual votes for that alternative is a Nash equilibrium. Further, restricting attention to actions that are not weakly dominated, which reduces the set of equilibria dramatically in two-alternative games, has little impact in many-alternative games. For an individual who is not indifferent among all the alternatives, voting for her least favored alternative is weakly dominated (by voting for her favorite alternative), but if there are at least four individuals, no other action is weakly dominated (Proposition 4.1). Despite the game's profusion of Nash equilibria and the limited number of weakly dominated actions, it is not obvious that every plurality rule voting game has a Nash equilibrium in which no individual's action is weakly dominated. However, Proposition 4.2 asserts that this result holds.

Section 4.2 considers plurality rule voting games in which the set of alternatives is an interval of numbers and each individual's payoff function is strictly concave. Proposition 4.3 shows that in any Nash equilibrium of such a game, at most two alternatives tie for first place. As for a general plurality rule voting game, for every alternative a the game has a Nash equilibrium in which every individual votes for a. Further, given the strict concavity of the payoff functions, every individual's least favored alternative is an endpoint of the interval of alternatives, so for any alternative a except these two, the game has a Nash equilibrium in which the sole winner is a and no individual's action is weakly dominated.

Here is a situation that might seem challenging for plurality rule. There are three alternatives, a, b, and c. Sixty percent of individuals prefer a and b to c, with about half of them preferring a to b and the remainder preferring b to a, while the remaining forty percent prefer c to both a and b, between which they are indifferent. The plurality rule voting game that models this situation has a Nash equilibrium in which every individual votes for her favorite alternative, so that c wins, even though it is the worst alternative for a majority of individuals. It also has Nash equilibria in which all the individuals who favor a or b vote for a, so that a wins, or all of these individuals vote for b, so that b wins. Is there any reason to think that one of these equilibria is more likely to occur than the others? Section 4.3 studies this question in a model in which each individual is uncertain of the other individuals' preferences. The analysis identifies circumstances in which, when the number of individuals is sufficiently large, in every equilibrium the individuals who favor a or b all vote for one of these two alternatives, so that it wins, as well as circumstances in which equilibria in which c wins persist even when the number of individuals is arbitrarily large.

4.1 Plurality rule voting games

Definition 4.1: Plurality rule voting game

The *plurality rule voting game* $\langle N, X, (u_i)_{i \in N} \rangle$, where

- N is a finite set (of individuals) with at least two members
- X is a set (of alternatives) that is either finite, with at least two members, or a nonempty interval of real numbers
- $u_i : X \to \mathbb{R}$ for each $i \in N$ (individual i's payoff function)

is the strategic game with the following components.

Players
The set N.

Actions
Each player's set of actions consists of *vote for a* for each $a \in X$ and *abstain*.

Payoffs

For any action profile x, denote by $W(x) \subseteq X$ the set of alternatives that receive the most votes (the *winning alternatives*): for each $a \in W(x)$, the number of players i for whom $x_i =$ *vote for a* is the same and, if $W(x) \neq X$, exceeds the number for whom $x_i =$ *vote for b* for every $b \in X \setminus W(x)$. The payoff of each player $i \in N$ for x is the average value of $u_i(w)$ over $W(x)$ (that is, $\sum_{w \in W(x)} u_i(w)/|W(x)|$ if $W(x)$ is finite, as it is unless no one votes and X is an interval of real numbers).

As for a two-alternative voting game, one rationale for the specification of the players' payoffs in case of a tie for first place is that every alternative in the winning set is selected with the same probability, and each player evaluates a lottery according to its expected payoff.

Consider an action profile in a plurality rule voting game in which some alternatives are tied for first place, so that every individual's payoff is her average payoff for these alternatives. An individual who abstains or votes for an alternative that is not a winner can, by deviating to vote for a winner, cause that alternative to win outright. Thus in a Nash equilibrium every such individual is indifferent among the winners. Also, each individual who votes for a winner can, by switching her vote to another winner, cause that alternative to win outright. So in an equilibrium her payoff for each winning alternative for which she does not vote is at most her average payoff for the winning alternatives. Hence she either prefers the alternative for which she votes to every other winner or is indifferent among all the winners. These observations are stated in the following result.

Lemma 4.1: Nash equilibrium of plurality rule voting game

Let $\langle N, X, (u_i)_{i \in N} \rangle$ be a plurality rule voting game, let x be a Nash equilibrium of this game, and let $W(x)$ be the set of winning alternatives for x.

a. If for any $i \in N$ the action x_i is either a vote for an alternative outside $W(x)$ or *abstain*, then $u_i(w)$ is the same for all $w \in W(x)$ (i is indifferent among all winning alternatives).

b. If for some $a \in W(x)$ the action x_i is *vote for a*, then

$$u_i(b) \leq \sum_{w \in W(x)} u_i(w)/|W(x)| \quad \text{for all } b \in W(x) \setminus \{a\}$$

and hence either $u_i(a) > u_i(b)$ for all $b \in W(x) \setminus \{a\}$ (i prefers a to every other winning alternative) or $u_i(a) = u_i(w)$ for all $w \in W(x)$ (i is indifferent among all the winning alternatives).

When the alternatives number three or more, a complete characterization of the set of Nash equilibria of a plurality rule voting game is complicated. The details are less significant than the message that if there are three or more individuals then regardless of the individuals' preferences, the game has many Nash equilibria. In particular, if there are three or more individuals then for any preference profile and *any* alternative, the game has Nash equilibria in which that alternative wins. For example, all individuals' voting for any given alternative is a Nash equilibrium, because no deviation by a single player affects the outcome. Thus the Nash equilibria are unrelated to the individuals' preferences.

Weakly dominated actions

When there are two alternatives, the only action of an individual in a plurality rule voting game that is not weakly dominated is a vote for her favorite alternative (Proposition 3.1). An application of the logic for this result shows that when there are three or more alternatives, an individual's voting for any alternative that she ranks lowest is weakly dominated (by her voting for one of her favorite alternatives). If there are four or more individuals, no other alternative is weakly dominated. To see why, suppose that z is an alternative that individual i ranks lowest and b is an alternative that she does not rank lowest. If b and z are tied for the highest number of votes among the other individuals and every other alternative has two or more fewer votes (which requires the total number of individuals to be at least four), then i's voting for b is better for her than voting for any other alternative: if she votes for b, then b wins, whereas if she votes for any other alternative, then either z wins or b and z tie. Thus no action weakly dominates voting for b.

Proposition 4.1: Weak domination in plurality rule voting game

Let $\langle N, X, (u_i)_{i \in N} \rangle$ be a plurality rule voting game for which X contains three or more alternatives.

a. Let $i \in N$ be an individual who is not indifferent among all the alternatives. Individual i's abstaining and her voting for any alternative z that she likes least (that is, $u_i(z) \leq u_i(y)$ for all $y \in X$) are both weakly dominated by her voting for any of her favorite alternatives.

b. Suppose that the number of individuals is at least four. If an individual is indifferent among all the alternatives then none of her actions are weakly dominated. Otherwise, her only weakly dominated actions are votes for one of the alternatives she likes least.

Proof

a. Let a be one of i's favorite alternatives and z be one of the alternatives she likes least.

First consider an arbitrary list of actions of the individuals other than i. Denote by W^0 the set of winning alternatives given these actions when i abstains. If i switches to vote for a then the set of winning alternatives either remains W^0, becomes $W^0 \cup \{a\}$, or becomes $\{a\}$, depending on the other individuals' actions. Each of these outcomes is at least as good for i as W^0.

Now suppose that everyone but i abstains. Then if i abstains, the set of winning alternatives is the set X of all alternatives, and if she votes for a it is $\{a\}$. So her payoff is higher when she votes for a.

Thus i's voting for a weakly dominates her abstaining.

Now let W^0 be the set of winning alternatives given the actions of the individuals other than i when i votes for z. If she switches to vote for a then the set of winning alternatives either remains W^0, becomes $W^0 \cup \{a\}$, becomes $W^0 \setminus \{z\}$, becomes $\{a\}$, or changes from $\{z\}$ to $\{z\} \cup Y$ for some set Y of alternatives. In each case, the outcome when i votes for a is as least as good for her as the outcome when she votes for z. If everyone but i abstains, the winning alternative is z if i votes for z and a if she votes for a, so her payoff when she votes for a is higher than it is when she votes for z. So i's voting for a weakly dominates her voting for z.

b. Let z be an alternative that individual i likes least, and let b be one that she prefers to z. I argue that i's voting for b is not weakly dominated.

Consider an action profile in which one individual votes for b, two vote for z, and the remainder, except for i, abstain. (Such an action profile is possible because there are four or more individuals.) If i votes for b the outcome is a tie between b and z and if she votes for any other alternative or abstains the outcome is z. Thus her payoff from voting for b exceeds her payoff from all her other actions, and hence no action weakly dominates voting for b.

Exercise 4.1: Weak domination in plurality rule voting game with three individuals

Consider a plurality rule voting game with three individuals. Show that an individual's voting for an alternative b is weakly dominated if and only if the individual is not indifferent between all alternatives and her payoff for

b is at most her payoff for a set of alternatives consisting of b, one of her favorite alternatives, and an alternative she likes least.

For any alternative a, the action profile in which every individual votes for a is a Nash equilibrium of a plurality rule voting game. For which alternatives a does such a game have a Nash equilibrium in which the winner is a and no individual's action is weakly dominated? If a is ranked last by some individuals who are not indifferent among all the alternatives then if every individual votes for a, the actions of the individuals who rank a last are weakly dominated by Proposition 4.1a. However, if the number of individuals who rank a last is at most $\frac{1}{2}(n-3)$, where $n \geq 5$ is the total number of individuals, the game does have a Nash equilibrium in which no individual's action is weakly dominated and the winner is a. In one such equilibrium, every individual who ranks a last votes for her favorite alternative and every other individual votes for a. The number of individuals who vote for a is at least three more than the number who vote for any other alternative, so that no change in any individual's vote affects the identity of the winner, and hence the action profile is a Nash equilibrium; by Proposition 4.1b, no individual's action is weakly dominated.

Even though, as this argument shows, restricting individuals to actions that are not weakly dominated has no impact in many plurality rule voting games on the set of alternatives that are winners in some Nash equilibrium, such a restriction still seems sensible simply because weakly dominated actions lack appeal. Imposing the restriction raises a question: does a plurality rule voting game with three or more alternatives necessarily have a Nash equilibrium game in which no individual's action is weakly dominated? The answer is affirmative.

Proposition 4.2: Existence of Nash equilibrium in weakly undominated actions of plurality rule voting game

Every plurality rule voting game has a Nash equilibrium in which no individual's action is weakly dominated.

For a game with two individuals, this result follows from Proposition 3.1. For a game with three individuals, you are asked to prove the result in the next exercise.

Exercise 4.2: Nash equilibrium in weakly undominated actions of plurality rule voting game with three individuals

Show that every plurality rule voting game with three individuals has a Nash equilibrium in which no individual's action is weakly dominated.

For a game with four or more individuals, each of whom has strict preferences, the proof is easy.

Proof of Proposition 4.2 for four or more individuals, strict preferences

Assume that no individual is indifferent between any two alternatives. Choose two alternatives arbitrarily; call them a and b. Consider the action profile in which each individual votes for the alternative in $\{a, b\}$ that she prefers. Suppose without loss of generality that the number of individuals who prefer a to b is at least the number who prefer b to a.

- If a receives three or more votes than b, no change in any individual's action affects the outcome.
- If a receives two votes more than b, the only change in an individual's action that affects the outcome is a switch by an individual voting for a to vote for b, which changes the outcome from a to a tie between a and b, and hence makes her worse off.
- If a receives one more vote than b, the only change in an individual's action that affects the outcome is a switch by an individual voting for a. If she switches to b, the outcome changes from a to b, so that she is worse off. If she switches to another alternative or to abstention, the outcome changes to a tie between a and b (given that the number of individuals is at least four), so that she is also worse off.
- If a and b receive the same number of votes, any change in an individual's action changes the outcome from a tie between a and b to a win for the alternative the individual likes less.

We conclude that the action profile is a Nash equilibrium. In the profile, no individual votes for the alternative she likes least, so that by Proposition 4.1b no individual's action is weakly dominated.

For a game with four or more individuals whose preferences are not necessarily strict, the only proof of which I am aware, due to Duggan and Sekiya (2009), is lengthy, and I omit it. This proof shows that the following procedure generates an equilibrium in which no individual's action is weakly dominated. At each step $t = 1, 2, \ldots$, an action profile x^t is defined. In the initial profile, x^1, every individual votes for an alternative selected arbitrarily from her set of favorite alternatives. If, at any step t, x^t is a Nash equilibrium, the procedure ends. Otherwise, some individual i, by voting for an alternative z different from x_i^t, can cause z to become a winner or tie for first place and thereby increase her payoff, given the other individuals' actions. (If there is more than one such individual, select one arbitrarily.) Define x^{t+1} to be the profile derived from x^t by changing x_i^t to a vote for an alternative that yields i the highest payoff among such alternatives z. (If there is more than one such alternative, select one arbitrarily.) Duggan and Sekiya (2009, Theorem 1) show that this procedure terminates and generates a Nash equilibrium in which no individual's action is weakly dominated. (Their model does not allow abstention, but adding that option does not affect the result.)

Exercise 4.3: Nash equilibrium in weakly undominated actions of plurality rule voting game

Use the procedure just described to find a Nash equilibrium in weakly undominated actions of the plurality rule voting game $\langle\{1,2,3,4,5,6\}, \{a,b,c,d\},(u_i)_{i\in N}\rangle$, in which the payoffs are given as follows.

	1	2	3	4	5	6
payoff 9:	c	d	c	a	a	b
payoff 6:	d	b	b	d	c	a
payoff 0:	a,b	a,c	a,d	b,c	b,d	c,d

If we eliminate from a plurality rule voting game the actions of each individual that are weakly dominated, some actions may be weakly dominated in the resulting game. For example, suppose that in a game with four individuals and three alternatives, a, b, and c, three individuals rank c last and one ranks it first. Then by Proposition 4.1, voting for c is weakly dominated for the individuals who rank it last, and if we eliminate this action for these individuals then for the remaining individual voting for c is weakly dominated by voting for her preferred alternative in $\{a,b\}$ because for no remaining actions of the other individuals is c a winning alternative. In some cases the procedure of iteratively eliminating weakly dominated actions leads to a single action profile, as you are asked to demonstrate in the following exercise.

Exercise 4.4: Iterated elimination of weakly dominated actions in a plurality rule voting game

Consider a plurality rule voting game in which there are three alternatives and at least four individuals, all with strict preferences. Show that if more than two-thirds of the individuals rank the same alternative last then after all the weakly dominated actions are removed from the game, a single action profile remains.

Exercise 4.5: Voting game under proportional rule

Consider a variant of a plurality rule voting game in which the alternatives are numbers, abstention is not an option, and the outcome is a vote-weighted average of the alternatives. The set of players (individuals) is $N = \{1, \ldots, n\}$ and the set of alternatives is $X = \{a_1, \ldots, a_k\}$, where each a_j is a number and $a_1 < a_2 < \cdots < a_k$. For an action profile x in which the number of votes for each alternative a_j is $v_j(x)$, the outcome is $O(x) = \sum_{j=1}^{k} v_j(x) a_j / n$ and the payoff of each individual i is $u_i(O(x))$, where each $u_i : \mathbb{R} \to \mathbb{R}_-$ is a single-peaked function.

a. Show that for any Nash equilibrium x^*, all individuals except those with favorite positions close to $O(x^*)$ vote for one of the extreme alternatives (a_1 or a_k). Specifically, every individual i whose favorite alternative is at most $O(x^*) - (a_k - a_1)/n$ votes for a_1 and every individual i whose favorite alternative is at least $O(x^*) + (a_k - a_1)/n$ votes for a_k.

b. Suppose that for some number z^* that is not the favorite alternative of any individual we have $z^* = [L(z^*)a_1 + G(z^*)a_k]/n$, where $L(z^*)$ is the number of individuals whose favorite alternatives are less than z^* and $G(z^*)$ is the number of individuals whose favorite alternatives are greater than z^*. Find a Nash equilibrium of the game in which every individual votes for one of the two extreme alternatives.

4.2 Spatial model with concave payoff functions

In any Nash equilibrium of a plurality rule voting game in which the set of alternatives is a closed interval of real numbers and the individuals' payoff functions are strictly concave, the number of winning alternatives is at most two. To see why, suppose that the number of winning alternatives in a Nash equilibrium is three or more. Denote the average of the winning alternatives by $\bar{a}$. Either at least two alternatives are at least $\bar{a}$, or at least two alternatives are at most $\bar{a}$. Suppose the latter, as in the example in Figure 4.1. Given the strict concavity of the individuals' payoff functions, no individual is indifferent among all the winning alternatives, so in every equilibrium every individual votes for one of the winning alternatives by Lemma 4.1*a* and prefers the alternative for which she votes to every other winning alternative by Lemma 4.1*b*. I claim that an individual, say i, who votes for the smallest winner, a_1, can increase her payoff by switching her

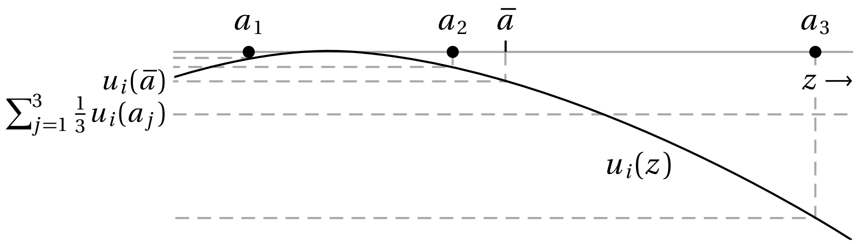


Figure 4.1 A strictly concave payoff function of individual i in a plurality rule voting game in which the alternatives are real numbers. If, when the individual votes for a_1, the set of winners is $\{a_1, a_2, a_3\}$, then her payoff is $\sum_{j=1}^{3} \frac{1}{3} u_i(a_j)$; if she switches her vote to a_2, the set of winners becomes $\{a_2\}$, yielding her the larger payoff $u_i(a_2)$.

vote to the next smallest winner, a_2, which causes that alternative to become the unique winner. The reason is that given $u_i(a_1) > u_i(a_2)$, the strict concavity of u_i, and the fact that $a_2 \leq \bar{a}$, we have $u_i(a_2) \geq u_i(\bar{a})$, and given the strict concavity of u_i, $u_i(\bar{a})$ is greater than the average of i's payoffs to the winners, which is her payoff for the action profile. Part *a* of the next result states this conclusion and the remaining parts state other properties of the Nash equilibria.

Proposition 4.3: Nash equilibria of plurality rule voting game in spatial setting with strictly concave payoff functions

Let $G = \langle N, X, (u_i)_{i \in N} \rangle$ be a plurality rule voting game in which $X \subseteq \mathbb{R}$ is a (nonempty) closed interval and u_i is strictly concave for each $i \in N$.

a. In any Nash equilibrium of G the number of winning alternatives is at most two.

b. If N contains at least three individuals, for every alternative $a \in X$ the game G has a Nash equilibrium in which a is the sole winner. If N contains at least four individuals, for every alternative $a \in X$ other than the boundary points of X, G has a Nash equilibrium in which a is the sole winner and no individual's action is weakly dominated.

c. In a Nash equilibrium of G with two winning alternatives, each individual who is not indifferent between these alternatives votes for the one she prefers.

Proof

a. This result is proved in the text.

b. For any $a \in X$, the action profile in which every individual votes for a is a Nash equilibrium because no change in any individual's action has any effect on the outcome. If a is not a boundary point of X, it is not the lowest ranked alternative for any individual, and hence by Proposition 4.1*b* if there are at least four individuals then voting for a is not weakly dominated for any individual.

c. This result follows from parts a and b of Lemma 4.1.

Exercise 4.6: Variant of plurality rule voting game in spatial setting

Consider a variant of a plurality rule voting game satisfying the conditions in Proposition 4.3 in which each individual's preferences are lexicographic: she is primarily concerned with the set of winning alternatives, but among actions that yield the same set of winning alternatives (given the other individuals' actions) she prefers to abstain. What is the set of Nash equilibria of this game?

4.3 Strategic and sincere voting: divided majority

When there are two alternatives, an individual's voting for her favorite alternative weakly dominates her voting for the other alternative (Proposition 3.1). In the argot of the field, her voting *sincerely* is a weakly dominant action. When there are three or more alternatives, voting sincerely is no longer a weakly dominant action: if, among the other individuals' votes, a and b are tied and every other alternative gets at least two fewer votes, an individual whose favorite alternative is neither a nor b and who is not indifferent between these alternatives is better off voting for whichever of a and b she prefers, making that alternative the winner, than for her favorite alternative, which results in a tie between a and b. Such an individual may be perfectly sincere in the everyday sense of the word ("proceeding from genuine feelings"), but her action is called *strategic*. In an equilibrium, are individuals' votes sincere or strategic?

In this section, I consider this question for an environment known as divided majority. There are three alternatives, say a, b, and c. A majority of individuals rank c last; among these individuals, n_A prefer a to b (type A) and n_B prefer b to a (type B). The remaining minority of individuals (type C, who number n_C)

	Type A n_A individuals	Type B n_B individuals	Type C n_C individuals
a	1	v	0
b	v	1	0
c	0	0	1
	majority		

Figure 4.2 The payoffs in a divided majority, with $\max\{n_A, n_B\} < n_C < n_A + n_B$ and $v \in (0,1)$.

prefer c to both a and b, between which they are indifferent. Although types A and B constitute a majority, each of these types is less populous than type C. That is, $\max\{n_A, n_B\} < n_C < n_A + n_B$. The payoffs are given in Figure 4.2.

The associated collective choice problem has a Condorcet winner. If $n_A > n_B$ then a is the unique Condorcet winner, if $n_A < n_B$ then b is the unique Condorcet winner, and if $n_A = n_B$ then both a and b are Condorcet winners. But if, in the associated plurality rule voting game, every individual votes sincerely, then c wins.

Voting for c and abstaining are weakly dominated for each individual of type A or B (by Proposition 4.1a) and voting for a or b and abstaining are weakly dominated for each individual of type C. So in any Nash equilibrium in which no individual uses a weakly dominated action, all individuals of type A or B vote for a or b and all individuals of type C vote for c. The set of such equilibria are of two types. In one type, the numbers of votes cast for a and for b are both at most $n_C - 2$, resulting in a win for c. Two such equilibria are illustrated in Figures 4.3a and 4.3b; in the first case each individual votes sincerely. In the other type of equilibrium, for either $x = a$ or $x = b$, $n_C + 1$ or more votes are cast for x, resulting in a win for x, as in Figures 4.3c and 4.3d.

If, for a given action profile, some change in an individual's vote would affect the outcome, we say that her vote is *pivotal* at that action profile. At the equilibrium shown in Figure 4.3c, where b gets one more vote than c, the vote of every individual of type A or B is pivotal: if an individual voting for a switches her vote to c or an individual voting for b switches her vote to a or switches to abstention then the outcome changes to a tie between b and c. (Also, if an individual voting for b switches her vote to c then the outcome changes to a win for c.) At none of the other equilibria shown in the figure is any individual's vote pivotal.

In this model, at any action profile an individual's vote is either pivotal or not. If we modify the model so that each individual is uncertain of how the other individuals will vote, a formulation that seems appropriate for many elections, an individual may believe that her vote is pivotal with positive probability less than

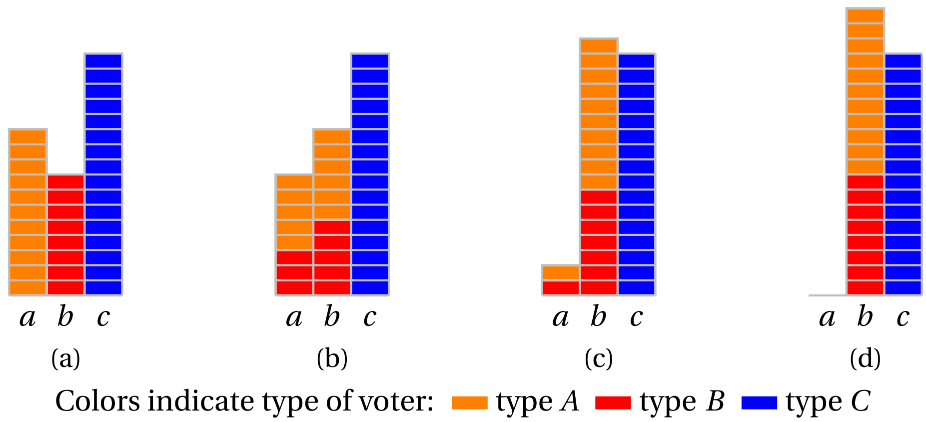


Figure 4.3 Some Nash equilibria in weakly undominated actions for a plurality rule voting game modeling a divided majority (Figure 4.2) in which $n_A > n_B$. Each block in each column represents an individual's vote for one of the alternatives. The colors of the blocks indicate the type of individual casting the vote: orange for type a, red for type b, and blue for type c. In panel a every individual votes sincerely, whereas in the other panels some individuals of types a and/or b vote strategically.

one, and even if this probability is small it may significantly affect her strategic calculations. If, for example, an individual of type B believes that, even though the pattern of votes is likely to be the one shown in Figure 4.3a, there is a small chance that enough of her fellow type B individuals will vote for a to make a and c tie, but no chance that enough type A individuals will vote for b to make b and c tie, then her optimal action may be to vote for a. The reason is that at any action profile for which the vote difference between the leading alternatives is two or more, her vote is not pivotal. The point is that the configurations of the other individuals' votes that determine an individual's optimal vote are the ones for which her vote is pivotal. If, at the only action profiles at which her vote is pivotal she is better off voting for a than for b, then she should vote for a.

One model that captures the individuals' uncertainty about each other's actions is a Bayesian game in which each individual knows her own type (preferences), but is uncertain of the other individuals' types. A strategy of each individual in such a game assigns an action (a vote for a, b, or c, or abstention) to each of her possible types. A strategy profile is a Nash equilibrium if the action of each type of every individual i is optimal given the distribution of the other individuals' actions implied by their strategies and i's belief about the distribution of their types. One interpretation of such an equilibrium is that every individual knows every other individual's strategy, possibly from her long experience of voting in similar elections, and, given her probabilistic belief about the distribution of the other individuals' types in the population, chooses her vote optimally given her own type.

In Nash equilibria of the Bayesian game, do individuals vote sincerely or strategically? The answer depends on the individuals' beliefs about each other's types. I start by assuming that for some numbers p_A, p_B, and p_C, each individual believes that the type of every other individual is t with probability p_t, for $t = A$, B, and C, independently of the type of every other individual. This assumption on the structure of the individuals' beliefs means that as the size of the population increases, every individual becomes increasingly certain of the proportions of the types in the population, and for this reason is not particularly plausible. However, it yields a striking result that is illuminating.

For simplicity, I assume that the number of individuals is odd and abstention is not an option for any individual.

Definition 4.2: Divided majority with independent beliefs

A *divided majority with independent beliefs* $\langle N, \{a, b, c\}, \{A, B, C\}, v, (p_A, p_B, p_C)\rangle$, where N is a finite set (of individuals) with an odd number of members that is at least five, a, b, and c are alternatives, A, B, and C are preference types, $v \in (0, 1)$, and p_A, p_B, and p_C are positive numbers with $p_A + p_B + p_C = 1$ and $\max\{p_A, p_B\} < p_C < p_A + p_B$, is the following Bayesian game.

Players

The set N.

States

The set of states is the set of profiles $(t_j)_{j\in N}$ of preference types, with $t_j \in \{A, B, C\}$ for every player $j \in N$.

Actions

The set of actions of each player is $\{$*vote for* a, *vote for* b, *vote for* $c\}$.

Signals

The signal function τ_i of each player i is given by $\tau_i((t_j)_{j\in N}) = t_i$ (every player knows her own preference type, but no other player's preference type).

Prior beliefs

Every player believes that the preference type of every individual is A with probability p_A, B with probability p_B, and C with probability p_C, independently of every other individual's preference type.

Payoffs

The Bernoulli payoff function of each player over the set of pairs of action profiles and states is defined as follows.

For each preference type $t \in \{A, B, C\}$ and each alternative $z \in \{a, b, c\}$, let $u_t(z)$ be the number given in the cell in column t and row z of Figure 4.2, and for any action profile x, denote by $W(x) \subseteq \{a, b, c\}$ the set of alternatives that receive the most votes. For any action profile x and any profile of preference types, the Bernoulli payoff for x of each player with each preference type $t \in \{A, B, C\}$ is $\sum_{w \in W(x)} u_t(w)/|W(x)|$.

In a divided majority with independent beliefs, voting for c is weakly dominated for an individual of type A or type B and voting for a or b is weakly dominated for an individual of type C, so the main question is whether individuals of types A and B vote for a or for b. The argument that, regardless of the values of p_a, p_B, and p_C, the game has an equilibrium in which all individuals of types A and B vote for a, and also one in which they vote for b, is straightforward.

Proposition 4.4: Nash equilibria of divided majority with independent beliefs

A divided majority with independent beliefs has a Nash equilibrium in which every individual votes for a if her preference type is A or B and votes for c if her preference type is C, and also one in which every individual votes for b if her preference type is A or B and votes for c if her preference type is C.

Proof

For the strategy profile in which every individual of type A or B votes for a and every individual of type C votes for c, the set of winners is $\{a\}$ or $\{c\}$, depending on the realized distribution of preference types. (A tie is not possible because the number of individuals is odd and abstention is not allowed.) The action of an individual of type A or B is pivotal at this strategy profile if and only if the the realized number of individuals of types A and B exceeds the realized number of individuals of type C by 1, an event with positive probability. For such a realization of types, the outcome of the strategy profile is a win for a. If an individual of type A or B changes her vote to b, the outcome changes to a tie between a and c (given that the number of individuals is at least five), and if she changes it to c, the outcome changes to a win for c. She prefers a win for a to each of these outcomes, so her voting for a is optimal.

Similar arguments apply to an individual of type C and to the strategy profile in which every individual of type A or B votes for b.

Is the strategy profile in which every individual votes sincerely also a Nash equilibrium of the game? We know that for some configurations of the other individuals' actions an individual is better off voting for an alternative different from her favorite, so that whether a vote for her favorite alternative is optimal depends on the probabilities of such configurations and her payoffs for the alternatives.

Denote the number of individuals by $n+1$. For any given individual i, the set of configurations of the other n individuals' preference types is illustrated in Figure 4.4. Each small disk represents one possible configuration; the three corners of the triangle represent configurations in which all n individuals have the same type, and the (brown) disk in the center represents the configuration in which the same number of individuals have each type. (The diagram assumes that n is divisible by both 2 and 3.) For the strategy profile in which each individual votes sincerely, the diagram represents also the possible configurations of the other individuals' votes for the three alternatives. Now, the alternative for which i optimally votes depends on the probabilities of vote configurations for the other individuals for which her vote is pivotal. These configurations are the ones in which there is a tie or near-tie (the margin of victory of the outright winner is one vote) for first place among the other individuals' votes. These close races are indicated in color in Figure 4.4.

Consider type A of individual i. If the close race is between a and b (blue disks) or between a and c (purple disks), then she optimally votes (sincerely) for a. But if it is between b and c (magenta disks) then, given that she prefers b to c, she optimally votes for b. To determine her optimal action we need to compare the probabilities of these events. The probability density over the configurations of types that the model generates for $n = 60$, $p_A = 0.25$, $p_B = 0.3$, and $p_C = 0.45$ is shown in Figure 4.5. In this case, if all of the n individuals other than i vote sincerely, close races between b and c are more likely than ones between a and b or between a and c, so there is a number v^* such that if $v \leq v^*$ then type A of individual i optimally votes sincerely, for a, and if $v \geq v^*$ then she optimally votes strategically, for b. (If $v = v^*$, both votes are optimal.) Whenever $p_A < p_B < p_C$, the conclusion is the same.

Now suppose that the number of individuals increases. As it does so, the probability density over the configurations of types of the n individuals other than i becomes increasingly concentrated around (np_A, np_B, np_C), and the probability of a close race decreases (even if $p_A = p_B = p_C = \frac{1}{3}$). For $p_A = 0.25$, $p_B = 0.3$, and $p_C = 0.45$, Figure 4.6a shows this probability as a function of n; for

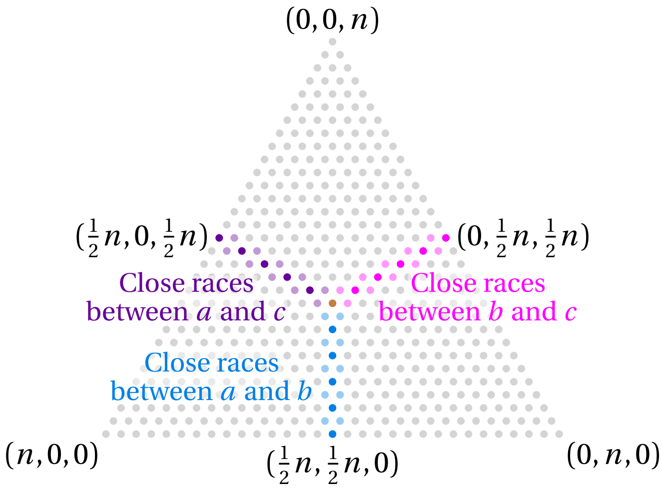


Figure 4.4 Possible configurations of the types of n individuals in a divided majority (Figure 4.2). Each disk represents one configuration (n_A, n_B, n_C). The corners of the triangle are the configurations in which all n individuals have the same type and the central disk (brown) is the configuration in which there are $\frac{1}{3}n$ individuals of each type. (The diagram assumes that n is divisible by 2 and by 3.) The colored disks represent close races when every type of every individual votes sincerely.

$n > 300$ it is less than 0.001, and for $n > 430$ it is less than 0.0001. If $p_A < p_B < p_C$ then as the number of individuals increases, the proportion of the probability of a close race attributable to a close race between b and c increases, approaching one. Figure 4.6b shows the proportions attributable to close races between each of the three pairs of alternatives, as a function of n, for the case in which $p_A = 0.25$, $p_B = 0.3$, and $p_C = 0.45$. (A tie between all three alternatives is also possible; the probability of this event rapidly decreases to zero.) In this case, for $n > 320$ the proportion attributable to a close race between b and c exceeds 0.99. The proof of the general result that if $p_A < p_B < p_C$ then the probabilities of a tie or near tie for the most populous type between A and B and between A and C become negligible compared with the probability of a tie or near-tie between B and C is routine but intricate; Lemma 2 of Palfrey 1989 is the result for a different tie-breaking rule. As a consequence of this result, for any given value of $v > 0$ there is a number N such that if $n > N$ then the strategy profile in which each type of each individual votes sincerely is not a Nash equilibrium, because almost every case in which the vote of an individual of type A is pivotal is a close race between b and c, so that such an individual optimally votes for B if each type of every other individual votes sincerely.

An implication of this result is that if $p_A < p_B < p_C$ then for any given value of $v > 0$, if the number of individuals is sufficiently large then if the game has a Nash equilibrium in which all individuals of the same type vote for the same

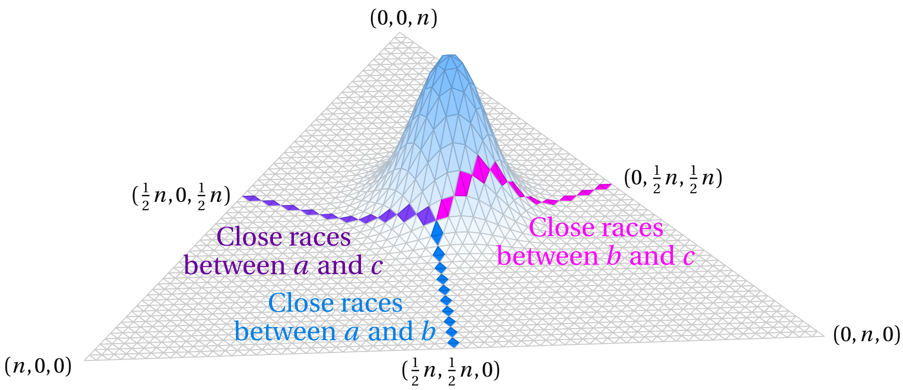
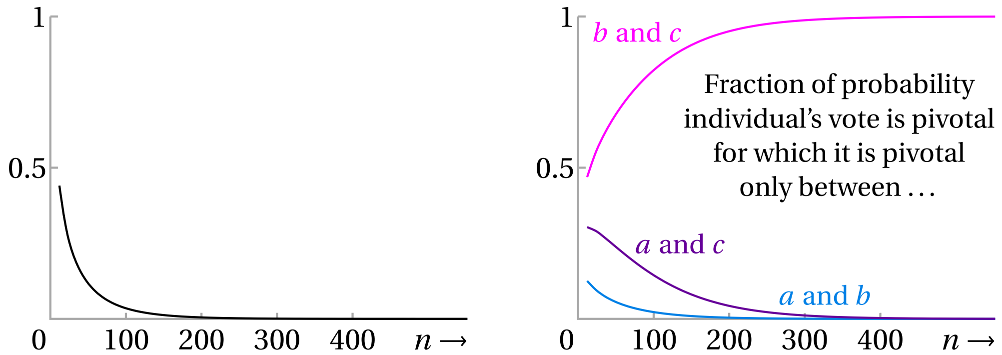


Figure 4.5 The probability density over the configurations of types for 60 individuals when each individual's type is a with probability 0.25, b with probability 0.3, and c with probability 0.45, independently of the types of the other individuals. (The distribution is discrete; its density is smoothed in the figure.) The regions indicated in dark blue, purple, and magenta are the closes races between two of the alternatives when every type of every individual votes sincerely.

alternative and no type uses a weakly dominated action, in any such equilibrium all individuals of types A and B vote for the same alternative, either a or b, and all individuals of type C vote for c. In particular, in any such equilibrium only two of the three alternatives receive votes.

These results concern the limit as the population size increases without bound. They do not mean that for any number that you or I might classify as large, the game has no equilibrium in which every type of every individual votes sincerely. The sufficiently large number depends on the parameters, and could be 100, 1 million, 1 billion, or any other number. Note also that the analysis implies that for any values of p_A, p_B, and p_C with $p_A < p_B < p_C$ and any given (finite) population size, there is a positive number v^* such that for $v < v^*$ the game has a Nash equilibrium in which every type votes sincerely. Finally, note that the analysis assumes that each individual chooses the alternative for which to vote by comparing the probabilities that votes for each of the alternatives changes the outcome, even though these probabilities are minuscule in a large population. If her motivation for voting is, instead, expressive (see Section 6.2), then she may vote sincerely regardless of the population size.

These conclusions depend on the model of the individuals' beliefs. The assumption that each individual believes that every other individual's type is t with probability p_t, independently of the other individuals' types, has two significant implications. First, when the number of individuals is large, every individual is almost certain of the proportions of the types in the population. Second, every individual's belief is the same. These implications seem implausible. Even in



(a) The probability that an individual's vote is pivotal when each of n other individuals votes for a with probability 0.25, for b with probability 0.3, and for c with probability 0.45, independently of the other individuals.

(b) The fraction of the probability that an individual's vote is pivotal under the conditions of panel (a) for which it is pivotal only between b and c (magenta curve), a and c (purple curve), and a and b (blue curve).

Figure 4.6

a large population—particularly in a large population?—each individual seems likely to be uncertain of the distribution of the other individuals' characteristics, and different individuals get information from different sources, so that their beliefs are likely to differ.

An alternative model assumes that the probabilities p_A, p_B, and p_C are themselves uncertain, so that the uncertainty about the proportion of individuals of each type in the population does not vanish as the population size increases without bound. In this case, every individual of each type may optimally vote sincerely if all the other individuals' votes are sincere, regardless of the population size. If, for example, an individual of type a believes that a tie between b and c is more likely than a tie between a and c, but thinks that the difference between the likelihoods is not large, then the expected loss from voting for b rather than a in the event of a tie between a and c may outweigh the expected gain from doing so in the event of a tie between b and c. For specific models of the uncertainty regarding p_A, p_B, and p_C, we may be able to say more about the circumstances under which the strategy profile in which every type of every individual votes sincerely.

If the individuals' beliefs differ, then their strategic calculations differ. Suppose that each individual of type a believes that in expectation, individuals of type a outnumber those of type b, and the reverse is true for individuals of type b. Then if all other individuals vote sincerely, an individual of type a may conclude that voting for a is optimal and an individual of type b may conclude that voting for b is optimal, so that again an equilibrium in which all individuals vote sincerely exists. If instead each individual of type a or type b believes that their type is outnumbered, then an equilibrium may exist in which every individual of type a votes for b and every individual of type b votes for a.

One way in which an individual may gather information about the voting intentions of the other individuals is through her interactions with those individuals. The next exercise asks you to analyze a model in which each individual bases her vote on the information obtained from a random sample of two other individuals.

Exercise 4.7: Sampling equilibrium in a divided majority

Consider the divided majority in Figure 4.2. Denote the fraction of the population consisting of individuals of type t by q_t for $t = A, B, C$. Each individual observes the voting intentions of two random-selected individuals. An individual of type A whose sample consists of one individual who intends to vote for B and one who intends to vote for C concludes that she should vote for B; for every other sample she votes for A. Symmetrically, an individual of type B whose sample consists of one individual who intends to vote for A and one who intends to vote for C concludes that she should vote for A, and for every other sample votes for B. An individual of type C votes for C regardless of her sample. For each type T, let $p_T(A)$, $p_T(B)$, and $p_T(C)$ be the fractions of the individuals of that type who vote for each alternative. In an equilibrium, $p_A(C) = p_B(C) = 0$, $p_C(C) = 1$, $p_A(B)$ is equal to the probability that the sample of an individual of type A consists of one individual who intends to vote for B and one who intends to vote for C, and $p_B(A)$ is equal to the probability that the sample of an individual of type B consists of one individual who intends to vote for A and one who intends to vote for C. Assume that the number of individuals is large enough that you can take the probability that a given member of the sample of an individual i of type T is an individual of type T other than i to be q_T. Find the equilibria. Why is there no equilibrium in which all individuals of types A and B vote for the same alternative?

Notes

Lemma 4.1 is based on Lemmas 1 and 2 of Feddersen et al. (1990). One source of Proposition 4.1b, which Duggan and Sekiya (2009, 879) say is well known, is Dhillon and Lockwood (2004, Lemma 1). The proof of Proposition 4.3a is based on the proof of Lemma A.3 in Feddersen et al. (1990). The main part of Section 4.3 is based on Palfrey (1989). Myatt (2007) and Bouton et al. (2017) study models in which the uncertainty regarding the proportions of the various preference types does not vanish as the population size grows without bound.

Exercise 4.3 is taken from Duggan and Sekiya (2009). Exercise 4.4 is based on Dhillon and Lockwood (2004), who study the iterated elimination of weakly dominated actions in plurality rule voting games more generally. Exercise 4.5 is based on De Sinopoli and Iannantuoni (2007). Voting under proportional rule is explored further by Indriðason (2011) and Cho (2014). The notion of equilibrium in Exercise 4.7 is taken from Osborne and Rubinstein (2003).

Solutions to exercises

Exercise 4.1

Let i be an individual, let a be one of i's favorite alternatives, let z be an alternative she likes least, and let u_i be her payoff function over alternatives.

If i is indifferent between all alternatives, for any alternative b her voting for b is not weakly dominated by her voting for any other alternative.

Now suppose that i is not indifferent between all alternatives, so that $u_i(a) > u_i(z)$. I argue that i's voting for b is weakly dominated if and only if her payoff to $\{b\}$ is at most her payoff for the set of winners $\{a, b, z\}$, in which case $u_i(a) > u_i(b)$.

First, i's voting for b is weakly dominated if and only if it is weakly dominated by her voting for a. I now consider the conditions under which her voting for b is weakly dominated by her voting for a.

- If the other two individuals vote for the same alternative, i's vote does not affect the set of winners.
- If the other two individuals vote for different alternatives, say x and y, neither of which is a or b, the set of winners when i votes for a is $\{a, x, y\}$ and the set of winners when she votes for b is $\{b, x, y\}$, so her payoff is higher when she votes for a.
- If one of the other individuals votes for a and the other votes for an alternative, say x, other than a or b, then the set of winners is $\{a\}$ if i votes for a and $\{a, b, x\}$ if she votes for b, so her payoff is higher if she votes for a.
- If one of the other individuals votes for b and the other votes for an alternative, say x, other than a or b, then the set of winners is $\{a, b, x\}$ if i votes for a and $\{b\}$ if she votes for b, so her payoff is at least as high if she votes for a if and only if her payoff for $\{a, b, x\}$ is at least as high as her payoff for $\{b\}$.

- If one of the other individuals votes for a and the other votes for b then the set of winners is $\{a\}$ if i votes for a and $\{b\}$ if she votes for b, so her payoff is higher if she votes for a.

In the second, third, and fifth cases, i's payoff is higher when she votes for a than it is when she votes for b, and in the first case it is the same, so her voting for b is weakly dominated by her voting for a if and only if for every alternative x other than a and b her payoff for $\{a,b,x\}$ is at least as high as her payoff for $\{b\}$, which is the case if and only if her payoff for $\{a,b,z\}$ is at least as high as her payoff for $\{b\}$.

Exercise 4.2

For each individual $i \in N$, denote by T_i the set of i's favorite alternatives. For no alternative a in T_i is voting for a weakly dominated for individual i.

Thus if some alternative a is a member of T_1, T_2, and T_3, then the action profile in which each individual votes for a is a Nash equilibrium in which no individual's action is weakly dominated.

If for some alternatives a and b, a is a member of two of the sets T_1, T_2, and T_3, say T_i and T_j, and b is a member of the remaining set, say T_k, then the action profile in which i and j vote for a and k votes for b is a Nash equilibrium in which no individual's action is weakly dominated.

The remaining possibility is that the sets T_1, T_2, and T_3 have no alternative in common. In this case, select one alternative from each set, say a from T_1, b from T_2, and c from T_3.

- If the action profile in which 1 votes for a, 2 votes for b, and 3 votes for c is a Nash equilibrium, we are done.
- If not, one of the individuals can increase her payoff by changing her vote. Without loss of generality, assume that individual 1 can do so.
    - If individual 1 changes her vote to another alternative in T_1, her payoff does not change.
    - If individual 1 changes her vote to an alternative, say z, outside T_1 and $z \notin \{b,c\}$, then the set of winners changes to $\{z,b,c\}$, so she is worse off.
    - Thus individual 1 must increase her payoff by changing her vote to either b or c, so that the set of winners becomes $\{b\}$ or $\{c\}$.
        - If b and c yield her the same payoff, then her payoff decreases.
        - Thus b and c yield her different payoffs. Let b be the one with the higher payoff. Given that individual 1 can increase her payoff

by changing her vote, her payoff from $\{b\}$ is greater than her payoff from $\{a,b,c\}$. Thus the action profile in which individuals 1 and 2 vote for b and individual 3 votes for c is a Nash equilibrium (because the best deviation for individual 1 is to vote for a, which generates the set of winners $\{a,b,c\}$) and by the result of Exercise 4.1 individual 1's action of voting for b is not weakly dominated.

Exercise 4.3

For the action profile x^1, in which every individual votes for her favorite alternative, the set of winners is $\{a,c\}$. This action profile is not a Nash equilibrium because individual 2, by changing her vote to b, can change the set of winners to $\{a,b,c\}$, thereby increasing her payoff from 0 to $\frac{1}{3}\cdot 6 = 2$. Individual 6 can also increase her payoff by changing her vote to a, which changes the set of winners to $\{a\}$ and hence increases her payoff from $\frac{1}{2}\cdot 6 = 3$ to 6.

If we select individual 2 then the list of votes in x^2 is (c,b,c,a,a,b) and the set of winners is $\{a,b,c\}$. This action profile is not a Nash equilibrium because individual 3, by changing her vote to b, can change the set of winners to $\{b\}$, thereby increasing her payoff from $\frac{1}{3}\cdot 9 + \frac{1}{3}\cdot 6 = 5$ to 6. Individual 5 can also increase her payoff by deviating to vote for c, and individual 6 can do so by deviating to vote for a.

If we select individual 3 then the list of votes in x^3 is (c,b,b,a,a,b) and the set of winners is $\{b\}$. This profile is a Nash equilibrium, so the procedure ends.

(If at the second step we choose individual 6 and set the list of votes in x^2 to be (c,d,c,a,a,a) then the set of winners is $\{a\}$. This profile is a Nash equilibrium, so the procedure ends. If at the third step we select individual 5, we reach the Nash equilibrium with votes (c,b,c,a,c,b), and if we select individual 6, we reach the Nash equilibrium with votes (c,b,c,a,a,a).)

Exercise 4.4

Denote the alternatives a, b, and c, and suppose that more than two-thirds of the individuals rank c below a and b. By Proposition 4.1, each individual's only weakly dominated action is a vote for her least preferred alternative. After eliminating this action for each individual, the set of winners generated by every remaining action profile is $\{a\}$, $\{b\}$, or $\{a,b\}$, because fewer than one-third of the individuals vote for c. Thus each individual's voting for whichever of a and b she prefers weakly dominates her other remaining actions, by the same argument as for a two-alternative game. We are left with the action profile in which every citizen votes for her favorite alternative in $\{a,b\}$.

Exercise 4.5

a. Let i be an individual and let x be an action profile in which i votes for a_l. Suppose that $a_l > a_1$. If i switches her vote from a_l to a_1 then the outcome decreases from $O(x) = \sum_{j=1}^{k} v_j(x)a_j/n$ to

$$\frac{1}{n}\left(\sum_{j=1}^{k} v_j(x)a_j - a_l + a_1\right) \geq \sum_{j=1}^{k} v_j(x)a_j/n - (a_k - a_1)/n = O(x) - (a_k - a_1)/n.$$

Thus if i's favorite alternative is at most $O(x) - (a_k - a_1)/n$ then she is better off voting for a_1 than for a_l and hence x is not a Nash equilibrium. Similarly if $a_l < a_k$ and i's favorite alternative is at least $O(x) + (a_k - a_1)/n$, she is better off voting for a_k than for a_l, so that x is not a Nash equilibrium.

b. The action profile in which every individual whose favorite alternative is less than z^* votes for a_1 and every individual whose favorite alternative is greater than z^* votes for a_k is a Nash equilibrium. The reason is that if an individual whose favorite alternative is less than z^* deviates to vote for an alternative other than a_1 then the outcome increases, which makes her worse off, and if an individual whose favorite alternative is greater than z^* deviates to vote for an alternative other than a_k then the outcome decreases, which makes her worse off.

Exercise 4.6

Every equilibrium of the modified game is an equilibrium of the original game, so by Proposition 4.3 in every equilibrium of the modified game the number of winning alternatives is at most two.

The modified game has no equilibrium with a single winning alternative unless every individual's favorite alternative is the same. To see why, consider an action profile that generates a single winning alternative, say a. Any individual who votes for an alternative other than a can switch to abstention without affecting the outcome, and if all votes are cast for a and more than one individual votes then any individual can switch to abstention without affecting the outcome. Thus in any equilibrium of the modified game one individual votes for a. But if some individual's favorite alternative b differs from a, an action profile with one vote for a is not an equilibrium because that individual can switch from abstention to voting for b and induce a tie between a and b, which she prefers to a.

Consider a two-alternative equilibrium of the original game in which at least one individual who votes is indifferent between the alternatives. If such an individual deviates to abstention then her payoff remains the same, so that such an equilibrium is not an equilibrium of the modified game.

Now consider a two-alternative equilibrium of the original game in which no individual who votes is indifferent between the alternatives. If an individual who votes deviates to voting for another alternative or to abstention, the resulting outcome is worse for her. Thus such an equilibrium is an equilibrium of the modified game.

Exercise 4.7

The equilibrium conditions are

$$p_A(B) = 2(q_A p_A(B) + q_B p_B(B))(1 - q_A - q_B)$$
$$p_B(A) = 2(q_A p_A(A) + q_B p_B(A))(1 - q_A - q_B).$$

The right-hand side of the first equation is the probability that an individual's sample consists of one individual who intends to vote for B and one who intends to vote for C, and the right-hand side of the second equation is the analogous expression for A and C. These equations have a unique solution,

$$p_A(B) = 2(1 - q_A)q_B - 2q_B^2$$
$$p_B(A) = 2(1 - q_B)q_A - 2q_A^2.$$

For the example in which $q_A = 0.25$, $q_B = 0.35$, and $q_C = 0.4$, $p_A(B) = 0.28$ and $p_B(A) = 0.2$.

The proportion of the population that votes for A in the equilibrium is

$$q_A p_A(A) + q_B p_B(A) = q_A(1 - 2(1 - q_A)q_B + 2q_B^2) + q_B(2(1 - q_B)q_A - 2q_A^2) = q_A$$

and the proportion that votes for B is q_B. Thus even though some individuals vote strategically, the total proportions who vote for each alternative are equal to the proportions of the types in the population.

There is no equilibrium in which all individuals of types A and B vote for the same alternative, say A, because for such a voting pattern an individual of type B who gets a sample consisting of two individuals who intend to vote for A is assumed to vote for B.

5 Sequential pairwise voting

5.1 Binary agendas 157
5.2 Successive agendas 168
5.3 Amendment agendas 170
5.4 Single-peaked preferences and convex agendas 178

One type of mechanism a group of individuals facing a collective choice problem can use to select an alternative involves a sequence of two-way votes. Such a mechanism is called a binary agenda. Suppose, for example, that the alternatives are a, b, and c. The individuals might first vote on a motion to drop a from consideration. If that motion passes, they might then vote between b and c, and if it fails, they might vote whether to choose c, with the failure of that motion leading to a vote between a and b. This procedure is illustrated in Figure 5.1. How does the outcome of a binary agenda depend on the sequence of choices and the individuals' preferences? Does the outcome have desirable properties?

Synopsis

We model a binary agenda as an extensive game and apply to it the solution concept of subgame perfect equilibrium with the restriction that no player's vote at any point in the game is weakly dominated. The resulting strategy profiles are referred to as the outcomes of sophisticated voting. Say that a finite collective choice problem with an odd number of individuals, each of whose preference relations is strict, is *odd-strict*. As you might suspect, for an odd-strict collective choice problem that has a strict Condorcet winner, that alternative is the unique outcome of sophisticated voting for any binary agenda (Proposition 5.1).

For collective choice problems with no strict Condorcet winner, a set of alternatives called the top cycle set is relevant. Recall that a strict Condorcet winner beats every other alternative in two-way comparisons. Say that an alternative x *indirectly* beats another alternative y if for some alternatives $z_1, \ldots, z_k$, alternative x beats z_1, z_1 beats z_2, $\ldots$, z_{k-1} beats z_k, and z_k beats y. The top cycle

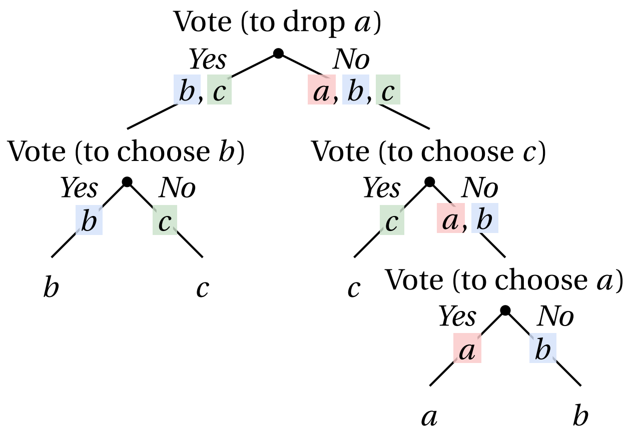


Figure 5.1 An example of a procedure in a binary agenda.

set consists of the alternatives that beat or indirectly beat every other alternative. Proposition 5.3 shows that for an odd-strict collective choice problem, every outcome of sophisticated voting in any binary agenda is in the top cycle set, and for any alternative in the top cycle set there is a binary agenda for which that alternative is the outcome of sophisticated voting. The members of the top cycle set, unlike a strict Condorcet winner, are not all reasonable outcomes of a collective choice problem. For example, a member of the top cycle set may be dominated in the sense that another alternative is preferred by every individual (Example 5.1).

Sections 5.2 and 5.3 consider binary agendas that take specific forms that model the procedures used in some legislatures. Section 5.2 considers successive agendas. In such agendas, the alternatives are considered in some order. A vote is taken on whether to choose the first alternative or to drop it from consideration; if it is dropped, a vote is taken on whether to choose the second alternative or to drop it from consideration; and so forth. Proposition 5.5 shows that for an odd-strict collective choice problem, for every alternative in the top cycle set there is a successive agenda for which the alternative is the outcome of sophisticated voting.

By contrast, for the amendment agendas considered in Section 5.3, the set of outcomes of sophisticated voting is a subset of the top cycle set. In an amendment agenda, as in a successive agenda, the alternatives are considered in some order. A vote is taken on whether to eliminate the first alternative or the second alternative from consideration; then a vote is taken on whether to eliminate the alternative not eliminated in the first round or the third alternative; and so forth. For amendment agendas, a subset of the top cycle set called the Banks set plays the role that the top cycle set plays for successive agendas. Unlike the top cycle set, the Banks set contains no alternative that is dominated (Exercise 5.6). Proposition 5.8 shows that for an odd-strict collective choice problem, the outcome of sophisticated voting in any amendment agenda is in the Banks set, and for every alternative in the Banks set we can find an amendment agenda for which the alternative is the outcome of sophisticated voting.

Section 5.4 considers an environment in which the individuals are not fully informed about each other's preferences, as they are in the model of a binary agenda. Each individual is assumed to know her own preferences but not the preferences of any other individual. The analysis is restricted to agendas for which the preference profile is single-peaked, so that the collective choice problem has a strict Condorcet winner, which means that under perfect information the outcome of sophisticated voting is the strict Condorcet winner (Proposition 5.1). It identifies a condition on the structure of an agenda, convexity, under which each individual's voting sincerely is optimal when every other individual votes sincerely, *regardless* of the other individuals' preferences, and the outcome of such voting is the strict Condorcet winner (Proposition 5.9). Thus for a convex agenda, the mutual optimality of the simple rule to vote sincerely is not sensitive to a lack of information about the other individuals' preferences.

5.1 Binary agendas

A binary agenda consists of a collective choice problem and a procedure for selecting an alternative that consists of a sequence of two-way votes in which every alternative is the outcome of at least one possible sequence. Let $\langle N, X, \succsim \rangle$ be a collective choice problem. The structure of the choices in a binary agenda for selecting an alternative for this problem is modeled as a set H of sequences. This set contains the empty sequence, representing the start of the procedure. Every element in every other member of H is either *Yes* or *No*; for every $h \in H$ either $(h, \mathit{Yes}) \in H$ and $(h, \mathit{No}) \in H$, in which case h is nonterminal, or there is no value of x for which $(h, x) \in H$, in which case h is terminal. Denote the set of terminal members of H by Z. Let O be a function that associates a member of the set X of alternatives (an outcome) with every terminal member of H and suppose that every alternative is the outcome of some sequence of votes: for every alternative $x \in X$ there is a terminal member h of H for which $O(h) = x$.

The binary agenda generated by $\langle N, X, \succsim \rangle$, Z, and O is the extensive game with perfect information and simultaneous moves in which the set of players is N, the set of terminal histories is Z, the player function assigns the set N of all players to every nonterminal history, the set of actions of each player is $\{\mathit{Yes}, \mathit{No}\}$ after every history, the action chosen by a majority of players is the next component of the history, and each player i likes the terminal history h at least as much as the terminal history h' if and only if $O(h) \succsim_i O(h')$.

In the example in Figure 5.1, the votes are given interpretations. For example, the first one is interpreted as a vote to drop a. These interpretations are not part of the formal description of the agenda. Everyone knows the structure of the agenda, and whenever a vote takes place, the options are to move down the branch on the left or to move down the branch on the right; these options may or may not have simple interpretations like dropping one of the alternatives from consideration or selecting one of the alternatives.

For simplicity I assume that the number of individuals is odd and that each individual's preferences are strict (no individual is indifferent between any two alternatives).

Definition 5.1: Odd-strict collective choice problem

An *odd-strict collective choice problem* is a finite collective choice problem in which the number of individuals is odd and every individual's preference relation is strict.

Definition 5.2: Binary agenda

Let $\langle N, X, \succsim \rangle$ be an odd-strict collective choice problem with $N = \{1, \ldots, n\}$. Let H be a set of sequences (*i*) that contains the empty sequence, (*ii*) in which every element is either *Yes* or *No*, and (*iii*) for which for every $h \in H$ either $(h, Yes) \in H$ and $(h, No) \in H$, in which case h is *nonterminal*, or there is no value of x for which $(h, x) \in H$, in which case h is *terminal*. Denote the terminal members of H by Z and let O be a function that assigns to each $h \in Z$ an alternative $O(h) \in X$. Assume that for every $x \in X$ we have $O(h) = x$ for some $h \in Z$.

The *binary agenda* $\langle \langle N, X, \succsim \rangle, Z, O \rangle$ is the following extensive game with perfect information and simultaneous moves.

Players

The set of players is the set N (of individuals).

Terminal histories

For any profile $(v_1, \ldots, v_n)$ with $v_i \in \{Yes, No\}$ for all $i \in N$ (a *vote profile*), denote by $M(v_1, \ldots, v_n)$ the member of $\{Yes, No\}$ such that $\{i \in N : v_i = M(v_1, \ldots, v_n)\}$ is a majority of N.

Terminal histories are sequences of vote profiles. Specifically, a sequence $((v_1^1, \ldots, v_n^1), \ldots, (v_1^q, \ldots, v_n^q))$ of vote profiles, where $q \geq 1$ and $v_i^l \in \{Yes, No\}$ for $l = 1, \ldots, q$ and all $i \in N$, is a terminal history if and only if $(M(v_1^1,\ldots,v_n^1),\ldots,M(v_1^q,\ldots,v_n^q)) \in Z$.

Player function

For every nonterminal history $h \in H$, the player function assigns the set N of all individuals to h.

Actions

Each player's set of actions after any nonterminal history is $\{Yes, No\}$.

Preferences

Each player i prefers the terminal history $((v_1^1,\ldots,v_n^1),\ldots,(v_1^q,\ldots,v_n^q))$ to the terminal history $((y_1^1,\ldots,y_n^1),\ldots,(y_1^q,\ldots,y_n^q))$ if and only if

$$O(M(v_1^1,\ldots,v_n^1),\ldots,M(v_1^q,\ldots,v_n^q)) \succ_i O(M(y_1^1,\ldots,y_n^1),\ldots,M(y_1^q,\ldots,y_n^q)).$$

Equilibrium

Assume that whenever an individual votes, she is forward-looking: she votes for the option that, given her expectation of the outcomes of future votes (which, in equilibrium, are correct), leads to the better outcome for her. Specifically, we look for a subgame perfect equilibrium: a strategy profile for which no change in any vote of any individual leads (ultimately) to an outcome that is better for her, given the other individuals' strategies.

We further restrict each individual to weakly undominated strategies. If there are three or more individuals then without this restriction, for every alternative $a \in X$ the game has a subgame perfect equilibrium in which the outcome is a. The reason is the same as the reason that for each alternative a in a two-alternative majority rule voting game the action profile in which every individual votes for a is a Nash equilibrium (Section 3.1.1). Take a terminal history with the outcome a and suppose that at every opportunity every individual votes for the option consistent with the history. Then given that there are three or more individuals, no change in any individual's strategy has any effect on the outcome.

For brevity, we refer to an alternative that is the outcome of a subgame perfect equilibrium in which no individual's strategy in any subgame is weakly dominated as an outcome of sophisticated voting.

Definition 5.3: Sophisticated voting

An alternative a^* is an *outcome of sophisticated voting* in a binary agenda B if it is the outcome of a subgame perfect equilibrium s^* of B in which for no individual i is the restriction of s_i^* to any subgame Γ of B weakly dominated in Γ: i has no strategy s_i' that, for some strategies of the other individuals, yields an outcome in Γ that she prefers to the outcome when she uses s_i^* and, for all strategies of the other individuals, yields an outcome in Γ that she likes at least as much as the outcome when she uses s_i^*.

Suppose that if the vote at the start of some subgame Γ is in favor of *Yes* then, given the individuals' strategies, the outcome of the game is ultimately a, whereas if the vote favors *No* then the outcome is ultimately b. Then it is reasonable that an individual who favors a votes *Yes* at the start of Γ and one who favors b votes *No*, so that the outcome of Γ is the member of $\{a, b\}$ preferred by a majority of individuals. The next result shows that any outcome of sophisticated voting has this property. It is expressed in terms of a game with a fictional single player who prefers one outcome to another if and only if a majority of the individuals do so.

Lemma 5.1: Sophisticated voting and subgame perfect equilibrium

An alternative is an outcome of sophisticated voting in the binary agenda $\langle\langle N, X, \succsim\rangle, Z, O\rangle$ if and only if it is a subgame perfect equilibrium outcome of the variant of an extensive game with perfect information with possibly nontransitive preferences in which there is one player, the set of terminal histories is Z, the player function assigns the player to every nonterminal history, and for any terminal histories $h \in Z$ and $h' \in Z$ the player prefers h to h' if and only if a majority of the members of N prefer $O(h)$ to $O(h')$.

The preferences of the single player in the extensive game with perfect information defined in this result are not transitive if, for example, the collective choice problem $\langle N, X, \succsim\rangle$ does not have a Condorcet winner. However, whenever the player chooses an action in the game she has exactly two options, so that she has a well-defined optimal action.

Proof of Lemma 5.1

Denote the binary agenda by B. The argument uses induction on the length of Γ (i.e. the length of its longest history).

Let Γ be a subgame of B of length 1 (so that Γ is at the end of B); denote the two possible outcomes of Γ by a and b. By Corollary 3.1, in a Nash equilibrium of Γ in which no individual's action is weakly dominated, every individual votes for the option that leads to the alternative in $\{a, b\}$ that she prefers. Thus the outcome of sophisticated voting in the subgame, itself a binary agenda, is the member of $\{a, b\}$ preferred by a majority of

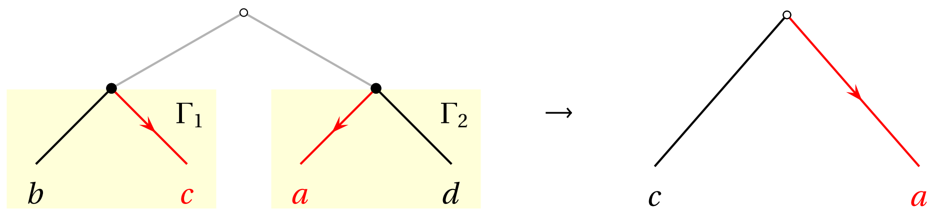


Figure 5.2 An illustration of the argument for Lemma 5.1 for a binary agenda in which a majority of individuals prefer c to b, a majority prefer a to d, and a majority prefer a to c. The agenda, shown on the left, has two subgames of length 1, Γ_1 and Γ_2. In the game shown on the right, each of these subgames is replaced by the outcome of sophisticated voting in the subgame.

individuals.

Now replace each subgame of length 1 with the outcome of sophisticated voting in the subgame. (See Figure 5.2 for an example.) In the resulting binary agenda, repeat the process. Continue in the same manner until reaching the start of the game.

This result makes the outcomes of sophisticated voting easy to find. Starting at the end of the game, we first find the alternative preferred by a majority of individuals among the outcomes of each subgame of length 1. Then we replace each of these subgames with the associated alternative and repeat the process for the resulting game, working our way to the start of the game.

If, for an agenda with the procedure shown in Figure 5.1, for example, a majority of individuals prefer a to b, a majority prefer b to c, and a majority prefer c to a, then the outcome of sophisticated voting in the subgame following the history *Yes* is b and the outcome in the subgame following the history (*No*, *No*) is a, so the outcome in the history following *No* is c, and hence the outcome in the whole game is b.

An application of this procedure shows that for any binary agenda for a collective choice problem with a Condorcet winner (which is strict, given that the number of individuals is odd and their preferences are strict), this alternative is the only outcome of sophisticated voting.

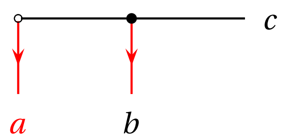
Proposition 5.1: Sophisticated voting in binary agenda with strict Condorcet winner

If an odd-strict collective choice problem $\langle N, X, \succsim \rangle$ has a Condorcet winner then for any binary agenda $\langle \langle N, X, \succsim \rangle, Z, O \rangle$ this alternative is the only outcome of sophisticated voting.

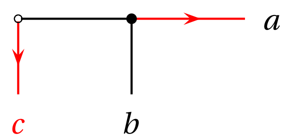
Proof

Denote the Condorcet winner by a^*. Given that the collective choice problem is odd-strict, this winner is strict. Let Γ be a subgame of the binary agenda of length one (i.e. at the end of the game) for which at least one of the outcomes is a^*. Then a^* is the outcome of sophisticated voting in the subgame by Lemma 5.1 and the fact that for every other alternative a a majority of individuals prefer a^* to a. But then in the subgame of length two that includes Γ, the option that leads to Γ, and hence ultimately a^* (or the other option if that also leads to a^*), wins. Working back to the start of the game, we conclude that the only outcome of sophisticated voting is a^*.

For collective choice problems without Condorcet winners, things are more interesting. In particular, the outcome of sophisticated voting depends on the agenda. For example, for the collective choice problem of Example 1.5, which is a Condorcet cycle in which a beats b beats c beats a, the outcome of sophisticated voting for the agenda



is a (b beats c, a beats b), whereas the outcome for the agenda



is c (a beats b, c beats a), and the outcome for the variant of this agenda in which c and b are interchanged is b.

In this example, every alternative is the outcome of sophisticated voting for some agenda. The same is not true for every collective choice problem. Recall that an alternative x is a Condorcet winner if it beats every other alternative y in two-way comparisons, in the sense that a majority of individuals prefer x to y. Say that x *indirectly* beats y if for some alternatives $z_1, \ldots, z_k$, alternative x beats z_1, z_1 beats $z_2, \ldots, z_{k-1}$ beats z_k, and z_k beats y. The set of alternatives that beat or indirectly beat every other alternative is called the top cycle set. A subsequent result, Proposition 5.3, shows that an alternative is the outcome of sophisticated voting for some binary agenda if and only if it is in this set.

Definition 5.4: Top cycle set

Let $\langle N, X, \succsim\rangle$ be an odd-strict collective choice problem and let $x \in X$ and $y \in X$ be alternatives. Then x *indirectly beats* y if for some $l \geq 1$ there are alternatives $z_1, \ldots, z_l$ such that x beats z_1, z_j beats z_{j+1} for $j = 1, \ldots, l-1$, and z_l beats y. The *top cycle set* of $\langle N, X, \succsim\rangle$ is the set of all alternatives x such that for every alternative $y \neq x$, x beats y either directly or indirectly.

For a collective choice problem with a strict Condorcet winner, the top cycle set consists solely of that alternative. For a collective choice problem without a Condorcet winner, it is nonempty and contains every Copeland winner, defined as follows.

Definition 5.5: Copeland winner

An alternative is a *Copeland winner* of a collective choice problem if it beats at least as many alternatives as does every other alternative.

Proposition 5.2: Nonemptiness of top cycle set and relation to Condorcet winner

The top cycle set of an odd-strict collective choice problem is nonempty and contains every Copeland winner. If the problem has a strict Condorcet winner then its top cycle set consists solely of that alternative.

Proof

Let $\langle N, X, \succsim\rangle$ be an odd-strict collective choice problem. By definition, its set of Copeland winners is nonempty. To show that every Copeland winner is in the top cycle set, I argue that if x is a Copeland winner and y is another alternative, then x either beats y or indirectly beats it in two steps. The reason is that if y beats x and there is no such alternative z such that x beats z and z beats y, then if x beats z, y beats z, so that y beats more alternatives than does x.

Now assume that $\langle N, X, \succsim\rangle$ has a strict Condorcet winner, x^*. Then no alternative beats x^* directly or indirectly, so no other alternative is in the top cycle set.

The top cycle set can alternatively be characterized as the smallest nonempty set of alternatives with the property that every member of the set beats every alternative not in the set.

Exercise 5.1: Characterization of top cycle set

Show that (*a*) every alternative in the top cycle set beats every alternative not in the set and (*b*) no proper subset of the top cycle set has this property. Deduce from (*a*) that any alternative that indirectly beats an alternative in the top cycle set is in the top cycle set .

For a problem without a strict Condorcet winner, the top cycle set contains at least three alternatives.

Exercise 5.2: Size of top cycle set

Show that for an odd-strict collective choice problem without a strict Condorcet winner, the top cycle set contains at least three alternatives.

For a Condorcet cycle, the top cycle set consists of all three alternatives (in Example 1.5, a beats b, b beats c, and c beats a). If we add an alternative to Example 1.5 that does not beat a, b, or c (for example, it could be ranked third by all individuals), then the top cycle set remains $\{a, b, c\}$.

The next result shows that the alternatives in the top cycle set may be ordered so that each alternative beats the next one and the last alternative beats the first one. This result justifies the word "cycle" in the name of the notion and is used in the proof of the next result.

Lemma 5.2: Top cycle set is a cycle

If the top cycle set of an odd-strict collective choice problem contains more than one alternative, for some ordering x_1, x_2, ..., x_k of its members x_1 beats x_2 beats ... beats x_k beats x_1.

Proof

Step 1 *For some subset* $\{x_1, \ldots, x_p\}$ *of the top cycle set with* $p \geq 2$, x_1 *beats* x_2 ... *beats* x_p *beats* x_1.

Proof. Let x_1 and x_2 be members of the top cycle set, where x_1 beats x_2. Given that x_2 is in the set, it indirectly beats x_1. ◁

Let $C = \{x_1, \ldots, x_p\}$ be a largest subset of the top cycle set such that x_1 beats x_2 ... beats x_p beats x_1 and suppose, contrary to the result, that C is not the whole top cycle set.

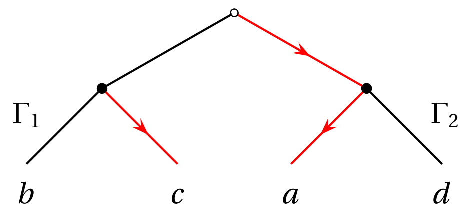


Figure 5.3 An illustration of the argument that an outcome of sophisticated voting beats every other alternative either directly or indirectly.

Step 2 *For each alternative y in the top cycle set outside C, either (i) y beats every member of C or (ii) every member of C beats y.*

Proof. If not, then for two consecutive members x_i and x_j of the sequence $(x_1, x_2, \ldots, x_p, x_1)$, x_i beats y and y beats x_j. But then the sequence can be extended by adding y between x_i and x_j, contradicting the maximality of C. ◁

Step 3 *Denote by C^+ the set of alternatives in (i) of Step 2 and by C^- the set in (ii). Both C^+ and C^- are nonempty.*

Proof. Either C^- or C^+ is nonempty. If C^- is empty, no member of C beats any member of C^+ directly or indirectly. If C^+ is empty, no member of C^- beats any member or C directly or indirectly. Thus both C^+ and C^- are nonempty. ◁

Now, take $y \in C^-$. Given that y is in the top cycle set and does not beat any alternative in C, it beats some alternative $z \in C^+$. But then C can be augmented by adding the alternatives y and z, contradicting its maximality.

To understand why every outcome of sophisticated voting in a binary agenda is in the top cycle set, look at Figure 5.3, in which the red branches indicate the outcomes of sophisticated voting in the game and its subgames. For the outcome of sophisticated voting in the game to be a, this alternative must beat d, against which it is pitted in the subgame Γ_2. It must beat also the winner in the subgame Γ_1, namely c, which beats b in the subgame. Thus a must beat c and d directly and b indirectly (a beats c beats b).

The next proposition establishes the general result and also its converse: for any alternative in the top cycle set, there is a binary agenda for which that alternative is the outcome of sophisticated voting,

Proposition 5.3: Sophisticated voting in binary agenda and top cycle set

Let $\langle N, X, \succsim \rangle$ be an odd-strict collective choice problem.

a. For any binary agenda $\langle\langle N, X, \succsim\rangle, Z, O\rangle$, every outcome of sophisticated voting is in the top cycle set of $\langle N, X, \succsim \rangle$.

b. For every alternative x in the top cycle set of $\langle N, X, \succsim \rangle$ there is a binary agenda $\langle\langle N, X, \succsim\rangle, Z, O\rangle$ for which x is the outcome of sophisticated voting.

Proof

a. Let a be an outcome of sophisticated voting. By Lemma 5.1, a is the outcome of a subgame perfect equilibrium of the one-player game G^* in which the set of terminal histories is Z, the player function assigns the player to every nonterminal history, and for any terminal histories $h \in Z$ and $h' \in Z$ the player prefers h to h' if and only if $O(h) \succ_i O(h')$ for a majority of $i \in N$.

I use induction. The top cycle set is nonempty, and every member of it is the outcome of at least one terminal history of G^* (by the assumption that every alternative is the outcome of some terminal history). Thus the outcome of sophisticated voting in a subgame of G^* of length 1 is in the top cycle set of $\langle N, X, \succsim \rangle$.

Now let ℓ be a positive integer less than the length of the longest subgame of G^* and suppose that Γ is a subgame of G^* of length ℓ for which the outcome of sophisticated voting, say x, is in the top cycle set of $\langle N, X, \succsim \rangle$. Let Γ' be the subgame of G^* of length $\ell + 1$ that contains Γ. I argue that the outcome of sophisticated voting in Γ' is in the top cycle set. The player has the option to choose the action at the start of Γ' that leads to the subgame Γ, and hence to the outcome x. Thus the subgame perfect equilibrium outcome of Γ' is either x or an alternative y that beats x. By Exercise 5.1a every alternative in the top cycle set beats every alternative outside the set, so y, like x, is in the top cycle set.

Thus by induction the outcome of sophisticated voting in G^* is in the top cycle set.

b. Denote the cycle among all the members of the top cycle set that is shown to exist by Lemma 5.2 by $(x, a_1, a_2, \ldots, a_k)$ and denote the remaining members of X (in an arbitrary order) by $(z_1, z_2, \ldots, z_l)$. The outcome of sophisticated voting for the binary agenda with the procedure shown in

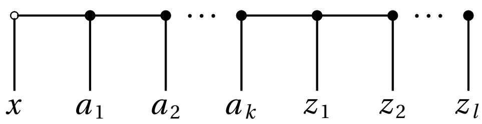


Figure 5.4 Binary agenda used in the proof of Proposition 5.3*b*.

Figure 5.4 is x: a_k beats every z_j, a_i beats a_{i+1} for $i = 1,\ldots,k-1$, and x beats a_1.

Part b of this result says that for any alternative in the top cycle set, a binary agenda can be designed so that the outcome of sophisticated voting is that alternative. If every alternative in the top cycle set were a reasonable outcome of the collective choice problem, that might not be bad. But unfortunately the top cycle may be large and include alternatives that are dominated by other alternatives, as the following example (a generalization of a Condorcet cycle) shows.

Example 5.1: Large top cycle set, containing dominated alternatives

Suppose that $N = \{1,2,3\}$, $X = \{a_1, a_2, \ldots, a_k\}$, and

$$\begin{aligned}
&a_k \succ_1 a_1 \succ_1 a_2 \succ_1 \cdots \succ_1 a_{k-2} \succ_1 a_{k-1} \\
&a_1 \succ_2 a_2 \succ_2 a_3 \succ_2 \cdots \succ_2 a_{k-1} \succ_2 a_k \\
&a_2 \succ_3 a_3 \succ_3 a_4 \succ_3 \cdots \succ_3 a_k \quad\ \succ_3 a_1.
\end{aligned}$$

Then a_i beats a_{i+1} for $i = 1,\ldots,k-1$ and a_k beats a_1, so that the top cycle set contains all k alternatives. However, all three individuals prefer a_2 to each alternative a_3, a_4, ..., a_{k-1}. (More generally, all individuals prefer a_i to a_j for $i = 2,\ldots,k-2$ and $j = i+1,\ldots,k-1$.)

Further, the outcome of sophisticated voting (a member of the top cycle set by Proposition 5.3a) may not respond positively to changes in the individuals' preferences, as you are asked to show in the following exercise.

Exercise 5.3: Outcome of sophisticated voting not positively responsive in binary agenda

Let $N = \{1,2,3\}$, $X = \{a,b,c,d\}$, $b \succ_1 a \succ_1 c \succ_1 d$, $c \succ_2 d \succ_2 b \succ_2 a$, and $d \succ_3 a \succ_3 c \succ_3 b$. Consider the binary agenda $\langle\langle N, X, \succsim\rangle, Z, O\rangle$ with the procedure shown in Figure 5.5, in which the individuals first vote on whether to decide in the order (a,b,c,d) or in the order (d,c,b,a), and then conduct three votes, first whether to select the first alternative in the order,

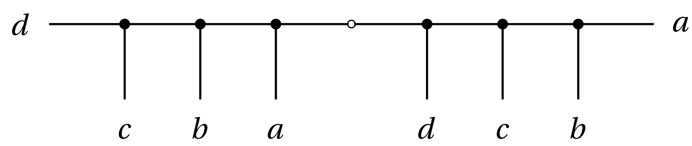


Figure 5.5 The procedure in the binary agenda in Exercise 5.3. (Note that the initial history is in the middle.)

then whether to select the second alternative, and then whether to select the third alternative. Find the outcome of sophisticated voting for this binary agenda and also for the binary agenda that differs only in that individual 1's preference between a and b is reversed. Check that the change is inconsistent with positive responsiveness.

5.2 Successive agendas

The procedural rules in many European and Latin American legislatures are approximated by a specific type of binary agenda known as a successive agenda. In a such an agenda the alternatives are considered in some order $(x_1,\ldots,x_k)$. First a vote is taken on whether to choose x_1 or to drop it from consideration; if it is dropped, then a vote is taken on whether to choose x_2 or to drop it from consideration; and so forth. The successive agenda for four alternatives is shown in Figure 5.6.

Definition 5.6: Successive agenda

Let $\langle N,X,\succsim\rangle$ be an odd-strict collective choice problem, denote by k the number of alternatives (members of X), and let $(x_1,\ldots,x_k)$ be an ordering of the alternatives. The *successive agenda* $\langle\langle N,X,\succsim\rangle,(x_1,\ldots,x_k)\rangle$ is the binary agenda $\langle\langle N,X,\succsim\rangle,Z,O\rangle$ in which the set Z consists of two terminal histories h' and h'' of length $k-1$, with $O(h')=x_{k-1}$ and $O(h'')=x_k$, and for each $l=1,\ldots,k-2$ one terminal history h of length l, with $O(h)=x_l$.

We can find the outcome of sophisticated voting in a successive agenda by using backward induction, starting from the single subgame of length 1 (a choice between x_{k-1} and x_k). Let $x_k^*=x_k$. The choice in the single subgame Γ_1 of length 1 is between x_{k-1} and x_k^*, so, using Lemma 5.1, the outcome of sophisticated voting in the subgame is x_k^* if x_k^* beats x_{k-1}, and x_{k-1} if x_{k-1} beats x_k^*. Denote this alternative x_{k-1}^* and replace Γ_1 with it. Continue to the subgame of length 1 in the resulting game, where the choice is between x_{k-1} and x_{k-1}^*, and repeat

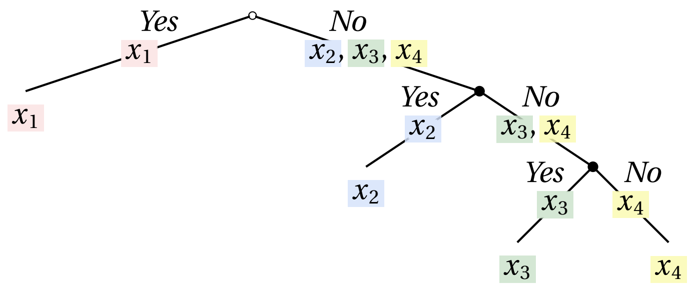


Figure 5.6 The procedure in a successive agenda for which $X = \{x_1, x_2, x_3, x_4\}$ and the ordering is (x_1, x_2, x_3, x_4).

the process. The sequence $(x_1^*, \ldots, x_k^*)$ thus created is called the sophisticated sequence for the agenda, and the conclusion of the argument is that x_1^* is the outcome of sophisticated voting in the agenda.

Definition 5.7: Sophisticated sequence for successive agenda

Let $\langle\langle N, X, \succsim\rangle, (x_1, \ldots, x_k)\rangle$ be a successive agenda. The *sophisticated sequence* for $\langle\langle N, X, \succsim\rangle, (x_1, \ldots, x_k)\rangle$ is the sequence $(x_1^*, \ldots, x_k^*)$ of alternatives defined iteratively as follows, starting with x_k^* and working backwards to x_1^*. First let $x_k^* = x_k$. Then for any j with $1 \le j \le k-1$ let

$$x_j^* = \begin{cases} x_j & \text{if } x_j \text{ beats } x_{j+1}^* \\ x_{j+1}^* & \text{otherwise.} \end{cases} \qquad (5.1)$$

Proposition 5.4: Outcome of sophisticated voting in successive agenda

The outcome of sophisticated voting in a successive agenda is the first alternative in the sophisticated sequence for the agenda.

One implication of this result is that the outcome of sophisticated voting in a successive agenda is positively responsive: if the outcome is x and then x rises in the preferences of some individual, x continues to beat every alternative it beat previously, and hence remains the outcome of sophisticated voting.

The next exercise concerns two more implications of the result. First, the last alternative in a successive agenda is the outcome of sophisticated voting only if it is a strict Condorcet winner. Second, an alternative remains the outcome of sophisticated voting in a successive agenda if it is moved earlier in the agenda.

Exercise 5.4: Effect of order of alternatives on outcome of sophisticated voting in successive agenda

Let $B = \langle\langle N, X, \succsim\rangle, (x_1, \ldots, x_k)\rangle$ be a successive agenda. Use Proposition 5.4 to show that (*a*) if x_k is the outcome of sophisticated voting in B then it is a strict Condorcet winner of $\langle N, X, \succsim\rangle$, and (*b*) for any $l \in \{2, \ldots, k\}$, if x_l is the outcome of sophisticated voting in B then it is also the outcome of sophisticated voting in the successive agenda $\langle\langle N, X, \succsim\rangle, (y_1, \ldots, y_k)\rangle$ in which $(y_1, \ldots, y_k)$ differs from $(x_1, \ldots, x_k)$ only in that x_l and x_{l-1} are interchanged.

The agenda used to prove Proposition 5.3*b* is a successive agenda, so we have the following result.

Proposition 5.5: Sophisticated voting in successive agenda and top cycle set

Let $\langle N, X, \succsim\rangle$ be an odd-strict collective choice problem. For every alternative x in the top cycle set of $\langle N, X, \succsim\rangle$ there is a successive agenda $\langle\langle N, X, \succsim\rangle, (x_1, \ldots, x_k)\rangle$ for which x is the outcome of sophisticated voting.

Although the outcome of sophisticated voting in a successive agenda, unlike the outcome in a general binary agenda, is necessarily positively responsive, Proposition 5.5 means that the outcome suffers from a drawback: for some collective choice problems, the top cycle set contains dominated alternatives (see for example Example 5.1), so the outcome of sophisticated voting in a successive agenda may be dominated.

5.3 Amendment agendas

The procedural rules in a few European legislatures, as well as in Canada and the United States, are approximated by agendas whose structures differ from those of successive agendas, and which generate sets of sophisticated outcomes smaller than the top cycle set, with better properties. Let $(x_1, \ldots, x_k)$ be an ordering of the alternatives. In an amendment agenda, first a vote is taken to eliminate x_2 from consideration (*Yes*) or to eliminate x_1 (*No*); then a vote is taken whether to eliminate x_3 (*Yes*) or to eliminate whichever of x_1 or x_2 was retained on the first round (*No*); and so on, until the remaining alternative is pitted against x_k. The procedure in an amendment agenda for four alternatives is shown in Figure 5.7. One interpretation of this agenda is that x_4 is the status quo, x_3 is a bill, x_2 is an amendment, and $x_1$is an amendment to the amendment. The first vote determines which version of the amendment is considered, the second vote de-

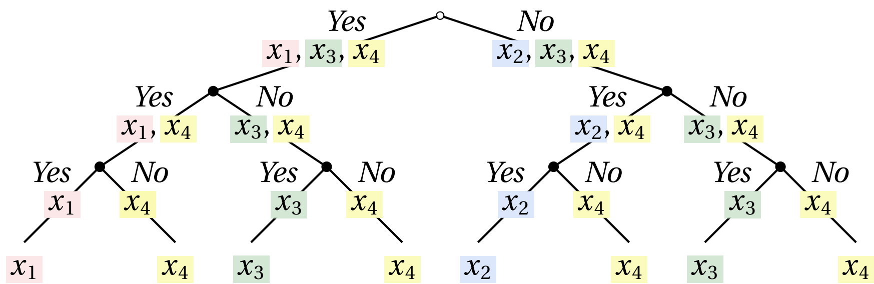


Figure 5.7 The procedure in an amendment agenda for $X = \{x_1, x_2, x_3, x_4\}$ and the ordering (x_1, x_2, x_3, x_4).

termines whether the bill or an amended version of it is considered, and the final vote determines whether the (possibly amended) bill passes.

Definition 5.8: Amendment agenda

Let $\langle N, X, \succsim \rangle$ be an odd-strict collective choice problem and let $(x_1, \ldots, x_k)$ be an ordering of the members of X. The *amendment agenda* $\langle \langle N, X, \succsim \rangle, (x_1, \ldots, x_k) \rangle$ is the binary agenda $\langle \langle N, X, \succsim \rangle, Z, O \rangle$ in which every terminal history (member of Z) has length $k-1$ and for any terminal history $(y^1, \ldots, y^{k-1})$ we have $O(y^1, \ldots, y^{k-1}) = x_{r+1}$, where r is the index of the last *No* in $(y^1, \ldots, y^{k-1})$, with $r = 0$ if $y^j = Yes$ for all $j = 1, \ldots, k-1$.

The set of alternatives that are the outcomes of sophisticated voting in an amendment agenda is the Banks set, named for its originator, Jeffrey S. Banks (1958–2000). Recall that an alternative is in the top cycle set if it beats every other alternative either directly or indirectly. To qualify for membership in the Banks set, an alternative x must satisfy a more stringent requirement: there must exist a sequence $(z_1, \ldots, z_l)$ of alternatives such that (*i*) each z_j beats every subsequent member of the sequence, $z_{j+1}, \ldots, z_l$, (*ii*) x beats every member of the sequence, and (*iii*) no alternative beats x and every member of the sequence. For example, x is in the Banks set if it beats some alternative z and no alternative beats both x and z, or if it beats some alternatives z_1 and z_2, z_1 beats z_2, and no alternative beats all three of the alternatives x, z_1, and z_2.

Definition 5.9: Banks set

Let $\langle N, X, \succsim \rangle$ be an odd-strict collective choice problem. An alternative $x \in X$ is in the *Banks set* of $\langle N, X, \succsim \rangle$ if there is a sequence $(z_1, \ldots, z_l)$ of alternatives such that

- for $j = 1, \ldots, l-1$, z_j beats each of the alternatives $z_{j+1}, \ldots, z_l$

- x beats each of the alternatives $z_1, \ldots, z_l$
- no alternative beats all of the alternatives $z_1, \ldots, z_l$ and x.

Suppose that x is in the Banks set, with the associated sequence $(z_1, \ldots, z_l)$. Then by the last of the three conditions, every alternative other than x or $z_1, \ldots, z_l$ is beaten by either x or some z_j, and hence, given that x beats every z_j, is beaten by x either directly or indirectly. Thus the Banks set is a subset of the top cycle set. For a problem that has a strict Condorcet winner, say a^*, the Banks set is $\{a^*\}$, because a^* beats every other alternative. For problems without a strict Condorcet winner, the Banks set is nonempty.

Proposition 5.6: Banks set is nonempty subset of top cycle set

The Banks set of any odd-strict collective choice problem is a nonempty subset of the top cycle set. In particular, for a problem with a strict Condorcet winner the Banks set consists solely of that alternative.

Proof

An argument that the Banks set is a subset of the top cycle set is given in the text.

To prove the nonemptiness of the Banks set I use an induction on the number of alternatives.

Let $\langle N, X, \succsim \rangle$ be an odd-strict collective choice problem and $X_l = \{x_1, \ldots, x_l\} \subset X$.

If $l = 2$ and the alternatives are labeled so that x_1 beats x_2, then the Banks set of the problem $\langle N, X, \succsim \rangle$ is $\{x_1\}$.

Now let $l \geq 2$ and assume that the Banks set of $\langle N, X_l, \succsim|_{X_l} \rangle$ is nonempty, with x_1 a member of the set and the associated sequence $(z_1, z_2, \ldots, z_p)$. Let $X_{l+1} = X \cup \{x_{l+1}\}$ for some $x_{l+1} \in X \setminus X_l$. Either x_{l+1} beats x_1 and every z_j, in which case x_{l+1} is in the Banks set of $\langle N, X_{l+1}, \succsim|_{X_{l+1}} \rangle$ with the associated sequence $(x_1, z_1, z_2, \ldots, z_p)$, or some z_j or x_1 beats x_{l+1}, in which case x_1 is in the Banks set of $\langle N, X_{l+1}, \succsim|_{X_{l+1}} \rangle$ with the associated sequence $(z_1, z_2, \ldots, z_p)$. Thus the Banks set of $\langle N, X_{l+1}, \succsim|_{X_{l+1}} \rangle$ is nonempty.

The second part of Proposition 5.2 implies that for a problem with a strict Condorcet winner the Banks set consists solely of that alternative.

Exercise 5.5: Uncovered set

The *uncovered set* consists of the alternatives x such that for every other alternative y either x beats y or x beats y indirectly in two steps. Thus the uncovered set is a subset of the top cycle set. Show that for an odd-strict collective choice problem it contains the Banks set.

For a Condorcet cycle, the Banks set, like the top cycle set, consists of all three alternatives. (In Example 1.5, a sequence of alternatives supporting a, for example, is (b): a beats b and c does not beat b.) But for many collective choice problems the Banks set is smaller than the top cycle set. An example is the problem in Example 5.1, in which the Banks set contains only the alternatives that are not dominated.

Example 5.2: Banks set for Example 5.1

For the collective choice problem in Example 5.1, the Banks set is $\{a_1, a_2, a_k\}$. A sequence of alternatives supporting a_1 is $(a_2, \ldots, a_{k-1})$ (a_k does not beat a_{k-1}), a sequence supporting a_2 is $(a_3, a_4, \ldots, a_k)$ (a_1 does not beat a_k), and a sequence supporting a_k is (a_1) (no alternative beats both a_1 and a_k). No alternative a_i with $3 \leq i \leq k-1$ is in the Banks set because a_2 beats all the alternatives that a_i beats.

In fact, no alternative in the Banks set of any odd-strict collective choice problem is dominated.

Exercise 5.6: No alternative in Banks set is dominated

Show that for any odd-strict collective choice problem, for no member a of the Banks set is there an alternative that every individual prefers to a.

To show that the set of outcomes of sophisticated voting for amendment agendas is the Banks set, I first define a procedure for finding the outcome of sophisticated voting in any amendment agenda. This procedure involves generating a sequence of alternatives defined as follows.

Definition 5.10: Sophisticated sequence for amendment agenda

The *sophisticated sequence* for the amendment agenda $\langle\langle N, X, \succsim\rangle, (x_1, \ldots, x_k)\rangle$ is the sequence $(x_1^*, \ldots, x_k^*)$ of alternatives defined iteratively as follows, starting with x_k^* and working backwards to x_1^*. First let $x_k^* = x_k$.

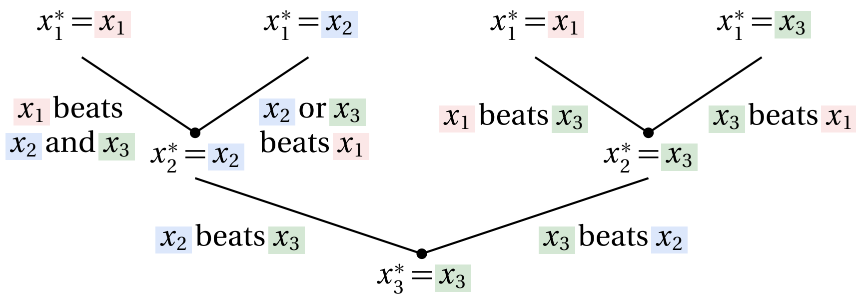


Figure 5.8 The sophisticated sequence for the amendment agenda for three alternatives and the ordering (x_1, x_2, x_3).

Then for any j with $1 \le j \le k-1$ let

$$x_j^* = \begin{cases} x_j & \text{if } x_j \text{ beats } x_l^* \text{ for } l = j+1, \dots, k \\ x_{j+1}^* & \text{otherwise.} \end{cases} \tag{5.2}$$

For $k = 2$, we have $(x_1^*, x_2^*) = (x_1, x_2)$ if x_1 beats x_2 and $(x_1^*, x_2^*) = (x_2, x_2)$ if x_2 beats x_1. An analysis of the case $k = 3$ is illustrated in Figure 5.8. The conclusion is

$$(x_1^*, x_2^*, x_3^*) = \begin{cases} (x_1, x_2, x_3) & \text{if } x_1 \text{ beats } x_2 \text{ and } x_3\text{, and } x_2 \text{ beats } x_3 \\ (x_2, x_2, x_3) & \text{if } x_2 \text{ or } x_3 \text{ beats } x_1\text{, and } x_2 \text{ beats } x_3 \\ (x_1, x_3, x_3) & \text{if } x_1 \text{ beats } x_3\text{, and } x_3 \text{ beats } x_2 \\ (x_3, x_3, x_3) & \text{if } x_3 \text{ beats } x_1 \text{ and } x_2. \end{cases}$$

The next result shows that the outcome of sophisticated voting is the first alternative in the sophisticated sequence.

Proposition 5.7: Outcome of sophisticated voting in amendment agenda

The outcome of sophisticated voting in an amendment agenda is the first alternative in the sophisticated sequence for the agenda.

Proof

Denote by k the number of alternatives and by $\Gamma(y_1, \dots, y_q)$, with $2 \le q \le k$, the subgame of the agenda in which the alternatives $y_1, \dots, y_q$ remain, in that order. This subgame is itself an amendment agenda.

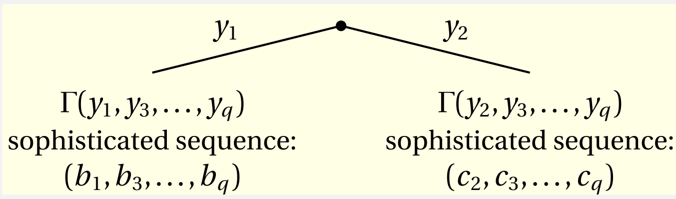
For $q = 2$, from (5.2) the sophisticated sequence is (x_1^*, x_2^*) where $x_2^* = y_2$

and

$$x_1^* = \begin{cases} y_1 & \text{if } y_1 \text{ beats } y_2 \\ y_2 & \text{if } y_2 \text{ beats } y_1. \end{cases}$$

By Lemma 5.1, x_1^* is the outcome of sophisticated voting in $\Gamma(y_1, y_2)$.

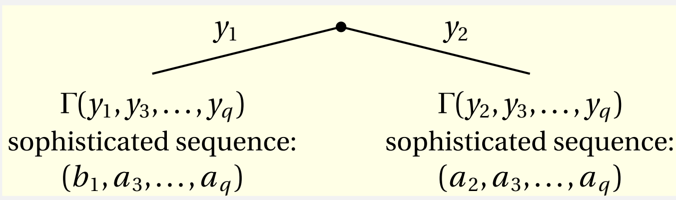
Now suppose that the outcome of sophisticated voting in every subgame in which at most $q-1$ alternatives remain, with $3 \le q \le k$, is the first alternative in the sophisticated sequence for the subgame. I argue that the same is true for the subgame $\Gamma(y_1, \ldots, y_q)$. This subgame starts with a vote on whether to keep y_1 and eliminate y_2, moving to the subgame $\Gamma(y_1, y_3, \ldots, y_q)$, or to keep y_2 and eliminate y_1, moving to the subgame $\Gamma(y_2, y_3, \ldots, y_q)$. Denoting by $(b_1, b_3, \ldots, b_q)$ and $(c_2, c_3, \ldots, c_q)$ the sophisticated sequences for the subgames $\Gamma(y_1, y_3, \ldots, y_q)$ and $\Gamma(y_2, y_3, \ldots, y_q)$, we can thus represent $\Gamma(y_1, \ldots, y_q)$ as follows.



Denote the sophisticated sequence for $\Gamma(y_1, \ldots, y_q)$ by $(a_1, \ldots, a_q)$. Given that $(y_1, y_3, \ldots, y_q)$ and $(y_2, y_3, \ldots, y_q)$ differ only in their first components, and $(y_1, \ldots, y_q)$ shares $(y_3, \ldots, y_q)$ with the first sequence and $(y_2, \ldots, y_q)$ with the second, from (5.2) we have

$$a_j = b_j = c_j \text{ for } j = 3, \ldots, q \quad \text{and} \quad a_2 = c_2.$$

Thus we can represent $\Gamma(y_1, \ldots, y_q)$ as



By assumption, the outcome of sophisticated voting in $\Gamma(y_1, y_3, \ldots, y_q)$ is b_1 and the outcome of sophisticated voting in $\Gamma(y_2, y_3, \ldots, y_q)$ is a_2, so the outcome of sophisticated voting in $\Gamma(y_1, \ldots, y_q)$ is

$$z = \begin{cases} b_1 & \text{if } b_1 \text{ beats } a_2 \text{ or } b_1 = a_2 \\ a_2 & \text{if } a_2 \text{ beats } b_1. \end{cases} \tag{5.3}$$

We need to show that z is equal to the value of a_1 given by (5.2):

$$a_1 = \begin{cases} y_1 & \text{if } y_1 \text{ beats } a_2, a_3, \ldots, a_q \\ a_2 & \text{if some member of } \{a_2, a_3, \ldots, a_q\} \text{ beats } y_1. \end{cases} \tag{5.4}$$

- Suppose that y_1 beats $a_2, \ldots, a_q$, so that $a_1 = y_1$ by (5.4). The fact that b_1 is the first member of the sophisticated sequence for $\Gamma(y_1, y_3, \ldots, y_q)$ and y_1 beats $a_3, \ldots, a_q$ implies, using (5.2), that $b_1 = y_1$. Since y_1 beats a_2, the value of z given by (5.3), namely b_1, is the value of a_1 given by (5.4).

- Suppose that a_j beats y_1 for some $j = 3, \ldots, q$, so that $a_1 = a_2$ by (5.4). The fact that b_1 is the first member of the sophisticated sequence for $\Gamma(y_1, y_3, \ldots, y_q)$ implies, using (5.2), that $b_1 = a_3$.

  - Suppose that y_2 beats $a_3, \ldots, a_q$. Then the fact that a_2 is the first member of the sophisticated sequence for $\Gamma(y_2, y_3, \ldots, y_q)$ implies, using (5.2), that $a_2 = y_2$. Given that y_2 beats a_3, $y_2 = a_2$, and $a_3 = b_1$, a_2 beats b_1. Thus the value of z given by (5.3), namely a_2, is the value of a_1 given by (5.4).

  - Suppose that a_j beats y_2 for some $j = 3, \ldots, q$. Then the fact that a_2 is the first member of the sophisticated sequence for $\Gamma(y_2, y_3, \ldots, y_q)$ implies, using (5.2), that $a_2 = a_3$. Given $b_1 = a_3$, we have $b_1 = a_2$, so thus the value of z given by (5.3) is the value of a_1 given by (5.4).

We can now show that the outcome of sophisticated voting in an amendment agenda is in the Banks set, and for any alternative in the Banks set there is an amendment agenda for which the alternative is the outcome of sophisticated voting.

Proposition 5.8: Sophisticated voting in amendment agenda and Banks set

Let $\langle N, X, \succsim \rangle$ be an odd-strict collective choice problem.

a. For any amendment agenda $\langle \langle N, X, \succsim \rangle, (x_1, \ldots, x_k) \rangle$, the outcome of sophisticated voting is in the Banks set of $\langle N, X, \succsim \rangle$.

b. For every alternative a in the Banks set of $\langle N, X, \succsim \rangle$ there is an amendment agenda $\langle \langle N, X, \succsim \rangle, (x_1, \ldots, x_k) \rangle$ for which a is the outcome of sophisticated voting.

Proof

a. By Proposition 5.7, the outcome of sophisticated voting is the first alternative in the sophisticated sequence for the agenda. By definition, this alternative beats all other alternatives in the sequence, every alternative in the sequence beats every later alternative in the sequence, and every alternative not in the sequence is beaten by some alternative in the sequence. Thus the first alternative in the sequence is in the Banks set.

b. Let a be in the Banks set and let $(y_1,\ldots,y_l)$ be a sequence of alternatives that supports it. Let $z_1, \ldots, z_p$ be the remaining alternatives (other than a and $y_1, \ldots, y_l$); the order of these alternatives does not matter. I claim that a is the outcome of sophisticated voting for the amendment agenda $B = \langle\langle N, X, \succsim\rangle, (z_1,\ldots,z_p,a,y_1,\ldots,y_l)\rangle$. By the definition of $(y_1,\ldots,y_l)$, y_j beats $y_{j+1},\ldots,y_l$ for $j = 1,\ldots,l-1$ and a beats every y_j. Also, every z_j is beaten by either a or some y_j. Thus the first alternative in the sophisticated sequence for B is a, so that by Proposition 5.7, a is the outcome of sophisticated voting in B.

Unlike the top cycle set, the Banks set has the property that no member of it is dominated. In addition, the outcome of sophisticated voting in an amendment agenda, which is a member of the Banks set by Proposition 5.8, is positively responsive.

Exercise 5.7: Outcome of sophisticated voting in amendment agenda is positively responsive

Use Proposition 5.8 to show that the outcome of sophisticated voting in an amendment agenda is positively responsive.

Although the outcome of sophisticated voting in an amendment agenda has these desirable properties, such an agenda does not treat the alternatives equally. The last alternative on the agenda is the outcome of sophisticated voting only if it is a strict Condorcet winner, but an earlier alternative may be the outcome of sophisticated voting even if it is not a strict Condorcet winner, and an alternative is never disadvantaged by being moved to an earlier position in the agenda.

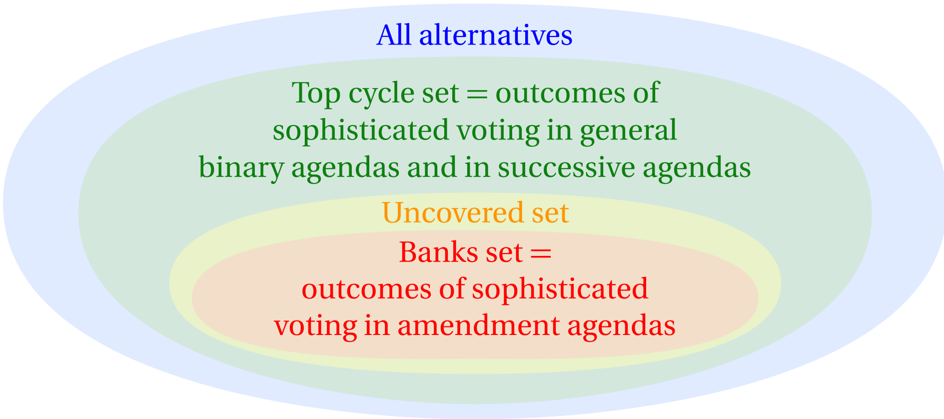


Figure 5.9 An illustration of the relations among the top cycle set, the Banks set, the uncovered set, and the outcomes of sophisticated voting in general binary agendas, successive agendas, and amendment agendas. The set of Copeland winners, like the Banks set, is a subset of the uncovered set; Laffond and Laslier (1991) show that it may be disjoint from the Banks set.

Exercise 5.8: Effect of order on outcome of sophisticated voting in amendment agenda

Show that the outcome of sophisticated voting in an amendment agenda satisfies the properties of the outcome of sophisticated voting in a successive agenda given in Exercise 5.4.

The relations among the top cycle set, Banks set, and outcomes of sophisticated voting in general binary agendas, successive agendas, and amendment agendas, as established in Propositions 5.3, 5.5, 5.6, and 5.8, are illustrated in Figure 5.9.

5.4 Single-peaked preferences and convex agendas

One feature of the model is that there is only one possible preference relation for each individual, so that every individual knows every other individual's preferences as well as her own. In this section I consider an environment in which each individual is uncertain of the other individuals' preference relations, and show that in a certain type of agenda the individuals' optimal choices are robust with respect to this imperfect information. Specifically, I show that if all of the preference relations that individuals might have are single-peaked with respect to the same linear order, then for a certain type of agenda—but not generally—an individual's voting sincerely is optimal when every other individual votes sincerely, *regardless* of the other individuals' (single-peaked) preferences.

To state the result precisely, I need to define a few concepts. Every vote in a binary agenda is a choice between two sets of alternatives, and notation for these sets is useful. For any binary agenda $\langle\langle N, X, \succsim\rangle, Z, O\rangle$, any nonterminal history h, and $p \in \{Yes, No\}$, let $A(h, p)$ be the set of alternatives that are outcomes of sequences of votes following a majority vote of p at h:

$$A(h,p) = \{O(z) : z = (h, p, h') \text{ for some sequence } h'\}.$$

For example, for a successive agenda with the procedure given in Figure 5.6, $A(\varnothing, Yes) = \{x_1\}$ and $A(\varnothing, No) = \{x_2, x_3, x_4\}$, and for an amendment agenda with the procedure given in Figure 5.7, $A(\varnothing, Yes) = \{x_1, x_3, x_4\}$ and $A(\varnothing, No) = \{x_2, x_3, x_4\}$.

For convenience, I restrict attention to binary agendas for which after every nonterminal history the set of outcomes possible following a majority vote of *Yes* is disjoint from the set of outcomes possible following a majority vote of *No*. Such agendas are called partitional. An agenda for which the procedure is given in Figure 5.6 is partitional, but one for which the procedure is given in Figure 5.7 is not—for example, x_3 and x_4 are members of both $A(\varnothing, Yes)$ and $A(\varnothing, No)$.

Definition 5.11: Partitional binary agenda

A binary agenda *partitional* if for every nonterminal history h, $A(h, Yes) \cap A(h, No) = \varnothing$.

The new concept central to the result in this section is that of a convex agenda. A binary agenda for a collective choice problem for which the preference profile is single-peaked is convex if for each nonterminal history h and each $p \in \{Yes, No\}$, whenever the set $A(h, p)$ of possible outcomes contains some alternatives x_r and x_s it contains also all alternatives between x_r and x_s according to the ordering of the alternatives. That is, at each stage, the alternatives that are possible outcomes of any given majority decision are adjacent. For example, in any vote, one option may be low funding for a project and the other may be either medium or high funding, but the individuals are never faced with a vote in which one option is medium funding and the other is either low or high funding. If the ordering with respect to which the preference profile is single-peaked is $x_1 \lhd x_2 \lhd \cdots \lhd x_4$, the successive agenda in Figure 5.6 is convex, for example, but the amendment agenda in Figure 5.7 is not (e.g. $A(\varnothing, Yes)$ contains x_1 and x_3 but not x_2).

Definition 5.12: Convex binary agenda

Let $\langle N, X, \succsim \rangle$ be an odd-strict collective choice problem for which the preference profile $\succsim$ is single-peaked with respect to a linear order $\trianglerighteq$. A binary agenda $\langle \langle N, X, \succsim \rangle, Z, O \rangle$ is *convex* if for every nonterminal history h, each $p \in \{Yes, No\}$, and any alternatives x_r and x_s, if $A(h, p)$ contains x_r and x_s then it contains every alternative x_t with $x_r \triangleleft x_t \triangleleft x_s$.

The last concept needed to state the result is that of sincere voting. An individual is said to vote sincerely at a history h in a binary agenda if she votes for the option p for which $A(h, p)$ contains the alternative she likes best in $A(h, Yes) \cup A(h, No)$. That is, we classify a vote as sincere if the best outcome to which it may lead is better than the best outcome to which the opposite vote may lead. For example, if a majority vote of *Yes* leads to either x_1 or x_3 whereas a majority vote of *No* leads to x_2, the sincere vote of an individual who prefers x_1 to x_2 to x_3 is *Yes*: the best outcome possible if the majority vote is *Yes* is x_1, which the individual prefers to the best outcome possible if the majority vote is *No*, namely x_3.

Definition 5.13: Sincere voting

An individual's strategy in a binary agenda for an odd-strict collective choice problem is *sincere* if for every nonterminal history h it assigns the option $p \in \{Yes, No\}$ for which $A(h, p)$ contains the alternative the individual likes best in $A(h, Yes) \cup A(h, No)$.

The next result says that for a binary agenda with single-peaked preferences that is partitional and convex, if the strategies of all individuals but one are sincere then a sincere strategy is optimal for the remaining individual, regardless of the other individuals' preferences. Further, if every individual votes sincerely, then the outcome is the strict Condorcet winner. Thus even in an environment in which each individual knows only her own preferences, not the preferences of any other individual, every individual's voting sincerely is mutually optimal and the outcome of such voting is that the strict Condorcet winner.

Proposition 5.9: Sincere voting in convex partitional binary agenda with single-peaked preferences

Consider a binary agenda with single-peaked preferences that is partitional and convex.

a. Let i be an individual. If the strategy of every individual other than i is sincere then i optimally votes sincerely in every ballot.

b. If every individual votes sincerely in every ballot, the outcome is the strict Condorcet winner.

Proof

a. Denote by k the number of alternatives and by $\trianglerighteq$ the linear order with respect to which the preferences are single-peaked. Label the alternatives so that $x_1 \triangleleft x_2 \triangleleft \cdots \triangleleft x_k$. Assume, contrary to the result, that i's optimal vote at some history h is not sincere. Specifically, suppose that i's sincere vote is *Yes* but she prefers the outcome of voting *No* to that of voting *Yes*, given the other individuals' strategies. Given that the agenda is partitional and convex, either $A(h, Yes) = \{x_q, \ldots, x_r\}$ and $A(h, No) = \{x_{r+1}, \ldots, x_s\}$ for some alternatives x_q, x_r, and x_s, or the roles of *Yes* and *No* are reversed. Suppose, without loss of generality, the former. (Refer to Figure 5.10.) The fact that i's sincere vote is *Yes* means that she prefers the best alternative in $\{x_q, \ldots, x_r\}$ to the best alternative in $\{x_{r+1}, \ldots, x_s\}$ and hence prefers x_r to every alternative in $\{x_{r+1}, \ldots, x_s\}$.

I argue that if, at h and later, i always votes for the option that contains x_r, the outcome is x_r, contradicting the supposition that i's voting *Yes* at h is not optimal for her.

Given our assumption that i's voting *Yes* at h is not optimal for her, her switching her vote from *Yes* to *No* must change the outcome. Thus the votes at h of the other individuals must be split equally between *Yes* and *No*. Given that these individuals are by assumption voting sincerely, the favorite alternatives of half of them are thus at most x_r.

Now suppose that i votes *Yes* at h, so that *Yes* wins. At the next vote, the options are $\{x_q, \ldots, x_t\}$ and $\{x_{t+1}, \ldots, x_s\}$ for some $t \in \{q, \ldots, s-1\}$. Given that the favorite alternatives of half of the other individuals are at most x_r, at most half of them are at most x_t, so that given that these individuals are voting sincerely, at least half of them vote *No* at (h, Yes). Thus if i votes for *No*, this option wins. Similarly, if i votes for the option that contains x_r in every subsequent ballot, that option wins, so that the ultimate outcome is x_r. But she prefers prefers x_r to every alternative in $\{x_{r+1}, \ldots, x_s\}$, contradicting the supposition that her voting *Yes* at h is not optimal.

b. At the initial history, the convexity of the agenda means that a majority of individuals vote sincerely for the option that contains the strict Condorcet winner. Thus this alternative is a member of one of the options in the next ballot, when the same argument implies that a majority votes for the option containing the strict Condorcet winner. Repeating this process

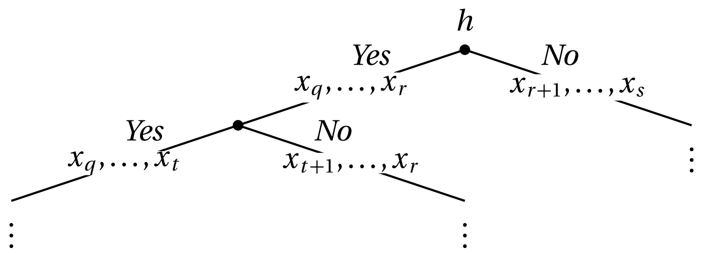


Figure 5.10 The part of an agenda following the history h used in the proof of Proposition 5.9.

leads to the conclusion that the outcome is the strict Condorcet winner.

Kleiner and Moldovanu (2017) show that this result does not depend on the assumption that the agenda is partitional. It does however depend on the assumption that the agenda is convex. Consider an agenda with the structure given in Figure 5.11 for a collective choice problem with preferences that are single-peaked with respect to the ordering $x_1 \lhd x_2 \lhd x_3$. Suppose there are three individuals, with individual 1's preferences given by $x_1 \succ_1 x_2 \succ_1 x_3$, individual 2's by $x_2 \succ_2 x_3 \succ_2 x_1$, and individual 3's by $x_3 \succ_3 x_2 \succ_3 x_1$. Suppose that individuals 2 and 3 vote sincerely. Then at the initial history individual 2 votes *No* and individual 3 votes *Yes*, and after the history *Yes* they both vote *No*. Thus if individual 1 votes *Yes* in the first ballot, the outcome is x_3 regardless of her vote in the second ballot, whereas if she votes *No*, the outcome is x_2. She prefers x_2 to x_3, so she optimally votes *No*, although her sincere vote is *Yes*.

Notes

The study of binary agendas was initiated by Black (1948a,b, 1958) and Farquharson (1969) (which was completed in 1958). In particular, Farquharson was the first person to study strategic behavior in agendas using tools from game theory; the names successive agenda and amendment agenda are his (Farquharson 1969, 61). My statements at the start of Sections 5.2 and 5.3 about the correspondence between agenda types and the rules of parliamentary procedure in various countries are based on evidence assembled by Rasch (2000, Table 1). This evidence has led some people to refer to successive agendas as Euro-Latin and amendment agendas as Anglo-American. It has also been questioned: see Schwartz (2008, 368) and Horan (2021, 236–237).

The top cycle set was first used to analyze voting by Ward (1961) (who calls it the "majority set"). Lemma 5.2 is due to Camion (1959). Proposition 5.3a is due to McKelvey and Niemi (1978, Corollary 2). Proposition 5.5, and thus

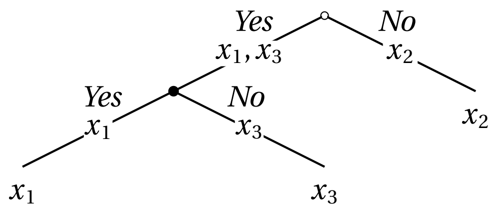


Figure 5.11 The structure of a nonconvex agenda for which the individuals' preferences are single-peaked with respect to the ordering $x_1 \lhd x_2 \lhd x_3$. An individual with preferences $x_1 \succ_1 x_2 \succ_1 x_3$ does not optimally vote sincerely when there are two other individuals, one with preferences $x_2 \succ_2 x_3 \succ_2 x_1$ and one with preferences $x_3 \succ_3 x_2 \succ_3 x_1$.

Proposition 5.3b, is due to Miller (1977, Proposition 6).

The Banks set, Proposition 5.6, and Proposition 5.8 are due to Banks (1985). Proposition 5.7 is due to Shepsle and Weingast (1984, Theorem 1); see also Moulin (1986, Theorem 3).

Section 5.4 is based on Kleiner and Moldovanu (2017); Proposition 5.9 is due to them.

The uncovered set (Exercise 5.5) was independently suggested by Miller (1977) and Fishburn (1977) (C_9, p. 473). Duggan (2013) analyzes several closely related notions, some of which were suggested before the work of Miller and Fishburn.

For the rationale for naming the concept of a Copeland winner (Definition 5.5), see the discussion of Example 1.7 in the Notes for Chapter 1. Example 5.1 is taken from Moulin (1986, 274) (see also Fishburn 1977, 478). Exercise 5.3 is taken from Moulin (1986, 284).

Solutions to exercises

Exercise 5.1

a. Denote the top cycle set by T. If $a \in T$ and b beats a, then b indirectly beats all alternatives that a indirectly beats, and hence $b \in T$. So if b is not in T then a beats it.

b. Now let T' be a proper subset of T with the property that every alternative in T' beats every alternative outside T'. Let $a \in T'$ and $b \in T \setminus T'$. Given that $b \in T$, b indirectly beats a; say b beats u_1 beats u_2 … beats u_{k-1} beats u_k beats a. Then $u_k \in T'$ because $a \in T'$ and hence beats every alternative outside T'. By the same argument, $u_{k-1} \in T'$, and hence $u_j \in T'$ for $j = 1, \ldots, k$. But then $b \in T'$ also, contrary to the assumption that $b \in T \setminus T'$. Hence no proper subset of T has the property that every alternative in it beats every alternative outside it.

Now let $a \in T$ and suppose that b beats a indirectly. Then for some alterna-

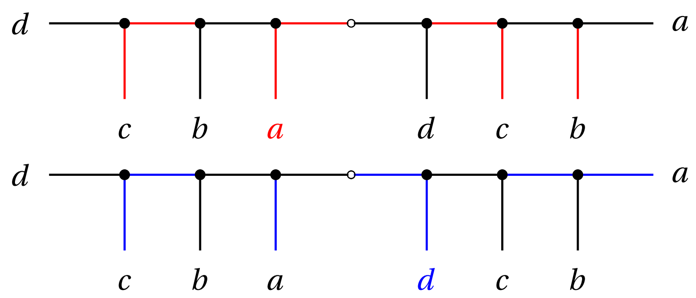


Figure 5.12 The agenda in Exercise 5.3. (Note that the initial history is in the middle.) The outcomes of the votes for the original preference profile are shown at the top, in red. Those for the modified preference profile are shown at the bottom, in blue.

tives $z_1, \ldots, z_p$, b beats z_1 beats … beats z_p beats a. If $z_p \notin T$ then by part (*a*), a beats it, so in fact $z_p \in T$. Similarly, $z_{p-1} \in T$ and hence … $z_1 \in T$ and hence $b \in T$.

Exercise 5.2

For a problem without a strict Condorcet winner, the top cycle set contains at least two alternatives. Suppose that x and y are both in the top cycle set. They cannot both beat each other (directly), so one must beat the other indirectly. Suppose that y beats z_1 beats … beats z_l beats x. Then every z_j beats (indirectly) every alternative that x beats, so that they are all in the top cycle set. Thus the top cycle set contains at least one alternative in addition to x and y.

Exercise 5.3

The agenda is shown in Figure 5.12. For the original preference profile, the outcome of sophisticated voting is a. After a rises in individual 1's preferences, positive responsiveness requires that the outcome remains a, but it changes to d (which is worse for individual 1).

Exercise 5.4

(*a*) By Proposition 5.4, if x_k is the outcome of sophisticated voting then every member of the sophisticated sequence is x_k, which means that x_k beats every other alternative, and hence is a strict Condorcet winner.

(*b*) The fact that x_l is the outcome of sophisticated voting in B means that the sophisticated sequence for B takes the form $(x_l, x_l, \ldots, x_l, x^*_{l+1}, \ldots, x^*_k)$. In particular, x_l beats x_{l-1}. Now let $(y^*_1, \ldots, y^*_k)$ be the sophisticated sequence for the agenda $B' = \langle\langle N, X, \succsim\rangle, (y_1, \ldots, y_k)\rangle$. If x_{l-1} beats x^*_{l+1} then $y^*_l = x_{l-1}$, and given that x_l beats x_{l-1}, $y^*_{l-1} = x_l$, so that $y^*_j = x_l$ for all $j = 1, \ldots, l-1$. If x^*_{l+1} beats x_{l-1} then $y^*_l = x^*_{l+1}$ and hence again $y^*_j = x_l$ for all $j = 1, \ldots, l-1$. Thus in both cases the outcome of sophisticated voting in B' is x_l.

Exercise 5.5

Suppose that x is not in the uncovered set. Then some other alternative y beats both x and every alternative that x beats. Thus no sequence z_1, z_2, ..., z_l satisfies the second and third points in the definition of the Banks set, so that x is not in the Banks set.

Exercise 5.6

If a is in the Banks set then there are alternatives x_1, ..., x_l such that (*i*) a beats each of these alternatives and (*ii*) no alternative beats a and all of these alternatives. If for some alternative b, every individual prefers b to a, then b beats a and also beats any alternative that a beats, violating (*ii*).

Exercise 5.7

Let a be the outcome of sophisticated voting for the amendment agenda $B = \langle\langle N, X, \succsim\rangle, (x_1, \ldots, x_k)\rangle$. By Proposition 5.8, a is in the Banks set, so that there exists a sequence $(x_1, x_2, \ldots, x_l)$ of alternatives satisfying the conditions in Definition 5.9. Let i be an individual, let $a \in X$, and let $\succsim'$ be a preference profile that differs from $\succsim$ only in that a is ranked higher by $\succsim'_i$ than it is by $\succsim_i$. Then for the problem $\langle N, X, \succsim'\rangle$, the sequence $(x_1, x_2, \ldots, x_l)$ satisfies the conditions in Definition 5.9, since a still beats every x_j, every x_j beats the same set of x_i's as it did for $\succsim$, and no alternative except possibly a beats any alternative that it did not beat for $\succsim$.

Exercise 5.8

a. By Proposition 5.7 and (5.2), if x_k is the outcome of sophisticated voting, then it beats every other alternative, and hence is the strict Condorcet winner.

b. Denote the sophisticated sequence for B by $(x_1^*, \ldots, x_k^*)$. Given that x_l is the outcome of sophisticated voting in B, by Proposition 5.7 we have $x_1^* = \cdots = x_l^* = x_l$.

	x_1	$\cdots$	x_{l-2}	x_{l-1}	x_l	x_{l+1}	$\cdots$	x_k
sophisticated sequence:	x_l	$\cdots$	x_l	x_l	x_l	x_{l+1}^*	$\cdots$	x_k^*

By (5.2), x_l beats every alternative x_{l+1}^*, ..., x_k^* and every alternative x_1, ..., x_{l-1} is beaten by some member of $\{x_l, x_{l+1}^* \ldots, x_k^*\}$.

Now use (5.2) to calculate the sophisticated sequence $(y_1^*, \ldots, y_k^*)$ for $\langle\langle N, X, \succsim\rangle$, $(y_1, \ldots, y_k)\rangle$. Given that $y_j = x_j$ for $j = l+1, \ldots, k$, we have $y_j^* = x_j^*$ for $j = l+1, \ldots, k$.

If x_{l-1} beats x_{l+1}^*, ..., x_k^*, so that $y_l^* = x_{l-1}$, then given that x_{l-1} is beaten by some member of $\{x_l, x_{l+1}^* \ldots, x_k^*\}$, x_l beats x_{l-1}, and consequently $y_j^* = x_l$ for $j = 1, \ldots, l-1$.

	x_1	$\cdots$	x_{l-2}	x_l	x_{l-1}	x_{l+1}	$\cdots$	x_k
sophisticated sequence	x_l	$\cdots$	x_l	x_l	x_{l-1}	x^*_{l+1}	$\cdots$	x^*_k

If x_{l-1} is beaten by one of the alternatives $x^*_{l+1}, \ldots, x^*_k$, then $y^*_l = x^*_{l+1}$, and again $y^*_j = x_l$ for $j = 1, \ldots, l-1$.

	x_1	$\cdots$	x_{l-2}	x_l	x_{l-1}	x_{l+1}	$\cdots$	x_k
sophisticated sequence	x_l	$\cdots$	x_l	x_l	x^*_{l+1}	x^*_{l+1}	$\cdots$	x^*_k

In both cases $y^*_1 = x_l$, so that x_l is the outcome of sophisticated voting in $\langle\langle N, X, \succsim\rangle, (y_1, \ldots, y_k)\rangle$.

6 Ethical voting and expressive voting

6.1 Ethical voting 188
6.2 Expressive voting 194

Each individual in the models of voting in the previous chapters is motivated by the possibility that her vote will increase the probability that the outcome is one that she likes. This chapter explores two other motivations. The first model retains the assumption that individuals vote because they believe that doing so may affect the outcome of the election, but assumes that rather than being self-interested, each individual chooses an action that, if chosen by everyone like her, would lead to the outcome that she believes is the best for society. This model is one possible formulation of the idea that individuals' voting decisions are driven by ethical concerns. The second model retains the assumption that individuals are self-interested but assumes that they vote because they derive satisfaction from expressing their opinions, regardless of whether doing so has any effect on the outcome of the election.

One reason why these models have interest is that in each case the motivation to vote is not related to the size of the electorate, and in particular turnout does not necessarily decline as the size of the electorate increases, as it does in the models in Sections 3.2 and 3.3.

Synopsis

In the model of ethical voting (Section 6.1), individuals differ both in the alternative they believe to be the best for society and in their voting cost. Each individual chooses a rule for casting a vote as a function of voting cost that, if adopted by all individuals who agree with her regarding the best alternative for society, would produce the best outcome for society given the rules used by individuals with other beliefs about the best outcome for society, taking into account everyone's voting cost. There is a continuum of individuals and two alternatives. Under these assumptions, no individual's vote affects the outcome of the election, so no individual motivated to vote by the possibility of her vote changing the outcome of the election has no incentive to do so. In any equilibrium, a positive mass of individuals vote. Example 6.1 shows one possible way in which the equilibrium may vary with the parameters of the model.

The idea that people vote to express their beliefs and to affirm their political identity has been subjected to little formal modeling. Section 6.2 briefly discusses the idea and presents a simple model of one facet of it: the benefit an individual derives from expressing her beliefs by voting for one candidate rather than another depends on the proportion of the population who support that candidate.

6.1 Ethical voting

A society selects alternative a or alternative b by casting votes, with the majority determining the winner. Each individual can vote for one of the alternatives or abstain. Some individuals believe that a is the best alternative for society, while others believe that b is the best alternative. For brevity, call the former a-individuals and the latter b-individuals. Any individual who votes incurs a cost. Apart from the differences in these costs, for each alternative z all z-individuals are identical.

The models in the previous chapters assume that each individual is self-interested: she chooses whether to vote by considering the probability that doing so affects the outcome in her favor, given the other individuals' strategies. The model here assumes that each individual chooses a voting rule that, *if used by all individuals like her*, generates what she believes is the best outcome for society, given the other individuals' behavior. This assumption is one possible formalization of the idea that individuals are motivated by ethical concerns. The voting rules chosen in the model may result in high turnout even in an arbitrarily large population.

For a model in which the individuals are self-interested, the discussion in Section 3.1.2 concludes that a deterministic model is unlikely to have an equilibrium. Similar considerations suggest the same conclusion when individuals act ethically. So we incorporate uncertainty in the model: as in a two-alternative voting game with uncertain voting costs, we assume that while each individual knows her own voting cost, she knows only the distribution from which the other individuals' costs are drawn. The amount of uncertainty that this assumption generates, however, converges to zero as the number of individuals increases. We want to model a large population, and a convenient way to do so is to assume that there are infinitely many individuals, with the set of individuals equal to the unit interval $[0,1]$. In this case, the assumption on costs leads to a deterministic outcome, and the problem of the nonexistence of an equilibrium reappears. To generate uncertainty that persists in a large population, we assume also

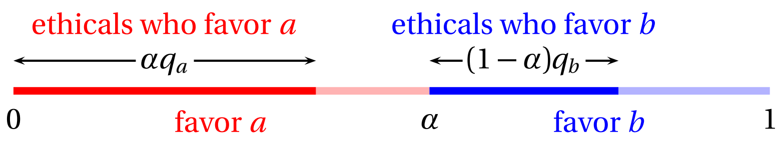


Figure 6.1 The structure of the set of individuals in a two-alternative plurality rule voting problem with ethical individuals. One realization of α (the fraction of individuals who favor a), q_a (the fraction of ethical individuals among a-individuals), and q_b (the fraction of ethical individuals among b-individuals), which are random variables, is shown.

that the fraction α of the population consisting of a-individuals is uncertain, and that within the sets of a-individuals and b-individuals, not all individuals behave ethically; the fractions q_a and q_b that do so are randomly-determined, with the remainder self-interested. (Refer to Figure 6.1.)

Given the continuum of individuals, no individual's vote affects the outcome of the election, and I assume that as a consequence no self-interested individual votes. A voting rule assigns an action (vote for a, vote for b, abstain) to each possible value of an ethical individual's voting cost. For each alternative z, each ethical z-individual chooses the voting rule that, if used by all ethical z-individuals, generates what she believes is the highest social welfare, given the voting rules chosen by the individuals who believe the other alternative is best, taking into account *all* individuals' voting costs.

In every equilibrium of a two-alternative voting game with uncertain voting costs studied in Section 3.2, every individual i uses a threshold voting rule: there is a number $\bar{c}_i$ such that i votes for her favored alternative if her voting cost is at most $\bar{c}_i$ and abstains otherwise. The logic behind this conclusion applies also in the current context, and for simplicity here I restrict individuals to such rules. That is, for each alternative z, each ethical z-individual chooses a number c_z such that if every ethical z-individual votes for z when her voting cost is at most c_z and abstains otherwise, then given the threshold chosen by the individuals who favor the other alternative, her evaluation of social welfare, taking into account the total cost of voting, is maximized.

Suppose that the fraction α of the population consisting of a-individuals is a draw from the distribution function H on $[0,1]$ and for each alternative z the fraction q_z of z-individuals who are ethical is an independent draw from the nonatomic distribution function G on $[0,1]$ that is also independent of α. Suppose also that for some number $\bar{c} > 0$, for each alternative z and each number $c \in [0,\bar{c}]$ the fraction of z-individuals with voting cost at most c is $F(c)$, where F is a nonatomic probability distribution function on $[0,\bar{c}]$ that has a density.

Then if the voting thresholds for ethical individuals are c_a and c_b, the probability that a wins is the probability that the fraction $\alpha q_a F(c_a)$ of individuals—

those who favor a, are ethical, and have a voting cost at most c_a—is greater than the fraction $(1-\alpha)q_b F(c_b)$—those who favor b, are ethical, and have a voting cost at most c_b—plus half the probability that these two fractions are equal. If $F(c_a) = F(c_b) = 0$ then the fractions are equal, and otherwise the probability that they are equal is zero, so the probability that a wins when the thresholds are c_a and c_b is

$$\Pr(a \text{ wins} \mid c_a, c_b) = \begin{cases} \frac{1}{2} & \text{if } F(c_a) = F(c_b) = 0 \\ \Pr\left(\alpha q_a F(c_a) \geq (1-\alpha) q_b F(c_b)\right) & \text{otherwise.} \end{cases} \tag{6.1}$$

The expression for the probability that b wins is analogous.

How does an individual evaluate social welfare? Assume that for each alternative z, each ethical z-individual believes that the welfare of every individual is w_z if the outcome is z and 0 if it is the other alternative, minus the individual's voting cost if she votes. An individual may construct an index of social welfare from these individual welfares in various ways. Section 1.8 discusses social welfare orderings in general, and the leximin and utilitarian orderings in particular. Here, following Feddersen and Sandroni (2006a), I assume that individuals use the utilitarian ordering, which ranks outcomes according to the sum of the individuals' welfares. Adapted to the current model, with a continuum of individuals of measure 1 and uncertainty, this assumption means that each ethical z-individual assigns to the pair (c_a, c_b) of voting thresholds the social welfare

$$u_z(c_a, c_b) = w_z \Pr(z \text{ wins} \mid c_a, c_b) - C(c_a, c_b), \tag{6.2}$$

where the probability is given by (6.1) and $C(c_a, c_b)$, the expected cost of voting, is the fraction $E(\alpha q_a)$ of individuals who are ethical and favor a times the expected cost of voting for these individuals, plus the analogous expression for individuals who favor b:

$$C(c_a, c_b) = E(\alpha q_a) \int_0^{c_a} c \, dF(c) + E((1-\alpha) q_b) \int_0^{c_b} c \, dF(c). \tag{6.3}$$

The theory is that each a-individual chooses c_a to maximize her evaluation $u_a(c_a, c_b)$ of social welfare ((6.2) for $z = a$) given c_b, and each b-individual chooses c_b to maximize $u_b(c_a, c_b)$ given c_a. That is, (c_a, c_b) is a Nash equilibrium of a two-player game in which the payoff functions are u_a and u_b. To specify the game precisely, I first collect the elements of the model in the following definition.

Definition 6.1: Two-alternative plurality rule voting problem with ethical individuals

A *two-alternative plurality rule voting problem with ethical individuals* $\langle[0,1],\{a,b\},H,G,F,(w_a,w_b)\rangle$ consists of

- $[0,1]$ (the set of individuals)
- $\{a,b\}$ (the set of alternatives)
- H, a probability distribution function with support $[0,1]$ (the distribution of the fraction α of individuals who favor a)
- G, a nonatomic probability distribution function with support $[0,1]$ (the distribution of the fractions q_a of a-individuals and q_b of b-individuals who are ethical)
- F, a nonatomic probability distribution function on some interval $[0,\bar{c}]$, where $\bar{c} > 0$, that has a density (the distribution of the individuals' voting costs)
- w_a and w_b, positive numbers (the weights assigned to outcomes by each type of ethical individual).

The strategic game associated with a two-alternative plurality rule voting problem with ethical individuals is defined as follows.

Definition 6.2: Strategic game for two-alternative plurality rule voting problem with ethical individuals

Let $\langle[0,1],\{a,b\},H,G,F,(w_a,w_b)\rangle$ be a two-alternative plurality rule voting problem with ethical individuals. The strategic game associated with this problem has the following components.

Players
The set of players is $\{a,b\}$.

Actions
The set of actions of each player is $[0,\bar{c}]$, the support of F (the set of possible cost thresholds for voting).

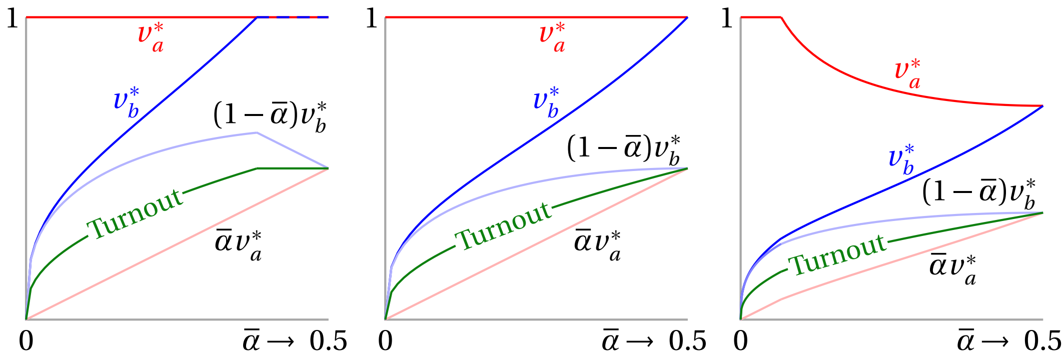
Payoffs
The players' payoff functions are u_a and u_b defined in (6.2), where $\Pr(z \text{ wins} \mid c_a, c_b)$, in which α, q_a, and q_b are independent draws from

H, G, and G respectively, is given by (6.1) and the function C is given in (6.3).

Not every such game has a Nash equilibrium. One approach to finding conditions for the existence of an equilibrium appeals to the conditions for the existence of a Nash equilibrium in a general strategic game in Proposition 16.4: each player's set of actions is a nonempty compact convex subset of a Euclidean space and each player's payoff function is continuous and quasiconcave in her action for any given action of the other player. This approach is taken by Feddersen and Sandroni (2006a), who apply it to a transformation of the game in Definition 6.2 in which the strategic variables are the fractions $F(c_a)$ and $F(c_b)$ of ethical individuals of each type who vote rather than the thresholds c_a and c_b. Given that F is one-to-one, a pair (c_a^*, c_b^*) is an equilibrium of the game in Definition 6.2 if and only if $(F(c_a^*), F(c_b^*))$ is an equilibrium of the transformed game. Each player's set of actions in the transformed game is $[0,1]$, a nonempty compact convex set. Each player's payoff function, however, is not continuous at $(0,0)$. If $F(c_a) = F(c_b) = 0$ (no one votes) then the election is a tie. But if one of these numbers, say $F(c_a)$, is positive, while the other is zero, then given the assumptions about the distributions F, G, and H, the probability that some individuals favor a, are ethical, and vote is 1, so that the probability that a wins is 1. Thus a straightforward application of Proposition 16.4 is not possible. One way to avoid the problem is to consider the existence of an equilibrium for a variant of the game in which the set of actions of each player is $[\varepsilon, 1]$ for some $\varepsilon > 0$ and then study the equilibria as ε approaches 0. The main remaining condition required by Proposition 16.4 is the quasiconcavity of each player's payoff function in her own action. Feddersen and Sandroni (2006a) show that if G, the common distribution function of q_a and q_b, is concave on its support, then this condition is satisfied, and this property of G is sufficient also for the existence of a Nash equilibrium in the game in which each individual's set of actions is $[0,1]$ rather than $[\varepsilon, 1]$.[1]

The fraction of the population that votes in an equilibrium depends on the distributions H, G, and F, and the numbers w_a and w_b. Here is an example.

[1] Feddersen and Sandroni (2006a) study a more general model in which the distributions of q_a and q_b may differ. My claim follows from their Proposition 4 combined with the observation that if the distributions of q_a and q_b are the same, their Assumption A is satisfied if (and only if) G is concave.



(a) $w/\bar{c} = 1$. (b) $w/\bar{c} = 0.5$. (c) $w/\bar{c} = 0.25$.

Figure 6.2 The fractions v_a^* and v_b^* of a- and b-individuals who vote and the expected turnout for the unique Nash equilibrium of the strategic game associated with the two-alternative plurality rule voting problem with ethical individuals in Example 6.1, as a function of $\bar{\alpha}$, the fraction of individuals favoring a, for various values of $w/\bar{c}$.

Example 6.1: Voting problem with ethical individuals

Consider a two-alternative plurality rule voting problem with ethical individuals $\langle [0,1], \{a,b\}, H, G, F, (w_a, w_b)\rangle$ in which H (the distribution of the fraction of individuals favoring a) assigns probability 1 to one value, denoted $\bar{\alpha}$, G (the distribution of the fractions q_a and q_b) is uniform on $[0,1]$, F (the distribution of voting costs) is uniform on $[0,\bar{c}]$, and $w_a = w_b = w$. The strategic game associated with this problem has a unique Nash equilibrium, which can be calculated explicitly. The fractions v_a^* and v_b^* of a- and b-individuals who vote in this equilibrium are illustrated in Figure 6.2 as a function of $\bar{\alpha}$ for $\bar{\alpha} \in [0, 0.5]$ and various values of $\bar{c}/w$, the ratio of the upper limit of the cost of voting to the weight in the payoff function on the probability of winning. In the cases shown, $v_a^* \geq v_b^*$: the fraction of a-individuals (a minority of all individuals) who vote is at least the fraction of b-individuals who vote. However, the number (measure) $\bar{\alpha}v_a^*$ of a-individuals who vote is less than the number $(1-\bar{\alpha})v_b^*$ of b-individuals who do so if $\bar{\alpha} \in (0, 0.5)$, so that b wins despite the higher turnout rate among a-individuals. As the fraction $\bar{\alpha}$ of a-individuals in the population declines to 0, the turnout rate among b-individuals, and hence the overall turnout rate, approaches 0. All of these properties are shared by the equilibrium of the strategic game associated with any two-alternative plurality rule voting problem with ethical individuals (Feddersen and Sandroni 2005, Propositions 4 and 5). Other properties of the equilibria in the example, like the fact that turnout is increasing in $w/\bar{c}$, do not hold generally.

6.2 Expressive voting

In the words of Schuessler (2000a, 88), "at least for some voters, voting is a means to express political beliefs and preferences and, in doing so, to establish or reaffirm their own political identity". Voting may convey expressive benefits even though it is not publicly observable, and is associated with the observable acts of wearing a campaign button, displaying a lawn sign, participating in campaign rallies, and expressing views to other people. Like spectators at a sporting event, individuals do not expect their actions (voting, cheering) to affect the outcome; they take these actions to express their support for a team or a candidate.

A simple model is based on the hypothesis that an individual's expressive benefit from voting for a candidate depends on the candidate's identity and the fraction of the population who support the candidate. The idea is that an individual's expressive benefit from supporting a candidate derives from being associated with the candidate's community of supporters and from being distinguished from the supporters of other candidates.

Suppose there are two candidates, 1 and 2, and denote individual i's (expressive) benefit from supporting (and voting for) candidate j by $u_i^j(q^j)$, where q^j is the fraction of the population that supports j. Individual i votes for candidate 1 if $u_i^1(q^1) > u_i^2(q^2)$ and $u_i^1(q^1)$ exceeds her voting cost c_i, with $q^2 = 1 - q^1$. In an equilibrium, q^1 is equal to the fraction of individuals for whom these conditions are satisfied:

$$q^1 = \text{fraction of individuals } i \text{ for whom } u_i^1(q^1) > u_i^2(1-q^1) \text{ and } u_i^1(q^1) > c_i.$$

As a simple example, suppose that the benefit from voting for a candidate j depends only on q^j, not on the identities of the candidate or the individual. Denote it $u(q^j)$. Then if $u(q^1) > u(1-q^1)$, all individuals vote for candidate 1, and if $u(q^1) < u(1-q^1)$ they all vote for candidate 2. Thus in an equilibrium $u(q^1) = u(1-q^1)$. In particular, one equilibrium is $q^1 = 0.5$.

If an individual's expressive benefit from supporting a candidate derives from being identified with the other individuals supporting the candidate and distinguished from those supporting the other candidate, then the expressive benefit from supporting a candidate is small or nonexistent if no one supports the candidate or everyone does so. Thus we might expect $u(q^j)$ to initially increase as q^j increases from 0 and ultimately decrease as q^j approaches 1, as in Figure 6.3. In the example shown, in which each individual's cost c_i is assumed to be less than the smallest value of $u(q^j)$, the model has two asymmetric equilibria in addition to the equilibrium $q = 0.5$.

If an individual's expressive benefit from voting for a candidate depends on the identities of the individual and the candidate, as well as the proportion of

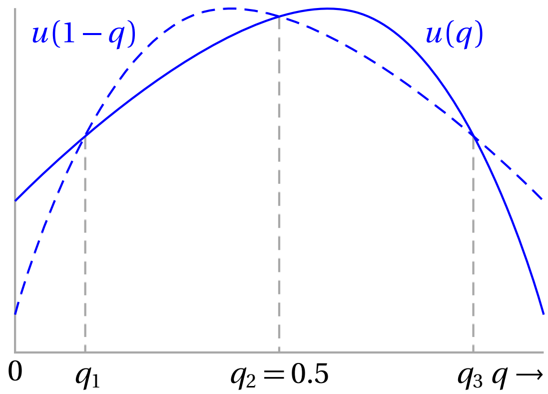


Figure 6.3 Equilibria in an example of a model of expressive voting. The value of $u(q)$ is the expressive benefit of each individual from supporting candidate 1 when the proportion of individuals who do so is q. Each individual i's cost c_i is assumed to be less than the smallest value of $u(q)$. If the proportion of individuals who support candidate 1 is q_1, q_2, or q_3, no individual has an incentive to switch her support to another candidate.

the population supporting the candidate, then 0.5 may not be an equilibrium. If some or all of the costs c_i exceed $u(q)$ for some values of q, then in an equilibrium some individuals may not vote. But there is no reason for the proportion of such individuals to increase with the size of the population, as it does in the models in Sections 3.2 and 3.3.

This model captures only one facet of the idea that people are motivated to vote, at least in part, by the desire to express their beliefs. To date, other facets of the idea have not been expressed in formal models.

Notes

Harsanyi (1977b, Section 7) first explored a model in which individuals choose a voting rule that, if adopted by everyone, would be best for society. (See also Harsanyi 1977a, 1980.) Section 6.1 is based on Feddersen and Sandroni (2005, 2006a). Example 6.1 is the subject of Feddersen and Sandroni (2006b). Coate and Conlin (2004) study a closely related model. The main differences between their model and that of Feddersen and Sandroni are that they make a specific assumption about the distribution of the fraction of individuals who favor each alternative and assume that the payoff of each side includes only the voting costs borne by that side, not the costs borne by the individuals who favor the other alternative. Thus in their model each ethical individual chooses the voting rule that, if adopted by all members of her group, is best for her group, given the rule used by the other group.

Section 6.2 is based on Schuessler (2000a,b). Hamlin and Jennings (2011,

2019) discuss the idea of expressive voting informally.

7 Voting with shared values and asymmetric information

7.1 Strategic abstention 198
7.2 Unanimity rule 220

The individuals in the models of voting in the previous chapters disagree about the desirability of the alternatives. A central question in these models is how well a voting system aggregates preferences. The individuals in the models in this chapter differ in their information about the state of society, but agree about the alternative that is best in each state. A central question is how well a voting system aggregates the individuals' information.

Synopsis

The model in Section 7.1.2 is intended to capture the idea that a poorly-informed individual may abstain because she thinks the decision is better left to well-informed individuals. In the model, there are two alternatives, a and b, and two states, α and β. Some individuals, called partisans, prefer one of the alternatives regardless of the state, and others, called independents, prefer a in state α and b in state β. Among the independents, some individuals know the state and others do not. Proposition 7.1 shows that in any equilibrium, every partisan votes for the alternative she favors, every informed independent votes for the alternative she favors given the state, and uninformed independents vote so as to cancel out, as far as possible, the partisans' votes. The behavior of the uninformed independents has the effect of putting the decision in the hands of the informed independents as much as possible, and results in the same outcome as does an equilibrium of the variant of the game in which every individual is informed of the state. That is, the equilibrium aggregates information perfectly.

The analysis in Section 7.1.3 shows that the implication of Proposition 7.1 that poorly-informed individuals may abstain does not depend on these individuals being completely uninformed and facing individuals who are perfectly informed. In the model, each individual observes a signal about the state; the quality of this signal is drawn randomly from a distribution, independently of the qualities of the other individuals' signals. Each individual observes her own signal and its quality, but not the other individuals' signals or signal qualities. Proposition 7.2 shows that if there are two individuals and the states are equally likely, the game has an equilibrium in which each individual votes if and only if the quality of her signal is at least equal to some threshold. Thus some individuals whose signals are informative abstain, ceding the decision to individuals with higher quality signals. If the individuals were to make a decision by pooling their signals, taking into account the quality of each signal, then every individual would optimally contribute her signal. But instead they are making a decision by voting, which provides no means by which to convey the quality of a signal. An individual with a signal that is only slightly informative does not vote because if she did so her signal would effectively be given the same weight as the other individual's signal, the quality of which is likely to be significantly higher than hers.

Section 7.2 studies the implications of unanimity rather than plurality rule. Consider the variant of the model in Section 7.1.2 in which the outcome is a unless all individuals vote for b. In this case, in every equilibrium every uninformed individual votes for b, because that is the only way she can hand the decision to the informed individuals: if she votes for a then the outcome is a regardless of the other individuals' votes. Similar considerations in a variant of the model in Section 7.1.3, in which the individuals are a priori identical, lead to the conclusion that when the number of individuals is large, the strategy profile in which every individual votes for the alternative that is more likely to be best according to her own signal is not an equilibrium. The reason is that when everyone else votes in this way, the remaining individual's vote affects the outcome only if all the other individuals vote for b, which happens only if all of them receive signals suggesting that b is the best outcome. But if all of them receive such a signal, the probability that b is the best outcome is high even if the signal of the remaining individual suggests that a is best, so that the remaining individual should vote for b regardless of her signal. The general point is that when deciding how to cast her vote, an individual should take into account the information implied by the fact that her vote affects the outcome.

7.1 Strategic abstention

Faced in a voting booth with lists of choices for mayor, city councillor, and school superintendent, you realize that although you believe you know the best candidate for mayor and are reasonably confident about the merits of the candidates for city councillor, you have no idea about the candidates for school superintendent. As a consequence, you decide to abstain on the ballot for that position, hoping that other, better-informed, voters endorse the best person for the job. This section presents a model that captures this idea.

7.1.1 Example

Suppose that there are two alternatives and two individuals, each of whom can vote for one of the alternatives or abstain. The society is in one of two states, α or β. The individuals agree that alternative a is best if the state is α and alternative b is best if the state is β. Each individual's payoff is 1 if the best alternative for the state is chosen, 0 if the other alternative is chosen, and $\frac{1}{2}$ if the alternatives tie.

The individuals differ in their information: individual 1 knows the state and individual 2 does not. Individual 2 believes that the state is α with probability 0.9 and β with probability 0.1. Each individual may vote for a, vote for b, or abstain. Neither individual incurs any cost when she votes.

An intuitive analysis of this situation suggests that individual 1, who knows the state, should vote for a in state α and b in state β. If individual 1 acts in this way, what should individual 2 do? If, absent individual 1, she were choosing an alternative by herself, she would choose a, because this alternative gives her an expected payoff of $0.9 \times 1 + 0.1 \times 0 = 0.9$, whereas b gives her an expected payoff of $0.9 \times 0 + 0.1 \times 1 = 0.1$. However, in the presence of individual 1, she can abstain, leaving the choice to individual 1, who votes for the best alternative in each state. In fact, abstention is her best option. If she votes for a then if the state is α, her vote makes no difference, and if the state is β, it changes the outcome from a win for b to a tie, making her worse off. By a similar argument, her voting for b also makes her worse off. In both cases, if her vote makes a difference, it leads to an outcome worse than abstention, an effect known as the swing voter's curse. So not only can she safely abstain, but she is better off doing so than voting.

We can model this situation as the following Bayesian game.

Players
The set of players is the set of individuals, $\{1,2\}$.

States
The set of states is $\{\alpha,\beta\}$.

Actions
The set of actions of each individual is $\{$*vote for a*, *vote for b*, *abstain*$\}$.

Signals
Individual 1 gets different signals in states α and β; individual 2 gets the same signal in both states.

Prior beliefs

Each individual assigns probability 0.9 to state α and probability 0.1 to state β.

Payoffs

The payoffs are given in the following tables, where ϕ stands for abstention and the actions a and b stand for voting for a and voting for b.

	a	b	ϕ
a	$1,1$	$\frac{1}{2},\frac{1}{2}$	$1,1$
b	$\frac{1}{2},\frac{1}{2}$	$0,0$	$0,0$
ϕ	$1,1$	$0,0$	$\frac{1}{2},\frac{1}{2}$

State α

	a	b	ϕ
a	$0,0$	$\frac{1}{2},\frac{1}{2}$	$0,0$
b	$\frac{1}{2},\frac{1}{2}$	$1,1$	$1,1$
ϕ	$0,0$	$1,1$	$\frac{1}{2},\frac{1}{2}$

State β

A player's strategy in a Bayesian game is a function that associates an action with each of her signals. So in this game, a strategy for player 1 specifies two actions: one associated with the signal generated by state α and one associated with the signal generated by state β. A strategy for player 2 is a single action (her signal conveys no information about the state). A Nash equilibrium of the game is a pair of strategies such that neither player has a strategy that increases her expected payoff, given the other player's strategy. The game has two Nash equilibria, illustrated in Figure 7.1.

a. Player 1 votes for a in state α and for b in state β, and player 2 abstains (highlighted in pink in Figure 7.1a).

If player 2 abstains, then player 1's voting for a in state α and for b in state β is better than any other strategy. If player 1 uses this strategy, then player 2's payoff to abstention is 1 whereas her payoff to voting for a is $0.9 \times 1 + 0.1 \times 0.5 = 0.95$ and her payoff to voting for b is $0.9 \times 0.5 + 0.1 \times 1 = 0.55$.

b. Player 1 abstains in state α and votes for b in state β, and player 2 votes for a (highlighted in blue in Figure 7.1b).

If player 2 votes for a, in state α player 1 can do no better than abstain (if she votes for a, her payoff remains 1, and if she votes for b, her payoff falls to $\frac{1}{2}$) and in state β she can do no better than vote for b. If player 1 abstains in state α and votes for b in state β then player 2's payoff to voting for a is $0.9 \times 1 + 0.1 \times 0.5 = 0.95$ whereas her payoff to abstention is $0.9 \times 0.5 + 0.1 \times 1 = 0.55$ and her payoff to voting for b is $0.9 \times 0 + 0.1 \times 1 = 0.1$.

In the second equilibrium, player 1's strategy is weakly dominated: her payoff from the strategy of voting for a in state α and b in state β is at least as high whatever strategy player 2 uses, and is higher if player 2 votes for b or abstains.

	a	b	ϕ
a	1, 1	$\frac{1}{2}, \frac{1}{2}$	1, 1
b	$\frac{1}{2}, \frac{1}{2}$	0, 0	0, 0
ϕ	1, 1	0, 0	$\frac{1}{2}, \frac{1}{2}$

State α (prob. 0.9)

	a	b	ϕ
a	0, 0	$\frac{1}{2}, \frac{1}{2}$	0, 0
b	$\frac{1}{2}, \frac{1}{2}$	1, 1	1, 1
ϕ	0, 0	1, 1	$\frac{1}{2}, \frac{1}{2}$

State β (prob. 0.1)

(a) Nash equilibrium in which no strategy is weakly dominated.

	a	b	ϕ
a	1, 1	$\frac{1}{2}, \frac{1}{2}$	1, 1
b	$\frac{1}{2}, \frac{1}{2}$	0, 0	0, 0
ϕ	1, 1	0, 0	$\frac{1}{2}, \frac{1}{2}$

State α (prob. 0.9)

	a	b	ϕ
a	0, 0	$\frac{1}{2}, \frac{1}{2}$	0, 0
b	$\frac{1}{2}, \frac{1}{2}$	1, 1	1, 1
ϕ	0, 0	1, 1	$\frac{1}{2}, \frac{1}{2}$

State β (prob. 0.1)

(b) Nash equilibrium in which the strategy of player 1 is weakly dominated.

Figure 7.1 The two Nash equilibria of the Bayesian game in Section 7.1.1.

The first equilibrium reflects the earlier informal analysis. Player 1, who is fully informed, votes for the appropriate alternative in each state. Player 2, who is uninformed, abstains, leaving the decision to player 1; if she votes and her vote makes a difference then it affects the outcome adversely.

7.1.2 Model with some informed and some uninformed individuals

What are the implications of the idea captured by the example in a more general model with many individuals who differ not only in their information, but also in their preferences?

I analyze a model with two alternatives, two states, and many individuals. Some individuals, called partisans, prefer one of the alternatives regardless of the state: some always prefer a, and some always prefer b. Others, called independents, prefer a if the state is α and b if it is β. Among the independents, some are informed of the state, and some are not. The state is irrelevant to the partisans, and I assume that they are not informed of it.

Definition 7.1: Plurality rule voting game with two alternatives and asymmetric information

A *plurality rule voting game with two alternatives and asymmetric information* $\langle\{a,b\},(n_a,n_b,n_i,n_u),\{\alpha,\beta\},\pi,(v_a,v_b)\rangle$, where a and b are alternatives, n_a, n_b, n_i, and n_u are nonnegative integers with $n_a+n_b+n_i+n_u \geq 3$, α and β are states, $\pi \in (0,1)$, and v_a and v_b are positive numbers, is the following Bayesian game.

Players

A set with $n_a + n_b + n_i + n_u$ members; n_a players are *a-partisans*, n_b are *b-partisans*, n_i are *informed independents*, and n_u are *uninformed independents*.

States

The set of states is $\{\alpha, \beta\}$.

Actions

The set of actions of each player is $\{$*vote for a*, *vote for b*, *abstain*$\}$.

Signals

Each informed independent gets a different signal in each state; every other player gets the same signal in both states.

Prior beliefs

Each player assigns probability π to state α.

Payoffs

The payoff of player j for an action profile in which a majority of players vote for $x \in \{a, b\}$ and the state is $s \in \{\alpha, \beta\}$ is

$$j \text{ is an } a\text{-partisan:} \quad \begin{cases} v_a & \text{if } x = a \\ 0 & \text{if } x = b \end{cases}$$

$$j \text{ is a } b\text{-partisan:} \quad \begin{cases} 0 & \text{if } x = a \\ v_b & \text{if } x = b \end{cases}$$

$$j \text{ is an independent:} \quad \begin{cases} 1 & \text{if } (x, s) = (a, \alpha) \text{ or } (b, \beta) \\ 0 & \text{otherwise.} \end{cases}$$

Her payoff for an action profile in which a and b are tied is the average of her payoffs for profiles in which a wins and b wins.

We refer to a player in a Bayesian game who has received a given signal as a *type* of the player or a *player-type*. In a voting game with asymmetric information, each informed independent has two types, one for the signal α and one for the signal β, and every other player has a single type. A strategy for a player is a function that assigns an action to each of her types. Thus a strategy of an informed independent is a pair of actions, one for state α and one for state β, and a strategy for every other player is a single action.

A plurality rule voting game with two alternatives and asymmetric information has many Nash equilibria, and even many in which no player uses a weakly dominated action. Suppose, for example, that there are no partisans ($n_a = n_b = 0$), some informed independents ($n_i \geq 1$), and at least three more uninformed independents than informed independents ($n_u \geq n_i + 3$). In one Nash equilibrium every uninformed independent abstains and every informed independent votes for a in state α and for b in state β. This equilibrium, illustrated in Figure 7.2a, generalizes the equilibrium of the example in the previous section in which no

(a) The Nash equilibrium σ^*, in which only informed independents vote, with the outcome a in state α and b in state β.

(b) The Nash equilibrium σ', in which informed and uninformed independents vote, with the outcome a in both states.

Figure 7.2 Two Nash equilibria of a plurality rule voting game with two alternatives and asymmetric information with no partisans and at least three more uninformed than informed independents .

player's strategy is weakly dominated. Call the equilibrium σ^*.

Consider the strategy profile σ' that differs from σ^* only in that every uninformed independent votes for a. This strategy profile is illustrated in Figure 7.2b. Like σ^*, it is a Nash equilibrium, with the outcome a in both states. It is an equilibrium because no change in any player's strategy affects the outcome: uninformed independents outnumber informed independents by at least three, so for any deviation by a player from σ' the alternative a still wins in both states. Further, no player's strategy in σ' is weakly dominated: every informed independent is voting for the alternative she prefers, given the state, and the strategy of every uninformed independent to vote for a would yield a payoff greater than the strategies of voting for b or abstaining if all other individuals were to abstain, so that it is not dominated by either of these strategies.

Even though σ' is a Nash equilibrium and no player's strategy is weakly dominated, the strategy of the uninformed independents to vote for a seems foolish. Suppose that $n_u = n_i + 3$, and put yourself in the shoes of one of the uninformed independents. If everyone votes according to σ', the outcome is a win for a in both states, regardless of your action. But if one of the other uninformed independents fails to vote (perhaps she is ill on election day, or her bicycle chain breaks on the way to the polling station), then by switching your vote to b, you change the outcome in state β from a win for a to a tie between a and b, and do not affect the outcome in state α, so that you are better off. If one of the informed independents fails to vote, then switching your vote from a to b does not affect the outcome. If two or more of the other individuals fail to vote, but at least two remain able to vote, your switching to vote for b does not affect the outcome in state α and either does not affect it in state β or changes it in state β either from a to b, from a to a tie between a and b, or from a tie between a and b to b, all of which are better for you. Only if all of the other individuals or all but one of them fail to vote may a change in your vote from a to b make you worse off (because it changes the outcome from a to either b or a tie between a and b in state α). You should reasonably believe that the chance that one or two, or a few, of the other individuals fail to vote is much greater than the chance that all of them or all but one of them fail to vote, so that your switching to vote for b is sensible: it makes no difference to the outcome if everyone else complies with σ', and it raises your expected payoff in the most likely cases of non-compliance.

More generally, if $n_u \geq n_i + 3$ then a switch from voting for a to voting for b by an uninformed independent does not effect the outcome if at most $n_u - n_i - 3$ individuals fail to vote and either has no effect on the outcome or improves it for her if more than that do so, as long as at least two individuals vote.

These arguments lead to a definition of equilibrium that assumes that a player deviates from a strategy not only if she has another strategy that yields a higher payoff given the other players' strategies, but also if she has another strategy that yields the same payoff given the other players' strategies and a higher payoff if one of the other players fails to vote, or the same payoff in both cases, but a higher payoff if two of the other players fail to vote, or the same payoff in all these cases, but a higher payoff if three of the other players fail to vote, and so on. I define a deviation to be desirable if it satisfies these conditions, and define an equilibrium to be a strategy profile from which no player has a desirable deviation.

Definition 7.2: Desirability of deviation from strategy profile in voting game with asymmetric information

Let σ be a strategy profile in a plurality rule voting game with two alternatives and asymmetric information, let t be a player-type, let x_t be the action of t specified by σ, let x'_t be an action of t different from x_t, and let m be the number of other player-types who vote according to σ. The *desirability* for t of a deviation from x_t to x'_t is determined by the following procedure.

Initialization

Set $k = 0$.

Step k

Assume that a randomly-determined group of exactly k of the other player-types who vote according to σ fail to do so, with each such group equally likely. Then if the expected payoff generated by x'_t is

- less than that generated by x_t, a deviation to x'_t is *not desirable*

- more than that generated by x_t, a deviation to x'_t is *desirable*
- the same as that generated by x_t and $k = m$, a deviation to x'_t is *not desirable*
- the same as that generated by x_t and $k \leq m-1$, continue to Step $k+1$.

Definition 7.3: Equilibrium of voting game with asymmetric information

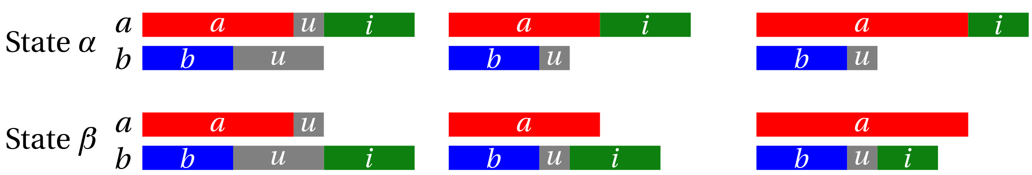
A strategy profile is an *equilibrium* of a plurality rule voting game with two alternatives and asymmetric information if no deviation from the strategy profile by any player-type is desirable.

Suppose that all individuals are partisans (there are no independents), so that the game is essentially a two-alternative voting game. I argue that in every equilibrium (in the sense of Definition 7.3) every individual votes for the alternative she favors. (By contrast, in some Nash equilibria some individuals vote for their less-favored alternative.) Consider a strategy profile in which r more players vote for a than for b, with $r \geq 1$.

If $r \leq 2$, then a player who deviates from voting for a to voting for b changes the outcome from a win for a to either a win for b or a tie, both of which are better for the player if she favors b and worse if she favors a. Thus such a strategy profile is an equilibrium only if every player who votes for a favors a and, similarly, if every player who favors b votes for b.

Now let $r \geq 3$. Consider a player i who votes for a but favors b. If she deviates to voting for b, the outcome does not change if up to $r-3$ of the other players whose strategies call for them to vote in fact fail to do so. Now suppose that a group of $r-2$ of the other players whose strategies call for them to vote fail to do so. If this group contains both players whose strategies call for them to vote for a and ones whose strategies call for them to vote for b, then i's deviation from voting for a to voting for b still does not affect the outcome. However, if the group contains only players whose strategies call for them to vote for a, the deviation changes the outcome from a win for a to a tie. Thus the procedure in Definition 7.2 stops at step $r-2$ with the conclusion that the deviation is desirable. Thus, as for $r \leq 2$, such a strategy profile is an equilibrium in the sense of Definition 7.3 only if every player votes for the alternative she favors.

The next result, illustrated in Figure 7.3, shows that in any equilibrium of a general plurality rule voting game with two alternatives and asymmetric information, every partisan (red and blue in the figure) votes for the alternative she



(a) $n_a - n_b < n_u$. More than enough uninformed independents to cancel out partisans' votes.

(b) $n_u < n_a - n_b < n_i + n_u$. Not enough uninformed independents to cancel out partisans' votes, but enough independents in total to do so.

(c) $n_i + n_u < n_a - n_b$. Not enough independents to cancel out partisans' votes.

Figure 7.3 Equilibria of a plurality rule voting game with two alternatives and asymmetric information (Proposition 7.1) for $n_a > n_b$.

favors, every informed independent (green) votes for the alternative she favors given the state, and uninformed independents (gray) vote so as to cancel out, as far as possible, the partisans' votes:

- if there are enough uninformed independents to cancel out the partisans' votes, uninformed independents vote so that in the absence of any votes by the informed independents, a and b are tied in both states, and hence the margin in favor of a in state α among all votes is the same as the margin in favor of b in state β (Figure 7.3a)

- if there are too few informed independents to cancel out the partisans' votes, all uninformed independents vote for the alternative favored by fewer partisans, maximizing the influence of the informed independents in case some partisans do not participate (Figures 7.3b and 7.3c).

Proposition 7.1: Equilibrium of voting game with asymmetric information

Consider a plurality rule voting game with two alternatives and asymmetric information $\langle \{a,b\}, (n_a, n_b, n_i, n_u), (\alpha, \beta), \pi, (v_a, v_b) \rangle$ with $n_i \geq 1$ (at least one individual is an informed independent). A strategy profile is an equilibrium if and only if

a. every a-partisan votes for a and every b-partisan votes for b

b. every informed independent votes for a in state α and for b in state β

c. the number n_u^a of uninformed independents who vote for a and the number n_u^b who vote for b satisfy

$$\begin{aligned} n_u^a - n_u^b &= \min\{n_b - n_a, n_u\} \quad \text{if } n_a \le n_b \\ n_u^b - n_u^a &= \min\{n_a - n_b, n_u\} \quad \text{if } n_a \ge n_b. \end{aligned} \tag{7.1}$$

If $n_a = n_b$ the outcome of an equilibrium is a in state α and b in state β, if $n_a > n_b$ it is

$$\begin{array}{ll} a \text{ in state } \alpha, b \text{ in state } \beta & \text{if } n_a - n_b < n_i + n_u \\ a \text{ in state } \alpha, \text{tie in state } \beta & \text{if } n_a - n_b = n_i + n_u \\ a \text{ in both states} & \text{if } n_a - n_b > n_i + n_u, \end{array} \tag{7.2}$$

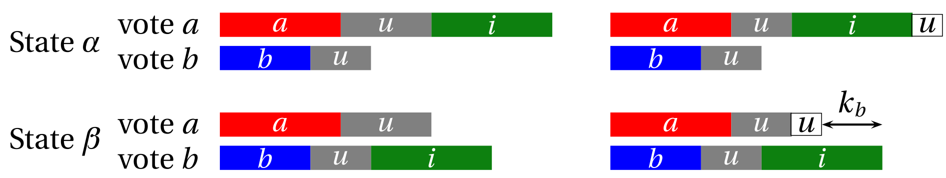
and if $n_a < n_b$ it is the variant of (7.2) in which a and b and interchanged and α and β are interchanged.

The outcome of an equilibrium is the same as the outcome of an equilibrium of the variant of the game in which every individual is informed of the state.

Comments

- If $|n_a - n_b| < n_u$ (Figure 7.3a if $n_a > n_b$), some uninformed independents may abstain; the only requirement on their behavior for equilibrium is that the difference between the number who vote for b and the number who vote for a is $|n_a - n_b|$.
- If a-partisans outnumber b-partisans ($n_a > n_b$) and α is the more likely state, the alternative for which most uninformed independents vote is b, which according to their prior belief is the *wrong* alternative. By doing so they nullify, as much as possible, the partisans' votes, leaving the decision to the informed independents.
- In each case, the outcome is the same as outcome of the equilibrium in the variant of the game in which all individuals are informed (that is, there are $n_i + n_u$ instead of n_i informed independents and no uninformed independents). For example, in the cases in Figures 7.3a and 7.3b the outcome is a in state α and b in state β, and in the case in Figure 7.3c the outcome is a in both states. One way to express this feature of equilibrium is to say that the equilibrium *fully aggregates information.*

In an equilibrium, every partisan and informed independent votes for her favorite alternative because voting for the other alternative has no possible benefit,



(a) Original strategy profile.

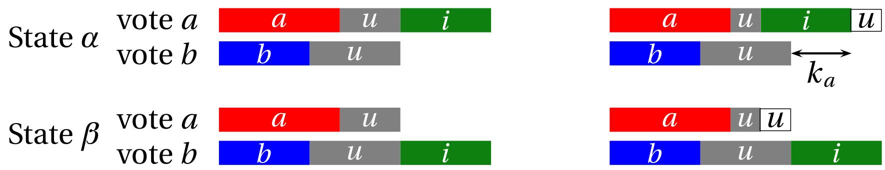
(b) Effect of uninformed independent's deviation to abstention.

Figure 7.4 The effect of an uninformed independent's deviation to abstention for a strategy profile in a plurality rule voting game with two alternatives and asymmetric information in which the winning margin for a in state α is larger than the winning margin for b in state β.

however the other players vote; doing so can only generate a worse outcome, if not when all the other players adhere to their strategies then when some of them fail to vote.

The argument that uninformed independents vote so that the margin in favor of a in state α is equal to the margin in favor of b in state β, if possible, is more involved. Suppose that there are more than enough uninformed independents to cancel out partisans' votes, but the uninformed independents vote in such a way that the margin in favor of a in state α is larger than the margin in favor of b in state β, as in Figure 7.4a. Suppose that an uninformed independent j who is voting for a deviates to abstention. (Refer to Figure 7.4b.) The smallest number of players whose failure to vote causes this change in j's strategy to affect the outcome is the original margin in favor of b in state β, say k_b. If k_b players whose strategies call for them to vote for b fail to do so, then the deviation by j changes the outcome from a tie to a win for b in state β and does not affect the outcome in state α. If any other k_b players fail to vote, the deviation by j does not affect the outcome. Thus the deviation is desirable, so that the strategy profile is not an equilibrium.

Now consider a strategy profile for which the margins in favor of a in state α and b in state β are the same, as in Figure 7.5a. If an uninformed independent j who is voting for a deviates to abstention, the smallest number of players whose failure to vote causes the outcome to change is the new margin in favor of a in state α, say k_a. (Refer to Figure 7.5b.) If k_a players whose strategies call for them to vote for a fail to vote, then the deviation by j changes the outcome from a win for a to a tie in state α, and does not affect the outcome in state β. If any other k_a players fail to vote, the deviation by j does not affect the outcome. Thus the deviation is undesirable, so that the strategy profile is consistent with an equilibrium.



(a) Original strategy profile.

(b) Effect of uninformed independent's deviation to abstention.

Figure 7.5 The effect of an uninformed independent's deviation to abstention for a strategy profile in a plurality rule voting game with two alternatives and asymmetric information in which the winning margins for a in state α and for b in state β are the same.

The cases in which the there are enough independents to cancel out the partisans' votes, but not enough uninformed independents to do so, and in which there are not enough independents to cancel out the partisans' votes, require similar arguments. The following proof contains the details.

Proof of Proposition 7.1

I first argue that a strategy profile satisfying the conditions in the result is an equilibrium. Let σ^* be a strategy profile satisfying the conditions in the result.

Step 1 *No change in the strategy in σ^* of an informed independent or partisan is desirable.*

Proof. If an informed independent changes her action in state α from a to either b or abstention, then either the outcome in state α does not change or it changes to b with positive probability ($\frac{1}{2}$ or 1), regardless of how many other players fail to vote, so the deviation is not desirable. A similar argument applies to an informed independent who changes her action in state β and to a partisan who changes her action. ◁

Step 2 *No change in the strategy in σ^* of an uninformed independent is desirable.*

Proof. Consider an uninformed independent, say j. Assume that $n_a \geq n_b$. There are three cases.

$n_a - n_b < n_i + n_u$
The outcome of σ^* is a in state α and b in state β. (For $n_a > n_b$, Figures 7.3a and 7.3b show examples.)

Suppose that j votes for a. If she deviates to abstention and the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, then the outcome changes only in state α, where it becomes a tie between a and b rather than a, decreasing j's expected payoff. Thus the deviation is not desirable. By a similar argument, a deviation by j to vote for b is not desirable.

A similar argument shows that if j votes for b then any deviation is not desirable.

Now suppose that j abstains, which happens only if $n_a - n_b < n_u$ (Figure 7.3a). Suppose that she deviates to vote for a. If the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, then the outcome changes only in state β, where it becomes a tie between a and b rather than a win for b, decreasing j's expected payoff. Thus the deviation is not desirable. A symmetric argument shows that j's deviation to vote for b is also not desirable.

$n_a - n_b > n_i + n_u$

Given $n_i \geq 1$, we have $n_a > n_b$, and the outcome of σ^* is a in both states (refer to Figure 7.3c). The strategy profile σ^* specifies that every uninformed independent, and in particular j, votes for b. If she deviates to abstention and the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, then the outcome changes only in state β, where it becomes a win for a rather than a tie between a and b, decreasing j's expected payoff. Thus the deviation is not desirable.

A similar argument shows that also a deviation by j to vote for a is undesirable.

$n_a - n_b = n_i + n_u$

The outcome of σ^* is a in state α and a tie between a and b in state β. As in the previous case, every uninformed independent, and in particular j, votes for b. If she deviates to abstention or to vote for a, the outcome in state β changes to a win for a and the outcome in state α does not change, so the deviation is not desirable.

The argument for the case $n_a \leq n_b$ is symmetric with this argument. ◁

Step 3 *The strategy profile σ^* is an equilibrium.*

Proof. The result follows from Steps 1 and 2. ◁

I now argue that every equilibrium satisfies the conditions in the result.

Step 4 *In every equilibrium, every informed independent votes for a in state α and for b in state β.*

Proof. Let σ be a strategy profile and let j be an informed independent. Suppose that σ_j specifies a vote for b in state α. If j deviates to vote for a in state α and the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, then the outcome changes only in state α, where it either becomes a win for a rather than a tie between a and b, or a win for a rather than a win for b, in both cases increasing j's expected payoff. Thus such a deviation is desirable.

A similar argument shows that if σ_j specifies abstention in state α then a deviation to vote for a is desirable.

Symmetric arguments show that if σ_j specifies a vote for a or abstention in state β then j has a desirable deviation. ◁

Step 5 *In every equilibrium, every a-partisan votes for a and every b-partisan votes for b.*

Proof. This conclusion follows from arguments like those in Step 4. ◁

Step 6 *In every equilibrium, the number n_u^a of uninformed independents who vote for a and the number n_u^b who vote for b satisfy (7.1).*

Proof. Suppose that $n_a \geq n_b$. There are two cases.

$n_a - n_b \geq n_u$

From (7.1), $n_u^b - n_u^a = n_u$, so $n_u^b = n_u$: all uninformed independents vote for b. Given $n_a - n_b \geq n_u$, if any uninformed independents exist we have $n_a > n_b$; Figures 7.3b and 7.3c are examples. Consider a strategy profile in which an uninformed independent j abstains. If she deviates to vote for b and the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, the outcome in state β changes either from a tie to a win for b, or from a win for a to a tie, and the outcome in state α remains a win for a (given $n_a > n_b$, $n_a - n_b \geq n_u$, and $n_i \geq 1$), so in both cases her expected payoff increases. A similar argument shows that if j's strategy calls for her to vote for a and she deviates to abstention then her expected payoff increases for any minimal set of players whose failure to vote affects the outcome. We conclude that in an equilibrium every uninformed independent votes for b, so that (7.1) is satisfied.

$n_a - n_b < n_u$

From (7.1), $n_u^b - n_u^a = n_a - n_b$. Figure 7.3a is an example. Consider a strategy profile in which $n_u^b - n_u^a > n_a - n_b$. Then $n_b + n_i + n_u^b > n_b + n_u^a + n_a - n_b = n_a + n_u^a$, so that b wins in state β, with a margin of victory of $n_b + n_i + n_u^b - n_a - n_u^a = n_i - (n_a - n_b) + (n_u^b - n_u^a) > n_i$.

- If a wins in state α, its margin of victory is $n_a + n_i + n_u^a - n_b - n_u^b = n_i + (n_a - n_b) - (n_u^b - n_u^a) < n_i$. Consider an uninformed independent j whose strategy calls for her to vote for b. If she deviates to abstention and the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, the outcome in state α changes from a tie to a win for a, and the outcome in state β, where the margin of victory is larger, does not change. Thus the deviation increases j's expected payoff, so the strategy profile is not an equilibrium.
- If b wins in state α, its margin of victory is $n_b + n_u^b - n_a - n_u^a - n_i$, which is less than b's margin of victory in state β, $n_b + n_u^b + n_i - n_a - n_u^a$. In this case, consider also an uninformed independent j whose strategy calls for her to vote for b. If she deviates to abstention and the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, the outcome in state α changes from a win for b to a tie, and the outcome in state β does not change. Thus the deviation increases j's expected payoff, so the strategy profile is not an equilibrium.

Now consider a strategy profile in which $n_u^b - n_u^a < n_a - n_b$. In this case, a wins in state α and the margin of victory of the winner in state β is less than the margin of victory of a in state α. By an argument symmetric with that in the previous case, an uninformed independent who switches from voting for a to abstention increases her expected payoff when the number of individuals who fail to vote is the smallest for which this deviation affects the outcome, so that the strategy profile is not an equilibrium.

The argument for the case $n_a \le n_b$ is similar. ◁

Step 7 *The outcome of an equilibrium is given in (7.2).*

Proof. If $n_a = n_b$ then by (7.1) we have $n_u^a = n_u^b$, so that given $n_i \ge 1$, a wins in state α and b wins in state β. Now suppose that $n_a > n_b$. (The case $n_a <$

n_b is symmetric.) Then by the characterization of an equilibrium, in state α alternative a gets $n_a + n_i + n_u^a$ votes and alternative b gets $n_b + n_u^b$ votes. If $n_a - n_b < n_u$ then by (7.1) we have $n_b + n_u^b = n_b + n_a - n_b + n_u^a = n_a + n_u^a$, so that given $n_i \geq 1$, alternative a wins. If $n_a - n_b \geq n_u$ then $n_u^b = n_u$ and $n_u^a = 0$, so that a gets $n_a + n_i$ votes and b gets $n_b + n_u < n_b + n_a - n_b = n_a$ votes, so that again a wins. In state β alternative a gets $n_a + n_u^a$ votes and alternative b gets $n_b + n_i + n_u^b$ votes. If $n_a - n_b < n_u$ then by (7.1) we have $n_u^b - n_u^a = n_a - n_b$, so that $n_b + n_i + n_u^b = n_a + n_u^a + n_i$ and hence b wins. If $n_a - n_b \geq n_u$ then $n_u^b = n_u$ and $n_u^a = 0$, so that a gets n_a votes and b gets $n_b + n_u + n_i$ votes, and hence a wins if $n_a - n_b > n_i + n_u$, a and b tie if $n_a - n_b = n_u + n_i$, and b wins if $n_a - n_b < n_i + n_u$. ◁

Step 8 *The outcome of an equilibrium is the same as the outcome of an equilibrium of the variant of the game in which every individual is informed of the state.*

Proof. From the characterization of an equilibrium, in a game in which the number of informed independents is $n_i + n_u$ and there are no uninformed independents, the number of votes for a is $n_a + n_i + n_u$ in state α and n_a in state β, and the number of votes for b is n_b in state α and $n_b + n_i + n_u$ in state β. Thus if $n_a = n_b$ then a wins in state α and b wins in state β, and if $n_a > n_b$ then a wins in state α and in state β alternative a wins if $n_a - n_b > n_i + n_u$, the alternatives tie if $n_a - n_b = n_i + n_u$, and b wins if $n_a - n_b < n_i + n_u$, as when n_u of the independents are uninformed. ◁

Note that the equilibria do not depend on the values v_a and v_b that partisans attach to their favorite alternatives. If $v_a = v_b = 1$ then the outcome of an equilibrium in each state maximizes the sum of the individuals' payoffs. Otherwise, it may not. For example, if the number of independents is not sufficient to cancel out the partisans' votes, and a-partisans outnumber b-partisans, then the outcome is a in both states, but if v_a and v_b are close enough to 0, alternative b maximizes the sum of the payoffs in state β.

7.1.3 Model with imperfectly-informed individuals

The result that some individuals abstain in equilibrium does not depend on these individuals being completely uninformed, facing individuals who are perfectly informed. Consider a variant of a plurality rule voting game with two alternatives and asymmetric information in which in each state the individuals are a priori identical. The processes generating signals for any individual i are shown

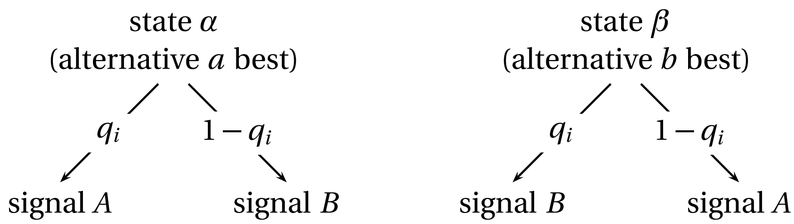


Figure 7.6 The processes generating signals for each individual i in a plurality rule voting game with two alternatives and uncertain signal qualities. The value of q_i is a draw from a distribution F with support $[\frac{1}{2}, 1]$.

in Figure 7.6. In state α each individual i gets the signal A with probability q_i and the signal B with probability $1-q_i$, and in state β she gets the signal B with probability q_i and the signal A with probability $1-q_i$, where q_i is a draw from a probability distribution with support $[\frac{1}{2}, 1]$ and a continuous density, independently of the draw of q_j for every other individual j. In particular, with probability 1 each individual's signal conveys some information. (Only the signal of an individual i for whom $q_i = \frac{1}{2}$, a value that occurs with probability 0, is completely uninformative.) Each individual i knows q_i and her signal, but not the value of q_j or the signal of any other individual j. To make the model symmetric, I assume that each individual believes that the prior probability of each state is $\frac{1}{2}$. Each individual votes for one of the alternatives or abstains; the outcome is the alternative that receives the most votes, and each individual's payoff is 1 if the outcome is a and the state is α or the outcome is b and the state is β, and is 0 otherwise.

Precisely, the model is the following Bayesian game. A state in this game, which captures all the uncertain features of the environment, is a triple consisting of the state of nature α or β, the signal qualities, and the signal realizations. Although "state" has this meaning in the following definition, outside the definition I continue to use the word to refer to α and β.

Definition 7.4: Plurality rule voting game with two alternatives and uncertain signal qualities

A *plurality rule voting game with two alternatives and uncertain signal qualities* $\langle \{a, b\}, n, (\alpha, \beta), \{A, B\}, F \rangle$, where a and b are alternatives, n is a positive integer, α and β are states of nature, A and B are signals, and F is a probability distribution function with support $[\frac{1}{2}, 1]$ and a continuous density is the following Bayesian game.

Players

A set N with n members.

States

The set of states is the set of triples $(\omega,(q_j)_{j\in N},(s_j)_{j\in N})$ where $\omega\in\{\alpha,\beta\}$ (the state of nature), $q_j\in[\frac{1}{2},1]$ for each $j\in N$ (j's signal quality), and $s_j\in\{A,B\}$ for each $j\in N$ (j's signal).

Actions

For each player the set of actions is $\{$*vote for a*, *vote for b*, *abstain*$\}$.

Signals

For each player i the set of signals is $\{A,B\}\times[\frac{1}{2},1]$ and the signal function τ_i is given by $\tau_i(\omega,(q_j)_{j\in N},(s_j)_{j\in N})=(s_i,q_i)$ for each state $(\omega,(q_j)_{j\in N},(s_j)_{j\in N})$.

Prior beliefs

Each player i believes that

- the value of ω is α with probability $\frac{1}{2}$ and β with probability $\frac{1}{2}$
- q_i is drawn randomly from F
- s_i is A with probability q_i and B with probability $1-q_i$ if $\omega=\alpha$, and B with probability q_i and A with probability $1-q_i$ if $\omega=\beta$ (as illustrated in Figure 7.6).

Every random draw is independent of every other random draw.

Payoffs

The payoff of each player for each pair consisting of an action profile and a state $(\omega,(q_j)_{j\in N},(s_j)_{j\in N})$ is

$$\begin{cases}1 & \text{if } \omega=\alpha\\ 0 & \text{if } \omega=\beta\end{cases}\quad \text{if more individuals vote for } a \text{ than for } b,$$

$$\begin{cases}0 & \text{if } \omega=\alpha\\ 1 & \text{if } \omega=\beta\end{cases}\quad \text{if more individuals vote for } b \text{ than for } a,$$

and $\frac{1}{2}$ (regardless of the value of ω) if the numbers of individuals who vote for a and for b are the same.

The signal of an individual i is more informative the higher is q_i, so it is reasonable to think that the game has an equilibrium in which each individual i uses a strategy with the following form: there is a number $q_i^*\in[\frac{1}{2},1]$, the threshold, such that type (A,q_i) of any individual i votes for a if $q_i\geq q_i^*$ and abstains if $q_i<q_i^*$, and type (B,q_i) of any individual i votes for b if $q_i\geq q_i^*$ and abstains if $q_i<q_i^*$. I refer to a strategy of this form as a quality threshold strategy.

Definition 7.5: Quality threshold strategy in voting game with uncertain signal qualities

A strategy of any individual i in a plurality rule voting game with two alternatives and uncertain signal qualities $\langle\{a,b\},n,(\alpha,\beta),\{A,B\},F\rangle$ is a *quality threshold strategy* if there is a number $q_i^* \in [\frac{1}{2},1]$, the *threshold*, such that the strategy is

$$\begin{cases} \textit{vote for a} & \text{if } i\text{'s signal is } A \text{ and } q_i \geq q_i^* \\ \textit{vote for b} & \text{if } i\text{'s signal is } B \text{ and } q_i \geq q_i^* \\ \textit{abstain} & \text{if } q_i < q_i^*. \end{cases}$$

Given the symmetry of the model, including the fact that the prior probability of each state is $\frac{1}{2}$, it is also reasonable to think that the game has a Nash equilibrium in which every individual uses a quality threshold strategy with the same threshold.

Definition 7.6: Quality threshold equilibrium of voting game with uncertain signal qualities

A strategy profile in a plurality rule voting game with two alternatives and uncertain signal qualities is a *quality threshold equilibrium* if it is a Nash equilibrium and every individual's strategy is a quality threshold strategy with the same threshold.

I show in the next result that if there are two individuals ($n = 2$), the game has a quality threshold equilibrium, and in every such equilibrium the threshold exceeds $\frac{1}{2}$, so that an individual with an informative but low-quality signal abstains. If each individual could report both her signal and its quality, and a decision were made by pooling this information, then each individual would optimally report her signal regardless of its quality. But voting provides no way for individuals to report the qualities of their signals. All they can do is vote or abstain, and an individual who votes based on a low-quality signal influences the outcome just as much as one who votes based on a high-quality signal. As a consequence, an individual with an informative but low-quality signal prefers to abstain than to vote, making the outcome depend on the vote of the other individual, who is likely to have received a more informative signal.

To understand in more detail why the equilibrium threshold exceeds $\frac{1}{2}$, consider the ingredients of an individual's decision of whether to abstain or vote. She should ponder the implications of her action for the outcome, given each possible action of the other individual, taking into account any information about the state that the other individual's action conveys.

Consider type (A, q_i) of individual i. Assume that the other individual, j, is using a quality threshold strategy with threshold q^*.

j votes for a The outcome is the same whether i votes for a or abstains, so this possibility is irrelevant to i's decision.

j abstains If individual i votes for a then the outcome is a, and if she abstains then it is a tie, so her gain from voting for a rather than abstaining is $\frac{1}{2}$ if the state is α and $-\frac{1}{2}$ if the state is β.

j votes for b If individual i votes for a then the outcome is a tie and if she abstains then it is b, so her gain for voting for a rather than abstaining is $\frac{1}{2}$ if the state is α and $-\frac{1}{2}$ if the state is β (as when j abstains).

Thus the expected gain of type (A, q_i) of individual i from voting for a rather than abstaining is

$$\begin{aligned}\Pr(j \text{ abstains} \mid i\text{'s signal } A)\big[&\Pr(\text{state } \alpha \mid j \text{ abstains \& } i\text{'s signal } A)\cdot\tfrac{1}{2}\\ &+\Pr(\text{state } \beta \mid j \text{ abstains \& } i\text{'s signal } A)\cdot(-\tfrac{1}{2})\big]\\ +\Pr(j \text{ votes } b \mid i\text{'s signal } A)\big[&\Pr(\text{state } \alpha \mid j \text{ votes } b \text{ \& } i\text{'s signal } A)\cdot\tfrac{1}{2}\\ &+\Pr(\text{state } \beta \mid j \text{ votes } b \text{ \& } i\text{'s signal } A)\cdot(-\tfrac{1}{2})\big].\end{aligned} \tag{7.3}$$

Now, j abstains if and only if her signal q_j is less than q^*, so the probability that i assigns to her abstaining is $F(q^*)$ (independent of i's signal). The fact that she abstains conveys no information about the likelihood that the state is α or β, so $\Pr(\text{state } \alpha \mid j \text{ abstains \& } i\text{'s signal } A)$ is the probability of α conditional on one signal of A with quality q_i, which, using Bayes' rule, is q_i (given that the prior probability of each state is $\frac{1}{2}$). Thus the expected gain in (7.3) is

$$\begin{aligned}&\tfrac{1}{2}F(q^*)(q_i-(1-q_i))\\ &\quad+\Pr(j \text{ votes } b \mid i\text{'s signal } A)\big[\Pr(\text{state } \alpha \mid j \text{ votes } b \text{ \& } i\text{'s signal } A)\cdot\tfrac{1}{2}\\ &\qquad\qquad+\Pr(\text{state } \beta \mid j \text{ votes } b \text{ \& } i\text{'s signal } A)\cdot(-\tfrac{1}{2})\big].\end{aligned} \tag{7.4}$$

For a quality threshold equilibrium with threshold q^*, this expected gain must be zero for type (A, q^*): type (A, q^*) of individual i must be indifferent between voting for a and abstaining. Now, individual j votes for b only if her signal is B, which is more likely if the state is β than if it is α. Thus j's voting for b is evidence in favor of state β. Further, j votes for b only if her signal quality is at least q^*, and hence at least as high as i's signal quality. So j's voting for b provides stronger evidence about the state than i's signal A, and hence conditional on this event, the probability that i assigns to state α is less than $\frac{1}{2}$: $\Pr(\text{state } \alpha \mid j \text{ votes } b \text{ \& } i\text{'s signal } A) < \frac{1}{2}$ and $\Pr(\text{state } \beta \mid j \text{ votes } b \text{ \& } i\text{'s signal } A) > \frac{1}{2}$. Thus for $q_i = q^*$ the second term in (7.4) is negative, so for the whole expression to be zero

j's action	i's gain from voting for a rather than abstaining if state is α	if state is β	posterior prob. of state α
vote for a	0	0	irrelevant
abstain	$\frac{1}{2}$	$-\frac{1}{2}$	q^*
vote for b	$\frac{1}{2}$	$-\frac{1}{2}$	$< \frac{1}{2}$

Figure 7.7 The ingredients of the decision of type (A, q^*) of individual i regarding whether to vote for a or abstain.

we need the first term to be positive, which requires $q^* > \frac{1}{2}$. The ingredients of i's reasoning that lead to this conclusion are summarized in Figure 7.7.

We conclude that in any quality threshold equilibrium, an individual with an informative but low-quality signal abstains. She does so because if her vote affects the outcome, then taking into account the information implied about the state by the other individual's action, her vote is more likely to change the outcome adversely than advantageously, given the superior quality of the other individual's signal. As in a plurality rule voting game with two alternatives and asymmetric information, in which some individuals are uninformed and others are perfectly informed, we can characterize her predicament by saying that she is subject to the swing voter's curse.

Proposition 7.2: Quality threshold equilibrium of voting game with uncertain signal qualities and two individuals

Let $\langle \{a, b\}, n, (\alpha, \beta), \{A, B\}, F \rangle$ be a plurality rule voting game with two alternatives and uncertain signal qualities for which $n = 2$. This game has a quality threshold equilibrium in which the threshold is in $(\frac{1}{2}, 1)$ and in every quality threshold equilibrium the threshold lies in this interval.

Proof

Denote the individuals by i and j, and suppose that j's strategy is a quality threshold strategy with threshold q^*. First note that if $q^* = 1$, so that j votes with probability zero, the outcome is a tie if i abstains and otherwise is the outcome for which she votes, so voting is optimal for i regardless of her signal quality. Thus the game has no quality threshold equilibrium with threshold 1.

Now suppose that $q^* < 1$ and consider type (A, q_i) of individual i. By the argument preceding the result, her expected gain from voting for a rather than abstaining is given in (7.4). To show that $q^* > \frac{1}{2}$ in any equilibrium, it it enough to show that the probability that type (A, q^*) of individual i assigns to state α in the case that j votes for b is less than $\frac{1}{2}$, as argued informally in the text preceding this result. But to show that an equilibrium exists, we need to study the expected gain in more detail.

We have

$$\begin{aligned}
&\Pr(j \text{ votes } b \mid i\text{'s signal } A)\Pr(\text{state } \alpha \mid j \text{ votes } b \;\&\; i\text{'s signal } A) \\
&= \Pr(j \text{ votes } b \mid i\text{'s signal } A)\frac{\Pr(j \text{ votes } b \;\&\; i\text{'s signal } A \mid \text{state } \alpha)\Pr(\text{state } \alpha)}{\Pr(j \text{ votes } b \;\&\; i\text{'s signal } A)} \\
&= \frac{\Pr(j \text{ votes } b \;\&\; i\text{'s signal } A \mid \text{state } \alpha)\Pr(\text{state } \alpha)}{\Pr(i\text{'s signal } A)} \\
&= \frac{\Pr(j \text{ votes } b \mid \text{state } \alpha)\Pr(i\text{'s signal } A \mid \text{state } \alpha)\Pr(\text{state } \alpha)}{\Pr(i\text{'s signal } A)} \\
&= \frac{\int_{q^*}^{1}(1-q_j)\,dF(q_j)\cdot q_i \cdot \frac{1}{2}}{\frac{1}{2}q_i + \frac{1}{2}(1-q_i)} \\
&= q_i \int_{q^*}^{1}(1-q_j)\,dF(q_j).
\end{aligned}$$

By a similar argument,

$$\begin{aligned}
\Pr(j \text{ votes } b \mid i\text{'s signal } A)\Pr(\text{state } \beta \mid j \text{ votes } b \;\&\; i\text{'s signal } A)& \\
= (1-q_i)\int_{q^*}^{1} q_j\,dF(q_j)&.
\end{aligned}$$

Substituting these expressions into (7.4), we conclude that the expected gain of type (A, q_i) of individual i from voting for a rather than abstaining, given that j's strategy is a quality threshold strategy with threshold q^*, is

$$\begin{aligned}
G(q_i, q^*) &= \tfrac{1}{2}F(q^*)(2q_i - 1) + \tfrac{1}{2}q_i \int_{q^*}^{1}(1-q_j)\,dF(q_j) - \tfrac{1}{2}(1-q_i)\int_{q^*}^{1} q_j\,dF(q_j) \\
&= \tfrac{1}{2}(q_i - (1-q_i)F(q^*)) - \tfrac{1}{2}\int_{q^*}^{1} q_j\,dF(q_j).
\end{aligned}$$

The expected gain of type (B, q_i) of individual i from voting for b rather than abstaining is given by the same expression. Define the function $H : [\frac{1}{2}, 1) \to \mathbb{R}$ by $H(q) = G(q, q)$. The game has a quality threshold equilibrium with threshold q^* if and only if $H(q^*) = 0$. Now, the integral in the expression for $G(q_i, q^*)$ exceeds $q^*(1 - F(q^*))$, so $H(\frac{1}{2}) < 0$. Further, $H(q^*) \to \frac{1}{2} > 0$ as $q^* \to 1$ and H is continuous (given that F has a continuous density). Hence there is a number $q^* \in (\frac{1}{2}, 1)$ such that $H(q^*) = 0$ and every number q^* for which $H(q^*) = 0$ is in $(\frac{1}{2}, 1)$.

7.2 Unanimity rule

For a decision made by unanimity rather than plurality rule, the inference that an individual makes about the state from the fact that her vote is pivotal has a simple and striking implication. Assume, as before, that there are two alternatives, a and b. Alternative a is the default; the outcome is b if and only if every individual votes for b. Given that for this rule the implication of abstention is the same as that of voting for a, assume that the only actions available to each individual are *vote for* a and *vote for* b.

Suppose first, as in a plurality rule voting game with two alternatives and asymmetric information, that some individuals are perfectly informed and others are uninformed. The strategy profile in which every informed individual votes for a in state α and for b in state β and every uninformed individual votes for b is a Nash equilibrium of the Bayesian game. The reason is that given the strategies of the other individuals, a change in the action specified by any informed individual's strategy in state α either does not affect the outcome or, if only one individual is informed, changes it to b, while a change in the action in state β changes the outcome from b to a, and a change in any uninformed individual's strategy does not affect the outcome in state α (given the presence of at least one informed individual) and changes the outcome from b to a in state β. In fact, this strategy profile is the only equilibrium in the sense of Definition 7.3, as you are asked to show in the next exercise.

Exercise 7.1: Equilibria of unanimity rule voting game

Show that in the variant of a plurality rule voting game with two alternatives and asymmetric information in which the decision is made by unanimity rule, with a the default, the only equilibrium in the sense of Definition 7.3 is the strategy profile in which every informed individual votes for a in state α and b in state β and every uninformed individual votes for b.

The point is that under unanimity rule the only way an uninformed individual can hand the decision to the informed individuals is by voting for the non-default alternative, because if she votes for the default alternative then that alternative is the outcome regardless of the other individuals' votes.

If the qualities of the individuals' signals are less extreme, similar considerations lead to the conclusion that when the number of individuals is large, the strategy profile in which every individual votes for the alternative that is more likely to be best according to her signal is *not* a Nash equilibrium. Suppose that, as in the model in Section 7.1.3, the individuals are a priori identical. Every individual believes initially that the state is α with probability π and then receives

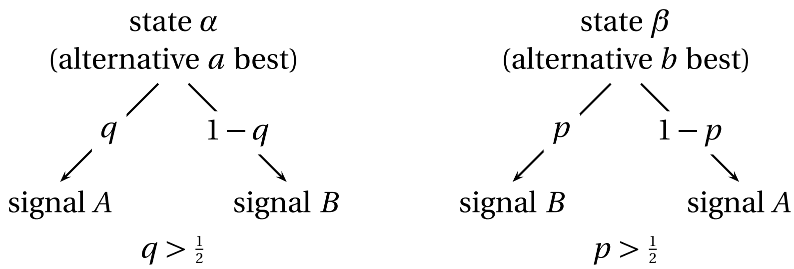


Figure 7.8 The processes generating signals in a model of unanimity rule in which each individual is a priori identical.

one of two signals, as shown in Figure 7.8. If the state is α, in which everyone agrees that alternative a is best, each individual independently gets the signal A with probability q and the signal B with probability $1-q$, where $\frac{1}{2} < q < 1$. If the state is β, in which everyone agrees that alternative b is best, each individual independently gets the signal B with probability p and the signal A with probability $1-p$, where $\frac{1}{2} < p < 1$. Given the asymmetry of the alternatives, I allow the payoffs to be asymmetric: each individual's payoff is

$$\text{state } \alpha: \begin{cases} v_a & \text{if outcome } a \\ -w_b & \text{if outcome } b \end{cases} \qquad \text{state } \beta: \begin{cases} v_b & \text{if outcome } b \\ -w_a & \text{if outcome } a \end{cases}$$

where $v_a > 0$, $v_b > 0$, $w_a > 0$, and $w_b > 0$.

We can model this situation as the following Bayesian game. As for the model of a plurality rule voting game with two alternatives and uncertain signal qualities in Section 7.1.3, a state in the game, which captures all the uncertain features of the environment relevant to the individuals, includes a specification of the profile of signals that they receive. However, outside of this definition I continue to refer to α and β as "states".

Definition 7.7: Unanimity rule voting game with two alternatives and asymmetric information

A *unanimity rule voting game with two alternatives and asymmetric information* $\langle \{a,b\}, n, (\alpha,\beta), \{A,B\}, \pi, (p,q), (v_a, v_b, w_a, w_b)\rangle$, where a and b are alternatives, $n \geq 2$ is an integer, α and β are states of nature, A and B are signals, $\pi \in (0,1)$, $p \in (\frac{1}{2}, 1)$, $q \in (\frac{1}{2}, 1)$, $v_a > 0$, $v_b > 0$, $w_a > 0$, and $w_b > 0$ is the following Bayesian game.

Players

A set N with n members.

States

The set of states is the set of pairs $(\omega, (s_j)_{j\in N})$ where $\omega \in \{\alpha, \beta\}$ (the state of nature) and $s_j \in \{A, B\}$ for each $j \in N$ (j's signal).

Actions

The set of actions of each individual is $\{\textit{vote for } a, \textit{vote for } b\}$.

Signals

The set of signals that each player may receive is $\{A, B\}$ and the signal function τ_i of each player i is defined by $\tau_i(\omega, (s_j)_{j\in N}) = s_i$ for each $\omega \in \{\alpha, \beta\}$ and each profile $(s_j)_{j\in N}$ of signals.

Prior beliefs

For each $k \in \{1, \ldots, n\}$, every individual assigns probability $\pi q^k(1-q)^{n-k}$ to each state $(\alpha, (s_j)_{j\in N})$ for which $s_j = A$ for k players and $s_j = B$ for the remaining $n-k$ players, and probability $(1-\pi)p^k(1-p)^{n-k}$ to each state $(\beta, (s_j)_{j\in N})$ for which $s_j = B$ for k players and $s_j = A$ for the remaining $n-k$ players.

Payoffs

The payoff of each player for an action profile in which all individuals vote for b is

$$\begin{cases} -w_b & \text{if the state of nature is } \alpha \\ v_b & \text{if the state of nature is } \beta \end{cases}$$

and her payoff for every other action profile is

$$\begin{cases} v_a & \text{if the state of nature is } \alpha \\ -w_a & \text{if the state of nature is } \beta. \end{cases}$$

I argue that if the number of individuals is sufficiently large, the strategy profile in which every individual votes for a if her signal is A and for b if her signal is B is not a Nash equilibrium of such a game. The reason derives from the fact that under unanimity rule, the vote of any individual i affects the outcome only if all the other individuals vote for b: if at least one of the other individuals votes for a, the outcome is a regardless of i's vote. Thus i's voting for b is optimal if and only if it yields her an expected payoff at least as high as the expected payoff from her voting for a, given the probabilities of the states that she infers from the fact that all the remaining individuals vote for b. Under the strategy profile we are considering, each remaining individual votes for b only if her signal is B, so conditional on all remaining individuals voting for b, the probability that the state is β is high if the number of individuals is large: given $p > \frac{1}{2}$ and $q > \frac{1}{2}$, the probability that every other individual receives a signal of B is larger if the state is β than if it is α, and the ratio of these probabilities approaches 1 as the number of individuals increases without bound. Thus if i's vote affects the outcome, when the number of individuals is large it is likely that the state is β, so i should vote for b.

Proposition 7.3: Voting according to signal not Nash equilibrium of unanimity rule voting game

For each integer $n \geq 2$ let $\Gamma(n) = \langle \{a,b\}, n, (\alpha,\beta), \{A,B\}, \pi, (p,q), (v_a, v_b, w_a, w_b)\rangle$ be a unanimity rule voting game with two alternatives and asymmetric information. (The parameters other than n are fixed.) There is a number n^* such that if $n > n^*$ then the strategy profile in which every individual votes for a if she receives the signal A and for b if she receives the signal B is not a Nash equilibrium of $\Gamma(n)$.

Proof

Consider an individual i who receives the signal A. Her vote affects the outcome only if every other individual votes for b and hence only if every other individual receives the signal B. In this case, if she votes for a the outcome is a and if she votes for b it is b, so her gain from voting for a rather than b is $v_a + w_b$ if the state is α and $-v_b - w_a$ if the state is β. Thus her expected gain from voting for a rather than b is

$$\begin{aligned}&\Pr(\text{state } \alpha \text{ \& } n-1 \text{ other signals } B \mid i\text{'s signal } A)(v_a + w_b)\\ &\qquad - \Pr(\text{state } \beta \text{ \& } n-1 \text{ other signals } B \mid i\text{'s signal } A)(v_b + w_a).\end{aligned}$$

Now,

$$\begin{aligned}&\Pr(\text{state } \alpha \text{ \& } n-1 \text{ other signals } B \mid i\text{'s signal } A)\\ &\qquad = \frac{\Pr(\text{state } \alpha \text{ \& } n-1 \text{ other signals } B \text{ \& } i\text{'s signal } A)}{\Pr(i\text{'s signal } A)}\\ &\qquad = \frac{\Pr(n-1 \text{ other signals } B \text{ \& } i\text{'s signal } A \mid \text{state } \alpha)\Pr(\text{state } \alpha)}{\Pr(i\text{'s signal } A)}\\ &\qquad = \frac{(1-q)^{n-1}q\pi}{q\pi + (1-p)(1-\pi)},\end{aligned}$$

and similarly

$$\Pr(\text{state } \beta \text{ \& } n-1 \text{ other signals } B \mid i\text{'s signal } A) = \frac{p^{n-1}(1-q)(1-\pi)}{q\pi + (1-p)(1-\pi)}.$$

Hence i's expected payoff from voting for a is at least her expected payoff from voting for b if and only if

$$\frac{(1-p)(1-\pi)}{q\pi}\left(\frac{p}{1-q}\right)^{n-1} \leq \frac{v_a+w_b}{v_b+w_a}.$$

Now, $1-q<\frac{1}{2}<p$, so the left-hand side of this inequality increases without bound as n increases. So for any given values of v_a, v_b, w_a, and w_b, for n sufficiently large, type A of individual i prefers to vote for b than for a if every other individual votes for a if her signal is A and for b if her signal is B. A similar argument shows that the same is true for type B of individual i. Thus if n is sufficiently large then the strategy profile in which every individual votes for a if she receives the signal A and for b if she receives the signal B is not a Nash equilibrium of $\Gamma(n)$.

What *is* a Nash equilibrium of the game? If every individual votes for a regardless of her signal then the action of any individual has no effect on the outcome. Thus this strategy profile is a Nash equilibrium. Under some conditions, the strategy profile in which every individual votes for b regardless of her signal is also a Nash equilibrium.

Exercise 7.2: Nash equilibria of unanimity rule voting game

Find conditions under which a unanimity rule voting game with two alternatives and asymmetric information has a Nash equilibrium in which every individual votes for b regardless of her signal.

In addition, for some values of the parameters the game has a mixed strategy equilibrium in which type A of each individual votes for both a and b with positive probability, the same for each individual, and type B of each individual votes for b with probability 1.

Proposition 7.4: Symmetric mixed strategy equilibrium of unanimity rule voting game

Let $\langle\{a,b\},n,(\alpha,\beta),\{A,B\},\pi,(p,q),(v_a,v_b,w_a,w_b)\rangle$ be a unanimity rule voting game with two alternatives and asymmetric information. Let

$$\sigma^*(A)=\frac{p-(1-q)X(n)}{qX(n)-(1-p)} \qquad \text{where } X(n)=\left(\frac{v_a+w_b}{v_b+w_a}\frac{q}{1-p}\frac{\pi}{1-\pi}\right)^{1/(n-1)}.$$

If $0<\sigma^*(A)<1$ then the game has a mixed strategy equilibrium in which the strategy of type A of every individual votes for b with probability $\sigma^*(A)$ and the strategy of type B of every individual votes for b with probability 1. For some number $\hat{n}$, if $n > \hat{n}$ the game has no other equilibrium in which type A of every individual uses the same strategy, type B of every individual uses the same strategy, and at least one of these strategies assigns positive probabilities to both voting for a and voting for b.

Proof

Consider a mixed strategy equilibrium in which type A of every individual votes for b with the same probability and type B of every individual votes for b with the same probability. Denote these probabilities by $\sigma(A)$ and $\sigma(B)$.

An individual's vote affects the outcome only if all the other individuals vote for b, so the gain of an individual of type $T \in \{A, B\}$ from voting for a rather than b is

$$\begin{aligned}&\Pr(\text{state } \alpha \ \&\ n-1 \text{ other individuals vote } b \mid i\text{'s signal } T)(v_a + w_b)\\ &\quad - \Pr(\text{state } \beta \ \&\ n-1 \text{ other individuals vote } b \mid i\text{'s signal } T)(v_b + w_a).\end{aligned}$$

Using the logic in the proof of Proposition 7.3 to transform the probabilities, this gain is positive or negative according to the following inequality:

$$\frac{\Pr(n-1 \text{ others vote } b \ \&\ i\text{'s signal } T \mid \text{state } \beta)\Pr(\text{state } \beta)}{\Pr(n-1 \text{ others vote } b \ \&\ i\text{'s signal } T \mid \text{state } \alpha)\Pr(\text{state } \alpha)} \lesseqgtr \frac{v_a + w_b}{v_b + w_a}.$$

For $T = A$, the left-hand side of this inequality is

$$\frac{1-p}{q}\,\frac{1-\pi}{\pi}\left(\frac{(1-p)\sigma(A) + p\sigma(B)}{q\sigma(A) + (1-q)\sigma(B)}\right)^{n-1}, \tag{7.5}$$

and for $T = B$ it is

$$\frac{p}{1-q}\,\frac{1-\pi}{\pi}\left(\frac{(1-p)\sigma(A) + p\sigma(B)}{q\sigma(A) + (1-q)\sigma(B)}\right)^{n-1}. \tag{7.6}$$

Given that $p > \frac{1}{2}$ and $q > \frac{1}{2}$, the first of these expressions is smaller than the second. Thus if type A of an individual is indifferent between voting for a and voting for b then type B prefers to vote for b, and if type B of an individual is indifferent between voting for a and voting for b then type A prefers to vote for a. Hence in an equilibrium in which $0 < \sigma(B) < 1$, every individual of type A votes for a, so that $\sigma(A)=0$, and in an equilibrium in which $0<\sigma(A)<1$, every individual of type B votes for b, so that $\sigma(B)=1$.

If $\sigma(A)=0$ then (7.6) is $p/(1-q)\cdot(1-\pi)/\pi\cdot(p/(1-q))^{n-1}$, which increases without bound as n increases, so that for any values of v_a, v_b, w_a, and w_b, type B prefers to vote for b when n is sufficiently large. Thus for sufficiently large values of n no equilibrium exists with $0<\sigma(B)<1$.

If $\sigma(B)=1$ then the equality of (7.5) with $(v_a+w_b)/(v_b+w_a)$ implies that $\sigma(A)=\sigma^*(A)$.

The condition $0<\sigma^*(A)<1$ is equivalent to $(1-p)/q<X(n)<p/(1-q)$ and $X(n)>1$, and the second of these conditions is equivalent to $q\pi(v_a+w_b)>(1-p)(1-\pi)(v_b+w_a)$. If this last condition is not satisfied, the game has an equilibrium in which all individuals vote for b regardless of their signals, as you know if you have done Exercise 7.2.

As n increases without bound, $X(n)$ approaches 1, so that $\sigma^*(A)$ approaches 1. Thus for large values of n, type A of each individual votes for b with high probability (and type B votes for b with probability 1). For $\pi=\frac{1}{2}$, $v_a=v_b=0$, $w_a=1-w_b$, and $p=q$, Feddersen and Pesendorfer (1998) calculate for this equilibrium the limits of the probabilities that a is selected in state β and b is selected in state α as the number of individuals increases without bound, and show that these limits are positive. Thus for this example, even in a large population, in each state the wrong alternative is selected with positive probability.

Note that this result considers only mixed strategy equilibria in which every individual's strategy is the same. The game may also have mixed strategy equilibria in which the individuals' strategies differ.

Finally, my comments about the difficulty of interpreting Nash equilibria of voting games at the end of Section 3.2 apply with equal if not more force to mixed strategy equilibria. Individuals do not typically engage repeatedly in similar voting games, so that the steady state interpretation of equilibrium does not fit such games well, and good interpretations of mixed strategy equilibria of one-off games are lacking. See Section 3.2 of Osborne and Rubinstein (1994) for an extended discussion of interpretations of mixed strategy equilibrium.

Notes

Sections 7.1.1 and 7.1.2 are based on Feddersen and Pesendorfer (1996). The notion of equilibrium is a variant of the one in Osborne and Turner (2010), which differs from the one used by Feddersen and Pesendorfer but retains the same spirit. (Feddersen and Pesendorfer's results are asymptotic in the number of individuals, in contrast to the results here, which hold for any number of individuals.) The proof of Proposition 7.1 is taken (in parts verbatim) from Osborne and Turner (2010, 182–184). Section 7.1.3 is based on McMurray (2013), who studies a model in which the number of individuals is random.

Section 7.2 is based on Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1998). This work presents unanimity rule as the decision-making process for juries in some jurisdictions, with the default outcome acquittal. I do not give the model this interpretation because juries that use a version of unanimity rule treat conviction and acquittal symmetrically, with the outcome a retrial if unanimity is not achieved. In addition, deliberation appears to be an essential feature of a jury's decision-making process.

Solutions to exercises

Exercise 7.1

I first argue that the strategy profile is an equilibrium in the sense of Definition 7.3. A change in any uninformed individual's strategy or the action specified by any informed individual's strategy in state β changes the outcome in state β to a, making her worse off. Now consider the effect of a change in the action specified an informed individual i's strategy in state α. If there are no other informed individuals, the outcome changes to b in that state, making i worse off. If there are other informed individuals, the outcome does not change. In this case, the smallest number of individuals whose failure to vote causes the outcome to change is the number of other informed individuals; their failure to vote would cause the change in i's strategy to change the outcome from a to b, making her worse off. Thus no change in any individual's strategy is desirable.

I now argue that no other strategy profile is an equilibrium.

Informed individual

Consider an informed individual who is voting for b in state α. Suppose she switches to voting for a. If none of the other individuals are voting for a, the outcome improves from b to a. If some of the other individuals are voting for a then the smallest number of individuals whose failure to vote affects the outcome is the number of such individuals, and their failure to vote would mean that the change in the informed individual's action improves the outcome from b to a. Thus a change in the individual's action from b to a in state α is desirable.

By a similar argument, the change to vote for b for an informed individual voting for a in state β is desirable.

We conclude that in every equilibrium every informed individual votes for a in state α and for b in state β.

Uninformed individual

If an uninformed individual votes for a, the outcome is a in state β, and if she switches to vote for b the outcome does not get worse in either state, regardless of how many individuals (if any) fail to vote (given that the informed individuals vote for a in state α) and improves in state β if either no other individual votes for a or all the individuals planning to vote for a fail to vote.

Exercise 7.2

Suppose that every individual other than i votes for b independently of her signal. Then i's vote determines the outcome (and the other individuals' votes convey no information about the state).

Suppose that i's signal is A. Then her expected payoff if she votes for a is

$$\Pr(\alpha \mid \text{signal } A)v_a - \Pr(\beta \mid \text{signal } A)w_a = \frac{q\pi v_a - (1-p)(1-\pi)w_a}{q\pi + (1-p)(1-\pi)}$$

(using Bayes' rule) and her expected payoff if she votes for b is

$$\Pr(\beta \mid \text{signal } A)v_b - \Pr(\alpha \mid \text{signal } A)w_b = \frac{(1-p)(1-\pi)v_b - q\pi w_b}{q\pi + (1-p)(1-\pi)},$$

so that she optimally votes for b if and only if $q\pi v_a - (1-p)(1-\pi)w_a \leq (1-p)(1-\pi)v_b - q\pi w_b$, or $q\pi(v_a + w_b) \leq (1-p)(1-\pi)(v_b + w_a)$.

Now suppose that i's signal is B. Then her expected payoff if she votes for a is

$$\Pr(\alpha \mid \text{signal } B)v_a - \Pr(\beta \mid \text{signal } B)w_a = \frac{(1-q)\pi v_a - p(1-\pi)w_a}{(1-q)\pi + p(1-\pi)}$$

and her expected payoff if she votes for b is

$$\Pr(\beta \mid \text{signal } B)v_b - \Pr(\alpha \mid \text{signal } B)w_b = \frac{p(1-\pi)v_b - (1-q)\pi w_b}{(1-q)\pi + p(1-\pi)},$$

so that she optimally votes for b if and only if $(1-q)\pi v_a - p(1-\pi)w_a \leq p(1-\pi)v_b - (1-q)\pi w_b$ or $(1-q)\pi(v_a + w_b) \leq p(1-\pi)(v_b + w_a)$.

Given that $p > \frac{1}{2}$ and $q > \frac{1}{2}$, the second inequality is satisfied whenever the first inequality is satisfied, so that i's voting for b regardless of her signal is optimal if and only if the first inequality is satisfied.

Thus the game has a Nash equilibrium in which every individual votes for b regardless of her signal if and only if $q\pi(v_a + w_b) \leq (1 - p)(1 - \pi)(v_b + w_a)$.

III Electoral competition

8 Electoral competition: two office-motivated candidates 233
9 Electoral competition: two policy-motivated candidates 287
10 Electoral competition: endogenous candidates 327
11 Distributive politics 369
12 Money in electoral competition 407
13 Two-period electoral competition with imperfect information 433

8 Electoral competition: two office-motivated candidates

8.1 General model 237
8.2 Spatial model 243
8.3 Candidates uncertain of citizens' preferences 248
8.4 Costly voting 256
8.5 Citizens with preferences over candidates 265
8.6 Electing a legislature 272
8.7 Interpreting Nash equilibrium 278

Collective decisions in some societies are made by legislatures consisting of relatively small groups of individuals. In a common procedure, some individuals are candidates for membership of the legislature and the members of a larger subset of the society, whom I call citizens, cast votes to determine which candidates are elected. Most of this chapter is devoted to models of a simple version of such a legislature, with a single member.

To specify a model, we have many options. For example, the set of candidates may be exogenous, or individuals may choose whether to be candidates; the citizens may or may not know the candidates' preferences; each candidate may be able to commit to act in the legislature according to given preferences, which may differ from her own, or may be unable to make such a commitment; the candidates may or may not know the citizens' preferences; the candidates may be motivated by the desire to win election, or they may have other motivations—for example, they may care about the policy ultimately chosen by the legislature; each candidate may make a decision not knowing the decisions of the remaining candidates, or the candidates may make their decisions sequentially, with each candidate observing the decisions of her predecessors.

When discussing models of electoral competition, I refer to an alternative as a *position*. A motivation for using this nomenclature is that legislatures decide multiple issues, some of which may be unknown at the time of an election. Candidates for legislative office may state principles that will guide their behavior if elected—political positions—rather than specifying the alternatives they will select if elected. However, formally the set of positions is the same as the set of alternatives in the earlier chapters.

In this chapter I first present a model that makes the following assumptions.

- The set of candidates
    - is given exogenously, distinct from the set of citizens
    - has two members.
- Each candidate
    - if elected, becomes the sole decision-maker (the legislature has a single member)
    - chooses a position, which she is committed to implement if she is elected
    - is motivated by the desire to win election
    - knows the citizens' preferences.
- Each citizen
    - cares about the position ultimately implemented
    - votes ("sincerely") for the candidate whose position she prefers.
- The candidate who receives the most votes is elected.

Subsequently I present variants of this model that retain the assumption of two exogenously given candidates whose aim is to win election. In the next chapter I consider models in which the candidates care about the policy ultimately implemented, rather than caring exclusively about winning election, and in Chapter 10 I present models in which individuals decide whether to become candidates. In Chapter 12 I present a model in which the candidates can affect their chances of winning by spending money on a campaign.

The analysis in this chapter and the next is restricted to two candidates not because most elections, or even most elections in which plurality rule determines the winner, involve two candidates, but because the analysis of models with many candidates differs significantly from the analysis of models with two candidates.

In some of the models I discuss, the outcome of an election is uncertain. In these models, a candidate's aim to win election is operationalized by the assumption that she aims to maximize the probability that she wins. Two alternatives to this assumption are that each candidate aims to maximize her expected vote share and that she aims to maximize her expected plurality. I do not present models that make either of these alternative assumptions mainly because they are generally inconsistent with the maximization of the probability of winning. Specifically, for some distributions α and β over electoral outcomes for which a candidate's expected vote share or expected plurality under α is greater than it is under β, the candidate's probability of winning under α is less than it is under β.

The originators of some of the models I present interpret the candidates as parties; because they are single decision-makers, I generally stick with the term "candidates".

Synopsis

In Section 8.1 I present a model with a finite number of citizens in which two candidates simultaneously choose positions from an arbitrary set and the candidate whose position is favored by more citizens wins. Proposition 8.1 shows that in any Nash equilibrium, both candidates choose a Condorcet winner of the underlying collective choice problem, so that in particular the game has a Nash equilibrium only if the collective choice problem has a Condorcet winner. An implication of this result combined with Propositions 1.4 and 1.5 is that if the citizens' preferences are single-peaked or single-crossing and the number of citizens is odd, then the game has a unique Nash equilibrium, and in this equilibrium each candidate's position is the median with respect to the ordering of the alternatives of the citizens' favorite positions (Corollary 8.2). In particular, the candidates are driven to choose the same position, a feature of the equilibria of most of the models in this chapter.

For the variant of this model in which the candidates move sequentially, Proposition 8.3 shows that for a collective choice problem that has a Condorcet winner, the outcome of a subgame perfect equilibrium is the same as the outcome of a Nash equilibrium of the simultaneous-move game, and for a collective choice problem without a Condorcet winner, the outcome of every subgame perfect equilibrium is that the second-mover wins.

Section 8.2 studies another variant of the model, in which the set of citizens is a continuum and the set of alternatives is an interval of real numbers. This variant is used as a component of several of the models in subsequent chapters. Proposition 8.4 shows that, as for the case in which the number of citizens is finite, the game has a unique Nash equilibrium, and in this equilibrium the position of each candidate is the median of the citizens' favorite positions.

In the models considered so far, the candidates know the citizens' preferences. Section 8.3 considers two models in which the candidates are uncertain of these preferences. In the first model, the candidates share a common belief about the distribution of the median of the citizens' favorite positions. The resulting game has a unique Nash equilibrium, in which the candidates' common position is the median of this distribution (Proposition 8.5). In the second model, each candidate gets a private signal about the median of the citizens' favorite positions. Proposition 8.6 characterizes any Nash equilibria that exist (there may be none).

An assumption common to the models in Sections 8.1 through 8.3 is that all citizens vote, even when the candidates' positions are the same, in which case no citizen's vote can possibly affect the outcome. Section 8.4 analyzes a model in which casting a vote is costly and abstention is an option. For simplicity, the model assumes there is one citizen, whose voting cost is known to her but not to the candidates. Her preferences over positions are also unknown to the candidates. For any pair of positions for the candidates, the citizen votes if her voting cost is less than a cutoff that depends on the difference between her payoffs for the positions of the two candidates. As a consequence, if one candidate's position becomes closer to the other candidate's position then the change in her probability of winning depends on the nature of the citizen's preferences and the distribution of her voting cost. Proposition 8.7 shows than in an equilibrium, the candidates' positions are the same, and characterizes the common equilibrium position. Given that the candidates' positions are the same in an equilibrium, the citizen votes only if her voting cost is zero, an event with probability zero.

Section 8.5 studies models in which the citizens have preferences over the candidates independently of the candidates' positions. In the main model, each candidate is uncertain of these preferences and thus is uncertain whether any given position will lead her to win, given the other candidate's position. Proposition 8.8 gives conditions under which in any equilibrium the candidates's positions are the same, and characterizes the common position. An example shows that the result is not vacuous—there are games for which an equilibrium exists—but no general result on the existence of an equilibrium is available.

Section 8.6 discusses models of legislatures with many members, each of whom is elected in a single district. Order the districts by the median of the favorite positions of the citizens in the district. Suppose that each candidate is associated with one of two parties and decisions in the legislature are made by the party whose candidates win the most districts. Then a model in which each party chooses a single position for all of its candidates has a unique Nash equilibrium, in which the position of each party is the median of the favorite positions of the citizens in the median district. The same is true for a model in which each candidate chooses her position independently and the position of a party is the average of its candidates' positions. However, if the first model is modified by assuming that the citizens' partisanships are uncertain, along the lines of the model in Section 8.5, and each party values winning an additional district even if doing so does not change is minority/majority status in the legislature, then equilibria in which the parties' positions are distinct are possible.

8.1 General model

8.1.1 Simultaneous decisions

The set of possible positions is an arbitrary set. Each of two candidates chooses a position not knowing the other candidate's position and then each of a finite set of citizens casts a vote for the candidate whose position she prefers. The electoral mechanism selects as the winner the candidate with the most votes. Each candidate cares only about whether she wins, preferring to win than to tie than to lose. One way to characterize these preferences is to say that each candidate is office-motivated, as opposed to policy-motivated.

I assume that if the candidates' positions are x_1 and x_2 and the number of citizens who prefer x_1 to x_2 is the same as the number who prefer x_2 to x_1 then the outcome of the election is a tie. This assumption is consistent with each citizen who is indifferent between the candidates' positions not voting, or splitting her vote, casting half a vote for each candidate. (The related assumption that each such citizen votes with equal probability for each candidate requires a formulation of the candidates' preferences regarding lotteries over outcomes.)

Definition 8.1: Electoral competition game with two office-motivated candidates

An *electoral competition game with two office-motivated candidates* $\langle\{1,2\},\langle I,X,\succsim\rangle\rangle$, where $\langle I,X,\succsim\rangle$ is a collective choice problem in which the set I (of *citizens*) is finite, is the strategic game with the following components.

Players
$\{1,2\}$ (*candidates*).

Actions
The set of actions of each player is X (the set of possible *positions*).

Preferences
For each $(x_1,x_2) \in X \times X$, denote by $O(x_1,x_2)$ be the electoral outcome when each citizen (member of I) votes for the position in $\{x_1,x_2\}$ that

she prefers and the position that receives the most votes wins:

$$O(x_1,x_2)=\begin{cases} \textit{win for } 1 & \text{if } |\{i\in I : x_1 \succ_i x_2\}| > |\{i\in I : x_2 \succ_i x_1\}| \\ \textit{tie} & \text{if } |\{i\in I : x_1 \succ_i x_2\}| = |\{i\in I : x_2 \succ_i x_1\}| \\ \textit{win for } 2 & \text{if } |\{i\in I : x_1 \succ_i x_2\}| > |\{i\in I : x_2 \succ_i x_1\}|. \end{cases}$$

The preference relation $\trianglerighteq_j$ of each player j over pairs of positions satisfies

$$(w_1,w_2) \triangleright_j (y_1,y_2) \triangleright_j (z_1,z_2)$$

whenever $O(w_1,w_2) = \textit{win for } j$, $O(y_1,y_2) = \textit{tie}$, and $O(z_1,z_2) = \textit{win for } k$, where k is the other player.

The Nash equilibria of such a game and the Condorcet winners of the associated collective choice problem are closely related. In particular, if the collective choice problem has a Condorcet winner then in any Nash equilibrium of the game each candidate's position is a Condorcet winner and the outcome is a tie.

Proposition 8.1: Nash equilibrium of electoral competition game with two office-motivated candidates and Condorcet winner

Let $\langle\{1,2\},\langle I,X,\succsim\rangle\rangle$ be an electoral competition game with two office-motivated candidates. The outcome of any Nash equilibrium of this game is a tie, and (x_1,x_2) is a Nash equilibrium if and only if both x_1 and x_2 are Condorcet winners of $\langle I,X,\succsim\rangle$, so that in particular the game has a Nash equilibrium if and only if $\langle I,X,\succsim\rangle$ has a Condorcet winner.

Proof

Suppose that (x_1,x_2) is a Nash equilibrium of the game. If $x_1 = x_2$, the outcome $O(x_1,x_2)$ is a tie. If $x_1 \neq x_2$ and either candidate deviates to the position of the other candidate, the outcome becomes a tie, so that $O(x_1,x_2)$ is at least as good as a tie for each candidate, and hence also is a tie.

Thus a pair (x_1,x_2) of positions is a Nash equilibrium if and only if

$$\begin{aligned}&\text{for all } x_1' \in X \text{ the outcome } O(x_1',x_2) \text{ is a tie or a loss for } 1\\ &\text{for all } x_2' \in X \text{ the outcome } O(x_1,x_2') \text{ is a tie or a loss for } 2,\end{aligned}$$

conditions that are satisfied if and only if x_1 and x_2 are Condorcet winners of the collective choice problem $\langle I,X,\succsim\rangle$.

If an alternative is a strict Condorcet winner then it is the only Condorcet winner, so if the collective choice problem has a strict Condorcet winner then in any Nash equilibrium each candidate chooses that position.

Corollary 8.1: Unique Nash equilibrium of electoral competition game with two office-motivated candidates and strict Condorcet winner

Let $\langle\{1,2\},\langle I,X,\succsim\rangle\rangle$ be an electoral competition game with two office-motivated candidates. If $\langle I,X,\succsim\rangle$ has a strict Condorcet winner x^* then (x^*,x^*) is the unique Nash equilibrium of the game.

If the number of citizens is odd and their preferences are single-peaked or single-crossing then the collective choice problem has a strict Condorcet winner. This position is the median of the citizens' favorite positions if the citizens' preferences are single-peaked (Proposition 1.4), and the favorite position of the median citizen if the citizens' preferences are single-crossing (Proposition 1.5), so the next result follows from Corollary 8.1.

Corollary 8.2: Median voter theorem for electoral competition game with two office-motivated candidates

Let $\langle\{1,2\},\langle I,X,\succsim\rangle\rangle$ be an electoral competition game with two office-motivated candidates in which the number of citizens (members of I) is odd.

- If $\langle I,X,\succsim\rangle$ has single-peaked preferences with respect to a linear order $\trianglerighteq$ on X, then the game has a unique Nash equilibrium, and in this equilibrium each candidate's position is the median with respect to $\trianglerighteq$ of the citizens' favorite positions.

- If $\langle I,X,\succsim\rangle$ has single-crossing preferences with respect to a linear order $\geq$ on I and the median individual with respect to $\geq$ has a unique favorite position, say x^*, then the game has a unique Nash equilibrium, and in this equilibrium each candidate's position is x^*.

A strict Condorcet winner is more than a Nash equilibrium action: it weakly dominates all other actions.

Proposition 8.2: Dominant actions in electoral competition game with two office-motivated candidates and strict Condorcet winner

Let $\langle\{1,2\},\langle I,X,\succsim\rangle\rangle$ be an electoral competition game with two office-motivated candidates. If $\langle I,X,\succsim\rangle$ has a strict Condorcet winner x^* then for each candidate the action x^* weakly dominates all other actions in the game.

Proof

Suppose that candidate i chooses x^*. If the other candidate chooses x^*, the outcome is a tie, and if the other candidate chooses any other action, i wins. Now suppose that i chooses an action other than x^*. Then if the other candidate chooses x^*, i loses. No outcome is better for i than her winning, so x^* weakly dominates all of her other actions.

		other candidate x^*	other candidate $\neq x^*$
candidate i	x^*	tie	i wins
	$\neq x^*$	i loses	?

This result is significant because it means that a candidate's choosing the strict Condorcet winner is optimal for her regardless of the other candidate's action. The fact that (x^*,x^*) is a Nash equilibrium means that a candidate's choosing x^* is optimal for her if she believes that the other candidate will choose x^*; the fact that x^* weakly dominates all other actions means that her choosing it is optimal for her regardless of her belief about the other candidate's action. If the other candidate chooses an action different from x^*, the action x^* remains optimal for i even though other actions may also be optimal for her in that case.

8.1.2 Sequential decisions

If the candidates choose alternatives sequentially, we can model their interaction as an extensive game. In the following game, candidate 1 chooses an alternative, candidate 2 observes this alternative, and then candidate 2 chooses an alternative.

Definition 8.2: Sequential electoral competition game with two office-motivated candidates

A *sequential electoral competition game with two office-motivated candidates* $\langle\{1,2\},\langle I,X,\succsim\rangle\rangle$, where $\langle I,X,\succsim\rangle$ is a collective choice problem in which the set I (of *citizens*) is finite, is the extensive game with perfect information with the following components.

Players
$\{1,2\}$ (*candidates*).

Terminal histories
The set of sequences (x_1,x_2) where $x_i \in X$ for each $i \in \{1,2\}$.

Player function
The function P given by $P(\varnothing)=1$ and $P(x_1)=2$ for all $x_1 \in X$.

Preferences
The preference relation of each player over terminal histories (pairs of positions) satisfies the conditions in Definition 8.1.

The set of electoral outcomes in a sequential electoral competition game with two office-motivated candidates is finite (*win for candidate* 1, *tie*, *win for candidate* 2), so every such game has a subgame perfect equilibrium. If the underlying collective choice problem has a strict Condorcet winner, then the outcome of every subgame perfect equilibrium is that both candidates choose this alternative. The following result gives the subgame perfect equilibrium outcomes also for problems that do not have strict Condorcet winners.

Proposition 8.3: Subgame perfect equilibrium of sequential electoral competition game with two office-motivated candidates and Condorcet winner

Let $G=\langle\{1,2\},\langle I,X,\succsim\rangle\rangle$ be a sequential electoral competition game with two office-motivated candidates.

a. If $\langle I,X,\succsim\rangle$ has no Condorcet winner then a pair (x_1,x_2) is the outcome of a subgame perfect equilibrium of G if and only if the electoral outcome of (x_1,x_2) is a win for candidate 2.

b. If $\langle I,X,\succsim\rangle$ has a Condorcet winner then a pair (x_1,x_2) is the outcome of a subgame perfect equilibrium of G if and only if x_1 is a Condorcet winner of $\langle I,X,\succsim\rangle$ and the electoral outcome of (x_1,x_2) is a tie.

c. If $\langle I, X, \succsim \rangle$ has a strict Condorcet winner x^* then the outcome of every subgame perfect equilibrium of G is (x^*, x^*).

Proof

a. The result follows from the observation that because $\langle I, X, \succsim \rangle$ has no Condorcet winner, for every alternative $x \in X$ there is an alternative $y \in X$ such that the electoral outcome $O(x, y)$ is a win for candidate 2.

b. First suppose that (x_1, x_2) is the outcome of a subgame perfect equilibrium of G. For any alternative chosen by candidate 1, the electoral outcome is a tie if candidate 2 chooses the same alternative, so the electoral outcome of (x_1, x_2) is either a tie or a win for candidate 2. If candidate 1 chooses a Condorcet winner of $\langle I, X, \succsim \rangle$, then for every alternative chosen by candidate 2 the electoral outcome is either a tie or a win for candidate 1. Thus the electoral outcome of (x_1, x_2) is a tie. Hence no alternative beats x_1, so that x_1 is a Condorcet winner of $\langle I, X, \succsim \rangle$.

Conversely, suppose that x_1 is a Condorcet winner of $\langle I, X, \succsim \rangle$ and the electoral outcome of (x_1, x_2) is a tie. Let (s_1^*, s_2^*) be the strategy pair in which $s_1^* = x_1$ and

$$s_2^*(z_1) = \begin{cases} x_2 & \text{if } z_1 = x_1 \\ z_1 & \text{if } z_1 \neq x_1 \text{ and } z_1 \text{ is a Condorcet winner} \\ y_2(z_1) & \text{if } z_1 \text{ is not a Condorcet winner} \end{cases}$$

where for every alternative z_1 that is not a Condorcet winner, $y_2(z_1)$ is an alternative that beats z_1. I argue that (s_1^*, s_2^*) is a subgame perfect equilibrium of G. The outcome of (s_1^*, s_2^*) is (x_1, x_2); by assumption $O(x_1, x_2)$ a tie. If candidate 1 deviates to another alternative z_1, then $O(z_1, s_2^*(z_1))$ is a tie if z_1 is a Condorcet winner and is a win for candidate 2 otherwise. Thus s_1^* is optimal for candidate 1 given s_2^*. The action prescribed by candidate 2's strategy s_2^* after each history is optimal because if z_1 is a Condorcet winner then candidate 2 can do no better than tie and if z_1 is not a Condorcet winner then the electoral outcome of $(z_1, s_2^*(z_1))$ is a win for candidate 2.

c. This result follows from *b* because if x_1 is the strict Condorcet winner of $\langle I, X, \succsim \rangle$ then the only alternative x_2 for which the electoral outcome of (x_1, x_2) is a tie is $x_2 = x_1$.

The answer to Exercise 1.5 shows that in case (*b*), the alternative x_2 chosen by candidate 2 is not necessarily a Condorcet winner of the collective choice

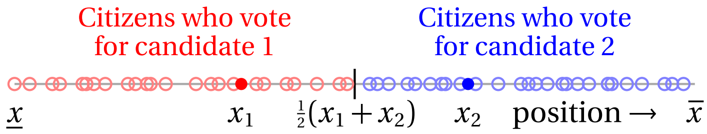


Figure 8.1 Two-candidate electoral competition in which the set of positions is an interval and each citizen's preference relation is symmetric about her favorite position. The favorite position of each citizen is indicated with a small circle, and the positions of the candidates are indicated with small disks.

problem.

8.2 Spatial model

8.2.1 One-dimensional set of positions

Suppose that the set X of positions is an interval of numbers, the number of citizens is odd, and the preference relation $\succsim_i$ of each citizen i is single-peaked with respect to $\geq$ (that is, if $x < y < x_i^*$ or $x_i^* < y < x$ then $x_i^* \succ_i y \succ_i x$, where x_i^* is i's favorite position). Then by Corollary 8.2 the electoral competition game with two office-motivated candidates has a unique Nash equilibrium, and in this equilibrium each candidate's position is the median of the citizens' favorite positions (with respect to $\geq$).

Now further assume that the preference relation of every citizen i is symmetric in the sense that $x_i^* - \delta \sim_i x_i^* + \delta$ for every $\delta > 0$. Then for any positions x_1 and x_2 for the candidates with $x_1 < x_2$, any citizen i votes for candidate 1 if $x_i^* < \frac{1}{2}(x_1+x_2)$ and for candidate 2 if $x_i^* > \frac{1}{2}(x_1+x_2)$. The division of votes between the candidates is illustrated in Figure 8.1, where the horizontal line represents the interval of positions and each citizen is identified with her favorite position. As x_1 increases, the dividing line $\frac{1}{2}(x_1+x_2)$ increases, so that the number of votes for candidate 1 increases and the number for candidate 2 falls.

In a variant of this model the set of citizens is a continuum, rather than being finite, the distribution of the citizens' favorite positions has a density, and the support of this distribution is an interval, so that the distribution has a unique median.

Definition 8.3: Median of distribution

Let X be an interval of real numbers and let $F : X \to [0,1]$ be a distribution function (a nondecreasing function with $F(\underline{x}) = 0$ for some $\underline{x} \in X$ and $F(\overline{x}) = 1$ for some $\overline{x} \in X$). A *median* of F is a number x such that $F(x) = \frac{1}{2}$.

This variant may be analyzed with a diagram like Figure 8.2. For any position

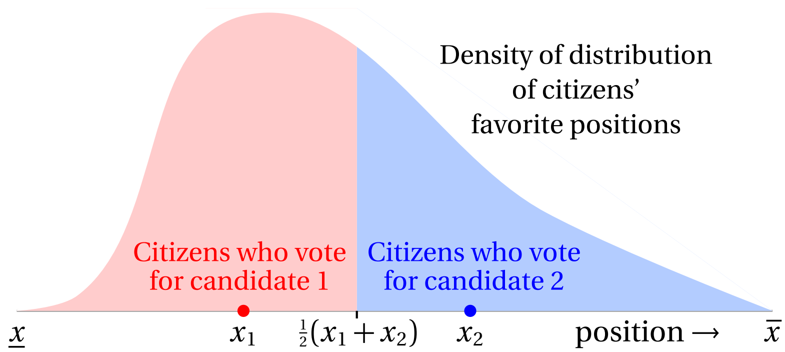


Figure 8.2 Two-candidate electoral competition in which the set of positions is the interval $[\underline{x},\overline{x}]$ and there is a continuum of citizens.

x, the height of the curve represents the density of citizens with favorite position x; the area under the curve between any two positions x and y represents the fraction of citizens with favorite positions between x and y. The fraction of citizens who prefer x_1 to x_2 is thus the area shaded pink and the fraction who prefer x_2 to x_1 is the area shaded blue.

Although the symmetry assumption on the citizens' preferences that this variant of the model entails is stronger than the assumptions of the model with finitely many citizens, the graphical analysis that it permits is appealing. Also, several models analyzed in later chapters take the variant as a starting point. For these reasons I give a precise definition of the variant.

In the definition, unlike in Definition 8.1, citizens do not appear explicitly. Instead, the electoral outcome for any pair (x_1,x_2) of the candidates' positions and any nonatomic distribution F of the citizens' favorite positions with a unique median is assumed to be

$$O_F(x_1,x_2) = \begin{cases} \textit{tie} & \text{if } x_1 = x_2 \text{ or } \frac{1}{2}(x_1+x_2) = \text{med}(F) \\ \textit{win for } j & \text{if } \begin{cases} \text{either } x_k < x_j \text{ and } \frac{1}{2}(x_1+x_2) < \text{med}(F) \\ \text{or } x_k > x_j \text{ and } \frac{1}{2}(x_1+x_2) > \text{med}(F), \end{cases} \end{cases} \tag{8.1}$$

where k is the player other than j and med(F) denotes the median of F. One rationale for this assumption is that each citizen's preference relation over positions is single-peaked and symmetric about her favorite position. Note that a compact way to characterize O_F is that the winner is the candidate favored by the median voter.

Definition 8.4: Electoral competition game with a continuum of citizens and two office-motivated candidates

An *electoral competition game with a continuum of citizens and two office-motivated candidates* $\langle\{1,2\}, X, F\rangle$, where X is a closed interval of real numbers and F is a nonatomic distribution with support X (and hence a unique median), is the strategic game with the following components.

Players
$\{1,2\}$ (*candidates*).

Actions
The set of actions of each player is X (the set of possible *positions*).

Preferences
The preference relation $\trianglerighteq_j$ of each player j over $X \times X$ satisfies

$$(w_1, w_2) \triangleright_j (y_1, y_2) \triangleright_j (z_1, z_2)$$

whenever $O_F(w_1, w_2) = \textit{win for } j$, $O_F(y_1, y_2) = \textit{tie}$, and $O_F(z_1, z_2) = \textit{win for } k$, where k is the other player and O_F is given by (8.1).

Any such game has a unique Nash equilibrium, in which each candidate's position is the median of F. Further, as for a two-candidate electoral competition game for an arbitrary collective choice problem, the action of each candidate in a Nash equilibrium weakly dominates all her other actions. These results do not follow from Corollary 8.2, because that result assumes a finite number of citizens, but the arguments are straightforward.

Proposition 8.4: Nash equilibrium of electoral competition game with continuum of citizens and two office-motivated candidates

Every electoral competition game with a continuum of citizens and two office-motivated candidates $\langle\{1,2\}, X, F\rangle$ has a unique Nash equilibrium, in which each candidate's position is the median of F, and for each candidate this action weakly dominates all her other actions.

Proof

Denote the median of F by m. We have $O_F(m, m) = \textit{tie}$, and if either candidate deviates from m she loses, so (m, m) is a Nash equilibrium.

Now let $(x_1, x_2) \neq (m, m)$. If $\frac{1}{2}(x_1 + x_2) = m$ then $O_F(x_1, x_2) = \textit{tie}$, and either candidate can win by deviating to m. Otherwise one of the candidates loses, and by moving to m she at least ties. Thus (x_1, x_2) is not a Nash equilibrium.

If candidate i's position is m, the outcome is a tie if the other candidate's position is m and a win for i otherwise. If i's position differs from m, she loses if the other candidate's position is m. Thus i's action m weakly dominates all her other actions.

Note that this result does not depend on the shape of the distribution F of the citizens' favorite positions. This distribution can be concentrated or dispersed, unimodal or multimodal, symmetric or skewed; in every case the action of choosing the median of the distribution weakly dominates all other actions, and the action pair in which both candidates choose the median is the only Nash equilibrium.

A notable feature of the unique Nash equilibrium is that both candidates choose the same position. Harold Hotelling, who in 1929 suggested that the model captures competition between parties, asserted that in the US at the time the convergence of positions was "strikingly exemplified" (Hotelling 1929, 54). Some observers of US politics claim that the exemplification of convergence is now less striking.

Exercise 8.1: Electoral competition with alienation

Consider a model that differs from an electoral competition game with a continuum of citizens and two office-motivated candidates only in that citizens whose favorite positions are more than some distance k from both candidates' positions do not vote. (Perhaps citizens' motivations for voting are expressive (Section 6.2), with alienation setting in when both candidates' positions are remote.) Characterize the Nash equilibria of the game when the distribution of the citizens' favorite positions is unimodal, with a differentiable density. Give an example in which the distribution of the citizens' favorite positions is not unimodal and a Nash equilibrium exists in which the candidates' positions differ.

The fact that a candidate's choosing m weakly dominates all her other actions means that regardless of her beliefs about the other candidates, m is an optimal action for her. However, it does not mean that choosing a position different from m is unambiguously irrational. If, for example, one candidate believes that the other will certainly choose some given position $x > m$, then any position closer to m than x is optimal for her. But should a candidate not believe that the other candidate is rational? And also that the other candidate believes she is rational?

And that the other candidate believes she believes that the other candidate is rational? And so forth .... For an electoral competition game with a finite set of possible positions, the next exercise asks you to show that the implication of this set of hypotheses is that a candidate chooses m.

Exercise 8.2: Rationalizable actions in two-candidate electoral competition

Consider a variant of an electoral competition game with a continuum of citizens and two office-motivated candidates in which the set of possible positions is finite: $\{z_1, \ldots, z_k\}$ with $z_1 < z_2 < \cdots < z_k$. Each citizen's favorite position is a member of the set and each candidate is restricted to choose a member of the set. Call a candidate's action *rational* if it is a best response to *some* belief about the other candidate's actions, or equivalently if it is not strictly dominated. (Take this equivalence as given; the argument for it is not simple.) Find a candidate's rational actions, her rational actions if she assumes that the other candidate is rational, her rational actions if she assumes that the other candidate assumes that she is rational, and so forth. An action that is rational under the union of these assumptions is *rationalizable*. Show that the only rationalizable action is the median of the citizens' favorite positions.

Just as an electoral competition game with a continuum of citizens and two office-motivated candidates is a variant of an electoral competition game with two office-motivated candidates, so we can define a variant with a continuum of citizens of a sequential electoral competition game with two office-motivated candidates. I postpone doing so until Section 10.3, where I study a version of the game with many office-seekers, each of whom has the option to become a candidate.

8.2.2 Two-dimensional positions

Proposition 8.1 shows that in a Nash equilibrium of an electoral competition game with two office-motivated candidates, each candidate's position is a Condorcet winner of the underlying collective choice problem. When the set of alternatives is two-dimensional, what do we know about the set of Condorcet winners? Section 1.6 shows that the character of the Condorcet winners depends on the nature of the individuals' preferences. Proposition 1.6 says that when the individuals have city block preferences, each component of a Condorcet winner is the median of the individuals' favorite values of that component, and Proposition 1.7 says that when the individuals have max preferences, a Condorcet win-

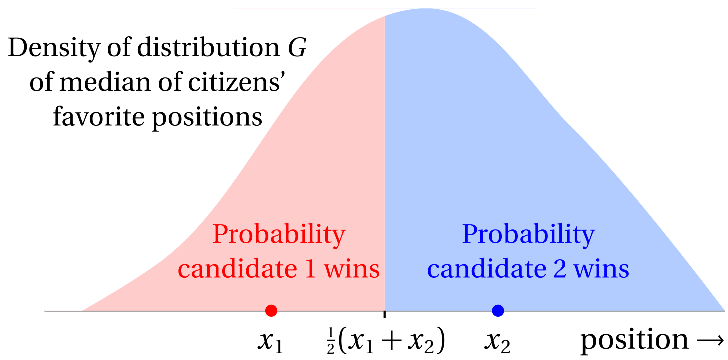


Figure 8.3 Two-candidate electoral competition in which median of the citizens' favorite positions is uncertain.

ner is a component-wise median of a 45 degree rotation of the individuals' favorite positions. For the case in which the individuals have Euclidean preferences, Proposition 1.8 says that a Condorcet winner exists only if the individuals' favorite positions possess a specific symmetry: for some position x, half of these favorite positions lie on each side of x, in which case x is a Condorcet winner.

8.3 Candidates uncertain of citizens' preferences

In the models of electoral competition discussed so far, the candidates know the distribution of the citizens' preferences. I now present two models in which they are uncertain about this distribution.

8.3.1 Common information

The analysis in Section 8.2.1 shows that when the set of positions is an interval of real numbers, the key feature of the citizens' preferences is the median of their favorite positions. In the model I now present, the candidates' uncertainty about the citizens' preferences directly concerns this median. The candidates are assumed to share the belief that this median has a nonatomic distribution G. Denote the candidates' positions by x_1 and x_2. If $x_1 = x_2$ then each candidate wins with probability $\frac{1}{2}$, and if $x_1 < x_2$ then candidate 1 wins with probability $G(\frac{1}{2}(x_1+x_2))$, the probability that the median favorite position is at most $\frac{1}{2}(x_1+x_2)$, and candidate 2 wins with probability $1-G(\frac{1}{2}(x_1+x_2))$. (Refer to Figure 8.3.) Each candidate cares about her probability of winning.

Definition 8.5: Electoral competition game with two office-motivated candidates and uncertain median

An *electoral competition game with two office-motivated candidates and uncertain median* $\langle\{1,2\},X,G\rangle$, where X is a closed interval of real numbers and G is a nonatomic distribution with support X (and hence a unique median), is the strategic game with the following components.

Players
$\{1,2\}$ (candidates).

Actions
The set of actions of each player is X.

Preferences
The preferences of each player j over $X \times X$ are represented by the function $u_j : X \times X \to \mathbb{R}$ defined by

$$u_j(x_1,x_2) = \begin{cases} G(\frac{1}{2}(x_1+x_2)) & \text{if } x_j < x_k \\ \frac{1}{2} & \text{if } x_1 = x_2 \\ 1 - G(\frac{1}{2}(x_1+x_2)) & \text{if } x_j > x_k, \end{cases}$$

where k is the other player.

The interpretation of this game differs from that of an electoral competition game with a continuum of citizens and two office-motivated candidates, but formally the games are similar, and their analyses are also similar. The game has a unique Nash equilibrium, in which each candidate's position is the median of the distribution G of the median of the citizens' favorite positions.

Proposition 8.5: Nash equilibrium of electoral competition game with two office-motivated candidates and uncertain median

An electoral competition game with two office-motivated candidates and uncertain median $\langle\{1,2\},X,G\rangle$ has a unique Nash equilibrium, in which each candidate's position is the median of G.

Proof

Denote the median of G by m. The outcome of the action pair (m,m) is that each candidate wins with probability $\frac{1}{2}$. If either candidate deviates to a position x, her probability of winning becomes $G(\frac{1}{2}(x+m)) < G(m) = \frac{1}{2}$

if $x < m$ and $1 - G(\frac{1}{2}(x+m)) < G(m) = \frac{1}{2}$ if $x > m$. Thus (m, m) is a Nash equilibrium.

Now let $(x_1, x_2) \neq (m, m)$. Assume without loss of generality that $x_1 \leq x_2$.

First suppose that $x_1 = x_2 < m$, so that each candidate's probability of winning is $\frac{1}{2}$. Let $x = x_1 = x_2$. Then either candidate can increase her probability of winning to $1 - G(\frac{1}{2}(x+m)) > \frac{1}{2}$ by deviating to m. A similar argument applies if $x_1 = x_2 > m$.

Now suppose that $x_1 < x_2$, so that candidate 1's probability of winning is $G(\frac{1}{2}(x_1+x_2))$ and candidate 2's is $1 - G(\frac{1}{2}(x_1+x_2))$. If $\frac{1}{2}(x_1+x_2) < m$ then $G(\frac{1}{2}(x_1+x_2)) < \frac{1}{2}$ and candidate 1 can increase her probability of winning to $\frac{1}{2}$ by deviating to x_2. If $\frac{1}{2}(x_1+x_2) > m$ then similarly candidate 2 can increase her probability of winning by deviating to x_1. If $\frac{1}{2}(x_1+x_2) = m$ then each candidate's probability of winning is $\frac{1}{2}$ and either candidate can increase this probability by deviating to m.

8.3.2 *Private information*

Now suppose that each candidate gets a private signal about the location of the median of the citizens' favorite positions (perhaps from a poll she conducts). Assume that for each candidate, this signal is drawn from the same finite set T, and the distribution of the median of the citizens' favorite positions depends only on the pair of signals the candidates receive, not on the identity of the candidate who received each signal. Denote by $G_{\{t_1,t_2\}}$ the distribution function of this median when one candidate's signal is t_1 and the other's is t_2. Assume that each distribution function $G_{\{t_1,t_2\}}$ has a density, $g_{\{t_1,t_2\}}$. An example of possible densities in a case in which T contains two signals, 1 and 2, is shown in Figure 8.4.

Each candidate observes only her own signal. Before receiving their signals, the candidates' beliefs about the probabilities of the pairs $(t_1, t_2) \in T \times T$ are the same: each candidate believes that the probability of (t_1, t_2) is $P(t_1, t_2)$. The function P is assumed to be symmetric: $P(t_1, t_2) = P(t_2, t_1)$ for each $(t_1, t_2) \in T \times T$. The signal each candidate receives determines, via P, her belief about the other candidate's signal.

Definition 8.6: Electoral competition game with two office-motivated candidates privately informed about citizens

An *electoral competition game with two office-motivated candidates privately informed about citizens* $\langle \{1,2\}, X, T, P, (G_{\{t_1,t_2\}})_{\{t_1,t_2\} \subset T} \rangle$, where

- $X \subset \mathbb{R}$ is a finite interval

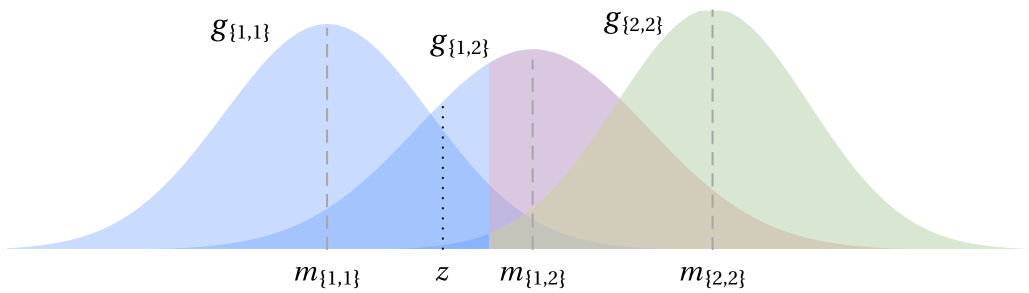


Figure 8.4 An example of the distributions over positions of the median of the citizens' favorite positions for each possible pair of signals in an electoral competition game with two office-motivated candidates privately informed about citizens. The set of signals in this example is $\{1,2\}$; $g_{\{t_1,t_2\}}$ is the density of the distribution function $G_{\{t_1,t_2\}}$ of the median of the citizens' favorite positions when candidate 1's signal is t_1 and candidate 2's is t_2, and $m_{\{t_1,t_2\}}$ is the median of $G_{\{t_1,t_2\}}$.

- T is a finite set
- P is a probability distribution over $T \times T$ with $P(t_1, t_2) > 0$ and $P(t_1, t_2) = P(t_2, t_1)$ for all $(t_1, t_2) \in T \times T$
- for each $\{t_1, t_2\} \subset T$, $G_{\{t_1,t_2\}}$ is a nonatomic probability distribution function for X that has a density and whose support is an interval

is a Bayesian game with the following components.

Players
$\{1,2\}$ (the candidates).

States
$T \times T$ (the set of pairs (t_1, t_2) with $t_i \in T$ for $i = 1, 2$).

Actions
The set of actions of each player is X (the set of positions).

Signals
For each player i, the set of signals is T and her signal function associates with each state (t_1, t_2) the signal t_i.

Prior beliefs
The players' common prior belief is that t_1 and t_2 are drawn from $T \times T$ according to P.

Payoffs
The payoff of each player j for the pair of actions (z_1, z_2) and state (t_1, t_2)

is her probability of winning when the distribution of the median is $G_{\{t_1,t_2\}}$:

$$u_j((z_1,z_2),(t_1,t_2)) = \begin{cases} G_{\{t_1,t_2\}}(\frac{1}{2}(z_1+z_2)) & \text{if } z_j < z_k \\ \frac{1}{2} & \text{if } z_j = z_k \\ 1 - G_{\{t_1,t_2\}}(\frac{1}{2}(z_1+z_2)) & \text{if } z_j > z_k \end{cases} \quad (8.2)$$

where k is the other player.

Given that the distributions $G_{\{t_1,t_2\}}$ depend only on the pair of signals, not the identity of the candidate who receives each signal, it is natural to consider the possibility that the game has a Nash equilibrium (x_1,x_2) (where $x_i : T \to X$ for $i = 1, 2$) in which the candidates choose the same position whenever their signals are the same: $x_1(t) = x_2(t)$ for all $t \in T$.

Suppose that there are two possible signals, with $T = \{1,2\}$. For each set $\{t_1,t_2\} \subset T$, denote the median of $G_{\{t_1,t_2\}}$ by $m_{\{t_1,t_2\}}$ and assume that the higher signal is associated with a larger value of the median: $m_{\{1,1\}} < m_{\{1,2\}} < m_{\{2,2\}}$, as in the example in Figure 8.4.

I first argue that the game has no Nash equilibrium in which both types of each candidate choose the same position: $x_1(1) = x_2(1) = x_1(2) = x_2(2)$. For such a strategy pair, the probability of each type of each candidate winning is $\frac{1}{2}$. Denote the candidates' common position by z. Suppose that $z < m_{\{1,2\}}$, as in Figure 8.4. Consider type 2 of candidate 1. She believes that the density of the median is either $g_{\{1,2\}}$ (if candidate 2's signal is 1) or $g_{\{2,2\}}$ (if candidate 2's signal is 2). In both cases, her probability of winning increases if she deviates from z to $m_{\{1,2\}}$. If candidate 2's signal is 1, this probability becomes equal to the area under $g_{\{1,2\}}$ for positions at least equal to the midpoint of z and $m_{\{1,2\}}$, shaded purple in Figure 8.4, and if candidate 2's signal is 2, it becomes the area under $g_{\{2,2\}}$ for positions at least equal to the midpoint of z and $m_{\{1,2\}}$, shaded green in Figure 8.4. Given that $\frac{1}{2}(z+m_{\{1,2\}}) < m_{\{1,2\}} < m_{\{2,2\}}$, both of these probabilities exceed $\frac{1}{2}$. Thus type 2 of candidate 1 gains by deviating from z to $m_{\{1,2\}}$, and hence the game has no equilibrium in which both types of both candidates choose the same position less than $m_{\{1,2\}}$. By a symmetric argument, it has no equilibrium in which both types of both candidates choose the same position greater than $m_{\{1,2\}}$. Finally, suppose that both types of both candidates choose the position $m_{\{1,2\}}$. Then if type 2 of candidate 1 deviates to a slightly larger position, she slightly reduces her probability of winning when candidate 2's signal is 1 and discretely increases it when candidate 2's signal is 2, so that she has a deviation that increases her probability of winning.

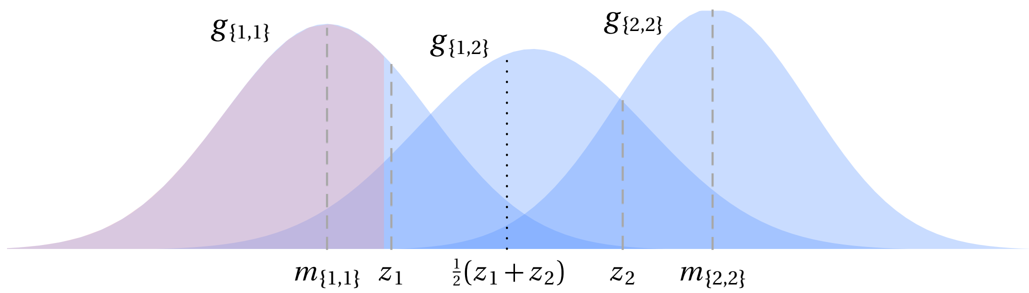


Figure 8.5 The effect of a deviation by type 1 of a candidate from a strategy pair (x_1, x_2) with $x_1(1) = x_2(1) = z_1$ and $x_1(2) = x_2(2) = z_2$ in an electoral competition game with two office-motivated candidates privately informed about citizens.

I now consider the possibility that the game has a ("symmetric") Nash equilibrium in which type 1 of each candidate chooses the same position, as does type 2 of each candidate, but the types choose different positions. Consider a strategy pair (x_1, x_2) with $x_1(1) = x_2(1) = z_1$, $x_1(2) = x_2(2) = z_2$, and $z_1 \neq z_2$. For this strategy pair, type 1 of each candidate i ties with the other candidate, j, when j's signal is 1 (both candidates choose z_1) and wins with probability $G_{\{1,2\}}(\frac{1}{2}(z_1+z_2))$ if $z_1 < z_2$ and with probability $1 - G_{\{1,2\}}(\frac{1}{2}(z_1+z_2))$ if $z_1 > z_2$ when j's signal is 2 (i chooses z_1 and j chooses z_2). I argue that if this strategy pair is a Nash equilibrium then $z_1 = m_{\{1,1\}}$ and $z_2 = m_{\{2,2\}}$. Suppose that $z_1 > m_{\{1,1\}}$, as in Figure 8.5. Then if type 1 of a candidate deviates to a slightly smaller position, her probability of winning is discretely larger than $\frac{1}{2}$ when the other candidate's signal is 1 (it is an area like the one shaded purple in Figure 8.5) and is close to what it was when her position was z_1 when the other candidate's signal is 2 (if $z_2 > z_1$, as in Figure 8.5, it is slightly less than it was before, and if $z_2 < z_1$ then it is slightly more than it was before). Thus the deviation increases the candidate's probability of winning. A symmetric argument shows that if $z_1 < m_{\{1,1\}}$ then a candidate's deviation to a slightly larger position increases her probability of winning.

We conclude that if the game has a symmetric equilibrium, then the position of each type t of each candidate is the median $m_{\{t,t\}}$ of the distribution of the citizens' favorite positions when the other candidate's signal is also t. In a sense, such an equilibrium amplifies the candidates' signals: although a candidate who receives a signal t assigns positive probability to the other candidate's receiving each possible signal, the distribution of the citizens' favorite positions that determines her equilibrium position is the one associated with both candidates' signals being t; her position is more extreme than the expected value of the median of the citizens' favorite positions given her own signal.

Proposition 8.6: Nash equilibrium of electoral competition game with two office-motivated candidates privately informed about citizens

Let $\langle\{1,2\}, X, T, P, (G_{\{t_1,t_2\}})_{\{t_1,t_2\}\subset T}\rangle$ be an electoral competition game with two office-motivated candidates privately informed about citizens. Suppose that $T = \{1,2\}$ and $m_{\{1,1\}} < m_{\{1,2\}} < m_{\{2,2\}}$, where $m_{\{t_1,t_2\}}$ is the median of $G_{\{t_1,t_2\}}$ for each set $\{t_1, t_2\} \subset T$. In any Nash equilibrium of the game, the position of each type $t \in T$ of each candidate is the median $m_{\{t,t\}}$ of $G_{\{t,t\}}$.

Proof

In the text I argue that in any Nash equilibrium (x_1, x_2) in which $x_1(t) = x_2(t)$ for all $t \in T$ we have $x_1(1) = x_2(1) = m_{\{1,1\}}$ and $x_1(2) = x_2(2) = m_{\{2,2\}}$. To complete the proof, suppose that (x_1, x_2) is a Nash equilibrium in which $x_1 \neq x_2$. Then given the symmetry of the game (in particular, the symmetry of P), (x_2, x_1) is also a Nash equilibrium. Now, the game is strictly competitive (an outcome that is better for one candidate is worse for the other candidate), so its Nash equilibria are interchangeable (see for example Osborne 2004, Corollary 369.3). Thus if (x_1, x_2) and (x_2, x_1) are Nash equilibria then so are (x_1, x_1) and (x_2, x_2). The argument in the text shows that the game has at most one symmetric equilibrium, so we conclude that it has no asymmetric equilibria.

This result asserts only that if the game has an equilibrium then it takes a certain form, not that the strategy pair given is necessarily a Nash equilibrium. I now give an example in which the strategy pair is in fact a Nash equilibrium. Suppose that each distribution $G_{\{t_1,t_2\}}$ is symmetric about its median and has the same form, differing only in its location, and the locations of $G_{\{1,1\}}$, $G_{\{1,2\}}$, and $G_{\{2,2\}}$ are equally spaced, as in Figure 8.6. That is, for some number $\alpha > 0$ and function $H : \mathbb{R} \to \mathbb{R}$ with a symmetric density we have $G_{\{1,1\}}(x) = H(x + \alpha)$, $G_{\{1,2\}}(x) = H(x)$, and $G_{\{2,2\}}(x) = H(x - \alpha)$ for each $x \in X$. Consider the strategy pair (x_1, x_2) for which $x_1(1) = x_2(1) = m_{\{1,1\}}$ and $x_1(2) = x_2(2) = m_{\{2,2\}}$. Suppose that type 1 of candidate i deviates to z_1. Denote the other candidate by j.

$z_1 < m_{\{1,1\}}$ or $z_1 > m_{\{2,2\}}$

Type 1 of candidate i's probability of winning falls regardless of j's type.

$m_{\{1,1\}} < z_1 < m_{\{2,2\}}$

If j's signal is 1, type 1 of candidate i's probability of winning decreases from $\frac{1}{2}$ to the area shaded purple in Figure 8.6, and if j's signal is 2, this probability

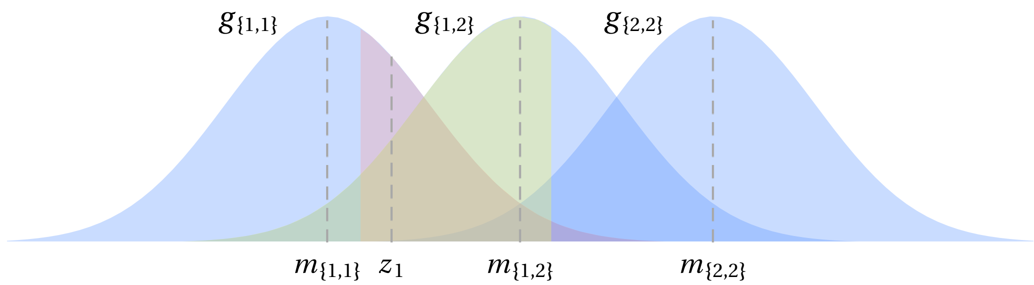


Figure 8.6 The effect of a deviation by type 1 of a candidate to z_1 from a strategy pair (x_1, x_2) with $x_1(1) = x_2(1) = m_{\{1,1\}}$ and $x_1(2) = x_2(2) = m_{\{2,2\}}$ in an electoral competition game with two office-motivated candidates privately informed about citizens

increases from $\frac{1}{2}$ to the area shaded green in the figure. Given the symmetry of the distributions, the decrease in the first case is equal to the increase in the second case. Thus if the probability that j's signal is 1 given that i's signal is 1 is at least $\frac{1}{2}$, then the deviation does not increase i's probability of winning.

$z_1 = m_{\{2,2\}}$
If j's signal is 1, type 1 of candidate i's probability of winning falls from $\frac{1}{2}$ to $1 - G_{\{1,1\}}(\frac{1}{2}(m_{\{1,1\}} + m_{\{1,2\}})) = 1 - G_{\{1,1\}}(m_{\{1,2\}})$, and if j's signal is 2 it remains $\frac{1}{2}$.

Symmetric arguments apply to deviations by type 2 of a candidate, so we conclude that the strategy pair (x_1, x_2) is a Nash equilibrium in this example if for each $t \in \{1, 2\}$ the probability that one candidate's signal is t given that the other candidate's signal is t is at least $\frac{1}{2}$.

If the number of possible signals is arbitrary, an extension of this logic implies that the game has a Nash equilibrium only if the probability that one candidate's signal is extreme given that the other candidate's signal is extreme is at least $\frac{1}{2}$ (Bernhardt et al. 2009, Theorem 2). When the number of signals is large, this condition is particularly restrictive. If the condition is violated, then a version of the model has a mixed strategy equilibrium in which the strategy of each type t of each candidate with a relatively moderate assigns probability 1 to $m_{\{t,t\}}$, like the strategy of each candidate in Proposition 8.6, and the support of the strategy of type t of each remaining candidate is an interval consisting of positions more moderate than $m_{\{t,t\}}$ (Bernhardt et al. 2009, Theorem 3). Thus the amplification of the candidates' private information in the strategy profile in Proposition 8.6 carries over to this mixed strategy equilibrium, although it is tempered for candidates with extreme signals.

Exercise 8.3: Previous electoral outcomes as information sources

Candidates may obtain information about the citizens' preferences from the outcomes of previous elections, leading citizens to consider the effect of their votes not only on the outcome of the current election but on the candidates' positions in future elections. Here is a simple example. Two candidates compete in a sequence of two elections, in each of which the set of possible positions is $[0,1]$. There is a single citizen, whose favorite position $\hat{x}$ is unknown to the candidates, who believe that its distribution function is H. For any outcomes x^1 in period 1 and x^2 in period 2, the citizen's payoff is $-|x^1 - \hat{x}| - |x^2 - \hat{x}|$. In the first period the candidates' positions are fixed at x_1 and x_2 with $x_1 < x_2$; the citizen chooses a cutoff position x^*, voting for candidate 1 if her favorite position is less than x^* and for candidate 2 if it is greater than x^*. If candidate 1 wins in the first period, in the second period, in line with Proposition 8.5, the candidates both choose the median of the distribution of the citizen's favorite position conditional on this position being in $[0,x^*]$. If candidate 2 wins in the first period, in the second period the candidates similarly both choose the median of the distribution of the citizen's favorite position conditional on this position being in $[x^*,1]$. If $x_1 = \frac{1}{2}$, $x_2 = 1$, and H is the uniform distribution on $[0,1]$, which position x^* does the citizen choose as her cutoff? How does this position compare with the one she would choose if there were no second period?

8.4 Costly voting

In the models of electoral competition I have discussed so far, every citizen votes even if the candidates' positions are the same. If voting is voluntary, the devotion to civic duty that this behavior requires seems excessive. In this section I discuss a model in which voting entails a cost, as in the models in Chapter 4, and a citizen votes only if she believes that the expected benefit of doing so outweighs this cost.

Assume that two candidates simultaneously choose positions and then the citizens simultaneously vote. The citizens differ in their preferences over positions and their voting costs. Each citizen votes (for the candidate whose position she prefers) only if her expected benefit from doing so, given the probability that her vote affects the outcome, is at least her voting cost. Then only if the candidates' positions differ do any citizens vote, and then only those whose voting costs are sufficiently small.

To simplify the analysis, assume that there is only one citizen. This assumption may seem extreme, but the model captures the main idea we want to study: that a citizen votes only if doing so sufficiently affects the outcome. In a model with many citizens, each citizen must consider the probability that her vote will affect the outcome, given the other citizens' behavior. This consideration complicates the analysis, but after establishing the main result I argue that it appears not to affect the key feature of an equilibrium.

I formulate the model as a Bayesian extensive game with observable actions. In such a game, each player may have many possible types. Each player knows her own type, but not the other players' types. Every player observes every other player's action, and holds the same probabilistic belief about the other players' types.

Specifically, in the the model here, two candidates simultaneously choose positions in a finite interval X and then a single citizen either votes for one of the candidates, in which case that candidate wins, or abstains, in which case the candidates tie. The candidates' characteristics are known (formally, each candidate has a single type), but the citizen's preferences over X and voting cost are not known to the candidates. Thus the citizen's type is a pair (θ, c), where $\theta \in \Theta$ parameterizes the citizen's preferences over X (the value of θ may be the citizen's favorite position, for example) and $c \in \mathbb{R}_+$ is her voting cost. The payoff of type (θ, c) of the citizen is $v(x_j, \theta) - c$ if she votes for candidate j (in which case j is the winner), where x_j is j's position, and $\frac{1}{2}(v(x_1, \theta) + v(x_2, \theta))$ if she abstains (in which case the outcome is a tie). Each candidate's belief about the citizen's type is given by a probability measure P over the set $\Theta \times \mathbb{R}_+$ of the citizen's types. Each candidate prefers to win than to tie than to lose. The game is illustrated in Figure 8.7.

Definition 8.7: Electoral competition game with two office-motivated candidates and costly voting

An *electoral competition game with two office-motivated candidates and costly voting* $\langle X, \Theta, P, v \rangle$, where

- $X \subset \mathbb{R}$ is a finite interval (the set of possible positions)
- $\Theta \subset \mathbb{R}$ (the set of possible preference parameters for the citizen)
- P is a probability measure on $\Theta \times \mathbb{R}_+$ (the candidates' beliefs about the citizen's preference parameter and voting cost)
- $v : X \times \mathbb{R} \to \mathbb{R}$ (the citizen's payoff function)

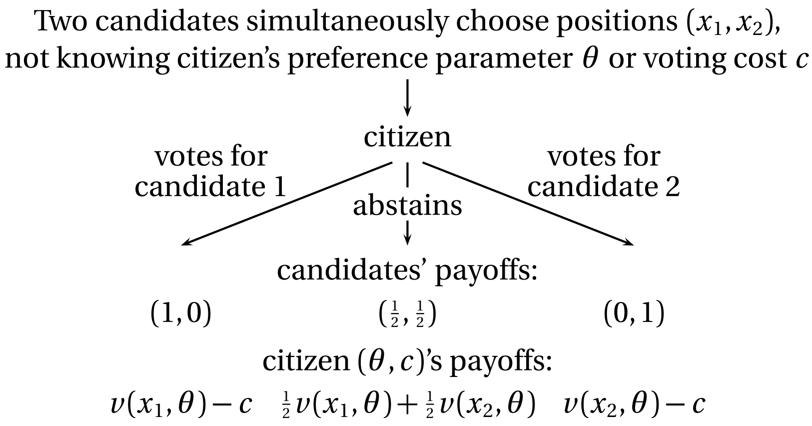


Figure 8.7 An electoral competition game with two office-motivated candidates and costly voting.

is a Bayesian extensive game with observable actions with the following components.

Players
$\{1,2\} \cup \{z\}$ (the candidates and a single citizen).

Terminal histories
The set of sequences $\{((x_1, x_2), b) : x_j \in X \text{ for } j = 1,2 \text{ and } b \in \{1,2,\phi\}\}$ (where x_j is the position of candidate j, $b \in \{1,2\}$ is the candidate for whom the citizen votes, and ϕ stands for abstention).

Player function
The function L with $L(\varnothing) = \{1,2\}$ (the candidates move (simultaneously) at the start of the game) and $L((x_1, x_2)) = \{z\}$ for each $(x_1, x_2) \in X \times X$ (the citizen moves after the candidates have chosen positions).

Actions
The set of actions of each candidate at the start of the game is X and the set of actions of the citizen following any pair of actions of the candidates is $\{1,2,\phi\}$ (a vote for one of the candidates or abstention).

Types
Each candidate has one possible type, known to the citizen. The set of types of the citizen is $\Theta \times \mathbb{R}_+$, the set of pairs consisting of a preference parameter and a nonnegative number (the citizen's cost of voting). The probability measure on the set $\Theta \times \mathbb{R}_+$ of the citizen's types is P.

Preferences

The preferences over lotteries over terminal histories of each candidate j (= 1, 2) are represented by the expected value of the function u_j defined by

$$u_j((\theta,c),((x_1,x_2),b)) = \begin{cases} 1 & \text{if } b=j \\ \frac{1}{2} & \text{if } b=\phi \\ 0 & \text{if } b=k, \end{cases}$$

where k is the other candidate. The preferences over terminal histories of the citizen of type (θ,c) are represented by the function u defined by

$$u((\theta,c),((x_1,x_2),b)) = \begin{cases} v(x_b,\theta)-c & \text{if } b\in\{1,2\} \\ \frac{1}{2}(v(x_1,\theta)+v(x_2,\theta)) & \text{if } b=\phi. \end{cases}$$

A strategy for a candidate in such a game is a position (member of X), and a strategy for the citizen is a function that assigns an action (member of $\{1,2,\phi\}$) with each type $(\theta,c)\in\Theta\times\mathbb{R}_+$ and each pair of actions (x_1,x_2) for the candidates.

A standard notion of equilibrium for a Bayesian extensive game with observable actions is perfect Bayesian equilibrium. The definition of this notion of equilibrium has features that are irrelevant to a two-candidate electoral competition game with office-motivated candidates and costly voting because of the simple structure of such a game: each player moves only once, and the citizen is perfectly informed when she does so. The following notion of equilibrium suffices.

Definition 8.8: Equilibrium of electoral competition game with two office-motivated candidates and costly voting

An *equilibrium* of an electoral competition game with two office-motivated candidates and costly voting is a strategy profile in which, for each pair of the candidates' positions, the action of each type of the citizen maximizes that type's expected payoff, and each candidate's position maximizes her expected payoff given the other candidate's position and the citizen's strategy.

If $\Theta = X$ and the parameter θ is the citizen's favorite position, we can illustrate the citizen's optimal decision in a diagram. Figure 8.8 shows the candidates' positions, x_1 and x_2, and the function $v(\cdot,\theta_0)$. If the citizen's type is (θ,c) then she

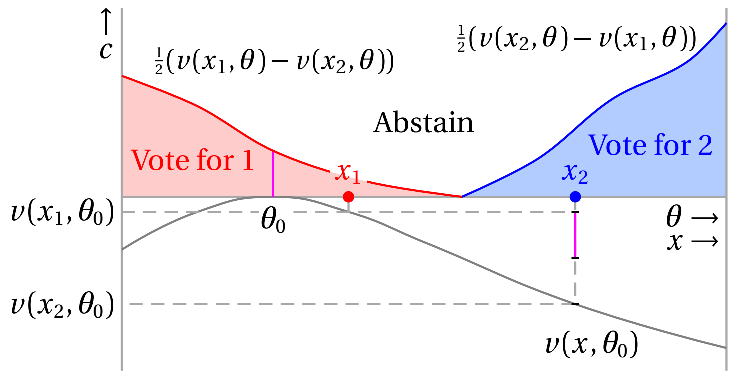


Figure 8.8 Payoffs in a two-candidate electoral competition with costly voting. The area shaded red is the set of types (θ, c) of the citizen who optimally vote for candidate 1 and the area shaded blue is the set of types who optimally vote for candidate 2. The citizen's function v is shown for the parameter value θ_0; the magenta lengths are equal.

optimally votes for candidate 1 if

$$v(x_1,\theta) > v(x_2,\theta) \quad \text{and} \quad v(x_1,\theta) - c > \tfrac{1}{2}(v(x_1,\theta) + v(x_2,\theta)).$$

The second condition is equivalent to $c < \frac{1}{2}(v(x_1,\theta) - v(x_2,\theta))$. The length of the magenta line segments in the figure is this cutoff $\frac{1}{2}(v(x_1,\theta_0) - v(x_2,\theta_0))$ for a citizen with preference type θ_0. The figure shows the result of a similar calculation for each preference type: the area shaded red is the set of types (θ, c) who optimally vote for candidate 1 and the area shaded blue is the set of types who optimally vote for candidate 2.

The effect of a change in candidate 1's position that reduces the difference between the candidates' positions is illustrated in Figure 8.9. As for a game in which every type of the citizen is assumed to vote, this change causes some types of the citizen to switch from voting for candidate 2 to voting for candidate 1. But it affects also the set of types of the citizen that vote, because for almost all types of the citizen it changes the expected gain from voting and thus the cost cutoff for voting rather than abstaining. For example, it reduces this cost cutoff for types of the citizen who prefer x_1 to x_1' to x_2 or who prefer x_2 to x_1' to x_1, and increases this cost cutoff for types who prefer both x_1 to x_2 and x_1' to x_1. Informally, it reduces the motivation to vote for types of the citizen with relatively extreme preferences and increases this motivation for types whose favorite alternative is close to x_1'. In a game in which every citizen is assumed to vote, the net effect on candidate 1's probability of winning of her moving her position closer to that of candidate 2 is necessarily positive. In the model here the net effect depends on the character of the citizen's preferences and the distribution of her types.

Consider the possibility that the game has an equilibrium in which the can-

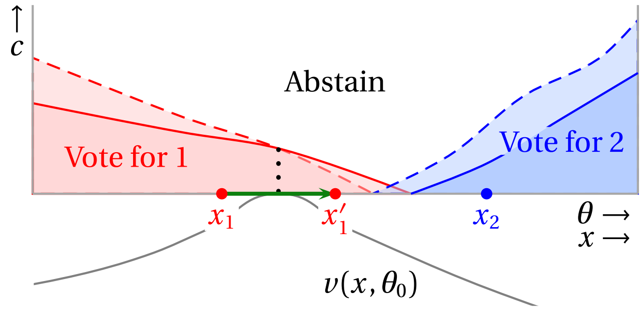


Figure 8.9 The decision to vote in a two-candidate electoral competition with costly voting. When candidate 1 moves from x_1 to x_1', the set of types who vote for each candidate changes from the areas shaded with light colors (and dashed boundaries) to those shaded with darker colors (and solid boundaries). For a citizen with preference type θ_0, who is indifferent between x_1 and x_1', the move does not affect the relative attractiveness of the candidates' positions.

didates' positions are the same, say equal to x^*. In such an equilibrium, the only types who vote are those for whom the voting cost is zero, and when the distribution of types is nonatomic, as the next result assumes, such types have measure zero. If one of the candidates, say i, deviates slightly from x^*, opening a small wedge between the positions, only citizen types with cost close to zero vote. For equilibrium, such a deviation must not increase i's payoff. That is, $x_i = x^*$ must locally maximize i's payoff, given that the other candidate's position is x^*. The next result shows that if for each value of the parameter $\theta \in \mathbb{R}$ the function $v(\cdot,\theta)$ is strictly concave and differentiable, and x^* is in the interior of X, then the condition for x^* to be a local maximizer implies that x^* maximizes the expected payoff, according to the candidates' belief, of the citizen types for which the voting cost is zero. Given the strict concavity of $v(\cdot,\theta)$ for each value of θ, there is only one such position x^*, so if the game has an equilibrium in which the candidates' positions are the same and in the interior of X then it has exactly one such equilibrium.

Now, for any given strategy of the citizen, the game between the candidates is strictly competitive (an outcome that is better for one candidate is worse for the other candidate), and the roles of the candidates are symmetric. So if (x_1, x_2) is an equilibrium pair of positions then so is (x_2, x_1), and hence by the interchangeability property of Nash equilibria of strictly competitive games so are (x_1, x_1) and (x_2, x_2). So by the interchangeability property of Nash equilibria of strictly competitive games (see for example Osborne 2004, Corollary 369.3), if the game has only one equilibrium in which the candidates' positions are the same then it has no equilibrium in which they differ.

Combining the arguments in the last two paragraphs, we conclude that if for each value of $\theta \in \mathbb{R}$ the function $v(\cdot,\theta)$ is strictly concave and differentiable and the game has an equilibrium in which the candidates' positions are in the interior of X then it has one such equilibrium, in which the candidates' positions are the same, equal to the position that maximizes the expected payoff of the citizen types for which the voting cost is zero.

Proposition 8.7: Equilibrium of electoral competition game with two office-motivated candidates and costly voting

Let $\langle X, \Theta, P, v\rangle$ be an electoral competition game with two office-motivated candidates and costly voting. If P is nonatomic and has a continuous density, for each $\theta \in \mathbb{R}$ the function $v(\cdot, \theta)$ is strictly concave and differentiable, and for each $x \in X$ the function $v(x, \cdot)$ is continuous, then in every equilibrium of $\langle X, \Theta, P, v\rangle$ in which the candidates' positions are in the interior of X these positions are the same, equal to the solution of

$$\max_{x\in X} \int_{\Theta} v(x,\theta)g(\theta,0)\,d\theta, \tag{8.3}$$

where g is the density of P, and the citizen abstains unless her cost is 0 (an event with probability zero).

Proof

First consider the citizen. For any pair (x_1, x_2) of the candidates' positions, type (θ, c) of the citizen chooses a solution of

$$\max_{b\in\{1,2,\phi\}} u((\theta,c),((x_1,x_2),b)).$$

For $j = 1, 2$, denote by $T_j(x_1, x_2)$ the set of pairs (θ, c) such that j (i.e. *vote for* j) is a solution of this problem, and by $T_\phi(x_1, x_2)$ the set of pairs such that ϕ (*abstain*) is a solution.

Now, j is a solution of the problem only if $v(x_j,\theta) \geq v(x_k,\theta)$ and $v(x_j,\theta) - c \geq \frac{1}{2}[v(x_1,\theta) + v(x_2,\theta)]$, where k is the other candidate. Given $c \geq 0$, these conditions are equivalent to $c \leq \frac{1}{2}[v(x_j,\theta) - v(x_k,\theta)]$. So

$$T_j(x_1,x_2) = \{(\theta,c) \in \Theta \times \mathbb{R}_+ : c \leq \tfrac{1}{2}[v(x_j,\theta) - v(x_k,\theta)]\}. \tag{8.4}$$

Thus for any position x, we have $P(T_j(x,x)) = 0$ for $j = 1, 2$, given that P is nonatomic: if the candidates' positions are the same, the set of types of the citizen that vote has measure zero, so the outcome is a tie, and hence each candidate's payoff is $\frac{1}{2}$. So for an equilibrium in which both candidates choose the same position, we need the payoff of a candidate who deviates to be at most $\frac{1}{2}$. For the pair of positions (x_1, x_2), the probability that candidate j wins is $P(T_j(x_1, x_2))$ and the probability she ties is $P(T_\phi(x_1, x_2))$, so her payoff is at most $\frac{1}{2}$ if and only if

$$P(T_j(x_1,x_2)) + \tfrac{1}{2}P(T_\phi(x_1,x_2)) \leq \tfrac{1}{2},$$

or, given that $P(T_1(x_1,x_2)) + P(T_2(x_1,x_2)) + P(T_\phi(x_1,x_2)) = 1$, $P(T_j(x_1,x_2)) - P(T_k(x_1,x_2)) \leq 0$, where k is the other candidate. So the game has an equilibrium in which both candidates' positions are x^* if and only if

$$P(T_1(x_1,x^*)) - P(T_2(x_1,x^*)) \leq 0 \text{ for all } x_1 \in X$$
$$P(T_2(x^*,x_2)) - P(T_1(x^*,x_2)) \leq 0 \text{ for all } x_2 \in X.$$

The left-hand side of each inequality is 0 for $x_1 = x_2 = x^*$, so equivalently x^* maximizes $P(T_1(x_1,x^*)) - P(T_2(x_1,x^*))$ and $P(T_2(x^*,x_2)) - P(T_1(x^*,x_2))$.

Now, from (8.4) we have

$$P(T_j(x_1,x_2)) = \int_\Theta \int_0^{\frac{1}{2}[v(x_j,\theta)-v(x_k,\theta)]} g(\theta,c)\,dc\,d\theta,$$

so that

$$P(T_1(x_1,x^*)) - P(T_2(x_1,x^*)) = \int_\Theta \left(\int_0^{\frac{1}{2}[v(x_1,\theta)-v(x^*,\theta)]} g(\theta,c)\,dc - \int_0^{\frac{1}{2}[v(x^*,\theta)-v(x_1,\theta)]} g(\theta,c)\,dc \right) d\theta.$$

A necessary condition for a position x^* interior to X to maximize this expression is that the derivative of the expression with respect to x_1 evaluated at x^* is zero, or

$$\int_\Theta \left(\tfrac{1}{2}v_1'(x^*,\theta)g(\theta,0) + \tfrac{1}{2}v_1'(x^*,\theta)g(\theta,0)\right) d\theta = \int_\Theta v_1'(x^*,\theta)g(\theta,0)\,d\theta = 0,$$

where v_1' denotes the derivative of v with respect to its first argument. Now, given that for any value of its second argument, v is strictly concave in its first argument,

$$\int_\Theta v_1'(x^*,\theta)g(\theta,0)\,d\theta = 0$$

if and only if x^* is the (unique) solution of (8.3). So for any equilibrium in which the positions of the candidates are the same and interior to X, the common position is the solution of this problem.

Now, the strategic game between the candidates, given the optimal action of each type of citizen for each pair of positions, is strictly competitive, and the candidates are symmetric. The argument in the text before the result shows that as a consequence the fact that the game has a unique equilibrium in which the candidates' positions are the same implies that it has no equilibrium in which they differ.

When there are many citizens, each of whom knows her own type but not the other citizens' types, the analysis is more complicated, because the last stage of the game is no longer simply a decision problem. However, an informal argument suggests that the candidates' equilibrium positions remain the same. When a candidate deviates slightly from a common position, only a vanishingly small measure of citizen types—those with very low voting cost—possibly find voting worthwhile, and their equilibrium actions are plausibly similar to the optimal action of the lone citizen in the model I have presented. Ledyard (1984, Theorem 1) shows that under some conditions an analog of Proposition 8.7 indeed holds.

The main qualitative assumption in the result is the strict concavity of the function $v(\cdot,\theta)$ for each preference type $\theta \in \mathbb{R}$. This assumption does not have any particular appeal; the assumption that this function is convex on each side of its maximizer (the type's favorite position), rather than concave, for example, seems equally plausible. As far as I know, no general analysis of the model without the concavity assumption exists, although under the assumptions of the following exercise, which include the symmetry of the citizen's payoff function and the uniformity of the probability measure on citizen types, the equilibria can be characterized.

Exercise 8.4: Electoral competition with costly voting

Consider an electoral competition game with two office-motivated candidates and costly voting $\langle X, \Theta, P, v\rangle$ in which the preferences of each type of the citizen have the same form, differing only in the favorite position. Specifically, assume that $\Theta = \mathbb{R}$ and $v(x,\theta) = \psi(x-\theta)$ for all $x \in X$ and $\theta \in \Theta$, where $\psi : \mathbb{R} \to \mathbb{R}$ is increasing for negative values of its argument and decreasing for positive values (so that its maximizer is 0) and $\psi(z) = \psi(-z)$ for all $z \in \mathbb{R}$ (so that it is symmetric about 0). Assume also that the measure P is uniform on $\Theta \times [0, \bar{c}]$ for some $\bar{c} > 0$. Show that the game has an equilibrium in which each candidate chooses the midpoint of X.

8.5 Citizens with preferences over candidates

The models so far in this chapter assume that each citizen cares about the policies proposed by the candidates, not about the candidates themselves. I now present a model in which each citizen cares directly about the candidates, as well as about the policies they propose. One motivation for the model is that a candidate has immutable personal characteristics that determine how energetically she will implement her proposed policies if she is elected. Another motivation is that each candidate is unable to credibly change her positions on some issues.

8.5.1 Candidates know citizens' preferences

Suppose that the preferences of each citizen i are represented by the function $u_i : X \times \{1,2\} \to \mathbb{R}$, where X is the set of possible positions and $\{1,2\}$ is the set of candidates, so that i prefers candidate 1 with position x_1 to candidate 2 with position x_2 if and only if $u_i(x_1,1) > u_i(x_2,2)$. Assume that $u_i(x,1) \neq u_i(x,2)$ for each citizen i and each position x, so that each citizen cares about the candidate as well as her position. An implication of this assumption is that a candidate does not necessarily have the option to tie with the other candidate by choosing the same position as she does, as was the case in the models discussed previously.

The following example illustrates possible forms for an equilibrium; I know of no general characterization of equilibria.

Example 8.1: Electoral competition when citizens have preferences over candidates

Suppose that the set X of possible positions is an interval of numbers and there are three citizens, with preferences represented by the functions shown in Figure 8.10. For each citizen there are two curves, a red one indicating the citizen's payoff as a function of the policy, x, if candidate 1 is elected, and a blue one indicating her payoff if candidate 2 is elected.

Is the pair (x^*, x^*) of positions a Nash equilibrium? For this pair of positions, candidate 2 wins, because citizens 2 and 3 vote for her; thus she has no deviation that generates an outcome she prefers. If candidate 1 deviates to a position $x < x^*$ then she becomes less desirable for citizens 2 and 3, and hence still loses. If she deviates to a position $x > x^*$ then she becomes less desirable for citizens 1 and 2 and more desirable for citizen 3. If for some such position x, citizen 3 prefers $(x,1)$ to $(x^*,2)$ and citizen 1 also still has this preference, then candidate 1 has a profitable deviation, and (x^*, x^*) is not a Nash equilibrium. In Figure 8.10, we need $x > x'$ for

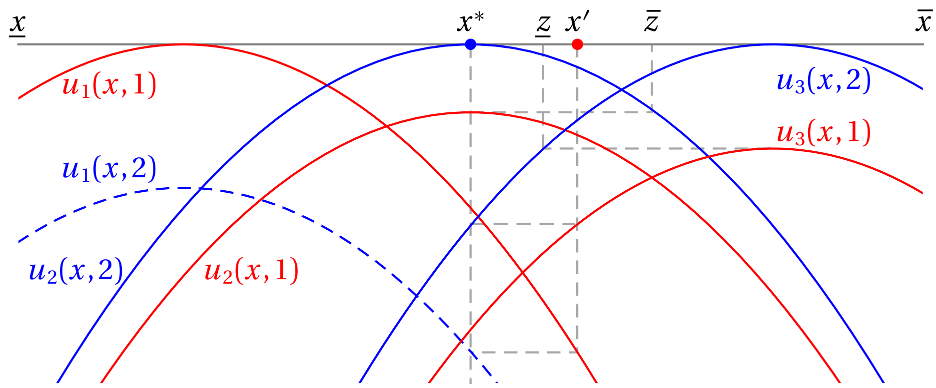


Figure 8.10 The functions u_i for the citizens in Example 8.1.

citizen 3 to prefer $(x,1)$ to $(x^*,2)$ and hence vote for candidate 1, and when $x = x'$ citizen 1 is indifferent between $(x,1)$ and $(x^*,2)$ if the graph of $u_1(x,2)$ is the dashed curve.

Thus if the graph of $u_1(x,2)$ lies above the dashed curve, for no position does candidate 1 attract the votes of citizens 1 and 3: if she moves far enough right to attract the vote of citizen 3, she loses the vote of citizen 1. Hence in this case (x^*,x^*) is a Nash equilibrium.

If the graph of $u_1(x,2)$ lies below the dashed curve, then for some position $x > x'$ with x close to x' candidate 1 attracts the votes of both citizen 1 and citizen 3, and hence wins, so that (x^*,x^*) is not a Nash equilibrium.

In both cases any pair (x_1,x_2) with $\underline{z} \le x_2 \le \bar{z}$ is a Nash equilibrium: for any value of x_1, citizens 2 and 3 prefer $(x_2,2)$ to $(x_1,1)$, so that candidate 2 wins and candidate 1 cannot stop her from doing so. (The game has also other Nash equilibria.)

Exercise 8.5: Electoral competition with an advantaged candidate

Suppose that every citizen prefers candidate 1 to candidate 2, in the sense that $u_i(x,1) > u_i(x,2)$ for all $x \in X$ for every citizen i. Assume specifically that (*i*) the set X of possible positions is an interval of numbers and (*ii*) for a single-peaked function $v : \mathbb{R} \to \mathbb{R}_-$ with $v(0) = 0$ and $v(z) < 0$ for all $z \neq 0$, a position $\hat{x}_i$ for each citizen i (i's favorite position), and a number $\delta > 0$, we have $u_i(x,1) = v(x - \hat{x}_i)$ and $u_i(x,2) = v(x - \hat{x}_i) - \delta$ for all $x \in X$ for each citizen i.

Assume that the number of citizens is finite and odd and that a citizen i for whom $u_i(x_1,1) = u_i(x_2,2)$, where x_1 and x_2 are the candidates' positions, votes for candidate 1. Denote by m the median of the citizens' favorite positions. Show that if $v(x_1 - m) \geq -\delta$ then (x_1, x_2) is a Nash equilibrium for any value of x_2. In any such equilibrium, candidate 1 wins.

Now suppose that each candidate's preferences are lexicographic: between two pairs of positions for which her probability of winning is the same, she prefers the one in which the number of votes she receives is larger. What can you say about the Nash equilibria in this case?

8.5.2 Candidates uncertain of citizens' preferences

Now suppose that the candidates are not perfectly informed about the citizens' preferences. Denote the set of possible positions by X and the payoff of any citizen i for the position x implemented by candidate j by $u_i(x, j)$. Assume that for each citizen i there is a function $v_i : X \to \mathbb{R}$ and numbers θ_i^1 and θ_i^2 such that

$$u_i(x, j) = v_i(x) + \theta_i^j \quad \text{for all } (x, j) \in X \times \{1, 2\},$$

so that $u_i(x, 1) \geq u_i(y, 2)$ if and only if $\theta_i \leq v_i(x) - v_i(y)$ where $\theta_i = \theta_i^2 - \theta_i^1$. Assume that each candidate knows each function v_i but not the number θ_i, which she believes is drawn from a nonatomic distribution F_i independently of every $\theta_{i'}$ for $i' \neq i$. Under these assumptions, if candidate 1's position is x_1 and candidate 2's is x_2 then each candidate believes that for each citizen i, the probability that i votes for candidate 1 is the probability that θ_i is at most $v_i(x) - v_i(y)$, namely $F_i(v_i(x_1) - v_i(x_2))$, independently of the other citizens' votes.

Assume that the number n of citizens is odd. Then if each citizen i votes for candidate 1 with probability p_i, independently of the other citizens, the probability $P(p_1, \ldots, p_n)$ that candidate 1 wins is the probability that the members of some set of more than $\frac{1}{2}n$ citizens all vote for candidate 1:

$$P(p_1, \ldots, p_n) = \sum_{\{S \subseteq I : |S| > n/2\}} \left(\prod_{i \in S} p_i \prod_{i \in I \setminus S} (1 - p_i) \right), \tag{8.5}$$

where I is the set of citizens. Thus the probability that candidate 1 wins as a function of the candidates' positions is

$$\Pr(1 \text{ wins}) = P\big(F_1(v_1(x_1) - v_1(x_2)), \ldots, F_n(v_n(x_1) - v_n(x_2))\big).$$

Definition 8.9: Electoral competition game with two office-motivated candidates and uncertain partisanship

An *electoral competition game with two office-motivated candidates and uncertain partisanship* $\langle I, X, (v_i)_{i \in I}, (F_i)_{i \in I}, \{1,2\}\rangle$, where

- I is a finite set (of citizens) with an odd number of members
- X is a set (of positions)

and, for each $i \in I$,

- $v_i : X \to \mathbb{R}$ (i's payoff function over positions)
- F_i is a nonatomic probability distribution function on $\mathbb{R}$

is the strategic game with the following components.

Players
$\{1,2\}$ (candidates).

Actions
The set of actions of each player is X.

Preferences
Letting $I = \{1,\ldots,n\}$, the preference relation of player 1 over action profiles (x_1, x_2) is represented by the function

$$P\big(F_1(v_1(x_1) - v_1(x_2)), \ldots, F_n(v_n(x_1) - v_n(x_2))\big)$$

(the probability that player 1 wins) and the preference relation of player 2 is represented by the function

$$1 - P\big(F_1(v_1(x_1) - v_1(x_2)), \ldots, F_n(v_n(x_1) - v_n(x_2))\big)$$

(the probability that player 2 wins), where P is given by (8.5).

Now assume that the set X of positions is a convex subset of a Euclidean space, every function v_i is differentiable and strictly concave, and all the functions F_i are equal to the same differentiable function. Then the next result shows that in any Nash equilibrium of the game in which each candidate's position is in the interior of X, the candidates' positions are the same, equal to the maximizer of the sum of $v_i(x)$ over all citizens.

Proposition 8.8: Nash equilibrium of electoral competition game with two office-motivated candidates and uncertain partisanship

Let $\langle I, X, (v_i)_{i\in I}, (F_i)_{i\in I}, \{1,2\}\rangle$ be an electoral competition game with two office-motivated candidates and uncertain partisanship. Assume that $I = \{1,\ldots,n\}$, X is a convex compact subset of a Euclidean space, each function v_i is differentiable and strictly concave, and each function F_i is differentiable, with $F_i'(\theta) > 0$ for all θ in the interior of its support, and this support includes $[\min_{i\in I, x\in X, y\in X}(v_i(x) - v_i(y)) - \underline{\varepsilon}, \max_{i\in I, x\in X, y\in X}(v_i(x) - v_i(y)) + \overline{\varepsilon}]$ for some $\underline{\varepsilon} > 0$ and $\overline{\varepsilon} > 0$.

a. If (x_1^*, x_2^*) is a Nash equilibrium of the game and x_1^* and x_2^* are in the interior of X then $x_1^* = x_2^*$.

b. If F_i is the same for all i then

$$x_1^* = x_2^* = \arg\max_{x\in X} \sum_{i=1}^{n} v_i(x).$$

Proof

a. For $j = 1, 2$, the position x_j^* of candidate j maximizes j's probability of winning, given the other candidate's position: x_1^* is a solution of

$$\max_{x_1\in X} P(F_1(v_1(x_1) - v_1(x_2^*)), \ldots, F_n(v_n(x_1) - v_n(x_2^*))) \tag{8.6}$$

and x_2^* is a solution of

$$\max_{x_2\in X} \big(1 - P(F_1(v_1(x_1^*) - v_1(x_2)), \ldots, F_n(v_n(x_1^*) - v_n(x_2)))\big). \tag{8.7}$$

If x_1^* is a solution of (8.6) in the interior of X then by Proposition 16.12 it satisfies the first-order condition

$$\sum_{i=1}^{n} P_i'(\pi(x_1^*, x_2^*)) F_i'(v_i(x_1^*) - v_i(x_2^*)) \nabla v_i(x_1^*) = 0, \tag{8.8}$$

where P_i' is the derivative of P with respect to its ith argument,

$$\pi(x_1^*, x_2^*) = \big(F_1(v_1(x_1^*) - v_1(x_2^*)), \ldots, F_n(v_n(x_1^*) - v_n(x_2^*))\big),$$

and ∇v_i is the gradient of v_i (the vector of its partial derivatives).

Define the function $W : X \to \mathbb{R}$ by

$$W(x) = \sum_{i=1}^{n} P_i'(\pi(x_1^*, x_2^*))F_i'(v_i(x_1^*) - v_i(x_2^*))v_i(x)$$

for all $x \in X$. Each function v_i is strictly concave and all the coefficients of $v_i(x)$ are positive ($P_i'(\pi(x_1^*, x_2^*))$ is the change in 1's probability of winning as citizen i becomes more likely to vote for her), so W is strictly concave. Thus W has a unique maximizer and, by Proposition 16.13, (8.8) is necessary and sufficient for x_1^* to be a maximizer in the interior of X. Hence if x_1^* is in the interior of X then it maximizes W.

A solution x_2^* of (8.7) that is in the interior of X satisfies the same condition, (8.8), and hence also is the unique maximizer of W. Thus $x_1^* = x_2^*$.

b. Given $x_1^* = x_2^*$,

$$W(x) = \sum_{i=1}^{n} P_i'(F(0), \ldots, F(0))F'(0)v_i(x),$$

where F is the common distribution. Thus given $P_i'(p, \ldots, p) = P_k'(p, \ldots, p)$ for all p and all i and k, the maximizers of $W(x)$ are the maximizers of $\sum_{i=1}^{n} v_i(x)$.

Note that this result does not restrict the set X of positions to be one-dimensional. If this set *is* one-dimensional and $v_i(x) = -(x - \hat{x}_i)^2$ for each citizen i then

$$\arg\max_{x \in X} \sum_{i=1}^{n} v_i(x) = \arg\max_{x \in X} \sum_{i=1}^{n} -(x - \hat{x}_i)^2 = \sum_{i=1}^{n} \hat{x}_i / n.$$

That is, in this case the common position chosen by the candidates is the mean of the citizens' favorite positions.

Note also that the result does not assert that an equilibrium exists—only that *if* an equilibrium exists, it has the claimed properties. I am not aware of a result that provides sufficient conditions for an equilibrium to exist, but the following example shows that the result is not vacuous.

Example 8.2: Electoral competition with two office-motivated candidates and uncertain partisanship

Let $\langle I, X, (v_i)_{i \in I}, (F_i)_{i \in I}, \{1, 2\} \rangle$ be a two-candidate electoral competition game with office-motivated candidates and uncertain partisanship. Suppose that there are three citizens, with $I = \{1, 2, 3\}$, and the set X of possible

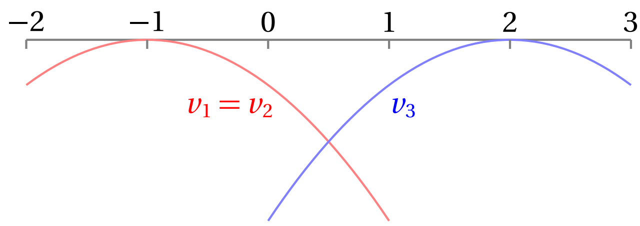


Figure 8.11 The functions v_i for the game in Example 8.2.

positions is the interval $[-3,3] \subset \mathbb{R}$. Citizens 1 and 2 have favorite position $\hat{x}_1 = \hat{x}_2 = -1$, citizen 3 has favorite position $\hat{x}_3 = 2$, and $v_i(x) = -(x - \hat{x}_i)^2$ for all $x \in X$ for every citizen i. (Refer to Figure 8.11.) The functions F_1, F_2, and F_3 are all equal to F, which for some number $\alpha > 0$ is uniform on $[-\alpha, \alpha]$.

The median favorite position is -1 and the maximizer of $\sum_{i=1}^{n} v_i(x)$ is 0, the mean favorite position. By Proposition 8.8, the only possible Nash equilibrium is $(x_1, x_2) = (0,0)$. Under what conditions is that pair of positions in fact a Nash equilibrium? We need each candidate's probability of winning to be at most $\frac{1}{2}$ for every position, given that the other candidate's position is 0. If candidate 2's position is 0 and candidate 1's position is x then

$$\begin{aligned}
\Pr(\text{citizen 1 votes for candidate 1}) &= \Pr(\theta < v_1(x) - v_1(0)) \\
&= \Pr(\theta < -(x-(-1))^2 + (0-(-1))^2) \\
&= \Pr(\theta < -x(x+2)) \\
&= [\max(-\alpha, \min(\alpha, -x(x+2))) - (-\alpha)]/(2\alpha).
\end{aligned}$$

(Refer to Figure 8.12.) The probability that citizen 2 votes for candidate 1 in this case is the same, and a similar calculation yields

$$\begin{aligned}
&\Pr(\text{citizen 3 votes for candidate 1}) \\
&\qquad = [\max(-\alpha, \min(\alpha, -x(x-4))) - (-\alpha)]/(2\alpha).
\end{aligned}$$

Using these expressions, we find that the probability a candidate wins (which happens if and only if two or three citizens vote for her) is given by the graphs at the top of Figure 8.13 for various values of α. We see that $(0,0)$ is a Nash equilibrium if α is large enough (the cutoff is about 2.4); that is, if there is enough uncertainty.

Proposition 8.8 implies that if α is less than the cutoff, the game has no Nash equilibrium.

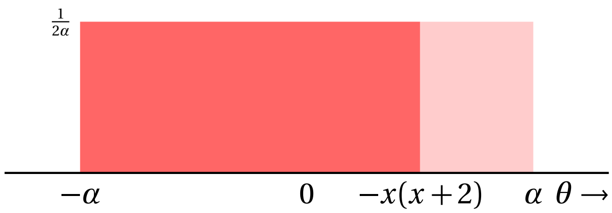


Figure 8.12 The values of θ for which citizen 1 votes for candidate 1 when candidate 1's position is x and candidate 2's position is 0 for the game in Example 8.2.

The pair of positions $(-1,-1)$, in which both candidates choose the median of the citizens' favorite positions, is a Nash equilibrium in the variant of the example in which there is no uncertainty, but not for any $\alpha > 0$. If the position of each candidate is -1, in the presence of any uncertainty one candidate can increase her probability of winning by deviating to a position slightly greater than -1, because the impact of such a deviation on the probability of her getting the votes of citizens 1 and 2, whose favorite positions are both -1, is almost zero, given that the functions u_1 and u_2 are differentiable, but the impact on the probability of her getting the vote of citizen 3, whose favorite position is 2, is bounded away from zero.

Exercise 8.6: Tent-shaped payoff functions

If the functions v_i are not differentiable, Proposition 8.8 does not apply. Suppose that X is an interval of numbers, the number n of citizens is odd, and for each citizen i, $v_i(x) = -|x - \hat{x}_i|$ for all $x \in X$. Show that (*a*) the position x^* that maximizes $\sum_{i=1}^{n} v_i(x)$ is the median of the numbers $\hat{x}_i$ for $i \in I$ and (*b*) if each distribution F_i is the same, equal to F, and the density of F is symmetric about 0, then (x^*, x^*) is a Nash equilibrium of the game.

8.6 Electing a legislature

In the models I have presented so far, a single candidate is elected. I now present a variant of an electoral competition game with a continuum of citizens and two office-motivated candidates in which the set of citizens is partitioned into an odd number of subsets, which may be interpreted as electoral districts. The distribution of the citizens' favorite positions may differ among districts. In each district, one candidate is elected to a legislature. I discuss two versions of the model.

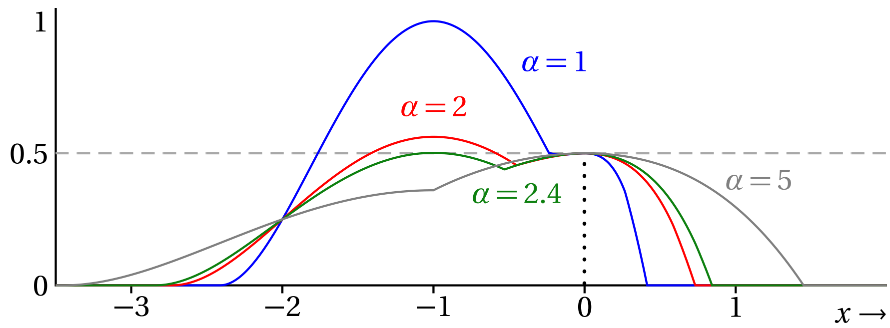


Figure 8.13 The probability of a candidate with position x winning when the other candidate's position is 0, for various values of α, for the game in Example 8.2.

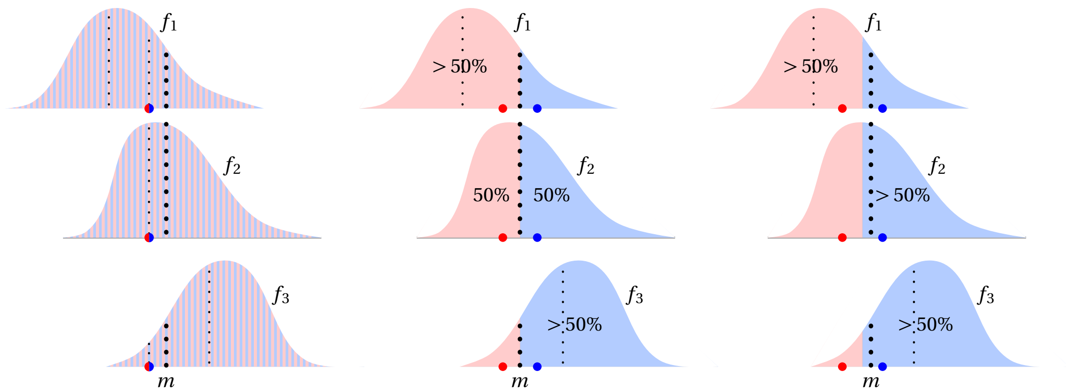
8.6.1 Each party chooses one position for all its candidates

Full information Assume that each of two parties fields a candidate in each district. Each party chooses a single position. In each district, each citizen votes for the candidate that represents the party whose position she prefers. Each party's objective is to win a majority of districts.

Denote the number of districts by l and the median of the favorite positions of the citizens in each district k by m_k. Order the districts so that $m_1 \leq m_2 \leq \cdots \leq m_l$ and denote the median of these medians m_k by m. That is, m is the median favorite position among the citizens in the median district.

If the parties choose the same position ($x_1 = x_2$), as in the example in Figure 8.14a, then the outcome is a tie in every district, and hence a tie in the legislature. If the parties' positions are symmetric around m, as in the example in Figure 8.14b, then the outcome is a tie in the median district, a win for party 1 in half of the remaining districts, and a win for party 2 in the remaining half, also resulting in a tie in the legislature. Otherwise suppose that $x_i < x_j$ and $\frac{1}{2}(x_1+x_2) < m$, as in the example in Figure 8.14c (for $i = 1$ and $j = 2$). Then party i wins in a majority of the districts. Thus for each pair (x_1, x_2) of the parties' positions the outcome is exactly the same as the outcome of (x_1, x_2) in an electoral competition game with a continuum of citizens and two office-motivated candidates. Hence that game models the electoral competition between the parties if we interpret each player as a party, the distribution F as the distribution of the citizens' favorite positions in the median district, and *win for j* to mean that j wins a majority of the districts.

We conclude from the analysis of Section 8.2.1 that the electoral competition between the parties has a unique Nash equilibrium, in which each party's position is m, the median favorite position among the citizens in the median district.



(a) $x_1 = x_2$; tie in every district.

(b) $\frac{1}{2}(x_1 + x_2) = m$; win for • in district 1, tie in district 2, and win for • in district 3.

(c) $x_1 < x_2$ and $\frac{1}{2}(x_1 + x_2) < m$; win for • in district 1 and win for • in districts 2 and 3.

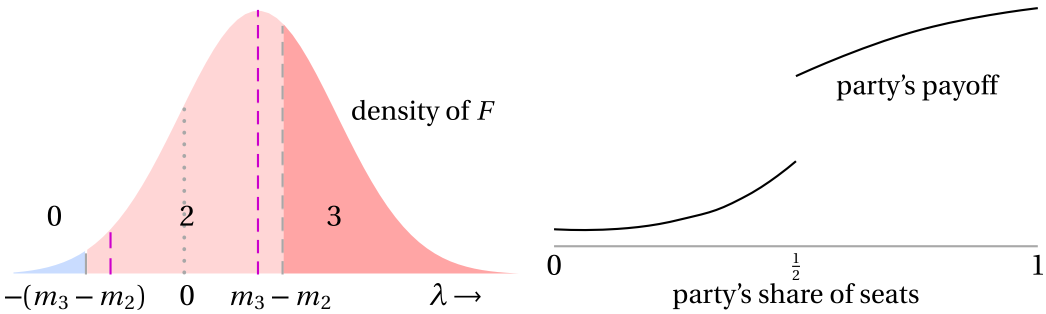
Figure 8.14 Electoral outcomes in a model in which each of two parties fields a candidate in each of several districts. In this example, there are three districts; the density of the distribution of the favorite positions in district k is f_k for $k = 1, 2, 3$.

Uncertain partisanship Now suppose that as in an electoral competition game with two office-motivated candidates and uncertain partisanship, each citizen's payoff depends not only on the policy implemented but also on the party that implements it, and when choosing positions the parties are uncertain of the degree of this partisanship (which may depend on events between the time the party commits to a policy and the election). Specifically, assume that the parties believe that the payoff for the policy x of a citizen with favorite policy $\hat{x}$ is

$$\begin{cases} v(x-\hat{x})+\lambda & \text{if } x \text{ is implemented by party 1} \\ v(x-\hat{x}) & \text{if } x \text{ is implemented by party 2,} \end{cases}$$

where v is a single-peaked function with its peak at 0 and λ is a random draw from a distribution F. An example of the density of such a distribution F is shown in Figure 8.15a. This distribution has a positive mean, so that each citizen, on average, favors implementation by party 1. One implication of these assumptions is that if the parties choose the same position then party 1 wins in every district with probability $1 - F(0)$ and party 2 wins in every district with probability $F(0)$.

Assume also that each party's payoff is increasing in the fraction of districts it wins, with an upward jump as this fraction passes $\frac{1}{2}$. That is, each party prefers to win than to lose overall, but conditional on winning it prefers to win more districts, and conditional on losing it also prefers to win more districts. Assume also that for any fraction less than $\frac{1}{2}$, each party's payoff is less than half its value for the fraction 1, so that the party prefers the outcome in which it wins all districts with probability $\frac{1}{2}$ to any outcome in which it wins a minority of districts. An



(a) The density of a distribution of party 1's advantage λ.

(b) A possible form for a party's payoff as a function of the proportion of seats it wins.

Figure 8.15

example of such a payoff function is shown in Figure 8.15b. These assumptions are intended to capture the possibility that the power of a party in a legislature, which is assumed to be its motivator, increases with its number of seats. The members of the minority party may receive more legislative committee assignments when their minority is larger, and may value the opportunity to make the majority exert more effort to implement its agenda.

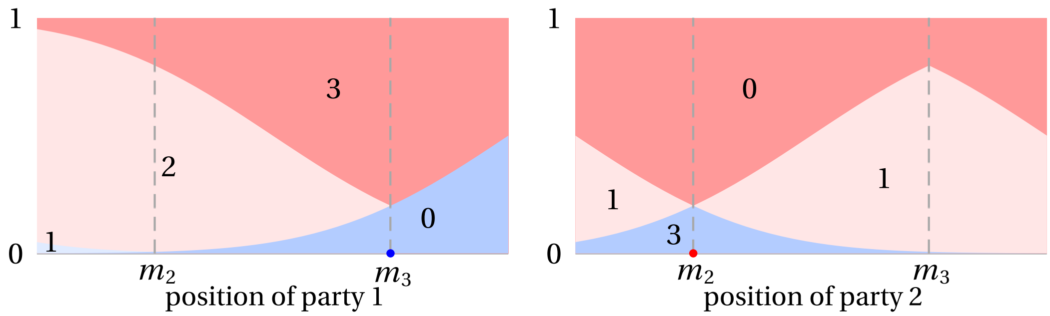
Adding these two ingredients to the model makes possible an equilibrium in which the parties' positions differ. I know of no general results, and present only an example.

Suppose there are three districts, each containing a continuum of citizens. Denote the median of the favorite positions of the citizens in district i by m_i and assume that $m_1 < m_2 < m_3$. Assume also that the median of λ is positive and that $v(z) = |z|$ for all z.

I look for conditions under which an equilibrium exists in which party 1's position is m_2 and party 2's position is m_3. If there were no partisanship, with λ fixed at 0, this pair of positions would not be an equilibrium, because by deviating to m_2 party 2 could win with probability $\frac{1}{2}$ rather than losing. In the presence of uncertain partisanship, the implications of such a deviation are mixed. The number of districts won by party 1 for the pair (m_2, m_3) of positions is

$$\begin{cases} 3 & \text{if } \lambda > m_3 - m_2\text{: probability } 1 - F(m_3 - m_2) \\ 2 & \text{if } -(m_3 - m_2) < \lambda < m_3 - m_2\text{: probability } F(m_3 - m_2) - F(-(m_3 - m_2)) \\ 0 & \text{if } \lambda < -(m_3 - m_2)\text{: probability } F(-(m_3 - m_2)). \end{cases}$$

(Given that $v(z) = |z|$, the payoff difference between the parties' positions is the same for a citizen with favorite position m_1 at it is for a citizen with favorite position m_2, so for no value of λ does party 1 win only district 1.) The probabilities of these three events are equal to the areas of the three regions shaded in Fig-



(a) The probabilities for party 1 given that party 2's position is m_3.

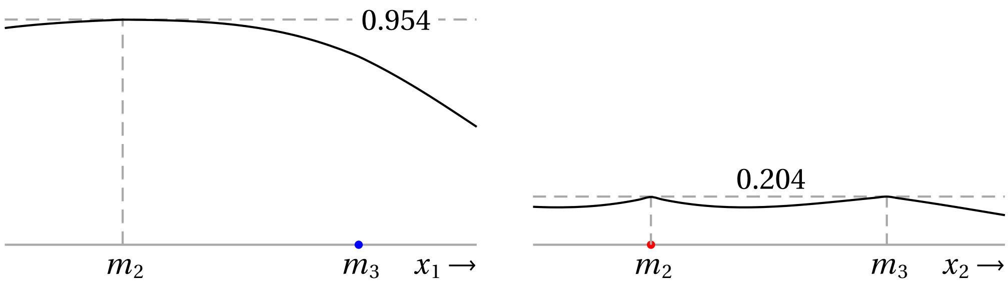
(b) The probabilities for party 2 given that party 1's position is m_2.

Figure 8.16 Representations of the probability distributions over the number of seats won by party 1 (left) and party 2 (right) as a function of the party's position, given the position of the other party, for the example discussed in the text, in which $m_1 = -1$, $m_2 = 0$, and $m_3 = 1$, and the distribution F is normal with mean 0.5 and standard deviation 0.6. The lengths of the segments of any vertical line in each of the colored regions are the probabilities that the party wins the indicated numbers of districts.

ure 8.15a, from right to left. In particular, party 2's probability of winning all three districts is $F(-(m_3 - m_2))$, which may be positive, and if it deviates to m_2 then its probability of such a win is $F(0)$, which is less than $\frac{1}{2}$ (given that the median of λ is positive). At the same time, a deviation by party 1 from m_2 to a position $x_1 \in (m_2, m_3)$ may be attractive. Such a deviation decreases the cutoff value of λ for party 1 to win all three districts to $1 - F(m_3 - x_1)$ and increases the cutoff for it to win two districts to $F(-(m_3 - x_1))$, as indicated by the dashed violet lines in Figure 8.15a. Depending on the magnitudes of these changes and the value party 1 attaches to winning two districts rather than three, such a deviation may increase its payoff.

Here is a specific example. Suppose that $m_1 = -1$, $m_2 = 0$, and $m_3 = 1$, and the distribution F is normal with mean 0.5 and standard deviation 0.6. The resulting probability distributions over the number of seats won as a function of a party's position, given the other party's position, are illustrated in Figure 8.15. Consider panel (a). For any position x_1 on the horizontal axis, draw a vertical line through x_1. The lengths of the segments of that line in each region are the probabilities that party 1 wins the indicated number of districts when its position is x_1 and party 2's position is m_2. For example, for $x_1 = m_1$, the probability that party 1 wins no districts is almost zero (it is 0.006), the probability that it wins two districts is about 0.8, and the probability it wins all three districts is about 0.2.

Denote each party's payoff when it wins k seats by w_k, and let $w_0 = 0$ and $w_1 = 1$. Figure 8.16b shows that if $w_1 = 0$—if winning one district has no value—then when party 1's position is m_2, party 2 optimally chooses the same position,



(a) Party 1's payoff as a function of its position x_1 when party 2's position is m_3.

(b) Party 2's payoff as a function of its position x_2 when party 1's position is m_2.

Figure 8.17 The parties' payoffs when they deviate from (m_2, m_3) for the case in which $m_1 = -1$, $m_2 = 0$, and $m_3 = 1$, the distribution F is normal with mean 0.5 and standard deviation 0.6, and $w_0 = 0$, $w_1 = 0.25$, $w_2 = 0.95$, and $w_1 = 1$.

which maximizes its probability of winning all three districts. Thus in this case (m_2, m_3) is not an equilibrium. But if party 2 derives some payoff from winning one district, then m_3 may be her best response to m_2. Specifically, if $w_1 \geq 0.25$ and $w_2 \geq 0.95$ then (m_2, m_3) is an equilibrium. Figure 8.17 shows each party's payoff as a function of its position, given the other party's position, for $w_1 = 0.25$ and $w_2 = 0.95$. (Given the symmetry of the model, the position m_1, which is not included in the figure, is, like m_3, optimal for party 2 given party 1's position m_2, so (m_2, m_1) is an equilibrium whenever (m_2, m_3) is an equilibrium.) If w_2 is greater than 0.95 then the decrease in party 1's payoff as it increases its position is more than slight, and it w_1 is greater than 0.25 then the advantage to party 2 from choosing m_3 or m_1 over m_2 is more significant.

This analysis shows that the addition of two elements to the basic model—uncertain partisanship and a positive value from a party's winning an additional district, even if doing so leaves it with a minority—enriches the set of equilibria to include, for some parameter values, ones in which the parties' positions differ. In these equilibria, one party finds it optimal to cater to the preferences of the citizens in a peripheral district, giving it a high probability of winning that district but a low probability of winning a majority of the districts. The alternative of competing head-to-head with the other party would raise its probability of winning a majority, but not enough to compensate for the reduction in its probability of capturing the peripheral district.

Bernhardt et al. (2020), to whom the model is due, study a different example, in which there is a continuum of districts and F is uniform. To generate nontrivial equilibria in their example, they need to add an element to the model: independent of the value of λ, each citizen's payoff for any given policy depends on the party implementing the policy, with the difference between the payoffs that result when party 1 implements it and when party 2 does so decreasing in the policy. Under this assumption, party 1 has an advantage in implementing policies on the left and party 2 has an advantage in implementing ones on the right. This element is not required in the example I describe.

8.6.2 *Each candidate chooses a position independently*

Now assume that each candidate chooses a position independently, and each party's position is the average of its candidates' positions. Each citizen votes for the candidate (in her district) whose party's position is closest to her favorite position, and the party that wins a majority of districts and hence acquires a majority in the legislature implements its position.

Assume that each candidate's preferences are lexicographic. She is concerned primarily with whether her party wins a majority of districts, ties, or loses a majority of districts; among outcomes in which this outcome for her party is the same, she prefers to win than to tie than to lose in her own district.

Suppose that the set of possible positions for a candidate is the whole real line. Then for any given positions of the other candidates for her party, a candidate can induce any position for the party by choosing an appropriate position for herself. This fact, combined with the fact that each candidate cares primarily about her party's fortunes, means that the analysis of the game is very similar to the analysis of an electoral competition game with a continuum of citizens and two office-motivated candidates. In particular, a profile of positions for the candidates is a Nash equilibrium if and only if each party's position is the median favorite position among the citizens in the median district (as for the game in the previous section).

Exercise 8.7: Candidates who care about their own electoral fortunes

Consider a variant of the model in this section in which each candidate cares primarily about her own electoral fortune (rather than her party's fortune). Show that if the median of the citizens' favorite positions is not the same in every district then the game has no Nash equilibrium.

8.7 Interpreting Nash equilibrium

Is it reasonable to expect that the candidates in an electoral game will choose their Nash equilibrium actions? In a Nash equilibrium, each player's action is optimal for her, given the other players' actions. When choosing an action, a player does not know the other players' actions, and one way of conceiving her decision-making process assumes that she formulates a belief about those actions. An assumption implicit in the notion of Nash equilibrium is that this belief is correct.

Where does a player's belief about the other players' actions come from? According to the leading interpretation of the notion of Nash equilibrium, this belief is based on the player's experience playing the game against a variety of opponents. For each player, we imagine a large population of decision-makers; each time the game is played, one decision-maker is drawn randomly from each of these populations to take the role of one of the players in the game. Over time, each decision-maker learns the action chosen by each of the other players, but does not gather information on the actions chosen by any specific decision-maker. For example, whenever I have found myself walking straight towards another pedestrian, she has almost always stepped to right (in the right-driving country in which I live). I do not know the history of any given pedestrian's actions, but my belief that most pedestrians I randomly encounter will step to the right to avoid a collision is correct.

This interpretation does not fit many elections well. At least, it needs some stretching. Perhaps the participants in electoral competitions have at least observed many elections, or have advisors who have done so, and this experience has taught them how their opponents are likely to behave.

The difficulty with interpreting a Nash equilibrium in an electoral game gives added significance to results concerning stronger notions of equilibrium, like the ones for a electoral competition game with a continuum of citizens and two office-motivated candidates that show that the median of the citizens' favorite positions weakly dominates all other positions (Proposition 8.4) and is the only rationalizable position (Exercise 8.2).

Notes

The model in Section 8.1 has its origins in Hotelling (1929), who developed a model of competition between two spatially-separated firms and suggested (pp. 54–55) that it applies also to competition between political parties. The idea was taken up by Downs (1957, Chapter 8), who discusses the model informally. The version of the model in Definition 8.4 is sometimes called Hotelling's model or the Hotelling-Downs model (although neither Hotelling nor Downs formulated exactly this game).

Proposition 8.1 is a version of Theorem 7.1 (p. 257) in Austen-Smith and Banks (2005). Proposition 8.5 is due to Calvert (1985, Theorem 4). Section 8.3.2 is based on Bernhardt et al. (2007, 2009). The model and results in Section 8.4 are due to Ledyard (1984). Models like the one in Section 8.5 in which the candidates treat the citizens' actions as probabilistic date back to Hinich et al. (1972). Proposition 8.8 is due to Duggan (2014, Theorem 10). The full-information models in Section 8.6 are due to Austen-Smith (1984) and the one with uncertain partisanship is due to Bernhardt et al. (2020).

Exercise 8.1 is based on Brennan and Hamlin (1998). Exercise 8.3 is Example 1 in Meirowitz and Shotts (2009). The model in Exercise 8.5 is a variant of the ones studied by Ansolabehere and Snyder (2000) and Aragonès and Palfrey (2002).

Solutions to exercises

Exercise 8.2

Denote the median of the citizens' favorite positions by m. First suppose that a single position is furthest from m. Without loss of generality let this position be z_1. The following table gives the outcomes of a candidate's actions z_1 and m for all the actions possible for the other candidate. We see that m strictly dominates z_1.

	z_1	z_2	$\ldots$	z_{l-1}	$z_l = m$	z_{l+1}	$\ldots$	z_k
z_1	tie	lose	lose	lose	lose	lose	lose	lose
m	win	win	win	win	tie	win	win	win

No other position is strictly dominated because every other position leads to a win for the candidate if the other candidate's position is more extreme.

Thus every position except the one furthest from m is rational for the candidate.

If two positions are furthest from m, at exactly the same distance from it, then a similar arguments shows that neither of them is rational for the candidate.

Now, the candidate's assuming that the other candidate is rational means that she assumes that the other candidate does not choose the position furthest from m. Under that assumption, the position second-furthest from m (or the two positions second-furthest from m, if there is a tie for that honor) is strictly dominated (by m) for the candidate.

Repeating this argument, we conclude that the only action that is rational if the candidate assumes that the other candidate is rational, that the other candidate assumes that she is rational, and so forth, is m.

Exercise 8.1

Denote the distribution function of the citizens' favorite positions by F and its density by f. I claim that if F is unimodal then a pair (x_1, x_2) of positions is a Nash equilibrium if and only if $x_1 = x_2$ and

$$F(x^*) - F(x^* - k) = F(x^* + k) - F(x^*), \tag{8.9}$$

where $x^* = x_1 = x_2$.

Step 1 *Any pair of positions that satisfies (8.9) is a Nash equilibrium.*

Proof. If x^* satisfies (8.9) then $f(x^* - k) < f(x^*)$ and $f(x^* + k) < f(x^*)$. Hence a candidate who deviates from x^* loses. ◁

Step 2 *In any Nash equilibrium the candidates tie.*

Proof. For a pair of positions at which they do not tie, the losing candidate can move to the position of the other candidate and tie. ◁

Step 3 *In any Nash equilibrium the candidates' positions are the same.*

Proof. Suppose that (x_1, x_2) is a Nash equilibrium and $x_1 < x_2$. By Step 2, the candidates tie, and given that F is unimodal, either $f(x_1 - k) < f(\frac{1}{2}(x_1 + x_2))$ or $f(x_2 + k) < f(\frac{1}{2}(x_1 + x_2))$. The two cases are symmetric; assume the former. The difference between candidate 1's share of the votes and candidate 2's share is

$$F(\tfrac{1}{2}(x_1 + x_2)) - F(x_1 - k) - F(x_2 + k) + F(\tfrac{1}{2}(x_1 + x_2)) \\ = 2F(\tfrac{1}{2}(x_1 + x_2) - F(x_1 - k) - F(x_2 + k).$$

The derivative of this expression with respect to x_1 is

$$f(\tfrac{1}{2}(x_1 + x_2) - f(x_1 - k),$$

which is positive. Thus by increasing x_1 slightly, candidate 1 wins rather than ties, contradicting the assumption that (x_1, x_2) is a Nash equilibrium. ◁

Step 4 *In any Nash equilibrium (x^*, x^*), the position x^* satisfies (8.9).*

Proof. If (8.9) is not satisfied, a candidate can win rather than tie by either decreasing or increasing her position slightly. ◁

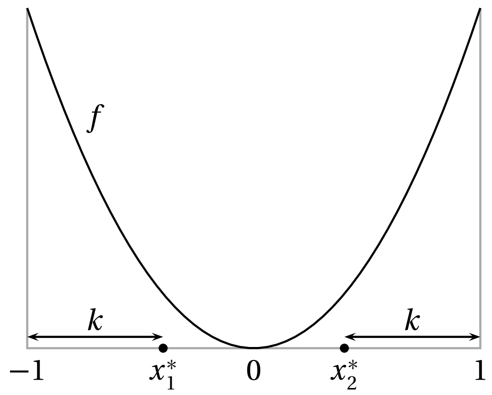


Figure 8.18 A Nash equilibrium of an electoral competition game with alienation in which the candidates' positions differ.

Here is an example of Nash equilibrium in which candidates' positions differ. Let $X = [-1,1]$ and let $f(x) = \frac{3}{2}x^2$. For any $k \leq 1$, the pair $(-1+k, 1-k)$ of positions is a Nash equilibrium. An example in which $k > \frac{1}{2}$ is shown in Figure 8.18. In this case, if candidate 1 increases her position she loses a large number of citizens with favorite positions close to -1 and gains a few with favorite positions close to 0, while candidate 2 loses these citizens with favorite positions close to 0. If she reduces her position then she loses a few citizens with favorite positions close to 0 and candidate 2 gains these citizens. If $k < \frac{1}{2}$ then citizens with favorite positions close to 0 do not vote. In this case, if candidate 1 increases her position she loses a large number of citizens with favorite positions close to -1 and gains a few with favorite positions close to $-1+2k$, and if she reduces her position she loses a few citizens with favorite positions close to $-1+2k$. In both cases, candidate 2 is unaffected.

Note, however, that this example depends on the symmetry of the distribution f. If f is slightly asymmetric the game appears not to have a Nash equilibrium.

Exercise 8.3

A citizen whose favorite position is x^* must be indifferent between voting for candidate 1 and for candidate 2 in the first period. If she votes for candidate 1 her payoff is $-\left|\frac{1}{2} - x^*\right| - \left|\frac{1}{2}x^* - x^*\right|$, and if she votes for candidate 2 her payoff is $-|1 - x^*| - \left|\frac{1}{2}(x^* + 1) - x^*\right|$. These payoffs are equal for $x^* = \frac{2}{3}$.

If there is no second period, the citizen votes for candidate 1 if her favorite position is less than $\frac{3}{4}$ and for candidate 2 if it is greater than $\frac{3}{4}$. Thus if the citizen's favorite position is between $\frac{2}{3}$ and $\frac{3}{4}$ she votes for candidate 1 if there is only one period but for candidate 2 if there are two periods and her vote is used as a signal about her favorite position. In this latter case, even though

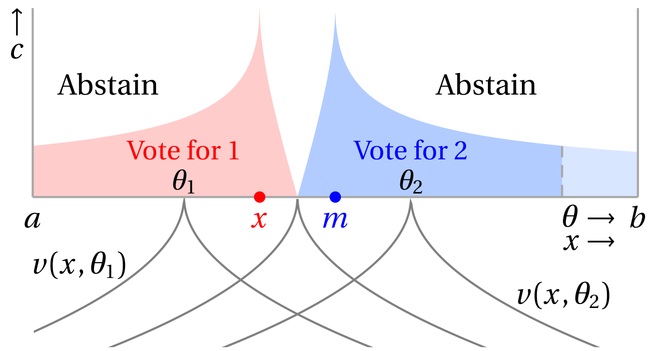


Figure 8.19 An example of a two-candidate electoral competition game with office-motivated candidates and costly voting that satisfies the assumptions of Exercise 8.4, with $\psi(z) = -\sqrt{|z|}$. For clarity, the vertical scale above the x-axis is exaggerated relative to the vertical scale below the axis. The set of types (θ, c) who vote for 1 when candidate 1's position is x and candidate 2's is m is shaded red, and the set who vote for 2 is shaded blue. The subset shaded dark blue is the reflection in the line $\theta = \frac{1}{2}(x+m)$ of the set shaded red, so that the area of the blue set exceeds the area of the red set by the area of the set shaded light blue.

the citizen prefers candidate 1's position, she votes for candidate 2 in the first period to move the candidates' second-period positions to the left.

Exercise 8.4

Let $X = [a, b]$ and $m = \frac{1}{2}(a+b)$. I argue that the game has an equilibrium in which each candidate's position is m, and it has no equilibrium in which either candidate chooses a position different from m.

Suppose that both candidates choose m. Then only types with zero cost vote; given that the distribution of types is nonatomic, these types have measure zero. Thus with probability 1 the candidates tie.

Now suppose that candidate 1 deviates to $x < m$. (Refer to Figure 8.19 for an example in which the function ψ is convex on each side of its maximizer.) Then given the symmetry of ψ, the set of types (θ, c) with $\theta \in [a, \frac{1}{2}(x+m)]$ who optimally vote for candidate 1 (shaded red in the figure) has the same measure as the set with $\theta \in [\frac{1}{2}(x+m), b-(m-x)]$ who optimally vote for candidate 2 (shaded dark blue in the figure). Candidate 2 gets, in addition, the votes of types with $\theta \in [b-(m-x), b]$ who optimally vote (the set shaded light blue in the figure); this set is nonempty because ψ is never constant, so that these types are not indifferent between the candidates. Thus candidate 1 wins with probability less than $\frac{1}{2}$, so that her deviation is not profitable.

If candidate 1 deviates to $x > m$, the same argument applies, so that this deviation is also not profitable.

Thus the game has an equilibrium in which each candidate's position is m.

The same argument shows that the game has no equilibrium in which both candidates choose a position different from m. If the common position is $x < m$, for example, then the candidates tie, and either candidate can win with probability greater than $\frac{1}{2}$ by moving to m.

Finally, the argument in the last paragraph of the proof of Proposition 8.7 shows that the game has no equilibrium in which the candidates' positions differ.

Exercise 8.5

Suppose that $x_1 \leq m$ satisfies the condition $v(x_1 - m) \geq -\delta$. I argue that if $x_2 \leq x_1$ then all citizens with favorite positions at least m, a majority, vote for candidate 1, and if $x_2 > x_1$ then all citizens with favorite positions at most m, a majority, vote for candidate 1.

First suppose that $x_2 \leq x_1$. Then for any citizen i with favorite position $\hat{x}_i \geq m$ the payoff from x_1 is $u_i(x_1, 1) = v(x_1 - \hat{x}_i)$ and the payoff from x_2 is $u_i(x_2, 2) = v(x_2 - \hat{x}_i) - \delta \leq v(x_1 - \hat{x}_i) - \delta < v(x_1 - \hat{x}_i)$, so i votes for candidate 1.

Now suppose that $x_2 > x_1$. Then for any citizen i with favorite position $\hat{x}_i$ for which $x_1 \leq \hat{x}_i \leq m$ the payoff from x_1 is $u_i(x_1, 1) = v(x_1 - \hat{x}_i) \geq v(x_1 - m) \geq -\delta$ and the payoff from x_2 is $u_i(x_2, 2) = v(x_2 - \hat{x}_i) - \delta \leq -\delta$, so i votes for candidate 1. For any citizen i with favorite position $\hat{x}_i < x_1$ the payoff from x_1 is $u_i(x_1, 1) = v(x_1 - \hat{x}_i)$ and the payoff from x_2 is $u_i(x_2, 2) = v(x_2 - \hat{x}_i) - \delta \leq v(x_1 - \hat{x}_i) - \delta$, so i votes for candidate 1.

Thus any pair (x_1, x_2) with $x_1 \leq m$ that satisfies the condition $v(x_1 - m) \geq -\delta$ is a Nash equilibrium. A symmetric argument shows that any pair (x_1, x_2) with $x_1 \geq m$ that satisfies the condition is a Nash equilibrium.

Now suppose that the candidates' preferences are lexicographic as described in the exercise. For any position of candidate 2, candidate 1 can obtain the vote of every citizen by choosing the same position, so in any best response to a position of candidate 2, candidate 1 obtains all the votes. Denote by $\underline{z}$ the smallest of the citizens' favorite positions and by $\bar{z}$ the largest. If there is a position x_1 such that $v(x_1 - \underline{z}) \geq -\delta$ and $v(x_1 - \bar{z}) \geq -\delta$, then for any such position the pair (x_1, x_2) is a Nash equilibrium for any position x_2. (For such a pair, candidate 1 gets all the votes and no position of candidate 2 yields candidate 2 any votes.) Otherwise, the game has no Nash equilibrium: a pair (x_1, x_2) in which candidate 1 gets all the votes is not an equilibrium because for some position of candidate 2, candidate 2 obtains some votes, and a pair (x_1, x_2) in which candidate 1 does not get all the votes is not an equilibrium because by deviating to x_2 candidate 1 can obtain all the votes.

Exercise 8.6 *a.* Let m be the median of the numbers $\hat{x}_i$ for $i \in I$. Divide I into three sets: a set M consisting of a single citizen j for whom $\hat{x}_j = m$, a set L consisting of citizens i for whom $\hat{x}_i < m$, and a set R consisting of citizens i for whom $\hat{x}_i > m$, so that the sets L and R have the same number of members. Then if $x < m$ we have

$$v_i(x) - v_i(m) \begin{cases} \leq m - x & \text{if } i \in L \\ = -(m-x) & \text{if } i \in R \cup M, \end{cases}$$

so that

$$\sum_{i \in I} v_i(x) - \sum_{i \in I} v_i(m) \leq -(m-x).$$

A symmetric argument shows the same result for $x > m$. Thus m is the unique maximizer of $\sum_{i \in I} v_i(x)$.

b. If the position of each candidate is x^*, then each candidates wins with probability $\frac{1}{2}$. Suppose that candidate 1 deviates to a position $x_1 < x^*$.

The number of citizens who prefer x^* to x_1 is at least $\frac{1}{2}(n+1)$, and for every such citizen i with $\hat{x}_i \geq x^*$, of which there are $\frac{1}{2}(n+1)$, we have $v_i(x^*) - v_i(x_1) = x^* - x_1$, so that the probability that each such citizen votes for candidate 2 is $F(x^* - x_1)$. Denote this probability by p^*. Given $x^* > x_1$ we have $p^* > \frac{1}{2}$.

The number of citizens who prefer x_1 to x^* is at most $\frac{1}{2}(n-1)$. For each such citizen i with $\hat{x}_i \leq x_1$ we have $v_i(x_1) - v_i(x^*) = x^* - x_1$, so that the probability that each such citizen votes for candidate 1 is $F(x^* - x_1) = p^*$, and for any of these citizens for whom $x_1 < \hat{x}_i \leq x^*$ we have $v_i(x_1) - v_i(x^*) < x^* - x_1$, so that the probability that she votes for candidate 1 is less than p^*.

Thus of the citizens who are more likely to vote for candidate 2, $\frac{1}{2}(n+1)$ vote for her with probability p^* and possibly additional citizens (with favorite positions between x_1 and x^*) vote for her with probability less than p^*. Of the citizens who are more likely to vote for candidate 1, who number at most $\frac{1}{2}(n-1)$, at most $\frac{1}{2}(n-1)$ vote for her with probability p^* and any remaining citizens (with favorite positions between x_1 and x^*) vote for her with probability less than p^*. Thus the probability that candidate 2 wins exceeds $\frac{1}{2}$.

Symmetric arguments apply if candidate 1 deviates to a position greater than x^* or if candidate 2 deviates. Thus (x^*, x^*) is a Nash equilibrium.

Exercise 8.7

Because any candidate can cause her party's position to take any value (given the positions of the other candidates for the party) by choosing an appropriate position, the argument in the proof of Proposition 8.4 applied to a given district shows that in any Nash equilibrium the parties' positions are the same, equal to the median of the favorite positions of the citizens in the district. Thus if this median favorite position varies across districts, no Nash equilibrium exists.

9 Electoral competition: two policy-motivated candidates

9.1 Basic model 289
9.2 Uncertain median 299
9.3 Candidates privately informed about policies 305

> "The peak of the campaign happened in Albuquerque, where a local reporter said to me, 'Dr. Commoner, are you a serious candidate or are you just running on the issues?'" (Barry Commoner, Citizens Party candidate in 1980 U.S. Presidential election, in interview in *New York Times*, 2007.6.19).

Some candidates for political office appear to be motivated by the possibility of implementing policies they like, rather than the possibility of winning per se, as the models in the previous chapter assume. We say that such candidates are policy-motivated.

Synopsis

Consider a variant of an electoral competition game with two office-motivated candidates in which each candidate cares about the position of the winner of the election, and not at all about whether she is the winner. Proposition 9.1 shows that if the underlying collective choice problem has a strict Condorcet winner, then the action pair in which each candidate chooses that alternative is a Nash equilibrium of the game. That is, the game has an equilibrium in which the candidates' actions are the same as they are in the unique Nash equilibrium of the game in which each candidate is office-motivated. The reason is simple: if either candidate deviates to a position different from the strict Condorcet winner then she loses, so that the outcome of the game remains the same.

Unlike the case in which the candidates are office-motivated, however, for some collective choice problems the game has Nash equilibria in which the policy outcome is not a Condorcet winner of the collective choice problem. For example, if every citizen prefers a to b but both candidates prefer b to a, then the action pair in which both candidates choose b is a Nash equilibrium. However, if the candidates are representative of the citizens in the sense that whenever all members of a majority of citizens prefer some alternative x to another alternative y, there is a candidate who prefers x to y, then in any Nash equilibrium in which the candidates' positions are the same, their common position is a Condorcet winner of the underlying collective choice problem (Proposition 9.2). In Nash equilibria in which the candidates' positions differ, however, these positions are not necessarily Condorcet winners.

Section 9.1.2 consider an analogue of an electoral competition game with a continuum of citizens and two office-motivated candidates in which the candidates are policy-motivated. Proposition 9.3 shows that the action pair in which each candidate's position is the median of citizens' favorite positions is a Nash equilibrium, and if the candidates are representative, the policy outcome of every Nash equilibrium is this position.

Suppose that we modify this model so that the candidates are uncertain about the median of the citizens' favorite alternatives. This model, unlike those I have discussed previously with the exception of the last model in Section 8.6.1, robustly has equilibria in which the candidates' positions differ. Suppose that the candidates' favorite positions differ, their common position is x, and candidate i's favorite position differs from x. Then if the probability that the median of the citizens' favorite positions lies between x and i's favorite position is positive, i can increase her payoff by deviating from x in the direction of her favorite position. If she does so, the worst that can happen for her is that the other candidate wins, in which case the outcome remains x; on the positive side, with positive probability she wins, in which case she is better off. Proposition 9.4 gives conditions under which the candidates' positions differ in every Nash equilibrium, and Proposition 9.5 gives conditions for the existence of a Nash equilibrium.

Section 9.3 explores models with two candidates and a single citizen in which the candidates are better informed than the citizen about the desirability of the policies. In Section 9.3.1 the candidates receive information about the desirability of the policies after they are elected. The game in this case may have equilibria in which each candidate offers an interval of policies, the winner choosing a specific policy after the uncertainty is resolved. In Section 9.3.2 the candidates are better informed than the citizen at the time of the election. In this case, the position taken by a candidate acts as a signal of her information. The game has equilibria in which the candidates choose the same position and this position does not depend on their information, and may also have equilibria in which the candidates choose the same position, equal to the citizen's favorite position given the candidates' information, and equilibria in which the candidates' positions differ and depend on their information.

Section 9.4 studies models of repeated elections. The same pair of policy-motivated candidates compete in a series of elections, in each period observing the outcomes in all previous periods. If in each period each candidate is free to choose any position, the features of the subgame perfect equilibria depend on the nature of the candidates' payoff functions. If these functions are concave—each candidate is more sensitive to changes in policies far from her favorite policy than she is to changes close to that policy—then only a repetition of the outcome in which both candidates choose the median m of the citizens' favorite positions is possible in a subgame perfect equilibrium. But if the candidates' payoff functions are convex on each side of their favorite positions and the discount factor is close enough to 1 then a subgame perfect equilibrium exists in which the outcome in each period differs from m. The assumption that each candidate may choose any position, unconstrained by her past positions, seems unreasonable. If it is replaced by the assumption that an incumbent in any period t is restricted to choose in period $t+1$ the policy she implemented in period t, then even if the candidates' payoff functions are concave the game has subgame perfect equilibria in which the winning policy differs from m in every period, the candidates alternating as winners.

9.1 Basic model

9.1.1 General set of positions

As in the models of the previous chapter, two candidates select positions, which I also refer to as policies, from a set X, then each member of a set of citizens votes for the candidate whose position she prefers, and the candidate who receives the most votes wins. As before, the election is decided by the citizens who have a strict preference between the candidates; if the number of citizens who prefer the position x_1 of candidate 1 to the position x_2 of candidate 2 is the same as the number who prefer x_2 to x_1, then the outcome of the election is a tie. (A citizen who is indifferent between the candidates' positions does not vote, or splits her vote, casting half a vote for each candidate.)

Each candidate has preferences over electoral outcomes. For any position $x \in X$, denote the outcome in which either a candidate with position x wins outright or x is the position of both candidates by $\{x\}$, and the outcome in which the candidates' positions are x and $y \neq x$ and the candidates tie by $\{x,y\}$. How do the candidates evaluate ties? I assume that (*i*) if a candidate likes the outcome $\{x\}$ at least as much as $\{y\}$ then she likes $\{x\}$ at least as much as $\{x,y\}$ and likes $\{x,y\}$ at least as much as $\{y\}$, and (*ii*) if she prefers $\{x\}$ to $\{z\}$ and likes $\{y\}$ at least as much as $\{z\}$ then she prefers $\{x,y\}$ to $\{z\}$. That is, the preference relation $\succsim_j^*$

over electoral outcomes of each candidate j satisfies

$$\begin{aligned} \{x\} \succsim_j^* \{y\} &\Rightarrow \{x\} \succsim_j^* \{x,y\} \succsim_j^* \{y\} \\ \{x\} \succ_j^* \{z\} \text{ and } \{y\} \succsim_j^* \{z\} &\Rightarrow \{x,y\} \succ_j^* \{z\}. \end{aligned} \tag{9.1}$$

These assumptions are consistent with the outcome of a tie $\{x,y\}$ being the lottery in which x and y each occur with probability $\frac{1}{2}$ and the candidates' preference relations over lotteries are vNM preference relations.

Definition 9.1: Electoral competition game with two policy-motivated candidates

An *electoral competition game with two policy-motivated candidates* $\langle\{1,2\},\langle I,X,\succsim\rangle,(\succsim_1^*,\succsim_2^*)\rangle$, where $\langle I,X,\succsim\rangle$ is a collective choice problem in which the set I (of *citizens*) is finite and $\succsim_j^*$ for $j=1,2$ is a preference relation over subsets of X containing one or two alternatives that satisfies (9.1), is the strategic game with the following components.

Players
$\{1,2\}$ (*candidates*).

Actions
The set of actions of each player is X (the set of possible *positions*).

Preferences
The preference relation $\trianglerighteq_j$ of each player j over pairs of positions satisfies

$$(x_1,x_2) \trianglerighteq_j (y_1,y_2) \iff W(x_1,x_2) \succsim_j^* W(y_1,y_2),$$

where for each pair of positions $(x_1,x_2) \in X \times X$, $W(x_1,x_2)$ is the set of members of $\{x_1,x_2\}$ preferred by a majority of citizens:

$$W(x_1,x_2) = \begin{cases} \{x_1\} & \text{if } |\{i \in I : x_1 \succ_i x_2\}| > |\{i \in I : x_2 \succ_i x_1\}| \\ \{x_1,x_2\} & \text{if } |\{i \in I : x_1 \succ_i x_2\}| = |\{i \in I : x_2 \succ_i x_1\}| \\ \{x_2\} & \text{if } |\{i \in I : x_1 \succ_i x_2\}| < |\{i \in I : x_2 \succ_i x_1\}|. \end{cases}$$

Suppose that the collective choice problem $\langle I,X,\succsim\rangle$ has a strict Condorcet winner, x^*. If both candidates in the electoral competition game choose this position, it is the policy outcome. If either candidate deviates from x^*, she loses, so that x^* remains the policy outcome. Hence (x^*,x^*) is a Nash equilibrium of the game, as it is when the candidates are office-motivated (Proposition 8.1).

Proposition 9.1: Nash equilibrium of electoral competition game with two policy-motivated candidates

Let $\langle\{1,2\},\langle I,X,\succsim\rangle,(\succsim_1^*,\succsim_2^*)\rangle$ be an electoral competition game with two policy-motivated candidates. If $\langle I,X,\succsim\rangle$ has a strict Condorcet winner, x^*, then (x^*,x^*) is a Nash equilibrium of the game.

Proof

We have $W(x^*,x^*)=\{x^*\}$ and $W(x,x^*)=\{x^*\}$ for any $x\neq x^*$, because x^* is a strict Condorcet winner of the collective choice problem. Thus neither candidate can profitably deviate from x^*.

If the number of citizens is odd and their preferences are single-peaked or single-crossing then the collective choice problem has a strict Condorcet winner and this position is the median of the citizens' favorite positions if the preferences are single-peaked (Proposition 1.4) and the favorite position of the median citizen if the preferences are single-crossing (Proposition 1.5), so the following result follows from Proposition 9.1. Note that this result is weaker than the corresponding result (Corollary 8.2) for the game with office-motivated candidates: it says only that a certain action pair is a Nash equilibrium, not that it is the only Nash equilibrium.

Corollary 9.1: Median voter theorem for electoral competition with two policy-motivated candidates

Let $\langle\{1,2\},\langle I,X,\succsim\rangle,(\succsim_1^*,\succsim_2^*)\rangle$ be an electoral competition game with two policy-motivated candidates for which the number of citizens (members of I) is odd.

- If $\langle I,X,\succsim\rangle$ has single-peaked preferences with respect to a linear order $\trianglerighteq$ on X, then the action pair (m,m) in which m is the median with respect to $\trianglerighteq$ of the citizens' favorite positions is a Nash equilibrium of the game.

- If $\langle I,X,\succsim\rangle$ has single-crossing preferences with respect to a linear order $\geq$ on I and the median individual with respect to $\geq$ has a unique favorite position, m, then the action pair (m,m) is a Nash equilibrium of the game.

When the candidates are office-motivated, a pair (x_1,x_2) is a Nash equilibrium if and only if both x_1 and x_2 are Condorcet winners (Proposition 8.1). The same is not true when the candidates are policy-motivated. If x^* is a Condorcet winner that is not strict then another alternative, say x, ties with x^*, so that a candidate who deviates from x^* to x induces a tie between x^* and x. If she prefers x to x^*, then by (9.1) she prefers a tie to x^* . So if at least one of the candidates prefers x to x^*, the action pair (x^*, x^*) is not a Nash equilibrium.

Further, if (x_1, x_2) is a Nash equilibrium, x_1 and x_2 are not necessarily Condorcet winners, even if $x_1 = x_2$. Consider a collective choice problem in which there are two alternatives, a and b, and all individuals prefer a to b, so that a is the strict Condorcet winner. If each candidate prefers b to a, then (b, b) is a Nash equilibrium of the associated electoral competition game with two policy-motivated candidate. Another example is based on the Condorcet cycle in Example 1.5, which has no Condorcet winner. Suppose that the preference relation $\succsim_j^*$ of each candidate j in the electoral competition game satisfies $\{a\} \succ_j^* \{b\} \succ_j^* \{c\}$. Then (a, a) is a Nash equilibrium: if either candidate deviates to b, she loses, so that the outcome remains $\{a\}$, and if either candidate deviates to c, she wins, so that the outcome changes to $\{c\}$, which is worse for her than $\{a\}$.

In these examples, the preferences of some sets of citizens containing a majority of individuals are not shared by any candidate: in the first case, all citizens prefer a to b but both candidates prefer b to a, and in the second example a majority of the citizens prefer c to a but both candidates prefer a to c. If for every majority of citizens who prefer some alternative x to some other alternative y there is a candidate with the same preference between x and y, I say that the candidates are representative.

Definition 9.2: Representative candidates in electoral competition game with two policy-motivated candidates

The candidates in an electoral competition game with two policy-motivated candidates are *representative* if, for all alternatives x and y, whenever every citizen in some set containing a majority of citizens prefers x to y, at least one candidate prefers $\{x\}$ to $\{y\}$.

For games in which the candidates are representative, their common position in any Nash equilibrium in which their positions are the same is a Condorcet winner of the collective choice problem.

Proposition 9.2: Nash equilibrium of electoral competition game with two representative policy-motivated candidates

Let $\langle\{1,2\},\langle I,X,\succsim\rangle,(\succsim_1^*,\succsim_2^*)\rangle$ be an electoral competition game with two policy-motivated candidates in which the candidates are representative. In any Nash equilibrium in which the candidates' positions are the same, the common position is a Condorcet winner of $\langle I,X,\succsim\rangle$.

Proof

Let $x \in X$ and consider the action pair (x,x). If x is not a Condorcet winner of $\langle I,X,\succsim\rangle$ then for some position, say x', a majority of citizens prefer x' to x. Given that the candidates are representative, at least one candidate thus prefers x' to x. If that candidate deviates to x', she wins, so that the outcome is $\{x'\}$, which she prefers to $\{x\}$. Hence (x,x) is not a Nash equilibrium.

What about Nash equilibria in which the candidates' positions differ? The game can have such equilibria in which neither candidate's positions is a Condorcet winner. Consider again the Condorcet cycle in Example 1.5. Suppose that candidate 1's preferences satisfy $\{b\} \succ_1^* \{a\} \succ_1^* \{c\}$ and candidate 2's satisfy $\{c\} \succ_2^* \{a\} \succ_2^* \{b\}$. The action pair (b,a), for which neither action is a Condorcet winner, is a Nash equilibrium by the following argument. Candidate 2 wins (a beats b), so that the outcome is $\{a\}$. If candidate 1 deviates to a, the outcome remains $\{a\}$; if she deviates to c, she wins and the outcome changes to $\{c\}$, which is worse for her than $\{a\}$. If candidate 2 deviates to b, the outcome changes to $\{b\}$, which is worse for her than $\{a\}$; if she deviates to c, candidate 1 wins and the outcome again changes to $\{b\}$.

A small change in the candidates' preferences eliminates this equilibrium. Assume that each candidate cares mainly about the policy outcome, but slightly about winning. Precisely, for any position x, among action pairs that generate the outcome $\{x\}$, each candidate prefers those in which she wins to those in which she ties to those in which she loses. That is, her preferences are lexicographic: if two action pairs have different policy outcomes, candidate i (= 1, 2) prefers the one that is better according to $\succsim_i^*$, while if they have the same policy outcome, she prefers winning to tieing to losing. Under this assumption, the game has no Nash equilibrium in which one candidate loses, because if the losing candidate deviates to the position of the winning candidate then the policy outcome remains the same and the deviating candidate ties rather than loses.

However, even if the candidates are representative and their preferences put some weight on winning, an electoral competition game may have equilibria in which the outcome is a tie and one of the possible outcomes is not a Condorcet winner, as you are asked to show in the following exercise.

Exercise 9.1: Nash equilibrium with policy-motivated candidates

Consider the collective choice problem with four individuals in which three have the preferences of the individuals in the Condorcet cycle in Example 1.5 and the fourth prefers b to c to a. Show that a is not a Condorcet winner of the collective choice problem but for some preferences of the candidates the action pair (a,b) is a Nash equilibrium of the associated electoral competition game with two policy-motivated candidates, even if the candidates are representative and their preferences lexicographically value winning.

If the number of citizens is odd and their preferences are strict, a tie when the candidates choose different positions is not possible. So if in this case the candidates are representative and their preferences lexicographically value winning, then in any Nash equilibrium the candidates choose the same position, and hence, by Proposition 9.2, this position is a Condorcet winner.

9.1.2 One-dimensional positions

Consider a variant of an electoral competition game with two policy-motivated candidates in which, as in an electoral competition game with a continuum of citizens and two office-motivated candidates, the set of positions is an interval of numbers and the set of citizens is a continuum. This variant does not include an explicit specification of the set of citizens, but like its cousin includes an outcome function that may be rationalized by an assumption about the citizens' preferences. The outcome relevant to policy-motivated candidates is the position of the winner, so the outcome function in this case specifies that position (or those positions, in the case of a tie), rather than the identity of the winner.

Assume specifically that the preference relation $\succsim_i$ of each citizen i is single-peaked with respect to $\geq$ and symmetric about i's favorite position x_i^* ($x_i^* - \delta \sim_i x_i^* + \delta$ for every $\delta > 0$), and the distribution F of the citizens' favorite positions is nonatomic, with support an interval. Under these assumptions, F has a unique median, say m, and the policy outcome of the pair (x_1, x_2) of the candidates' positions is

$$Y_F(x_1,x_2) = \begin{cases} \{x\} & \text{if } x_1 = x_2 = x \\ \{x_1,x_2\} & \text{if } x_1 \neq x_2 \text{ and } \frac{1}{2}(x_1+x_2) = m \\ \{x_j\} & \text{if } \begin{cases} \text{either } x_k < x_j \text{ and } \frac{1}{2}(x_1+x_2) < m \\ \text{or } x_k > x_j \text{ and } \frac{1}{2}(x_1+x_2) > m, \end{cases} \end{cases} \tag{9.2}$$

where $j \in \{1,2\}$ and k is the other candidate. (The function Y_F is the analog of the function O_F defined in (8.1) for the model with office-motivated candidates.)

Regarding the candidates, we assume that each candidate j has a preference relation $\succsim_j^*$ over policy outcomes (one- or two-member subsets of the set X of positions) that is single-peaked in the sense that

$$\text{for some } \hat{x}_j \in X: \quad x < y < \hat{x}_j \text{ or } \hat{x}_j < y < x \quad \Rightarrow \quad \{\hat{x}_j\} \succ_j^* \{y\} \succsim_j^* \{x\}. \tag{9.3}$$

The position $\hat{x}_j$ is the favorite position of candidate j.

Definition 9.3: Electoral competition game with continuum of citizens and two policy-motivated candidates

An *electoral competition game with a continuum of citizens and two policy-motivated candidates* $\langle\{1,2\}, X, F, (\succsim_1^*, \succsim_2^*)\rangle$, where X is a closed interval of real numbers, F is a nonatomic distribution with support X, and $\succsim_1^*$ and $\succsim_2^*$ are preference relations over subsets of X containing one or two positions that satisfy (9.1) and (9.3), is the strategic game with the following components.

Players

$\{1,2\}$ (*candidates*).

Actions

The set of actions of each player is X (the set of possible *positions*).

Preferences

The preference relation $\trianglerighteq_j$ of each player j over $X \times X$ satisfies

$$(x_1,x_2) \trianglerighteq_j (y_1,y_2) \quad \Leftrightarrow \quad Y_F(x_1,x_2) \succsim_j^* Y_F(y_1,y_2),$$

where Y_F is given by (9.2).

For the model with a finite number of citizens, the candidates are representative if whenever a majority of citizens prefer one position to another, so does at least one candidate (Definition 9.2). The analog of this definition for the model with a continuum of citizens is that the candidates' preferences are symmetric about their favorite positions, with one candidate's favorite position on each side of the median of the citizens' favorite positions.

Definition 9.4: Representative candidates in electoral competition game with continuum of citizens and two policy-motivated candidates

The candidates in an electoral competition game with a continuum of citizens and two policy-motivated candidates $\langle\{1,2\}, X, F, (\succsim_1^*, \succsim_2^*)\rangle$ are *representative* if the preference relation $\succsim_i^*$ of each candidate i is symmetric about her favorite position (she is indifferent between positions equidistant from her favorite position), the favorite position of one candidate is at most m, and the favorite position of the other candidate is at least m, where m is the median of the distribution F of the citizens' favorite positions.

The action pair in which each candidate's position is the median of the citizens' favorite positions is a Nash equilibrium, an analogue of Corollary 9.1. In addition, if the candidates are representative then the outcome of every Nash equilibrium is this position.

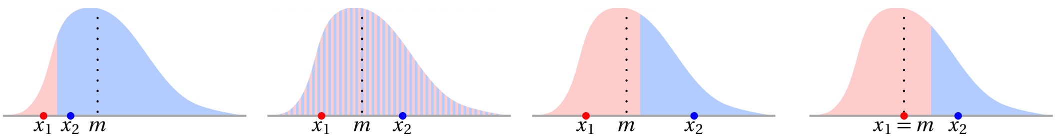
Proposition 9.3: Nash equilibrium of electoral competition game with continuum of citizens and two policy-motivated candidates

Let $\langle\{1,2\}, X, F, (\succsim_1^*, \succsim_2^*)\rangle$ be an electoral competition game with a continuum of citizens and two policy-motivated candidates. Denote the candidates' favorite positions, defined by (9.3), by $\hat{x}_1$ and $\hat{x}_2$. The action pair in which each candidate's position is the median m of the distribution F of the citizens' favorite positions is a Nash equilibrium of the game. If the candidates are representative then the outcome of every Nash equilibrium is $\{m\}$, and if in addition $\hat{x}_1 < m < \hat{x}_2$ then (m, m) is the only Nash equilibrium.

Proof

The fact that the support of F is an interval means that it has a unique median m. The outcome of the action pair (m, m) is $Y_F(m, m) = \{m\}$. If either candidate deviates from m, she loses, and the outcome remains $\{m\}$. Thus (m, m) is a Nash equilibrium.

Now assume that the candidates are representative, and let (x_1, x_2) be a



(a) $x_1 \le x_2 < m$ (b) $x_1 < m < x_2$ and $\frac{1}{2}(x_1+x_2) = m$ (c) $x_1 < m < x_2$ and $\frac{1}{2}(x_1+x_2) > m$ (d) $x_1 = m < x_2$

Figure 9.1 The four cases in the proof of Proposition 9.3.

Nash equilibrium. Assume without loss of generality that $x_1 \le x_2$. Given that the candidates are representative, $\hat{x}_1 \le m \le \hat{x}_2$.

First suppose that $\hat{x}_1 < m < \hat{x}_2$. Refer to Figure 9.1 for illustrations of the following four cases.

- If $x_1 \le x_2 < m$ then $Y_F(x_1, x_2) = \{x_2\}$. If candidate 2 deviates to m, the outcome changes to $\{m\}$, which she prefers to $\{x_2\}$. A symmetric argument applies if $m < x_1 \le x_2$.
- If $x_1 < m < x_2$ and $\frac{1}{2}(x_1 + x_2) = m$ then $Y_F(x_1, x_2) = \{x_1, x_2\}$. If candidate 1 deviates to $x_1 + \varepsilon$ with $0 < \varepsilon < x_2 - x_1$ then she wins and the outcome changes to $\{x_1 + \varepsilon\}$, which she prefers if ε is sufficiently small. (Candidate 2 has an analogous profitable deviation.)
- If $x_1 < m < x_2$ and $\frac{1}{2}(x_1 + x_2) > m$ then $Y_F(x_1, x_2) = \{x_1\}$. If candidate 2 deviates to m, the outcome changes to $\{m\}$, which she prefers to $\{x_1\}$. A symmetric argument applies if $\frac{1}{2}(x_1 + x_2) < m$.
- If $x_1 = m < x_2$ then $Y_F(x_1, x_2) = \{x_1\}$. If candidate 1 deviates to $m - \varepsilon$ with $0 < \varepsilon < x_2 - m$ then she wins and the outcome changes to $\{m - \varepsilon\}$, which she prefers if ε is sufficiently small. A symmetric argument applies if $x_1 < m = x_2$.

Thus no pair (x_1, x_2) other than (m, m) is a Nash equilibrium.

Now suppose that $\hat{x}_1 = m \le \hat{x}_2$. If the outcome of (x_1, x_2) is not $\{m\}$, then $x_1 \ne m$ and candidate 1 can induce the outcome $\{m\}$, her favorite outcome, by deviating to m. Thus in any Nash equilibrium the outcome is $\{m\}$. A symmetric argument applies if $\hat{x}_1 \le m = \hat{x}_2$.

Note that for the case that the candidates are representative, the result says only that the outcome of every Nash equilibrium is $\{m\}$, not that the only Nash equilibrium is (m, m). If, for example, candidate 1's favorite position is m and candidate 2's is greater than m, then the action pair $(m, \hat{x}_2)$ is also a Nash equi-

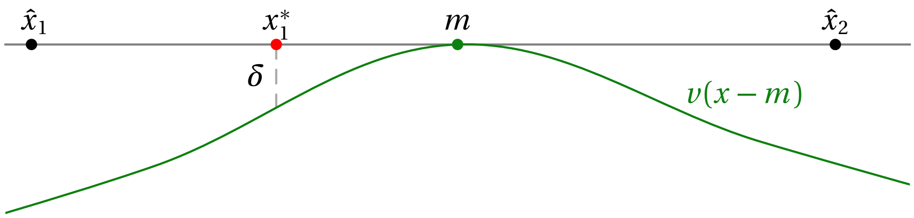


Figure 9.2 The policy x_1^* in the model in Exercise 9.3.

librium (for all positions of candidate 2, the outcome is $\{m\}$). If we modify each candidate's preferences so that they lexicographically value winning in the way discussed at the end of Section 9.1.1, then (m, m) is the only Nash equilibrium.

Exercise 9.2: Candidates' favorite positions both less than the median

Find the Nash equilibria of an electoral competition game with a continuum of citizens and two policy-motivated candidates in which both candidates' favorite positions are less than the median of the citizens' favorite positions (so that in particular the candidates are not representative).

For the case in which the candidates are office-motivated, Section 8.5.1 presents a model in which the citizens have preferences over both candidates and policies. The next exercise asks you to study an analogue of this model for the case in which the candidates are policy-motivated.

Exercise 9.3: Electoral competition with an advantaged candidate

Suppose that the citizens have preferences over the candidates independently of the candidates' positions, as in Section 8.5.1. Specifically, consider the model that differs from the one in Exercise 8.5 only in that each candidate is policy-motivated rather than office-motivated, with preferences that satisfy (9.1) and (9.3). Denote the favorite position of each citizen i by $\hat{z}_i$, the median of these favorite positions by m, and the candidates' favorite positions by $\hat{x}_1$ and $\hat{x}_2$. Assume that $\hat{x}_1 < m < \hat{x}_2$ and $v(\hat{x}_1 - m) < -\delta$. Let x_1^* be the position for which $x_1^* < m$ and $v(x_1^* - m) = -\delta$, so that a citizen whose favorite position is m is indifferent between the candidates when candidate 1's position is x_1^* and candidate 2's is m, and hence votes for candidate 1 in this case, given the tie-breaking assumption. (Refer to Figure 9.2.) Show that (x_1^*, m) is a Nash equilibrium of the game.

The models in this section, like those in the previous chapter, assume that after a candidate is elected, she implements the policy she chose as her platform in the election. The need to periodically face re-election may provide office-holders with an incentive not to deviate from their stated policies, and the existence of parties may reinforce this incentive. But a candidate may be unable to fully commit that, if she wins, she will implement the policy she chose in the election. One way to model this inability to commit is to assume that if a candidate j who chooses a position x_j wins, then the outcome is x_j with some probability $p_j < 1$ and j's favorite policy with probability $1 - p_j$. Suppose specifically that the number of citizens is odd, the citizen's favorite positions are distinct, and each citizen i has preferences on the set of lotteries over positions represented by the expected value of a function $u_i : X \to \mathbb{R}$ defined by $u_i(x) = v(x - \hat{z}_i)$, where v is a single-peaked function v with maximizer 0. Suppose also that there are two candidates, with favorite positions $\hat{x}_1$ and $\hat{x}_2$ that satisfy $\hat{x}_1 < m < \hat{x}_2$, where m is the median of the citizens' favorite positions. Finally, suppose that given p_1 and p_2, the citizen with favorite position m prefers candidate 1 when the pair of positions chosen by the candidates is (m, m) and candidate 2 when this pair is $(\hat{x}_1, m)$, and votes for candidate 1 when indifferent between the candidates. This model is closely related to the one in Exercise 9.3, and similar arguments lead to the conclusion that (x_1^*, m) is a Nash equilibrium, where x_1^* is the position in $(\hat{x}_1, m)$ for which the citizen with favorite position m is indifferent between the candidates:

$$p_1 v(x_1^* - m) + (1 - p_1) v(\hat{x}_1 - m) = p_2 v(0) + (1 - p_2) v(\hat{x}_2 - m).$$

The conclusion of Proposition 9.3 that the outcome of any equilibrium is the citizens' median favorite position when the candidates are representative, policy-motivated, viewed as interchangeable by the citizens, and committed to the policies they announce, even if their favorite positions are extreme, may seem surprising. The incentive for office-motivated candidates to cater to the median voter is clear, but you might think that a policy-motivated candidate faces a tradeoff: moving from her favorite position towards her rival's position increases her probability of winning, but results in a less desirable position if she wins. The models I have defined do not capture this tradeoff because they are deterministic: a candidate's probability of winning is either 0 or 1, or the outcome is a tie. I now specify and analyze a model that does capture the tradeoff.

9.2 Uncertain median

Suppose that the candidates are uncertain about the citizens' preferences. Specifically, consider a variant of the game in Section 8.3.1 in which the candidates are policy-motivated rather than office-motivated. Each candidate believes that the median of the citizens' favorite positions has the distribution function G, which

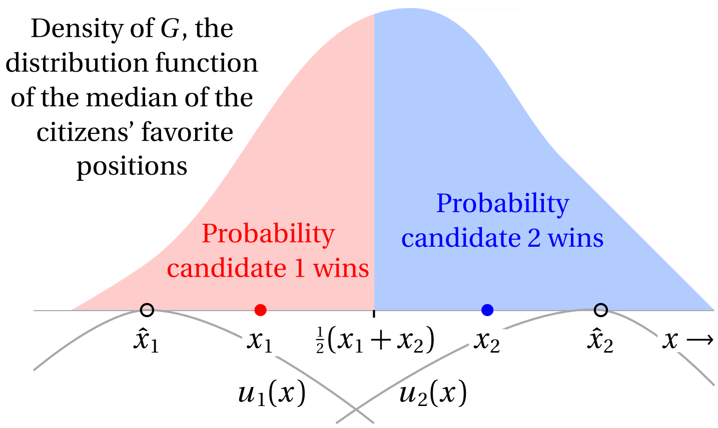


Figure 9.3 An illustration of the components of an electoral competition game with two policy-motivated candidates and uncertain median.

is nonatomic, with support an interval. Suppose that the candidates' positions are x_1 and x_2, with $x_1 < x_2$. Then if the median of the citizens' favorite positions is less than $\frac{1}{2}(x_1+x_2)$, an event with probability $G(\frac{1}{2}(x_1+x_2))$, candidate 1 wins and the policy outcome is x_1; if the median is greater than $\frac{1}{2}(x_1+x_2)$, an event with probability $1-G(\frac{1}{2}(x_1+x_2))$, candidate 2 wins and the policy outcome is x_2. Thus each candidate faces a lottery in which the outcome is x_1 with probability $G(\frac{1}{2}(x_1+x_2))$ and x_2 with probability $1-G(\frac{1}{2}(x_1+x_2))$.

Suppose that the preferences of candidate j $(= 1, 2)$ regarding probability distributions over positions are represented by the expected value of a single-peaked function u_j, with favorite position $\hat{x}_j$. (Refer to Figure 9.3.) If $x_1 < x_2$ then candidate j's expected payoff is

$$G(\tfrac{1}{2}(x_1+x_2))u_j(x_1)+\big(1-G(\tfrac{1}{2}(x_1+x_2))\big)u_j(x_2).$$

If $x_1 > x_2$ then $u_j(x_1)$ and $u_j(x_2)$ are interchanged in this expression, and if $x_1 = x_2 = x$ then j's payoff is $u_j(x)$.

Definition 9.5: Electoral competition game with two policy-motivated candidates and uncertain median

An *electoral competition game with two policy-motivated candidates and uncertain median* $\langle\{1,2\},X,G,(u_1,u_2)\rangle$, where X is a closed interval of real numbers, G is a nonatomic distribution with a density and support X (so that it has a unique median), and $u_j : X \rightarrow \mathbb{R}$ for $j = 1, 2$ is a single-peaked function, is the strategic game with the following components.

Players

$\{1,2\}$ (*candidates*).

Actions

The set of actions of each player is X (the set of possible *positions*).

Preferences

The preferences of each player j are represented by the function $v_j : X \times X \to \mathbb{R}$ defined by

$$v_j(x_1,x_2) = G(\tfrac{1}{2}(x_1+x_2))u_j(\min\{x_1,x_2\}) + \left(1 - G(\tfrac{1}{2}(x_1+x_2))\right)u_j(\max\{x_1,x_2\}).$$

If the candidates' positions in such a game are the same, equal to x, then the outcome is x. If the candidates' favorite positions differ, x is not the favorite position of at least one candidate, and that candidate's deviating from x to her favorite position causes the outcome to change to one in which her favorite position occurs with positive probability and x occurs with the complementary probability, which she prefers to x. Thus in any Nash equilibrium of the game the candidates' positions differ. The next result shows also that these positions lie between the candidates' favorite positions, and each candidate's equilibrium position is closer to her favorite position than is the other candidate's equilibrium position.

Proposition 9.4: Nash equilibrium of electoral competition game with two policy-motivated candidates and uncertain median

Consider an electoral competition game with two policy-motivated candidates and uncertain median $\langle\{1,2\},X,G,(u_1,u_2)\rangle$. Denote the candidates' favorite positions, the maximizers of u_1 and u_2, by $\hat{x}_1$ and $\hat{x}_2$, and suppose that $\hat{x}_1 < \hat{x}_2$. Then in every Nash equilibrium (x_1^*,x_2^*) we have $\hat{x}_1 \le x_1^* < x_2^* \le \hat{x}_2$. If G is differentiable, the density of G is positive on the interior of X, u_1 and u_2 are differentiable, and $\hat{x}_1$ and $\hat{x}_2$ are in the interior of X, then $\hat{x}_1 < x_1^* < x_2^* < \hat{x}_2$.

To prove this result, I first establish the following lemma, which is used also in the proof of a later result.

Lemma 9.1: Best responses in electoral competition game with two policy-motivated candidates and uncertain median

Consider an electoral competition game with two policy-motivated candidates and uncertain median $\langle\{1,2\}, X, G, (u_1, u_2)\rangle$. For any $x_2 \in X$, every best response of candidate 1 to x_2 is in $(x_2, \hat{x}_1]$ if $x_2 < \hat{x}_1$ and in $[\hat{x}_1, x_2)$ if $x_2 > \hat{x}_1$, where $\hat{x}_1$ is candidate 1's favorite position.

Proof

First suppose that $x_2 < \hat{x}_1$.

a. If $x_1 \le x_2$ then candidate 1's payoff is at most $u_1(x_2)$. If she deviates to $\hat{x}_1$ then the outcome is x_2 with positive probability less than 1 and $\hat{x}_1$ with the complementary probability. Given that $u_1(\hat{x}_1) > u_1(x_2)$, the position x_1 is thus not a best response to x_2.

b. If $x_1 > \hat{x}_1$ then candidate 1's payoff is

$$G(\tfrac{1}{2}(x_1 + x_2))u_1(x_2) + \big(1 - G(\tfrac{1}{2}(x_1 + x_2))\big)\,u_1(x_1). \tag{9.4}$$

A deviation by candidate 1 to $\hat{x}_1$ changes her payoff to

$$G(\tfrac{1}{2}(\hat{x}_1 + x_2))u_1(x_2) + \big(1 - G(\tfrac{1}{2}(\hat{x}_1 + x_2))\big)\,u_1(\hat{x}_1).$$

The difference between this payoff and (9.4) is

$$\begin{aligned}\big(G(\tfrac{1}{2}(x_1 + x_2)) - G(\tfrac{1}{2}(\hat{x}_1 + x_2))\big)&(u_1(\hat{x}_1) - u_1(x_2)) \\ &+ \big(1 - G(\tfrac{1}{2}(x_1 + x_2))\big)(u_1(\hat{x}_1) - u_1(x_1)),\end{aligned}$$

which is positive because $G(\tfrac{1}{2}(x_1 + x_2)) > G(\tfrac{1}{2}(\hat{x}_1 + x_2))$, $u_1(\hat{x}_1) > u_1(x_1)$, and $u_1(\hat{x}_1) > u_1(x_2)$. Thus x_1 is not a best response to x_2.

We conclude that if $x_2 < \hat{x}_1$ then every best response of candidate 1 to x_2 is in $(x_2, \hat{x}_1]$.

Now suppose that $x_2 > \hat{x}_1$. If $x_1 < \hat{x}_1$ or $x_1 \ge x_2$ then a deviation by candidate 1 to $\hat{x}_1$ increases her payoff; the argument for $x_1 < \hat{x}_1$ is symmetric with that for case *b* for $x_2 < \hat{x}_1$, and the argument for $x_1 \ge x_2$ is symmetric with that for case *a*.

Proof of Proposition 9.4

Lemma 9.1 implies that in every Nash equilibrium (x_1^*, x_2^*) we have $\hat{x}_1 \leq x_1^* < x_2^* \leq \hat{x}_2$. Now assume that G is differentiable, the density of G is positive on the interior of X, u_1 and u_2 are differentiable, and $\hat{x}_1$ and $\hat{x}_2$ are in the interior of X. We need to prove that $x_1^* \neq \hat{x}_1$ and $x_2^* \neq \hat{x}_2$. Given that u_1 is differentiable and $\hat{x}_1$ is in the interior of X, $u_1'(\hat{x}_1) = 0$, so the derivative of candidate 1's payoff with respect to x_1 at $(\hat{x}_1, x_2)$ for $x_2 > \hat{x}_1$ is

$$\tfrac{1}{2}G'(\tfrac{1}{2}(\hat{x}_1 + x_2))(u_1(\hat{x}_1) - u_1(x_2)).$$

This expression is positive given that the density G' of G is positive on the interior of X. Thus candidate 1's best response to any position $x_2 > \hat{x}_1$ is greater than $\hat{x}_1$ and hence $x_1^* > \hat{x}_1$ in any Nash equilibrium. A symmetric argument applies to x_2^*.

If the candidates' payoff functions are not differentiable at their favorite positions then the game may have an equilibrium in which the candidates choose these positions.

Exercise 9.4: Nash equilibrium with policy-motivated candidates and uncertainty

Consider an electoral competition game with two policy-motivated candidates and uncertain median $\langle\{1,2\}, X, G, (u_1, u_2)\rangle$ in which $X = [-k, k]$ for some $k > 0$, G is uniform on X, and for some positions $\hat{x}_1 \in X$ and $\hat{x}_2 \in X$ with $\hat{x}_1 < \hat{x}_2$ we have $u_j(x) = -|x - \hat{x}_j|$ for $j = 1, 2$ for all $x \in X$. Show that this game has a Nash equilibrium in which the position of each candidate j is $\hat{x}_j$, her favorite position.

Proposition 9.4 does not assert that the game necessarily has a Nash equilibrium. The next result gives sufficient conditions for the existence of a Nash equilibrium when G and the payoff functions u_1 and u_2 are differentiable.

Proposition 9.5: Existence of Nash equilibrium for electoral competition game with two policy-motivated candidates and uncertainty

Consider an electoral competition game with two policy-motivated candidates and uncertain median $\langle\{1,2\}, X, G, (u_1, u_2)\rangle$ for which G is differentiable and the candidates' favorite positions, the maximizers $\hat{x}_1$ and $\hat{x}_2$

of u_1 and u_2, differ and are in the interior of X. If $\log G$ is concave and u_1 and u_2 are concave and twice-differentiable then the game has a Nash equilibrium.

Proof

Assume without loss of generality that $\hat{x}_1 < \hat{x}_2$ and denote the players' payoff functions by v_1 and v_2, as in Definition 9.5.

Step 1 *For any $x_2 \in [\hat{x}_1, \hat{x}_2]$, candidate 1 has a unique best response to x_2, which is in $[\hat{x}_1, \hat{x}_2]$.*

Proof. First suppose that $x_2 = \hat{x}_1$. If candidate 1 chooses the same position, then the outcome is $\hat{x}_1$ with certainty. If she chooses any other position x_1, then given that the support of G is X, the outcome is x_1 with positive probability and $\hat{x}_1$ with the complementary probability. Thus her unique best response to x_2 is $\hat{x}_1$.

Now suppose that $x_2 \in (\hat{x}_1, \hat{x}_2]$. By Lemma 9.1, every best response of candidate 1 to x_2 is in $[\hat{x}_1, x_2)$. Her payoff to a pair (x_1, x_2) with $x_1 < x_2$ is

$$G(\tfrac{1}{2}(x_1 + x_2))u_1(x_1) + \left(1 - G(\tfrac{1}{2}(x_1 + x_2))\right)u_1(x_2).$$

This function is differentiable in x_1, so a best response of candidate 1 to x_2, which is in the interior of X by Lemma 9.1 and the assumption that $\hat{x}_1$ is in the interior of X, satisfies

$$\tfrac{1}{2}G'(\tfrac{1}{2}(x_1 + x_2))u_1(x_1) + G(\tfrac{1}{2}(x_1 + x_2))u_1'(x_1) - \tfrac{1}{2}G'(\tfrac{1}{2}(x_1 + x_2))u_1(x_2) = 0$$

or

$$\tfrac{1}{2}G'(\tfrac{1}{2}(x_1 + x_2))(u_1(x_1) - u_1(x_2)) + G(\tfrac{1}{2}(x_1 + x_2))u_1'(x_1) = 0,$$

which implies

$$\frac{G'(\tfrac{1}{2}(x_1 + x_2))}{2G(\tfrac{1}{2}(x_1 + x_2))} = \frac{-u_1'(x_1)}{u_1(x_1) - u_1(x_2)}.$$

The left-hand side of this equation is the derivative of $\log G(\tfrac{1}{2}(x_1 + x_2))$, which is positive and nonincreasing in x_1 by the assumption that $\log G$ is concave. The right-hand side of the equation is 0 for $x_1 = \hat{x}_1$ and the sign of its derivative with respect to x_1 is the sign of

$$(u_1(x_1) - u_1(x_2))(-u_1''(x_1)) + (u_1'(x_1))^2,$$

which is positive for $\hat{x}_1 < x_1 < x_2$ because $u_1(x_1) > u_1(x_2)$, $u_1''(x_1) \leq 0$ by the concavity of u_1, and $u_1'(x_1) < 0$. Thus candidate 1's best response to x_2, which is less than x_2 by Lemma 9.1, is unique and greater than $\hat{x}_1$. ◁

Step 2 *The game has a Nash equilibrium.*

Proof. By Step 1 and the analogous result for candidate 2, each candidate has a unique best response to each position of the other candidate in $[\hat{x}_1, \hat{x}_2]$. So the continuity of the payoffs implies that the best response of each candidate is continuous in the other candidate's action. By Proposition 9.4, any Nash equilibrium of the game is a Nash equilibrium of the game that differs only in that each candidate's set of actions is $[\hat{x}_1, \hat{x}_2]$. Denote candidate i's best response function in this game by $B_i : [\hat{x}_1, \hat{x}_2] \to [\hat{x}_1, \hat{x}_2]$. The action pair (x_1, x_2) is a Nash equilibrium of the game if and only if $x_1 = B_1(x_1, x_2)$ and $x_2 = B_2(x_1, x_2)$. Given the convexity and compactness of $[\hat{x}_1, \hat{x}_2]$ and the continuity of B_1 and B_2, Brouwer's fixed point theorem implies that these equations have a solution, so that the game has a Nash equilibrium. ◁

The model nicely captures the tradeoff a candidate faces when choosing a position: moving her position away from her favorite position, towards that of her rival, makes her worse off if she wins, but increases her probability of winning. If the model has an equilibrium, then the candidates' positions differ. A limitation of the model is that the conditions under which an equilibrium is known to exist are relatively restrictive.

9.3 Candidates privately informed about policies

Suppose that the candidates are privately informed about the appropriateness of the possible policies. We can model the private information by assuming that the candidates' preferences over policies depend on an unknown state of the world, and that each candidate gets a signal that depends probabilistically on the state. In this section I discuss informally two models that differ in the timing of the information.

In each model, two candidates, 1 and 2, simultaneously choose positions and then a single citizen votes for one of the candidates. The set of possible positions is denoted X and the set of states Θ. The preferences of each candidate j regarding probability distributions over X are represented by the expected value of a function $u_j : X \times \Theta \to \mathbb{R}$ and the preferences of the citizen are represented by the expected value of a function $v : X \times \Theta \to \mathbb{R}$.

9.3.1 Candidates privately informed after choosing positions

Suppose that the winning candidate receives information about the desirability of the possible policies after she is elected. Specifically, suppose that the desirability of each policy depends on the state, which is observed only by the winning candidate. Then one possibility is that rather than committing to a fixed policy, each candidate specifies a mapping between states and policies and commits to carry out the policy specified by the mapping after she learns the state, in the event she wins. But I assume that such a commitment is not possible. Instead, each candidate chooses a subset of X and commits to select a policy in this subset in the event she wins. That is, the model I analyze is the extensive game in which the candidates simultaneously choose subsets of X (*platforms*), rather than single positions, the citizen votes for a candidate, and then the winning candidate observes the state and chooses a policy in her platform.

To start with an extreme example, suppose that the preferences of the candidates and the citizens are identical. Then the game has a subgame perfect equilibrium in which each candidate's platform is the set X of all positions. The citizen votes for either of the candidates and the winning candidate chooses the policy that is best for her given the state. Given that the candidates' and citizen's preferences are the same, neither candidate can do better by deviating to a subset of X. Following such a deviation, voting for the other (non-deviating) candidate remains optimal for the citizen. By the same logic, in every equilibrium the winning candidate's platform contains all the policies that are optimal for some state.

If the candidates' and citizen's preferences differ, whether the candidates' equilibrium platforms contain one policy or many depends on the nature of the preferences and the distribution of the state. To get an idea of the factors involved, suppose that the set X of available policies is a compact interval $[\underline{x}, \overline{x}]$ that contains 0, the set Θ of states coincides with X, and there is a single-peaked function $u : \mathbb{R} \to \mathbb{R}$ with maximizer 0 and numbers b_1 and b_2 such that for each candidate j we have $u_j(x, \theta) = u(x - b_j - \theta)$ for all $(x, \theta) \in X \times \Theta$ and for the citizen we have $v(x, \theta) = u(x - \theta)$ for all $(x, \theta) \in X \times \Theta$. Under these assumptions, in state θ the policy optimal for the citizen is θ whereas the policies optimal for the candidates are $\theta + b_1$ and $\theta + b_2$.

Suppose that both candidates choose the platform $\{x\}$ consisting of the single policy x. Then the outcome is x regardless of the state. The difference between this outcome and the one best for the citizen, namely θ in each state θ, is indicated by the area shaded gray in Figure 9.4. Suppose that $b_2 > 0$ and for some $x' \in X$ with $x' > x$ candidate 2 deviates to the platform $[\underline{x}, x']$. Then if candidate 2 wins she chooses the policy $\theta + b_2$ if $\theta \in [\underline{x}, x' - b_2]$ and x' if $\theta \in [x' - b_2, \overline{x}]$, in-

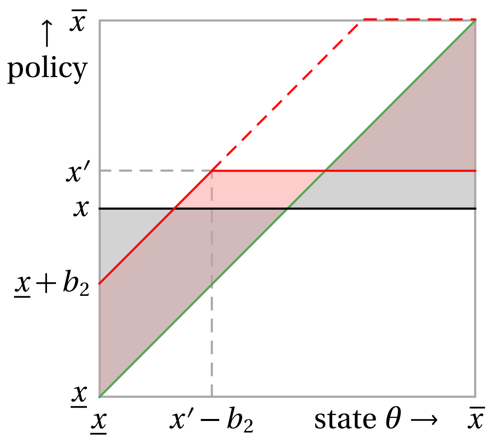


Figure 9.4 Policies as a function of the state in the model in Section 9.3.1. The policy optimal for the citizen in each state is given by the green line, and the policy induced when candidate 2 wins with the platform $[\underline{x}, x']$ is given by the red line.

dicated by the solid red line in the figure. The difference between this outcome and one best for the citizen is indicated by the area shaded pink. We see that the outcome when the winner is candidate 2 with the platform $[\underline{x}, x']$ is better for the citizen than the constant outcome x when the state is small or large, and is worse when the state takes intermediate values. Thus depending on the distribution of the state and the citizen's payoff function v, the citizen may prefer the outcome induced by candidate 2's deviation to the constant policy x. If she does, then she optimally votes for candidate 2, who prefers the resulting outcome in every state.

This argument suggests that for some specifications of the distribution of the state and the candidates' and citizen's preferences in which these preferences differ from each other, the game may have an equilibrium in which each candidate's platform is an interval of policies. Kartik et al. (2017) study an example in which such equilibria exist.

9.3.2 Candidates privately informed before choosing positions

Now assume that both candidates receive information about the desirability of the possible policies before they choose positions. Specifically, consider the extensive game with imperfect information in which chance determines the state, each candidate observes chance's move, the candidates simultaneously choose positions, and then the citizen observes the candidates' positions, but not the move of chance, and votes for one of the candidates. To make the structure of the game clear, an example for the case in which there are two states and two possible policies is given in Figure 9.5. In this figure, the initial move of chance is not shown explicitly; instead, the small circles at the top and bottom indicate

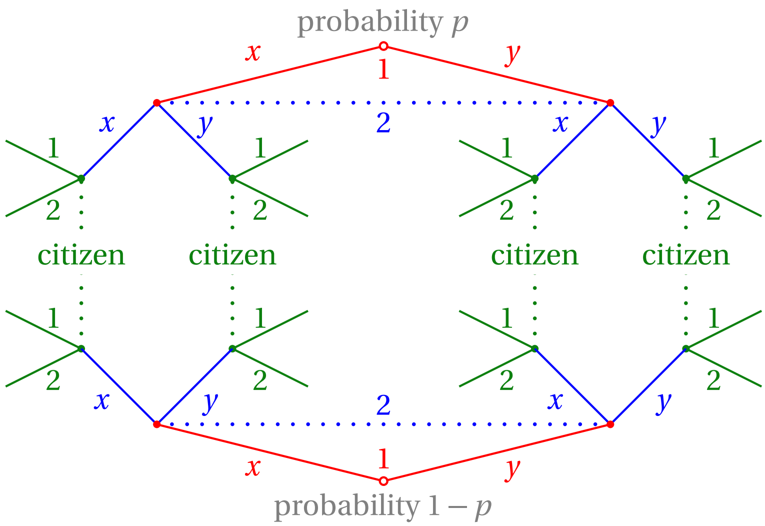


Figure 9.5 An example of the game in Section 9.3.2, in which the state is observed by the candidates before they choose positions. In this example, there are two possible policies, x and y, and two possible states. Candidate 1's actions are red, candidate 2's blue, and the citizen's green. The payoffs are not shown.

the two possible results of this move. (Payoffs are not shown in the figure.)

For the rest of the section I consider an example of the game in which, as in Figure 9.5, there are two states, but the set X of possible positions is the set of all real numbers. The states are -1 and 1: $\Theta = \{-1, 1\}$. The prior probability of state -1 is $p \in (0,1)$. For each candidate $j \in \{1,2\}$, the function $u_j : X \times \Theta \to \mathbb{R}$ whose expected value represents j's preferences is single-peaked in its first argument for each $\theta \in \Theta$. The favorite position of each candidate j in each state θ (i.e. the position x that maximizes $u_j(x,\theta)$) is denoted $\hat{x}_j(\theta)$. Assume that

$$\hat{x}_1(\theta) < \theta < \hat{x}_2(\theta) \text{ for each state } \theta \in \Theta.$$

The function $v : X \times \Theta \to \mathbb{R}$ whose expected value represents the citizen's preferences is also single-peaked in its first argument for each $\theta \in \Theta$; in each state θ, the citizen's favorite position is θ. For any positions x_1 and x_2, denote by $I(x_1, x_2)$ the citizen's information set that is reached when candidate 1's position is x_1 and candidate 2's is x_2.

Equilibria with full convergence

I first consider the possibility that the game has equilibria in which the positions chosen by the candidates are the same, and do not depend on the state, so that the citizen can make no inference regarding the state from those positions.

For any position $x^* \in [-1, 1]$, the game has weak sequential equilibria in which each candidate's position in each state is x^*. In one such equilibrium, the belief system assigns probability p to state -1 at $I(x^*, x^*)$ (as required by weak consistency of the strategies and beliefs, given that this information set is reached with probability 1 regardless of the state), and the citizen's strategy selects candidate 1. At every information set reached when one candidate deviates from her strategy to some position x, the belief system assigns probability 1 to the state in which the citizen prefers x^* to x, and the citizen votes for the non-deviating candidate, so that the deviation does not affect the outcome. The beliefs and citizen's actions at the remaining information sets, in which both candidates' positions differ from the ones prescribed by their strategies, do not affect the equilibrium status of the assessment as long as the citizen's action at each such information set is optimal given the belief at the information set, because none of these information sets is reached when a single candidate deviates.

The next exercise asks you to show that if candidate 2's favorite position in state 1 is at least -1 and candidate 1's favorite position in state -1 is at most 1 then the game has no equilibrium in which the candidates' common position in each state is outside $[-1, 1]$. Thus under this assumption, equilibria with full convergence are limited to $[-1, 1]$.

Exercise 9.5: Weak sequential equilibria with full convergence when candidates are privately informed

Show that if $\hat{x}_2(1) \geq -1$ and $\hat{x}_1(-1) \leq 1$ then for any position x^* with $x^* < -1$ or $x^* > 1$ the game has no weak sequential equilibrium in which each candidate's position in each state is x^*.

Equilibria with partial convergence

The game also has equilibria in which the candidates' positions depend on the state. Given the equilibria of the games with perfect information studied in Section 9.1, it is reasonable to consider the possibility of a weak sequential equilibrium in which each candidate's strategy selects the citizen's favorite position in each state: -1 in state -1 and 1 in state 1. To do so, we need to consider whether there is a belief system and a strategy for the citizen such that the belief system is weakly consistent with the players' strategies and each player's strategy is optimal given the belief system and the other players' strategies.

Weak consistency requires that at each of candidate 2's information sets, the belief system assigns probability 1 to the action specified by candidate 1's strategy. In all the assessments I discuss, I take as given that the belief system has

#	Information set	Prob. of state -1	Cand. chosen
1	$I(-1,-1)$	1	1
2	$I(1,1)$	0	1
3	$I(-1,x)$ for any $x \notin \{-1,1\}$	1	1
4	$I(x,-1)$ for any $x \notin \{-1,1\}$	1	2
5	$I(1,x)$ for any $x \notin \{-1,1\}$	0	1
6	$I(x,1)$ for any $x \notin \{-1,1\}$	0	2
7	$I(-1,1)$	0	2
8	$I(1,-1)$	0	1
9	$I(x,y)$ for $x \notin \{-1,1\}$ and $y \notin \{-1,1\}$	1	$c(x,y)$

Table 9.1 The probabilities assigned by the belief system to the citizen's information sets and the citizen's strategy in a weak sequential equilibrium of the game in Section 9.3.2. The candidate $c(x,y)$ is 1 if $v(x,-1) \geq v(y,-1)$ and 2 if $v(x,-1) < v(y,-1)$. The numbers in the first column are for reference.

this property. The more significant features of an assessment concern the beliefs and actions at the citizen's information sets. Consider the assessment in which each candidate selects position -1 in state -1 and position 1 in state 1 and the belief system and citizen's strategy are given in Table 9.1. I argue that if $u_2(-1,-1) \geq u_2(1,-1)$, a condition illustrated in Figure 9.6, then this assessment is a weak sequential equilibrium of the game. The beliefs at information sets of types 1–6 are specified so that deviations by either candidate to a position other than -1 and 1 are not profitable. For example, if candidate 1 deviates in state -1 (information set type 4) then the citizen continues to believe that the state is -1 and switches her vote to candidate 2, so that the policy outcome remains -1. The information set $I(-1,1)$ (type 7) is reached both if candidate 1 deviates to the position -1 in state 1 and if candidate 2 deviates to 1 in state -1. As a consequence, the belief at that information set cannot be specified in such a way that regardless of the candidates' payoff functions no deviation is profitable. The same is true for the information set $I(1,-1)$ (type 8), and checking that deviations that lead to these two information sets are not profitable is the most significant part of the argument. (No deviation by a single player leads to an information set of type 9, so the citizen's behavior at such an information set is not significant except that it must be optimal given the belief system.)

Weak consistency of beliefs with strategies

Information sets 1 and 2 in Table 9.1 are reached if the players follow their strategies, and the probabilities of state -1 are implied by Bayes' rule.

None of the remaining information sets are reached if the players follow their strategies, so weak consistency imposes no restriction on the beliefs at these

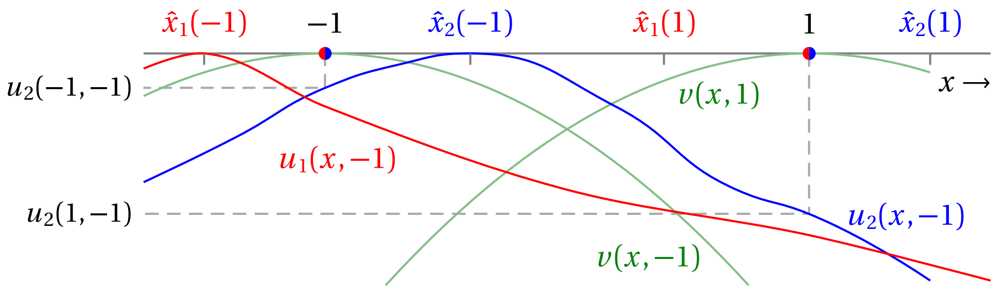


Figure 9.6 An illustration of the conditions for a weak sequential equilibrium of the game in Section 9.3.2. In state -1 each candidate's position is -1 and in state 1 it is 1. The candidates' payoff functions in state -1 are indicated in red for candidate 1 and in blue for candidate 2. The citizen's payoff function in each state is indicated in green.

sets.

Sequential rationality for citizen

At information sets 1 and 2 the candidates' positions are the same, so voting for candidate 1 is optimal for the citizen.

At information sets of types 3–8 the position of the candidate for whom the citizen votes is the citizen's favorite position in the state to which the belief system assigns probability 1, so the citizen's voting for that candidate is optimal for her.

At an information set of type 9 the citizen believes the state is -1 and votes for the candidate whose position is best for her in that state.

Sequential rationality for candidates

If candidate 1 deviates in state -1 to a position different from 1, an information set of type 4 of the citizen is reached, and the citizen votes for candidate 2, so that the outcome does not change.

Similarly, a deviation by candidate 1 in state 1 to a position different from -1 (and from 1) and deviations by candidate 2 in either state to positions different from -1 and 1 have no effect on the outcome.

If candidate 1 deviates in state -1 to the position 1 then information set 8 is reached and the citizen votes for candidate 1, changing candidate 1's payoff from $u_1(-1,-1)$ to $u_1(1,-1)$, and hence making her worse off given that $\hat{x}_1(-1) < -1$ and her payoff function is single-peaked.

If candidate 1 deviates in state 1 to the position -1 then information set 7 is reached and the citizen votes for candidate 2, so that the outcome remains policy 1.

If candidate 2 deviates in state -1 to the position 1 then information set 7 is reached and the citizen votes for candidate 2, changing candidate 2's payoff from $u_2(-1,-1)$ to $u_2(1,-1)$, and hence making her no better off given the assumption that $u_2(-1,-1) \geq u_2(1,-1)$.

If candidate 2 deviates in state 1 to the position -1 then information set 8 is reached and the citizen votes for candidate 1, so that the outcome remains policy 1.

This equilibrium is not the only one in which both candidates choose the position -1 in state -1 and the position 1 in state 1. In other such equilibria the belief system assigns a positive probability to state -1 at information sets 7 and 8; the conditions on the payoff functions for such equilibria differ from those for the equilibrium I have discussed.

Under some conditions the game has related equilibria in which in each state the candidates choose the same position, but these positions differ from the citizens' favorite positions -1 and 1. In the equilibrium I have presented, the fact that each candidate's position in each state is the citizen's favorite position in that state makes deviations by candidates easy to deter. In an equilibrium in which the candidates' common position in each state differs from the citizen's favorite position in that state, the beliefs need to be designed carefully to deter deviations.

Suppose, for example, that the candidates' common position is x^- in state -1 and x^+ in state 1, with $-1 < x^- < x^+ < 1$. If candidate 1 deviates to -1 in state -1 and the citizen continues to believe that the state is -1 then the citizen optimally votes for candidate 1, leading to the policy outcome -1, which candidate 1 prefers to x^-. To deter the deviation, the citizen must believe at her information set $I(-1,x^-)$ that the state is 1, so that she optimally votes for candidate 2. This belief is weakly consistent with the strategy profile because the information set $I(-1,x^-)$ is not reached if the candidates follow their strategies.

Exercise 9.6: Weak sequential equilibrium with partial convergence in game with privately-informed candidates

For positions x^- and x^+ with $-1 < x^- < x^+ < 1$, find a belief system and a strategy for the citizen that combined with the strategy for each candidate that selects x^- in state -1 and x^+ in state 1 is a weak sequential equilibrium of the game in this section if $u_2(x^-,-1) \geq u_2(x^+,-1)$.

The citizen's belief at her information set $I(-1,x^-)$ that the state is 1, though weakly consistent with the strategy profile, does not seem plausible. If $I(-1,x^-)$ is reached then the citizen knows that candidate 1, who prefers -1 to x^- in state -1, has deviated to -1. If this deviation induces the citizen to vote for candidate 1

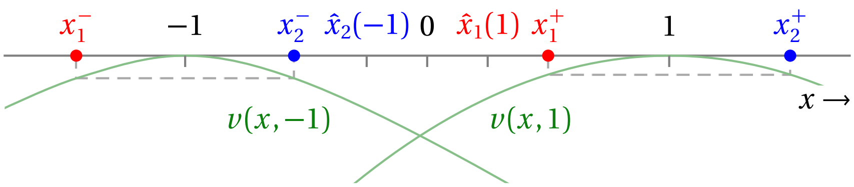


Figure 9.7 The candidates' positions in a weak sequential equilibrium of the game in Section 9.3.2. Candidate 1 chooses x_1^- in state -1 and x_1^+ in state 1 and candidate 2 chooses x_2^- in state -1 and x_2^+ in state 1.

then both candidate 1 and the citizen are better off, whereas if it induces the citizen to believe that the state is 1 and vote for candidate 2 then it does not affect candidate 1's payoff. So why the deviation should lead the citizen to believe that the state is 1 is unclear. Notions of equilibrium that impose conditions on the belief system that are stronger than the conditions imposed by weak sequential equilibrium and rule out beliefs that, like this one, seem implausible, have been proposed, but none has unqualified appeal, and I do not discuss them.

Equilibria with dispersed positions

Does the game have an equilibrium in which the candidates' positions depend on the state and differ from each other in each state? Suppose that, as shown in Figure 9.7, candidate 1 chooses x_1^- in state -1 and x_1^+ in state 1, and candidate 2 chooses x_2^- in state -1 and x_2^+ in state 1, where $x_1^- < -1 < x_2^- \le \hat{x}_2(-1)$, $\hat{x}_1(1) \le x_1^+ < 1 < x_2^+$, and the citizen is indifferent between x_1^- and x_2^- in state -1 and between x_1^+ and x_2^+ in state 1.

To determine whether an equilibrium exists in which the candidates choose such positions, we need to determine whether for some weakly consistent belief system the citizen's optimal response to any deviation by a candidate deters the deviation. Weak consistency requires that when the candidates' positions are x_1^- and x_2^- the citizen believes that the state is -1. I first argue that in an equilibrium the citizen votes for candidate 2 in this case. Suppose instead that she votes for candidate 1. If candidate 2 deviates in state -1 to a position $x_2 \in [-1, x_2^-)$, then given that the citizen prefers x_2 to x_1^- in both states, regardless of her belief she optimally switches her vote to candidate 2, making candidate 2 better off. Does a belief system exist for which the citizen optimally votes for candidate 2 when the candidates' positions are x_1^- and x_2^-? Suppose that candidate 1 in state -1 deviates to a position $x_1 \in (x_1^-, -1]$. If the citizen responds by switching her vote to candidate 1 then candidate 1 is better off, so for an equilibrium the citizen must continue to vote for candidate 2. For her to optimally do so, her belief following the deviation must assign sufficiently high probability to state 1, in which she prefers x_2^- to x_1. That is, the deviation by candidate 1 to x_1, which the citizen prefers to x_2^- in state -1, must induce the citizen to switch from assigning probability 1 to state -1 to assigning a significant probability to state 1. This change in beliefs is consistent with the requirements of a weak sequential equilibrium, but its intuitive rationale is unclear, to say the least.

The conclusion of this argument is that a weak sequential equilibrium in which the candidates' positions are those indicated in Figure 9.7 may exist, but the belief system that such an equilibrium entails, like the belief system for an equilibrium in which the candidates' common position is x^- in state -1 and x^+ in state 1, with $-1 < x^- < x^+ < 1$, is at best difficult to interpret. The next exercise invites you to fill in the details of an equilibrium.

Exercise 9.7: Weak sequential equilibrium with dispersed positions in game with privately-informed candidates

Find a belief system and a strategy for the citizen that combined with the strategies for the candidates that select the positions illustrated in in Figure 9.7 is a weak sequential equilibrium of the game in this section if $u_2(x_2^-, -1) \geq u_2(x_2^+, -1)$.

9.4 Repeated elections

Suppose that two policy-motivated candidates contest a sequence of elections. In each period $t = 1, 2, \ldots$, they choose positions in a strategic game G that is closely related to an electoral competition game with a continuum of citizens and two policy-motivated candidates. The players in G are the candidates, 1 and 2, the set of actions of each candidate is $X \subset \mathbb{R}$, a compact interval, and the payoff of each candidate j to the pair (x_1, x_2) of positions is

$$v_j(x_1, x_2) = \begin{cases} u_j(x_1) & \text{if } F(\frac{1}{2}(x_1 + x_2)) > \frac{1}{2} \\ \frac{1}{2}(u_j(x_1) + u_j(x_2)) & \text{if } F(\frac{1}{2}(x_1 + x_2)) = \frac{1}{2} \\ u_j(x_2) & \text{if } F(\frac{1}{2}(x_1 + x_2)) < \frac{1}{2}, \end{cases}$$

where F is a nonatomic probability distribution function with a density and support X, m is the median of F, and $u_j : X \to \mathbb{R}$ is a single-peaked function. Denote by $\hat{x}_j$ the maximizer of u_j and assume that $\hat{x}_1 < m < \hat{x}_2$.

I model the sequence of elections as an infinitely repeated game. In every period $t = 1, 2, \ldots$, the candidates choose positions in the game G after observing the positions chosen in every previous period. The payoff of each candidate j in the repeated game is the discounted average of her payoffs in the sequence of

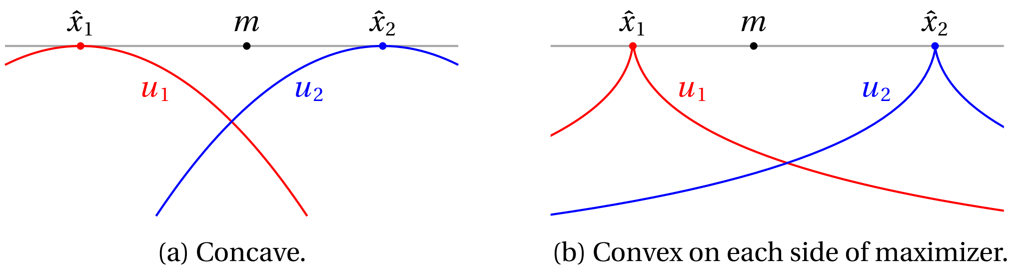


Figure 9.8 Payoff functions for policy-motivated candidates.

games, with discount factor $\delta \in (0,1)$: that is, $(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} v_j(x_1^t, x_2^t)$, where (x_1^t, x_2^t) is the pair of positions chosen in period t.

If candidate j chooses the position m in any given period t then for any position of the other candidate in period t her payoff is at least $u_j(m)$ in that period. Thus in every Nash equilibrium of the infinitely repeated game her (discounted average) payoff is at least $u_j(m)$ (for any value of δ).

The character of the equilibria of the repeated game depend on whether the functions u_j are concave, as in Figure 9.8a, or convex on each side of their maximizers, as in Figure 9.8b.

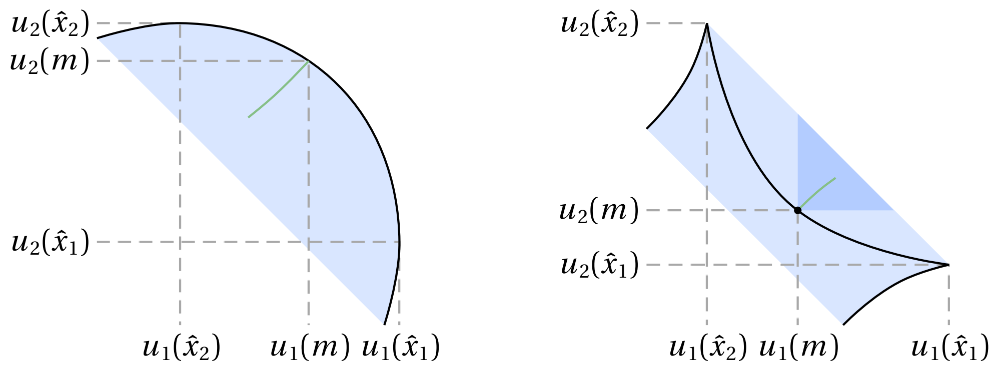
The set P of payoff pairs possible in G is the union of $\{(u_1(x), u_2(x)) : x \in X\}$, attainable when one candidate wins outright, and

$$\{(\tfrac{1}{2}(u_1(x) + u_1(x')), \tfrac{1}{2}(u_2(x) + u_2(x'))) : x \in X \text{ and } x' \in X\},$$

attainable when the candidates tie. The former set is the black curve in each panel of Figure 9.9 and the latter set is the green curve. Every sequence z^1, z^2, ... of winning positions in G generates a sequence w^1, w^2, ... of points in P. The resulting pair of payoffs in the repeated game is $(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} w_j^t$. This pair of payoffs is a weighted average of w^1, w^2, ... and hence lies in the area shaded blue in each panel.

Thus if each function u_j is concave (Figure 9.9a) then (for any value of δ) no sequence of positions yields a pair of payoffs in the repeated game for which each component j is larger than $u_j(m)$. Hence in every Nash equilibrium of the repeated game the outcome in every period is m and the payoff of each candidate j is $u_j(m)$.

If each function u_j is convex on each side of $\hat{x}_j$, so that j is more sensitive to changes in the position around $\hat{x}_j$ than she is to changes of the same size around positions distant from $\hat{x}_j$, the story is different. Suppose that $w = (1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} w^t$ lies in the area shaded dark blue in Figure 9.9b, with $w_j > u_j(m)$ for $j = 1, 2$. For each $t = 1, 2, \ldots$ let x^t be a pair of positions for the candidates for which $(v_1(x^t), v_2(x^t)) = w^t$. Consider the strategy pair in the repeated game in which each candidate j chooses x_j^1 in period 1 and, in each period $t \geq 2$,



(a) Candidates with concave payoff functions.

(b) Candidates with payoff functions convex on each side of their maximizers.

Figure 9.9 Possible payoffs in an infinitely repeated game with policy-motivated candidates.

chooses x_j^t after the history $(x^1, x^2, \ldots, x^{t-1})$ and m after every other history. If the discount factor δ is close enough to 1, this strategy pair is a subgame perfect equilibrium of the repeated game. The game satisfies the condition in Proposition 16.9, so it suffices to show that the strategy profile satisfies the one-deviation property. If in any period t following a history in which each candidate has adhered to her strategy, a candidate j chooses a position other than x_j^t and subsequently adheres to her strategy then, given the other candidate's strategy, the outcome in every subsequent period is (m, m), so that in each period after t her payoff is less than w_j, and hence for δ sufficiently close to 1 her payoff in the repeated game is less than w_j. In any period following any other history, both candidates choose m, and no deviation by either candidate affects the policy outcome in any future period, given the other candidate's strategy.

Thus if each function u_j is convex on each side of $\hat{x}_j$ and the candidates' discount factor is close to 1, the repeated game has subgame perfect equilibria in which the outcome in each period differs from m. In some of these equilibria the outcome in every period is either $\hat{x}_1$ or $\hat{x}_2$. Such an outcome arises if, for example, in some periods candidate 1 chooses the position $\hat{x}_1$ and candidate 2 chooses a position more extreme than $2m - \hat{x}_1$, and in the remaining periods candidate 1 chooses a position more extreme than $2m - \hat{x}_2$ and candidate 2 chooses the position $\hat{x}_2$.

The model of a repeated game assumes that a candidate can change her position arbitrarily from period to period. Even a candidate who implements the policy x while in office in period t can commit to a position radically different from x in the election in period $t + 1$. Such a metamorphosis is implausible: would citizens believe that a candidate who had previously espoused one policy

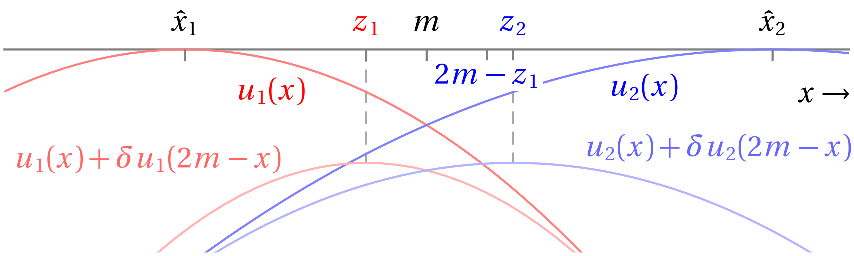


Figure 9.10 The positions relevant in a repeated election with persistent policies.

is now committed to a wildly different one?

If we assume that candidates can change their positions only when they are out of office, equilibria in which the winning policy in each period differs from the favorite position of the median voter are possible even if the candidates' payoff functions are strictly concave (as in Figure 9.8a).

Consider an extensive game that differs from the repeated game specified earlier in two respects. First, in each period, only one of the candidates is free to choose a position. In period 1, candidate 2's position is fixed; the idea is that she was the incumbent in the previous, unmodeled, period. In every subsequent period, only the challenger is free to choose a position. The winner in any period is constrained to adopt the same policy in the next period: the set of actions available to a candidate i in any period t following a period in which she won is $\{x_i^{t-1}\}$, where x_i^{t-1} is her policy in period $t-1$. Second, if the vote is tied in period 1 then candidate 1 is the winner, and if it is tied in any subsequent period then the challenger in that period is the winner.

Denote the game starting with candidate 2's position fixed at x_2^1 by $\Gamma(x_2^1)$. Assume that each payoff function u_j is strictly concave, and for simplicity assume that the set of positions from which candidate 1 can choose is $[\hat{x}_1, m]$ and the set from which candidate 2 can choose is $[m, \hat{x}_2]$. Let z_1 be the maximizer of $u_1(x_1) + \delta u_1(2m - x_1)$ for $x_1 \in [\hat{x}_1, m]$ and let z_2 be the maximizer of $u_2(x_2) + \delta u_2(2m - x_2)$ for $x_2 \in [m, \hat{x}_1]$. Given that $\delta < 1$, we have $z_1 < m$ and $z_2 > m$. Assume that $z_1 \geq 2m - z_2$. (Figure 9.10 shows an example.)

Let $x_2^* \in [m, 2m - z_1]$ and $x_1^* = 2m - x_2^*$, so that $x_1^* \in [z_1, m]$. I claim that the game $\Gamma(x_2^*)$ has a subgame perfect equilibrium in which the policy alternates between x_1^*, implemented by candidate 1, and x_2^*, implemented by candidate 2. If in any period the challenger deviates to a position x closer to m, the challenger in the next period reciprocates, choosing the position $2m - x$, and subsequently the outcome alternates between these two positions. Each candidate prefers an alternation between x_1^* and $x_2^* = 2m - x_1^*$ to one between x and $2m - x$, so such a deviation is not advantageous.

Specifically, the following strategy pair is a subgame perfect equilibrium of

$\Gamma(x_2^*)$: candidate 1 chooses x_1^* in period 1 and $\max\{2m - x_2^{t-1}, z_1\}$ in each subsequent period t in which she is the challenger, and candidate 2 chooses $\min\{2m - x_1^{t-1}, z_2\}$ in each period t in which she is the challenger. You are invited to verify this claim in the next exercise.

Exercise 9.8: Repeated elections with persistent policies

Show that strategy pair specified in the text is a subgame perfect equilibrium of the variant of a repeated game with persistent policies defined in the text.

Notes

Austen-Smith and Banks (2005, Section 7.7) study a variant of the model in Section 9.1 in which each candidate cares slightly about winning and the set of alternatives is convex and compact. Versions of Proposition 9.3 are established by Wittman (1977, Proposition 5), Calvert (1985, Theorems 1 and 2), and Roemer (1994, Theorem 2.1). The model in Section 9.2 is due to Wittman (1983) and Calvert (1985, Section 4); Proposition 9.4 is based on Duggan (2014, Theorem 22). Proposition 9.5 is due to Roemer (1997, Theorem 3.2) and Duggan (2014, Theorem 22), who credits unpublished joint work with Mark Fey. Section 9.3.1 is based on Kartik et al. (2017) and Section 9.3.2 is based on Schultz (1996) and Martinelli and Matsui (2002). The game with persistent policies discussed at the end of Section 9.4 is a variant of the one studied by Forand (2014).

The model with imperfect commitment discussed after Exercise 9.3 is due to Jean Guillaume Forand.

Solutions to exercises

Exercise 9.1

Alternative a is not a Condorcet winner because a majority of individuals prefer c to a.

Suppose that candidate 1 prefers $\{a\}$ to $\{b\}$ to $\{c\}$ and candidate 2 prefers $\{b\}$ to $\{c\}$ to $\{a\}$, and also $\{a,b\}$ to $\{c\}$. These candidates are representative. Then (a,b) is a Nash equilibrium of the game by the following argument.

- The outcome of (a,b) is a tie, $\{a,b\}$.
- If candidate 1 deviates to b, then she ties and the outcome is $\{b\}$, which she likes less than the outcome $\{a,b\}$ by the second part of (9.1).

- If candidate 1 deviates to c, then she loses and the outcome is $\{b\}$, which she likes less than $\{a, b\}$ by the second part of (9.1).
- If candidate 2 deviates to a, then she ties and the outcome is $\{a\}$, which she likes less than the outcome $\{a, b\}$ by the second part of (9.1).
- If candidate 2 deviates to c, then she wins and the outcome is $\{c\}$, which she likes less than the outcome $\{a, b\}$.

All of the deviations lead to outcomes different from the equilibrium outcome, so none of the arguments change if each candidate lexicographically favors winning.

Exercise 9.2

Assume without loss of generality that $\hat{x}_1 \leq \hat{x}_2$.

Denote the candidates' positions by x_1 and x_2. I first argue that $x_2 \geq \hat{x}_2$. If $x_2 < \hat{x}_2$ then

- $x_1 < \hat{x}_2 \Rightarrow$ outcome is $\{\max\{x_1, x_2\}\}$; by moving to $\hat{x}_2$, candidate 2 changes the outcome to $\{\hat{x}_2\}$, which she prefers
- $x_1 = \hat{x}_2 \Rightarrow$ outcome is $\{\hat{x}_2\}$; by moving to $\hat{x}_1$, candidate 1 changes the outcome to $\{x_2\}$, which she prefers
- $x_1 > \hat{x}_2$ and candidate 1 wins $\Rightarrow$ outcome is $\{x_1\}$; by moving to $\hat{x}_2$, candidate 1 changes the outcome to $\{\hat{x}_2\}$, which she prefers
- $x_1 > \hat{x}_2$ and candidate 1 loses or ties for first place $\Rightarrow$ outcome is $\{x_2\}$ or $\{x_1, x_2\}$; by moving to $\hat{x}_2$, candidate 2 changes the outcome to $\{\hat{x}_2\}$, which she prefers.

I now argue that $x_2 \leq m$. If $x_2 > m$ then

- $x_1 < \hat{x}_2 \Rightarrow$ outcome is $\{x_1\}$, $\{x_2\}$, or $\{x_1, x_2\}$; by moving to $\hat{x}_2$, candidate 2 changes the outcome to $\{\hat{x}_2\}$, which she prefers
- $x_1 \geq \hat{x}_2$ and candidate 1 wins $\Rightarrow$ outcome is $\{x_1\}$; by reducing x_1 slightly, candidate 1 reduces the value of the winning position, which she prefers
- $\hat{x}_2 \leq x_1 < x_2$ and candidate 2 wins or ties $\Rightarrow$ outcome is $\{x_2\}$ or $\{x_1, x_2\}$; by moving to x_1, candidate 2 changes the outcome to $\{x_1\}$, which she prefers
- $x_1 \geq x_2 \Rightarrow$ candidate 2 wins or ties and outcome is $\{x_2\}$; by moving to m, candidate 2 changes the outcome to $\{m\}$, which she prefers.

Thus $\hat{x}_2 \leq x_2 \leq m$.

Suppose that $x_2 > \hat{x}_2$. Then $x_1 = x_2$, otherwise the winning candidate, if one candidate wins outright, or else the rightmost candidate who ties for first place, can increase her payoff by moving slightly to the left.

Now suppose that $x_2 = \hat{x}_2$. Then $x_1 \leq x_2$, otherwise candidate 1 can increase her payoff by moving to $\hat{x}_2$

Finally, any pair (x_1, x_2) for which $\hat{x}_2 \leq x_1 = x_2 \leq m$ or $x_1 \leq x_2 = \hat{x}_2$ is an equilibrium.

Thus for $\hat{x}_1 \leq \hat{x}_2$ the set of Nash equilibria is the set of pairs (x_1, x_2) such that $\hat{x}_2 \leq x_1 = x_2 \leq m$ or $x_1 \leq x_2 = \hat{x}_2$; in all of the equilibria the outcome is $\{x_2\}$.

Exercise 9.3

Every citizen i with favorite position at most m prefers x_1^* implemented by candidate 1 to m implemented by candidate 2, so the outcome of the pair (x_1^*, m) of positions is that candidate 1 wins and implements x_1^*. (The payoff for a citizen i with favorite position $\hat{z}_i \in [x_1^*, m]$ is $v(x_1^* - \hat{z}_i) \geq v(x_1^* - m) = -\delta$ for candidate 1's position x_1^* and $v(m - \hat{z}_i) - \delta \leq -\delta$ for candidate 2's position m.)

Consider a deviation by candidate 1. If she deviates to a position less than x_1^* then for some $\varepsilon > 0$ every citizen with favorite position at least $m - \varepsilon$ votes for candidate 2, so that candidate 2 wins. Candidate 1 prefers x_1^* to m, so the deviation makes her worse off. If she deviates to a position greater than x_1^* then she either wins, in which case she is no better off, or she loses and the position of the winner, candidate 2, is m, so she is also no better off.

Now consider a deviation by candidate 2. If she deviates to a position at most x_1^* then all citizens with favorite positions at least x_1^* vote for candidate 1, so that candidate 1 continues to win. If she deviates to a position greater than x_1^* then all citizens with favorite positions at most m vote for candidate 1, so that candidate 1 continues to win. Thus no deviation makes candidate 2 better off.

We conclude that (x_1^*, m) is a Nash equilibrium of the game.

Exercise 9.4

By Proposition 9.4 in any Nash equilibrium (x_1^*, x_2^*) we have $\hat{x}_1 \leq x_1^* < x_2^* \leq \hat{x}_2$.

Suppose that $\hat{x}_1 < x_1 \leq \hat{x}_2$. We have $G(z) = (z + k)/2k$ for $z \in [-k, k]$, so candidate 1's payoff is

$$\begin{aligned}
-(x_1 - \hat{x}_1)((\tfrac{1}{2}(x_1 + x_2) + k)/2k) - (x_2 - \hat{x}_1)(1 - (\tfrac{1}{2}(x_1 + x_2) + k)/2k) \\
= -(1/2k)\big[(x_1 - \hat{x}_1)(\tfrac{1}{2}(x_1 + x_2) + k) + (x_2 - \hat{x}_1)(k - \tfrac{1}{2}(x_1 + x_2))\big] \\
= -(1/2k)\big[\tfrac{1}{2}x_1^2 + kx_1 + C\big],
\end{aligned}$$

Information set	Prob. of state −1	Cand. chosen
$I(x^-, x^-)$	1	1
$I(x^+, x^+)$	0	1
$I(x^-, x)$ for $x < x^-$	0	1
$I(x^-, x)$ for $x > x^-$, $x \neq x^+$	1	1
$I(x, x^-)$ for $x < x^-$	0	2
$I(x, x^-)$ for $x > x^-$, $x \neq x^+$	1	2
$I(x^+, x)$ for $x < x^+$, $x \neq x^-$	0	1
$I(x^+, x)$ for $x > x^+$	1	1
$I(x, x^+)$ for $x < x^+$, $x \neq x^-$	0	2
$I(x, x^+)$ for $x > x^+$	1	2
$I(x^-, x^+)$	0	2
$I(x^+, x^-)$	0	1
$I(x, y)$ for $x \notin \{x^-, x^+\}$ and $y \notin \{x^-, x^+\}$	1	$c(x, y)$

Table 9.2 The probabilities assigned by the belief system to the citizen's information sets and the citizen's strategy in a weak sequential equilibrium of the game in Section 9.3.2 in which the candidates' strategies are the ones given in Exercise 9.6. The candidate $c(x, y)$ is 1 if $v(x, -1) \geq v(y, -1)$ and 2 if $v(x, -1) < v(y, -1)$.

where C is a constant (independent of x_1). This payoff is decreasing in x_1.

Thus candidate 1's best response to $\hat{x}_2$ is $\hat{x}_1$. The same argument with the roles of candidates 1 and 2 interchanged shows that $\hat{x}_1$ is a best response to $\hat{x}_2$. Thus $(\hat{x}_1, \hat{x}_2)$ is a Nash equilibrium.

Exercise 9.5

Suppose that $x^* < -1$ and each candidate in each state chooses x^*. Then if candidate 2 deviates to the position -1 the citizen votes for her regardless of her belief about the state, in which case she is elected. She prefers the position -1 to x^*, given that $\hat{x}_2(1) \geq -1$ and her preferences are single-peaked, so she benefits from the deviation. Similarly, candidate 1 benefits from a deviation from x^* to 1 if $x^* > 1$.

Exercise 9.6

The assessment in which each candidate chooses the position x^- in state -1 and the position x^+ in state 1 and the belief system and strategy for the citizen are given in Table 9.2 is a weak sequential equilibrium of the game if $u_2(x^-, -1) \geq u_2(x^+, -1)$. This condition on candidate 2's payoffs is required so that in state -1 she does not benefit from deviating to x^+.

Exercise 9.7

The assessment in which the belief system and strategy for the citizen are

Information set	Prob. of state -1	Cand. chosen
$I(x_1^-, x_2^-)$	1	2
$I(x_1^+, x_2^+)$	0	1
$I(x_1^-, x)$ for $x < x_2^-$	0	2
$I(x_1^-, x)$ for $x > x_2^-$, $x \neq x_2^+$	1	1
$I(x, x_2^-)$ for $x \leq x_2^-$	0	2
$I(x, x_2^-)$ for $x > x_2^-$, $x \neq x_1^+$	1	2
$I(x_1^+, x)$ for $x < x_1^+$, $x \neq x_2^-$	0	1
$I(x_1^+, x)$ for $x \geq x_1^+$	1	1
$I(x, x_2^+)$ for $x < x_1^+$, $x \neq x_1^-$	0	2
$I(x, x_2^+)$ for $x > x_1^+$	1	1
$I(x_1^-, x_2^+)$	0	2
$I(x_1^+, x_2^-)$	0	1
$I(x, y)$ for $x \notin \{x_1^-, x_2^+\}$ and $y \notin \{x_1^-, x_2^+\}$	1	$c(x, y)$

Table 9.3 The probabilities assigned by the belief system to the citizen's information sets and the citizen's strategy in a weak sequential equilibrium of the game in Section 9.3.2 in which the candidates' strategies are the ones given in Exercise 9.7. The candidate $c(x, y)$ is 1 if $v(x, -1) \geq v(y, -1)$ and 2 if $v(x, -1) < v(y, -1)$.

given in Table 9.3 is a weak sequential equilibrium of the game if $u_2(x_2^-, -1) \geq u_2(x_2^+, -1)$. This condition is required so that in state -1 candidate 2 does not benefit from deviating to x_2^+. (Other belief systems and strategies for the citizen are consistent with equilibrium.)

The belief system is weakly consistent with the strategy profile because at the two information sets of the citizen that are reached with positive probability given the strategies, $I(x_1^-, x_2^-)$ and $I(x_1^+, x_2^+)$, it assigns probabilities derived from the prior via Bayes' rule.

The citizen's strategy is sequentially rational because at her information sets $I(x_1^-, x_2^-)$ and $I(x_1^+, x_2^+)$ she is indifferent between the candidates' positions and at every other information set she votes for the candidate she prefers, given the belief system.

The candidates' strategies are sequentially rational given this belief system and strategy for the citizen by the following arguments.

If candidate 1 changes her position in state -1 from x_1^- to a position other than x_1^+, the citizen votes for candidate 2, so that the outcome does not change.

Similarly, if candidate 2 changes her position in state 1 from x_2^+ to a position other than x_2^-, the citizen votes for candidate 1, so that the outcome does not change.

If candidate 1 changes her position in state 1 from x_1^+ to a position less than x_1^+ other than x_1^-, the citizen votes for candidate 2, so that the outcome does not change. If she changes her position in state 1 from x_1^+ to a position greater than x_1^+, the citizen votes for her, so that she is worse off given that $x_1^+ \geq \hat{x}_1(1)$.

Similarly, if candidate 2 changes her position in state -1 from x_2^- to a position greater than x_2^- other than x_2^+, the citizen votes for candidate 1, so that the outcome does not change. If she changes her position in state -1 from x_2^- to a position less than x_2^-, the citizen votes for her, so that she is worse off given that $x_2^- \leq \hat{x}_2(-1)$.

If candidate 1 deviates in state -1 to the position x_1^+ then the citizen votes for her, so that she is worse off, given that $\hat{x}_1(-1) < -1$. If she deviates in state 1 to the position x_1^- then the citizen votes for candidate 2, so the outcome does not change.

If candidate 2 deviates in state -1 to the position x_2^+ then the citizen votes for her, so that she is not better off, given that if $u_2(x_2^-, -1) \geq u_2(x_2^+, -1)$. If she deviates in state 1 to the position x_2^- then the citizen votes for candidate 1, so the outcome does not change.

Exercise 9.8

The game satisfies the condition in Proposition 16.9, so a strategy pair is a subgame perfect equilibrium if and only if it satisfies the one-deviation property.

First consider deviations by candidate 1 (in periods in which she is the challenger). For each of the following cases, the table gives the outcomes induced by candidate 1's adhering to her strategy and deviating from it in the first period of the subgame, given candidate 2's strategy.

Note that the function $u_1(x_1) + \delta u_1(2m - x_1)$ is concave in x_1, increasing up to z_1 and decreasing thereafter.

Subgame following history ending with $x_2 \in [m, 2m - z_1]$

	period 1	period $1+t$ $t \geq 1$ odd	period $1+t$ $t \geq 2$ even
adheres	$2m - x_2$	x_2	$2m - x_2$
deviates to $x_1 \in [\hat{x}_1, 2m - x_2)$	x_2	$2m - x_2$	x_2
deviates to $x_1 \in (2m - x_2, m]$	x_1	$2m - x_1$	x_1

She prefers $2m - x_2$ to x_2, so the first deviation makes her worse off. The second deviation makes her worse off because $u_1(x) + \delta u_1(2m - x)$ is decreasing in x for $x > z_1$.

The case $x_2 = 2m - x_1^*$ covers the subgame following the empty history (the start of the game).

Subgame following history ending with $x_2 \in (2m - z_1, \hat{x}_2]$

	period 1	period 2	period $1+t$, $t \geq 2$ even	period $2+t$, $t \geq 2$ even
adheres	z_1	$2m - z_1$	z_1	$2m - z_1$
deviates to $x_1 \in [\hat{x}_1, 2m - x_2)$	x_2	z_1	$2m - z_1$	z_1
deviates to $x_1 \in [2m - x_2, z_1)$	x_1	$2m - x_1$	z_1	$2m - z_1$
deviates to $x_1 \in (z_1, m]$	x_1	$2m - x_1$	x_1	$2m - x_1$

The first deviation makes her worse off because she prefers both z_1 and $2m - z_1$ to x_2, the second one does so because she prefers z_1 to x_1 and $2m - z_1$ to $2m - x_1$, and the third one does so because $u_1(x_1) + \delta u_1(2m - x_1)$ is decreasing in x_1 for $x_1 \geq z_1$.

Now consider deviations by candidate 2.

Subgame following history ending with $x_1 \in [z_1, m]$

	period 1	period $1+t$, $t \geq 1$ odd	period $1+t$, $t \geq 2$ even
adheres	$2m - x_1$	x_1	$2m - x_1$
deviates to $x_2 \in (2m - x_1, \hat{x}_1]$	x_1	$2m - x_1$	x_1
deviates to $x_2 \in [m, 2m - x_1)$	x_2	$2m - x_2$	x_2

She prefers $2m - x_1$ to x_1, so the first deviation makes her worse off. The second deviation makes her worse off because $x_2 < 2m - x_1 < z_2$ and $u_2(x) + \delta u_2(2m - x)$ is increasing in x for $x < z_2$.

Subgame following history ending with $x_1 \in [2m - z_2, z_1)$

	period 1	period 2	period $1+t$, $t \geq 2$ even	period $2+t$, $t \geq 2$ even
adheres	$2m - x_1$	z_1	$2m - z_1$	z_1
deviates to $x_2 \in (2m - x_1, \hat{x}_2]$	x_1	$2m - x_1$	z_1	$2m - z_1$
deviates to $x_2 \in (2m - z_1, 2m - x_1)$	x_2	z_1	$2m - z_1$	z_1
deviates to $x_2 \in [m, 2m - z_1]$	x_2	$2m - x_2$	x_2	$2m - x_2$

The first deviation makes her worse off because she prefers both $2m - x_1$ and z_1 to x_1, the second one does so because she prefers $2m - x_1$ to x_2 (given $x_2 < 2m - x_1 \leq z_2$), and the third one does so because she prefers $2m - x_1$ to x_2 and $u_2(x) + \delta u_2(2m - x)$ is increasing in x for $x < z_2$.

Subgame following history ending with $x_1 \in [\hat{x}_1, 2m - z_2)$

	period 1	period 2	period $1+t$ $t \geq 2$ even	period $2+t$ $t \geq 2$ even
adheres	z_2	z_1	$2m - z_1$	z_1
deviates to $x_2 \in (2m - x_1, \hat{x}_2]$	x_1	z_2	z_1	$2m - z_1$
deviates to $x_2 \in (z_2, 2m - x_1]$	x_2	z_1	$2m - z_1$	z_1
deviates to $x_2 \in [2m - z_1, z_2)$	x_2	z_1	$2m - z_1$	z_1
deviates to $x_2 \in [m, 2m - z_1)$	x_2	$2m - x_2$	x_2	$2m - x_2$

The first deviation makes her worse off because she prefers z_2, z_1, and $2m - z_1$ to x_1, the second and third ones do so because she prefers z_2 to x_2, and the fourth one does so because she prefers z_2 to x_2 and $u_2(x) + \delta u_2(2m - x)$ is increasing in x for $x < z_2$.

10 Electoral competition: endogenous candidates

10.1 Simultaneous entry with sincere voting 329
10.2 Simultaneous entry with strategic voting 331
10.3 Sequential entry 340
10.4 Citizen-candidates 346

In the models in the previous two chapters, the number of candidates is fixed, equal to two. In this chapter I present models in which each individual from a given set decides whether to become a candidate; the number of candidates is determined as part of an equilibrium.

Synopsis

Section 10.1 presents a straightforward extension to many candidates of an electoral competition game with a continuum of citizens and two office-motivated candidates. There are three or more office-seekers rather than two, each of whom has the option of running as a candidate, and each office-seeker prefers to stay out of the competition than to enter and lose. I argue that for almost any distribution of the citizens' favorite positions, the resulting model has no Nash equilibrium, so that it is not a useful vehicle to study multicandidate electoral competition.

In this model, when a set S of two or more candidates choose the same position, the votes of the citizens who prefer that position to the position of every other candidate are split equally among the members of S. Section 10.2 explores a variant of the model in which the candidate for whom each citizen votes is specified as part of an equilibrium. (That is, voting is "strategic".) The model is an extensive game in which the candidates first simultaneously choose positions and then the citizens simultaneously cast their votes. Proposition 10.1 shows that in every subgame perfect equilibrium in which each citizen's vote is weakly undominated, the position chosen by every office-seeker who becomes a candidate is the median m of the citizens' favorite positions. If one of these candidates deviates to a position x different from m, all the citizens who prefer m to x vote for one of the remaining candidates, who consequently wins outright. In the model in Section 10.1, such a deviation splits the votes of the citizens who prefer m to x among the candidates remaining at m, but in this model the deviation can lead these citizens to rally around one of these remaining candidates.

Section 10.3 presents a variant of the model in Section 10.1 in which the office-seekers move sequentially rather than simultaneously. Proposition 10.4 shows that when there are three office-seekers, in the unique subgame perfect equilibrium the first one to move chooses the median m of the citizens' favorite positions, the second one stays out of the competition, and the third one, like the first one, chooses m. The reason the second office-seeker to move stays out is that for every position x_2 at which she enters, there is a position x_3 at which the third office-seeker can win outright. If $x_2 = m$, then a position x_3 close to m is winning—the votes of the citizens who prefer m to x_3 are split between the first two entrants—and if x_2 differs from m then a position x_3 closer to m on the other side is winning. For the case of an arbitrary number n of office-seekers, you may find yourself conjecturing that in the unique subgame perfect equilibrium the first office-seeker enters at m, the next $n-2$ stay out, and the last one enters at m. The veracity of this conjecture is not known.

The model in Section 10.4 differs more significantly from the one in Section 10.1. There is no set of candidates distinct from the set of citizens. Rather, any citizen can choose to become a candidate. A citizen who does so implements her favorite position if she is elected; she cannot commit to a different position. Each citizen cares about the position implemented by the winner of the election. In addition, if she becomes a candidate she incurs a cost and, if she wins, she receives a benefit. I present two versions of the model, one in which voting is modeled as sincere (each citizen votes for the candidate whose favorite position she likes best) and one in which it is modeled as strategic. Both models have various types of equilibria that differ qualitatively from the equilibria of the models I have discussed previously. In one type of equilibrium, the candidates are two citizens with favorite positions symmetric about the median, m, of all citizens' favorite positions (Proposition 10.7). In another type of equilibrium, a single citizen becomes a candidate; her favorite position is either m or close to m (Proposition 10.5). Depending on the benefit and cost of running as a candidate, the model with strategic voting also has equilibria in which many citizens with favorite position m enter as candidates, for much the same reason that the model in Section 10.2 has such equilibria (Proposition 10.6*c*). And for certain ranges of the parameter values, both models have equilibria in which three or more citizens run as candidates (Exercises 10.7 and 10.8).

10.1 Simultaneous entry with sincere voting

I begin by arguing that a straightforward extension to many candidates of an electoral competition game with a continuum of citizens and two office-motivated candidates is not a useful vehicle for exploring multicandidate elections because for most distributions of the citizens' favorite positions it has no Nash equilibrium.

Suppose that each of many office-seekers chooses whether to run as a candidate and, if she runs, the position to take. As in the two-candidate model, assume that the set of possible positions is the real line and the distribution F of the citizens' favorite positions is nonatomic, with support an interval. I refer to an office-seeker who chooses a position as a candidate. Each citizen votes for a candidate whose position is closest to her favorite position; if a position is occupied by several candidates, then these candidates share equally the votes for that position. Each office-seeker is motivated by the possibility of winning; she prefers to stay out of the election than to enter and lose, to win outright than to tie for first place, and to tie for first place with one other candidate than to stay out of the election.

Suppose that the number of office-seekers is at least three. Any Nash equilibrium has the following properties.

1. At least two office-seekers become candidates. If none do so, any one of them can enter and win outright, and if one does so, another one can enter at the same position and tie for first place.

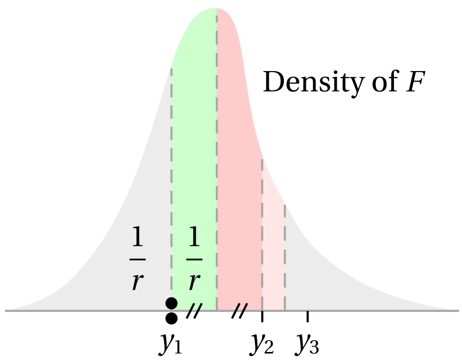
2. All candidates tie. If not, one of them loses and is better off withdrawing.

3. At most two office-seekers choose each occupied position. If more than two choose the same position then they all tie for first place, by property 2, and any one of them can deviate slightly, obtaining at least almost half the votes for the position and hence winning outright.

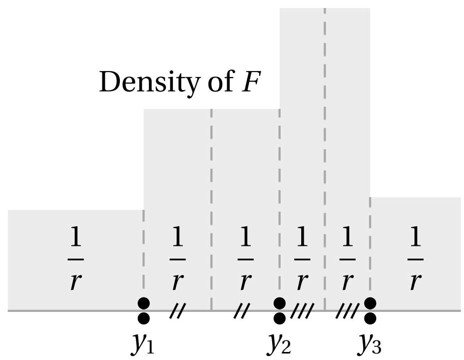
4. Exactly two office-seekers choose the smallest occupied position. By property 3, the only other possibility is that one office-seeker chooses this position, in which case she can increase her vote share and hence win rather than tie by increasing her position slightly. Similarly exactly two office-seekers choose the largest occupied position.

5. The number of candidates is at least four. This conclusion follows from property 4.

6. For any position y chosen by two office-seekers, the fraction of votes the position attracts from citizens with favorite positions less than y is equal to the



(a) If one candidate chooses y_2 then the mass of voters she attracts is the sum of the areas shaded dark and light pink.



(b) An example of a distribution F for which there is an equilibrium with six candidates.

Figure 10.1 The conditions required for an equilibrium with r candidates in a model with simultaneous entry and sincere voting. At most two office-seekers choose each occupied position, and exactly two choose the leftmost occupied position, y_1.

fraction it attracts from citizens with favorite positions greater than y, and the common fraction is $1/r$, where r is the total number of candidates. By property 2, the candidates at y tie for first place. If the fractions differ then either of the candidates with position y can deviate slightly in the direction of the larger fraction and win outright rather than tying.

7. Exactly two office-seekers choose each occupied position. Denote by y_1 the smallest occupied position. By properties 4 and 6, we have $y_1 = F^{-1}(1/r)$, as in Figure 10.1a. The value y_2 of the next smallest occupied position is determined by the condition that y_1 attracts the fraction $2/r$ of the votes: the midpoint of $[y_1, y_2]$ must be $F^{-1}(2/r)$. By property 3, at most two candidates occupy y_2. Suppose that one does so. Then she attracts the votes of all citizens with favorite positions between the midpoint of $[y_1, y_2]$ and the midpoint of $[y_2, y_3]$, the mass of which is the sum of the areas shaded dark and light pink in the figure. But if she deviates to a position $y_1 + \varepsilon$ for some small number $\varepsilon > 0$, she attracts the votes of all citizens with favorite positions between $y_1 + \frac{1}{2}\varepsilon$ and the midpoint of $[y_1 + \varepsilon, y_3]$, the mass of which exceeds the area shaded green, and hence exceeds $1/r$, so that the configuration is not an equilibrium. Thus two candidates occupy y_2. Repeating the argument leads to the conclusion that two candidates occupy every occupied position.

Properties 6 and 7 applied to y_2 imply that the area shaded dark pink in Figure 10.1a is $1/r$. That is, $y_2 = F^{-1}(2/r) + (F^{-1}(2/r) - F^{-1}(1/r)) = F^{-1}(3/r)$, or $F^{-1}(2/r) - F^{-1}(1/r) = F^{-1}(3/r) - F^{-1}(2/r)$. Applying the same argument to each occupied position, we conclude that an equilibrium with r candidates exists only if

$$F^{-1}((k+1)/r) - F^{-1}(k/r) = F^{-1}((k+2)/r) - F^{-1}((k+1)/r)$$
$$\text{for every odd number } k \text{ with } 1 \le k \le r-3. \qquad (10.1)$$

An example of a distribution F for which these conditions are satisfied for $r = 6$ is given in Figure 10.1b. (The conditions are not sufficient for an equilibrium, but the configuration of positions shown in this figure is an equilibrium.)

If you try to construct a distribution that satisfies (10.1) for some integer $r \ge 4$, I think you will conclude that such distributions are few and far between. An implication of such a conclusion is that for most distributions the game has no Nash equilibrium.

A variant of the game in which the winner is determined by plurality rule with a runoff does have Nash equilibria, which you are invited to study in the next exercise.

Exercise 10.1: Nash equilibria of electoral competition game under plurality rule with runoff

Consider a variant of the game studied in this section in which the winner is determined by plurality rule with a runoff. In this system, there may be one or two rounds of voting. Assume that in each round, each citizen votes for the candidate whose position she likes best. If one candidate obtains the votes of more than half the citizens in the first round, she wins and there is no second round. Otherwise, the two candidates who obtain the most votes in the first round compete in a second round; the one who obtains the most votes in the second round wins. All ties are broken equi-probably. Assume that each candidate's payoff is her probability of winning. (Note that each candidate chooses a single position; she is not allowed to change her position between the rounds of voting.)

Denote the number of candidates by n and the median of the citizens' favorite positions, which is assumed to be unique, by m. Are there values of k such that the game has a Nash equilibrium in which k candidates choose the position m and the remainder do not enter? If $n \ge 4$, are there values of k and $\delta > 0$ such that the game has a Nash equilibrium in which k candidates choose the position $m - \delta$, k choose $m + \delta$, and the remainder do not enter?

10.2 Simultaneous entry with strategic voting

The model in the previous section assumes that if two or more candidates occupy the same position, the votes for that position are split equally among them.

Each citizen is indifferent among candidates with the same position, so the optimality of the action of a citizen who prefers that position to every other occupied position requires only that she vote for one of these candidates. The conclusion that the candidates share the votes for the position equally follows, at least approximately, from the additional assumptions that the number of citizens is large and each citizen chooses the candidate for whom she votes randomly from those among whom she is indifferent, independently of all the other citizens. An alternative formulation, explored in this section, assumes that the candidate for whom each citizen votes is determined as part of the equilibrium. The electoral competition is modeled as a two-stage game, in which first each office-seeker chooses whether to become a candidate, and if so the position to take, and then each citizen selects the candidate for whom to vote. Each candidate who receives the highest number of votes wins with the same probability. In an equilibrium, the action of each office-seeker is optimal for her given the citizens' strategies, and the vote of each citizen is optimal for her given the candidates' positions.

The set of possible positions is the set of real numbers. There are a finite number of office-seekers and a finite number of citizens. Each office-seeker incurs a cost if she becomes a candidate and receives a benefit if she wins. Each citizen receives a payoff that depends on the position of the candidate who wins; her payoff function over positions is single-peaked. If no office-seeker becomes a candidate, each citizen gets a fixed negative payoff.

Definition 10.1: Electoral competition game with office-motivated candidates and strategic voting

An *electoral competition game with office-motivated candidates and strategic voting* $\langle n, h, (u_1, \ldots, u_h), b, c, L\rangle$, where

- $n \geq 2$ is an integer (the number of office-seekers)
- $h \geq 3$ is an odd integer (the number of citizens)
- $u_i : \mathbb{R} \to \mathbb{R}_-$ for $i = 1, \ldots, h$ is a single-peaked function (citizen i's payoff function over positions)
- $b > 0$ (each candidate's benefit from winning)
- $c > 0$ (each office-seeker's cost of running as a candidate)
- $L \in \mathbb{R}$ (each citizen's loss if no office-seeker runs as a candidate)

is an extensive game with perfect information and simultaneous moves with the following components.

Players

The set $N \cup I$, where $N = \{1, \dots, n\}$ (office-seekers) and $I = \{1, \dots, h\}$ (citizens).

Terminal histories

The set of sequences (x, v), where $x = (x_1, \dots, x_n)$ and $v = (v_1, \dots, v_h)$ with $x_j \in \mathbb{R} \cup \{Out\}$ for $j = 1, \dots, n$ and $v_i \in \{j \in N : x_j \in \mathbb{R}\} \cup \{Abstain\}$ for $i = 1, \dots, h$. (The value of x_j for $x_j \in \mathbb{R}$ is j's position; $v_i \in N$ is the candidate for whom i votes.)

Player function

The function P with $P(\varnothing) = N$ (every office-seeker moves at the start of the game) and $P(x_1, \dots, x_n) = I$ for all $(x_1, \dots, x_n) \in (\mathbb{R} \cup \{Out\})^n$ (every citizen moves after the office-seekers have moved).

Preferences

For any terminal history (x, v), define $W(x, v)$ to be the set of winning candidates:

$$W(x,v) = \{j \in N : x_j \in \mathbb{R} \text{ and } |\{i \in I : v_i = j\}| \geq |\{i \in I : v_i = j'\}| \text{ for all } j' \notin N \setminus \{j\}\}.$$

The preference relation of each office-seeker $j \in N$ over terminal histories (x, v) is represented by the payoff function

$$\begin{cases} 0 & \text{if } x_j = Out \\ -c & \text{if } x_j \in \mathbb{R} \text{ and } j \notin W(x,v) \\ b/|W(x,v)| - c & \text{if } x_j \in \mathbb{R} \text{ and } j \in W(x,v). \end{cases}$$

The preference relation of each citizen $i \in I$ over terminal histories (x, v) is represented by the payoff function

$$\begin{cases} -L & \text{if } \{j \in N : x_j \in \mathbb{R}\} = \varnothing \\ \sum_{j \in W(x,v)} u_i(x_j)/|W(x,v)| & \text{if } \{j \in N : x_j \in \mathbb{R}\} \neq \varnothing. \end{cases}$$

In this game, a strategy for each office-seeker is a position or *Out*, and a strategy for a citizen is a function that assigns to each possible profile of actions for the office-seekers either one of the candidates (an office-seeker whose action is a position) or *Abstain*. I argue that if the citizens' payoff functions are strictly concave and $b > c$ then in every subgame perfect equilibrium in which the citizens' actions in every subgame are weakly undominated, the number k of candidates satisfies $1 \leq k \leq b/c$, every candidate's position is the median of the citizens' favorite positions, and the outcome of the election is a tie among the candidates. Conversely, for any number $k \leq n$ with $1 \leq k \leq b/c$ the game has a subgame perfect equilibrium in which k office-seekers become candidates, each candidate chooses the median of the citizens' favorite positions, the candidates tie for first place, and the citizens' actions in every subgame are undominated.

In these equilibria, no candidate can benefit from deviating to a slightly different position because the citizens' strategies specify that after such a history, every citizen who prefers the median to the deviator's position votes for the *same* candidate. That is, after a history in which the position of every candidate but one is the median of the citizens' favorite position, the citizens who prefer the median to the position of the remaining candidate coordinate their votes on one of the candidates whose position is the median. By contrast, the model in the previous section assumes that such a configuration of positions leads these citizens to divide their votes equally among the candidates whose position is the median, so that the candidate whose position differs from the median wins if her position is close enough to the median.

The condition $k \leq b/c$ is required because for a strategy profile in which k office-seekers become candidates and choose the median of the citizens' favorite positions, the payoff of each candidate is $b/k - c$, whereas the payoff of a candidate who choose *Out* is 0.

The citizens' behavior in some of the equilibria with $k = 1$ is plausible: every citizen votes for the single candidate; if an office-seeker deviates to become a candidate at the same position, no citizen changes her vote (she has no positive incentive to do so), whereas if an office-seeker deviates to become a candidate at a different position, each citizen votes for the candidate whose position she prefers. The citizens' behavior in equilibria with $k \geq 2$ is less plausible: the citizens' votes are split equally among the candidates, but if an office-seeker deviates to become a candidate at a different position, the citizens rally around *one* of candidates at the median. How such coordination could occur is unclear.

If the citizens' payoff functions are strictly concave, why does the game have no equilibria in which the candidates' positions are dispersed? Here are the key points of the argument.

- In an equilibrium, all candidates tie, otherwise one of them loses and can increase her payoff by deviating to *Out*. Consequently every citizen's vote is pivotal: any change in any citizen's vote changes the set of winners.

- Thus by Lemma 4.1 each citizen either votes for a candidate whose position she likes best among all the candidates' positions or abstains, and if she abstains, she is indifferent among all the candidates' positions.

- The strict concavity of the citizens' payoff functions means that by the argument in the proof of Proposition 4.3*a* at most two positions are occupied by candidates.

- In any equilibrium with two occupied positions, say y and y', only one candidate occupies each position: if j and j' both occupy y, a citizen who is voting for j can, by switching her vote to j', change the outcome from a tie among the candidates to an outright win for j', an outcome she prefers unless she is indifferent between y and y'. If all citizens who vote for a candidate with the position y are indifferent between y and y' then they all prefer positions between y and y' to both y and y', and any candidate who deviates from y or y' to a position in the interval (y, y') attracts all their votes and thus wins.

- If two positions are occupied, each by one candidate, each citizen who is not indifferent between the positions must vote for the position she prefers and the candidates must tie. But then either candidate can deviate to the median of the citizens' favorite positions and win, because in an equilibrium of the resulting subgame each citizen votes for the candidate whose position she prefers, and a majority prefer the median to any other position.

Proposition 10.1: SPE of electoral competition game with office-motivated candidates and strategic voting

Let $\langle n, h, (u_1, \ldots, u_h), b, c, L\rangle$ be an electoral competition game with office-motivated candidates and strategic voting in which $b > c$ and each function u_i is strictly concave and has a maximizer (i's favorite position). Then $(x_1, \ldots, x_n)$ is the list of the office-seekers' strategies in a subgame perfect equilibrium of $\langle n, h, (u_1, \ldots, u_h), b, c, L\rangle$ in which every citizen's action in each subgame is weakly undominated if and only if for an integer $k \leq n$ with $1 \leq k \leq b/c$ the number of candidates $(|\{j \in N : x_j \in \mathbb{R}\}|)$ is k and the position x_j of every candidate j is the median of the citizens' favorite positions.

In an equilibrium, all candidates occupy the same position, so that every citizen is indifferent between voting for any of them and abstaining. For equilibrium, the candidates must tie, so that the same number of citizens must vote for each of them. If the number of citizens is not divisible by the number of candidates, the number of votes for each candidate cannot be the same unless some citizens abstain. One possibility is that they all abstain; I use this equilibrium in the proof because it is easy to specify.

Proof

I first show the "if" direction. Let $k \leq n$ be an integer with $1 \leq k \leq b/c$, let K be a set of office-seekers with k members, and denote the median of the citizens' favorite positions by m (which is unique, because the number of citizens is odd).

Consider a strategy profile in which the strategy x_j of each office-seeker $j \in K$ is m, the strategy of each remaining office-seeker is *Out*, and the citizens' actions after each history are given as follows, where $j^* \in K$ and, if $k \geq 2$, $\gamma(j) \in K \setminus \{j\}$ for each $j \in K$.

- History x. Every citizen chooses *Abstain*.
- History that differs from x only in that $x_j = Out$ for some $j \in K$. If $k = 1$, each citizen choose *Abstain*, her only option. If $k \geq 2$, every citizen votes for $\gamma(j)$.
- History that differs from x only in that $x_j \in \mathbb{R}$ with $x_j \neq m$ for some $j \in K$. If $k = 1$, each citizen votes for j. If $k \geq 2$, every citizen who prefers m to x_j votes for $\gamma(j)$ and every remaining citizen votes for j.
- History that differs from x only in that $x_j = m$ for some $j \notin K$. Every citizen votes for j^*.
- History that differs from x only in that $x_j \in \mathbb{R}$ with $x_j \neq m$ for some $j \notin K$. Every citizen who prefers m to x_j votes for j^* and every remaining citizen votes for j.
- History that differs from x in the actions of two or more office-seekers. By Proposition 4.2 the subgame that follows such a history has a Nash equilibrium in which each citizen's action is weakly undominated. Select one of these equilibria arbitrarily.

I claim that this strategy profile is a subgame perfect equilibrium in which every citizen's action in every subgame is weakly undominated.

In each case, the citizens' actions in each subgame constitute a Nash equilibrium. In the first, second, and fourth cases, the outcome is m regardless of the citizens' actions, so no citizen's action is weakly dominated. In the third and fifth cases, every citizen votes for a candidate whose position she likes best, so that her action is weakly undominated (Proposition 3.1).

To complete the proof, I argue that for these strategies of the citizens, no office-seeker can increase her payoff by deviating. If all office-seekers follow their strategies, the k who become candidates tie, each receiving the payoff $b/k - c$, while each remaining office-seeker receives the payoff 0. If $j \in K$ deviates to *Out*, then her payoff changes from $b/k - c$ to 0, so she is not better off. If $j \in K$ deviates to a position $z \neq m$, then if $k = 1$ she continues to be the winner, and is not better off, and if $k \geq 2$ then all citizens who prefer m to z, a majority, vote for $\gamma(j)$, one of j's ex-companions at m, so that she loses and her payoff changes to $-c$. If $j \notin K$ deviates to enter at m, then all citizens vote for the same member j^* of K, so that j loses and her payoff changes from 0 to $-c$. Finally, if $j \notin K$ deviates to enter at a position $z \neq m$, all citizens who prefer m to z, a majority, vote for j^*, so that j loses and her payoff changes to $-c$.

I now show the "only if" direction. Let (x, v) be the terminal history generated by a subgame perfect equilibrium in which every citizen's action in every subgame is weakly undominated. Denote by I the set of citizens and by $C(x)$ the set of office-seekers who choose to become candidates for this terminal history: $C(x) = \{j \in N : x_j \in \mathbb{R}\}$.

Step 1 *At least one office-seeker becomes a candidate* ($|C(x)| \geq 1$).

Proof. If no office-seeker becomes a candidate, each office-seeker's payoff is 0. Any office-seeker can increase her payoff by deviating to run as a candidate at any position, in which case she wins (regardless of the citizens' strategies) and obtains the payoff $b - c > 0$. ◁

Step 2 *Every candidate receives the same number of votes.*

Proof. If not, then at least one candidate loses, obtaining the payoff $-c$. Such a candidate can increase her payoff to 0 by deviating to *Out*. ◁

Step 3 *In the subgame following x, a citizen who abstains is indifferent among all the candidates. The payoff of a citizen who votes for a candidate is at least her payoff from any of the other candidates' positions.*

Proof. The subgame following x is the plurality rule voting game $\langle I, C(x), (V_i)_{i \in I} \rangle$ where for each $i \in I$ the function $V_i : C(x) \to \mathbb{R}$ is defined by $V_i(j) = u_i(x_j)$ for all $j \in C(x)$. Given that by Step 2 every candidate is a winner, the result follows from Lemma 4.1 applied to this game. ◁

Step 4 *The number of distinct values of x_j for $j \in C(x)$ is at most two.*

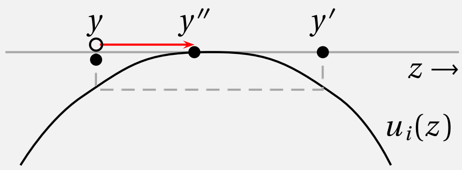
Proof. By Step 2, the set of winners of the voting subgame following x is the set $C(x)$ of candidates. The result follows from an argument like that in the proof of Proposition 4.3*a*. (That result does not apply directly because it assumes that the alternatives are distinct. Here, several candidates may choose the same position.) ◁

Step 5 *If two positions are occupied in x, exactly one candidate occupies each position.*

Proof. Suppose that the positions y and y' are occupied in x and two or more candidates occupy y. By Step 2 every candidate gets the same number of votes and hence is one of the winning candidates. Let i be a citizen who votes for a candidate at y. By Step 3, she either prefers y to y' or is indifferent between them.

If i prefers y to y' then her deviating to vote for another candidate at y induces the outcome in which that candidate wins outright. She prefers y to the outcome of (x, v), which includes y' with positive probability, so (x, v) is not a subgame perfect equilibrium.

Thus every citizen who votes for a candidate at y is indifferent between y and y'. By Step 3, every citizen who abstains is also indifferent between these positions. Now, by the strict concavity of each citizen's payoff function, positions between y and y' are better than y for every citizen who is indifferent between y and y'. Suppose that one of the candidates whose position is y deviates to a position y'' between y and y'.



By Proposition 4.1*a*, voting for the candidate at y'' is the only weakly undominated action of any citizen who is indifferent between y and y', so in the subgame following the candidate's deviation, every citizen who was voting for a candidate at y or abstaining votes for the candidate at y''. If one candidate occupies y', the candidate at y'' consequently wins outright. If more than one candidate occupies y', then by the same argument as for y, every citizen who votes for a candidate at y' is indifferent between y and y', and hence, like the citizens who were voting for a candidate at y, votes for the candidate at y'', so that in this case also the candidate at y'' wins outright. Hence (x, v) is not a subgame perfect equilibrium. ◁

Step 6 *The position of every candidate is the same, equal to the median m of the citizens' favorite positions.*

Proof. By Steps 4 and 5, the only other possibility is that the number of candidates is 2 and they choose different positions. In this case, by Step 2 the candidates tie and by Step 3 each citizen either votes for her favorite candidate or, if she is indifferent between the candidates, abstains. Thus neither candidate's position is m. Suppose that one of the candidates, say j, deviates to m. Then the only weakly undominated action of every citizen who prefers m to the position of the other candidate is to vote for j, so that j wins outright. Thus the game has no subgame perfect equilibrium in which two office-seekers become candidates.

If the position of every candidate is $x \neq m$, a candidate who deviates to m wins outright because by Proposition 4.1*a* the only weakly undominated action of every citizen who prefers m to x is to vote for her, and these citizens constitute a majority. ◁

Step 7 *The number of candidates is at most b/c.*

Proof. If the number of candidates exceeds b/c then each candidate's payoff is negative, so she is better off deviating to *Out*. ◁

Steps 1, 6, and 7 imply that in any subgame perfect equilibrium in which every citizen's action in every subgame is weakly undominated there is at least one and at most b/c candidates, and each candidate's position is m.

Exercise 10.2: Policy-motivated candidates and strategic voting

Consider a game that differs from an electoral competition game with office-motivated candidates and strategic voting $\langle n, h, (u_1, \dots, u_h), b, c, L\rangle$ only in the payoff functions of the office-seekers, who value both policies and winning. Specifically, the payoff of each office-seeker j to the strategy profile (x, v) is $-D < 0$ if no office-seeker becomes a candidate and otherwise is her payoff in the game with office-motivated candidates plus $\sum_{l \in W(x,v)} U_j(x_l)/|W(x,v)|$, where $U_j : \mathbb{R} \to \mathbb{R}$ is a single-peaked function. Study the subgame perfect equilibria of this game in which every citizen's action in every subgame is weakly undominated.

10.3 Sequential entry

In Section 8.1.2 I present a model of electoral competition in which two candidates act sequentially. I now present a model that differs in two main respects. First, there is an arbitrary finite number of players, rather than two. Second, each player, referred to as an office-seeker, chooses whether to become a candidate or to stay out of the election; if she chooses to become a candidate, she selects a position. The office-seekers move one at a time, and when choosing an action, each office-seeker observes the actions chosen by her predecessors. In one environment that the model fits, the opportunity to act arises randomly for each individual in a finite set; the individual selected in period i becomes office-seeker i. The significant assumptions are that each individual gets one opportunity to act, and when it occurs she knows the actions of the individuals who moved before her.

As in the model in Section 10.1, the set of possible positions is the real line, and in the background is a continuum of citizens, whose votes determine the winning candidate(s). Each citizen has single-peaked preferences and votes for a candidate whose position is closest to her favorite position; if a position is occupied by several candidates, these candidates share equally the votes for that position. The winning candidates are the ones who tie for the most votes. Each candidate prefers to tie with $k-1$ than with k other candidates, for any $k = 2, \ldots, n$, prefers to tie with all the other candidates than to stay out of the competition, and prefers to stay out of the competition than to lose.

Definition 10.2: Sequential electoral competition game with a continuum of citizens and office-motivated candidates

A *sequential electoral competition game with a continuum of citizens and office-motivated candidates* $\langle F, n, \trianglerighteq \rangle$, where

- n is a positive integer (the number of office-seekers)
- F is a nonatomic distribution with support an interval of real numbers (the distribution of the citizens' favorite positions)
- $\trianglerighteq$ is a preference profile over $(\mathbb{R} \cup \{Out\})^n$ (the profile of the office-seekers' preferences over action profiles)

is the following extensive game with perfect information.

Players
The set $N = \{1, \ldots, n\}$ (of office-seekers).

Terminal histories

The set of sequences $(x_1, \ldots, x_n)$ where $x_i \in \mathbb{R} \cup \{Out\}$ for each $i \in N$.

Player function

The function P given by $P(\varnothing) = 1$ and $P(x_1, \ldots, x_k) = k + 1$ for $k = 1, \ldots, n-1$ and any history $(x_1, \ldots, x_k)$.

Preferences

For each player j and terminal history x, let $w_j(x)$ be the number of candidates (including j) with whom j ties for first place, with a value of 0 meaning that j is not one of the candidates tied for first place. The preference relation $\trianglerighteq_j$ of each player j over terminal histories satisfies $x \triangleright_j y$ if any of the following conditions holds:

a. $1 \le w_j(x) < w_j(y)$ (tying with fewer candidates is preferred)

b. $w_j(x) = n$ and $y_j = Out$ (tying with all the other candidates is preferred to staying out of the competition)

c. $x_j = Out$ and $w_j(y) = 0$ (staying out of the competition is preferred to entering and losing).

Proposition 10.2: Existence of SPE of sequential electoral competition game

Every sequential electoral competition game with a continuum of citizens and office-motivated candidates has a subgame perfect equilibrium.

Proof

Denote the number of players by n. Each player's preferences are represented by a payoff function that takes at most $n+2$ values (for winning outright, tying with k of the other players for $k = 1, \ldots, n-1$, losing, and staying out of the competition). Thus the result follows from Proposition 16.8.

If there are two office-seekers then the game has a unique subgame perfect equilibrium outcome, in which both office-seekers enter at the median of the distribution of the citizens' favorite positions. Thus the outcome is the same as the outcome of the unique Nash equilibrium of the game in which the office-seekers move simultaneously (Proposition 8.4). (This result is closely related to Proposition 8.3, given Proposition 1.4.)

Proposition 10.3: SPE of sequential electoral competition game with two office-seekers

Every subgame perfect equilibrium of a sequential electoral competition game with a continuum of citizens and two office-motivated candidates $\langle F, 2, \trianglerighteq \rangle$ generates the terminal history in which each office-seeker is a candidate with position equal to the median of F.

Proof

Denote the median of F by m. If player 1 chooses m then player 2's only optimal response is m, because she loses if she enters at any other position. If player 1 chooses a position other than m, then player 2 wins outright by entering at m. Thus all of player 2's optimal actions lead her to win outright, and hence player 1 to lose. So player 1's only optimal action in a subgame perfect equilibrium is to enter at m, leading player 2 to do the same.

Note that although the two-player game has only one subgame perfect equilibrium outcome, it has many subgame perfect equilibria. In the subgame following player 1's entry at any position x different from m, any position closer to m than x results in player 2's winning, and hence is an optimal action for her. Thus any strategy pair in which player 1 chooses m and player 2's strategy s_2 satisfies $s_2(m) = m$ and $|s_2(x) - m| < |m - x|$ for all $x \neq m$ is a subgame perfect equilibrium.

Exercise 10.3: Sequential electoral competition with two policy-motivated candidates

Consider a variant of a sequential electoral competition game with a continuum of citizens and office-motivated candidates in which the players are policy-motivated. Specifically, for each player i there is a single-peaked function $U_i : \mathbb{R} \to \mathbb{R}$, and i's payoff to a terminal history in which the positions $x^1, \ldots, x^l$ are tied for first place is the expected value of U_i when each x^j occurs with the same probability, $1/l$. Assume that if no player enters then the outcome is an arbitrary position x_0. Denote by m the median of the distribution F of the citizens' favorite positions. Assume that the favorite position of at least one player is less than m and the favorite position of at least one player is greater than m. Consider the game for $n = 2$. Some subgames of this game do not have (exact) subgame perfect equilibria. For any small $\varepsilon > 0$, find a strategy profile for which player 1's strategy is optimal and player 2 cannot increase her payoff by more than ε by changing her action after any history.

For a game with three office-seekers, the equilibrium outcome is more interesting: the first and last player to move enter at the median of the distribution of the citizens' favorite positions and the second player stays out. Thus the outcome is consistent with a much-studied claim, often called "Duverger's Law", that plurality rule tends to lead to there being two candidates (or two parties if you are willing to equate a candidate in the model with a political party).

Proposition 10.4: SPE of sequential electoral competition game with three office-seekers

Every subgame perfect equilibrium of a sequential electoral competition game with a continuum of citizens and three office-motivated candidates $\langle F, 3, \trianglerighteq \rangle$ generates the terminal history in which the first and third office-seekers become candidates at the median of F and the second office-seeker chooses *Out*.

Proof

Denote the median of F by m.

First consider the subgame following player 1's entry at m. In this subgame, for every position of player 2 there is a position of player 3 such that player 3 wins outright:

- if player 2 enters at m, player 3 wins outright at any position sufficiently close to m
- if player 2 enters at a position different from m, player 3 wins at a position closer to m on the opposite side of m.

So in every subgame perfect equilibrium of the subgame, player 2 stays out and player 3 enters at m, and hence player 1 ties for first place.

Now consider the subgame following player 1's entry at a position different from m.

- The subgame has no subgame perfect equilibrium in which player 1 wins outright. In such an equilibrium players 2 and 3 must stay out (because entering and losing is worse than staying out), in which case player 3 can deviate to entering at m and win outright.

- The subgame has no subgame perfect equilibrium in which player 1 ties only with player 3. In such an equilibrium, the positions of players 1 and 3 are either the same, in which case player 3 can deviate to m and win, or symmetric around m, in which case player 3 can deviate slightly closer to m and win.

- The subgame has no subgame perfect equilibrium in which player 1 ties only with player 2. In such an equilibrium, the positions of players 1 and 2 are either the same, in which case player 3 can enter at m and win outright, or symmetric about m. If they are symmetric about m, with say $x_1 = m - \delta$ and $x_2 = m + \delta$, then player 3 loses if she enters at any position that is at most x_1 or at least x_2, and δ has to be small enough that player 3 does not win or tie with the other two candidates if she enters at any point in (x_1, x_2). (Such values of δ exist.) Now suppose that player 2 deviates to $x_2 - \varepsilon$ for some small $\varepsilon > 0$. Then player 3 still cannot win or tie at any position, and player 2 wins outright if player 3 does not enter, so that player 1 loses.

- The subgame has no subgame perfect equilibrium in which all three players tie. In such an equilibrium either every player's position is the same, in which case player 3 can deviate slightly and win outright, or at least one player's position is occupied only by her, in which case she can deviate to a position slightly closer to the other players and win outright.

We conclude that in every subgame perfect equilibrium of the subgame player 1 loses.

Player 1 prefers to tie with one other player than to lose or to stay out of the competition, so in every subgame perfect equilibrium of game, player 1 enters at m, player 2 stays out, and player 3 enters at m.

Exercise 10.4: Sequential electoral competition with three policy-motivated candidates

Consider the variant with policy-motivated candidates of a sequential electoral competition game with a continuum of citizens and office-motivated candidates in Exercise 10.3 in which there are three players. Denote the favorite position of player i by $\hat{x}_i$ and assume that $\hat{x}_1 < m < \hat{x}_2 = \hat{x}_3$ and $F(\hat{x}_2) < \frac{2}{3}$. What are the subgame perfect equilibrium outcomes of the game?

What if the number of office-seekers exceeds three? Based on partial analyses of games with four or more players in collaboration with Amoz Kats, I conjectured in the mid-1980s that for any finite number n of players the game $\langle F, n, \trianglerighteq \rangle$ has a unique subgame perfect equilibrium outcome, in which the first player enters at the median of the citizens' favorite positions, the next $n-2$ players choose *Out*, and the last player joins the first one at the median. This conjecture remains unproved and uncontradicted.

A variant of the model has a unique subgame perfect equilibrium for any number of office-seekers. Suppose that if two or more candidates are tied for first place, the one who entered first wins outright, rather than each of the tied candidates winning with the same probability. Then given the assumption that each candidate prefers to stay out than to enter and lose, in any equilibrium at most one office-seeker enters. The following exercise asks you to determine which one does so.

Exercise 10.5: Sequential electoral competition game with priority for early entrants

Consider a variant of a sequential electoral competition game with a continuum of citizens and office-motivated candidates in which each player's preferences are represented by a function that assigns 1 to every terminal history in which she is the player with the smallest index among those tied for the highest number of votes, -1 to every other terminal history in which she enters, and 0 to every terminal history in which she chooses *Out*. Find the subgame perfect equilibrium outcome (outcomes?) of this game.

The next exercise invites you to study the subgame perfect equilibrium outcomes of a game with three office-seekers in which the winner is determined by plurality rule with a runoff, rather than ordinary plurality rule.

Exercise 10.6: Sequential electoral competition with three office-motivated candidates and a runoff

Consider the variant of a sequential electoral competition game with a continuum of citizens and three office-motivated candidates in which the winner of the election is determined by plurality rule with a runoff, as described in Exercise 10.1. Show that in every subgame perfect equilibrium of the game all three players enter as candidates.

10.4 Citizen-candidates

The models I present in this section differ in two main respects from those in the previous sections. First, each office-seeker chooses only whether to run for election; she does not choose a position. If she runs and is elected, she implements her favorite policy, which the voters know. One motivation for this assumption is that each office-seeker is self-interested and there is no mechanism by which she can commit to a policy different from her favorite policy. Second, office-seekers care both about policy and about winning.

The set of office-seekers is assumed to coincide with the set of citizens: any citizen can run as a candidate. Each citizen first chooses whether to run as a candidate and then votes for a candidate (one of the citizens who chose to run). The candidate receiving the most votes wins and implements her favorite policy. If several candidates are tied for the most votes, each of them is selected with the same probability to be the winner.

Regarding the citizens' voting behavior, one option is to assume that each citizen votes ("sincerely") for the candidate whose position she likes best, as in Section 10.1. The resulting model is a strategic game in which each citizen chooses only whether to run as a candidate. Another option is to assume that voting is "strategic". If the citizens are perfectly informed, the resulting model is an extensive game with perfect information and simultaneous moves in which each citizen first chooses whether to run as a candidate, then chooses the candidate for whom to vote, as for the model in Section 10.2. Whether the model with sincere voting or the one with strategic voting under perfect information better captures the reality of imperfectly informed citizens who choose how to cast their votes is unclear.

In both models, the set of positions is the set of real numbers and the set of players is the set $N = \{1, \ldots, n\}$ of citizens. The preferences of each citizen $i \in N$ regarding lotteries over positions are represented by the expected value of a Bernoulli payoff function u_i. I assume that the payoff $u_i(z)$ of each citizen i for any position z depends only on the distance between z and i's favorite position, and the form of the relationship is the same for all individuals. Specifically, for some decreasing function $u : \mathbb{R}_+ \to \mathbb{R}_-$ with $u(0) = 0$, for each citizen i the function $u_i : \mathbb{R} \to \mathbb{R}_-$ is defined by $u_i(z) = u(|\hat{x}_i - z|)$ for all z, where $\hat{x}_i$ is i's favorite position, as illustrated in Figure 10.2.

A candidate is one of the winners if she obtains at least as many votes as every other candidate. In the model with sincere voting, a strategic game, each citizen votes for a candidate whose position is closest to her favorite position. If the positions of several candidates are equally close, the citizen's vote is divided equally among those candidates. (For simplicity, I allow fractional votes. You can alter-

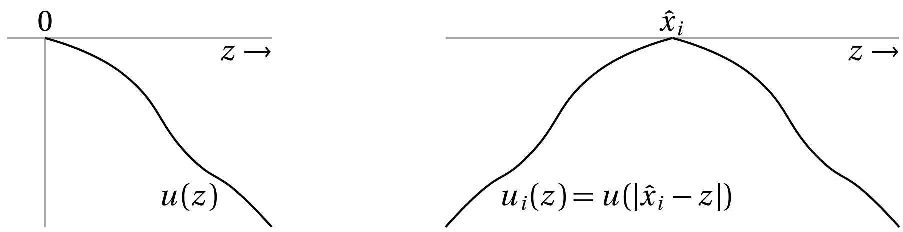


Figure 10.2 The Bernoulli payoff function of citizen i over positions in the citizen-candidate models in Section 10.4.

natively imagine that a large number of citizens have any given favorite position z, and their votes are divided equally among the candidates whose positions are closest to z. This alternative model is silent about how such coordination could be achieved.) The actions available to each citizen are *Run* (become a candidate) and *Out*. Let $a = (a_1, \ldots, a_n)$ be an action profile for which $a_j = Run$ for at least one citizen $j \in N$. For any citizen $i \in N$, the set of candidates whose positions are closest to i's favorite position is

$$C(i,a) = \{j \in N : a_j = Run \text{ and } |\hat{x}_j - \hat{x}_i| \leq |\hat{x}_l - \hat{x}_i| \text{ for all } l \in N\}$$

and the fraction of i's vote that goes to any citizen j for whom $a_j = Run$ is

$$v_i(j,a) = \begin{cases} 1/|C(i,a)| & \text{if } j \in C(i,a) \\ 0 & \text{otherwise,} \end{cases}$$

so that the total number of votes obtained by citizen j is

$$V_j(a) = \sum_{i \in N} v_i(j,a)$$

and the set of winners of the election is

$$W(a) = \{i \in N : a_i = Run \text{ and } V_i(a) \geq V_j(a) \text{ for all } j \text{ with } a_j = Run\}. \tag{10.2}$$

If $a_i = Out$ for all $i \in N$ (no citizen runs as a candidate), let $W(a) = \varnothing$.

In the model with strategic voting, an extensive game with perfect information and simultaneous moves, the citizens' strategies determine the number of votes for each candidate. The set of terminal histories consists of the action profile $(Out, \ldots, Out)$, in which no citizen runs as a candidate, and every sequence $((a_1, \ldots, a_n), (v_1, \ldots, v_n))$ for which $a_j \in \{Run, Out\}$ for each $j \in N$, $a_j = Run$ for at least one $j \in N$, and v_j is a citizen $i \in N$ for whom $a_i = Run$ (the candidate for whom j votes). For a terminal history (a, v), the set of winners of the election is the set of citizens who run as candidates and obtain at least as many votes as any other candidate:

$$W(a,v) = \{i \in N : a_i = Run \text{ and } |\{j \in N : v_j = i\}| \geq |\{j \in N : v_j = l\}| \text{ for all } l \in N\}. \tag{10.3}$$

For the terminal history $a = (Out, \dots, Out)$, let $W(a) = \varnothing$.

In both models, each citizen values the policy of the winner of the election. In addition, if she runs as a candidate she incurs a cost $c > 0$, and, if she wins election, obtains a benefit $b \geq 0$. These amounts appear as linear terms in her payoff. For example, it citizen i runs as a candidate and wins outright, her payoff is $u_i(\hat{x}_i) + b - c$ (which is equal to $b - c$ given that $u_i(\hat{x}_i) = u(0) = 0$). I assume that if no citizen runs as a candidate then a fixed position x_0 is realized.

Precisely, for any set $\mathcal{W} \subseteq N$ of winning candidates, the payoff of any citizen i is

$$V_i^O(\mathcal{W}) = \begin{cases} u_i(x_0) & \text{if } \mathcal{W} = \varnothing \\ \sum_{j \in \mathcal{W}} u_i(\hat{x}_j)/|\mathcal{W}| & \text{if } \mathcal{W} \neq \varnothing \end{cases} \tag{10.4}$$

if she does not run as a candidate and

$$V_i^R(\mathcal{W}) = \begin{cases} \sum_{j \in \mathcal{W}} u_i(\hat{x}_j)/|\mathcal{W}| - c & \text{if } i \notin \mathcal{W} \\ \left(\sum_{j \in \mathcal{W}} u_i(\hat{x}_j) + b\right)/|\mathcal{W}| - c & \text{if } i \in \mathcal{W} \end{cases} \tag{10.5}$$

if she does run (in which case $\mathcal{W} \neq \varnothing$).

Definition 10.3: Electoral competition game with citizen-candidates who vote sincerely

An *electoral competition game with citizen-candidates and sincere voting* $\langle n, (\hat{x}_1, \dots, \hat{x}_n), u, b, c, x_0 \rangle$, where

- n is an odd positive integer (the number of citizens)
- $\hat{x}_i \in \mathbb{R}$ for $i = 1, \dots, n$ (the favorite position of citizen i)
- $u : \mathbb{R}_+ \to \mathbb{R}_-$ is a decreasing function with $u(0) = 0$
- b is a nonnegative number (the benefit from winning)
- c is a positive number (the cost of running as a candidate)
- $x_0 \in \mathbb{R}$ (the policy realized if no citizen runs as a candidate)

is a strategic game with the following components.

Players

The set $N = \{1, \dots, n\}$ (citizens).

Actions

The set of actions of each citizen is $\{Run, Out\}$.

Preferences

The preferences of each citizen $i \in N$ over action profiles a are represented by the payoff function that assigns the payoff $V_i^O(W(a))$ if $a_i = Out$ and the payoff $V_i^R(W(a))$ if $a_i = Run$, where $W(a)$ is given by (10.2), V_i^O is given by (10.4), V_i^R is given by (10.5), and $u_i(z) = u(|\hat{x}_i - z|)$ for all z.

Definition 10.4: Electoral competition game with citizen-candidates who vote strategically

An *electoral competition game with citizen-candidates and strategic voting* $\langle n, (\hat{x}_1, \ldots, \hat{x}_n), u, b, c, x_0 \rangle$, where the variables have the same meanings as in Definition 10.3, is an extensive game with perfect information and simultaneous moves with the following components.

Players

The set $N = \{1, \ldots, n\}$ (citizens).

Terminal histories

The action profile $(Out, \ldots, Out)$ plus all sequences $((a_1, \ldots, a_n), (v_1, \ldots, v_n))$ with $a_i \in \{Run, Out\}$ for $i = 1, \ldots, n$, $a_j = Run$ for some $j \in N$, and $v_i \in \{j \in N : a_j = Run\}$ for $i = 1, \ldots, n$ (v_i is the citizen-candidate for whom i votes).

Player function

The function P with $P(\varnothing) = N$ and $P(a_1, \ldots, a_n) = N$ for all $(a_1, \ldots, a_n) \in \{Run, Out\}^n$ with $\{j \in N : a_j = Run\} \neq \varnothing$ (all citizens move simultaneously at the start of the game and again, as voters, after any profile of initial actions for which at least one citizen chooses to run).

Preferences

The preferences of each citizen $i \in N$ over terminal histories are represented by the payoff function that assigns to the terminal history $(Out, \ldots, Out)$ the payoff $V_i^O(\varnothing)$, and to any terminal history (a, v) with $a_j = Run$ for some $j \in N$, the payoffs $V_i^O(W(a, v))$ if $a_i = Out$ and $V_i^R(W(a, v))$ if $a_i = Run$, where $W(a, v)$ is given by (10.3), V_i^O is given by (10.4), V_i^R is given by (10.5), and $u_i(z) = u(|\hat{x}_i - z|)$ for all z.

In the game with strategic voting, each voting subgame has many Nash equilibria (see Section 3.1). I restrict to equilibria in which each citizen's action in every voting subgame is weakly undominated.

Definition 10.5: Equilibrium of electoral competition game with citizen-candidates

In an electoral competition game with citizen-candidates in which voting is sincere, an *equilibrium* is a Nash equilibrium. In a game in which voting is strategic, an *equilibrium* is a subgame perfect equilibrium in which each citizen's vote in every subgame following the citizens' decisions to run as candidates is weakly undominated.

Suppose that no citizen enters as a candidate. Then the payoff of each citizen i is $u_i(x_0)$. If citizen i deviates to become a candidate, in both games she wins and obtains the payoff $b-c$. Thus an equilibrium in which no citizen enters as a candidate exists if and only if $b-c \leq u_i(x_0)$ for the citizen i whose favorite position is furthest from x_0. I now consider some more interesting equilibria.

One-candidate equilibria

For some ranges of the parameter values, both games have equilibria in which exactly one citizen runs as a candidate. In such equilibria, that citizen has to be no better off deviating to *Out*, and every other citizen has to be no better off deviating to *Run*. If the candidate's favorite position differs from m then a citizen with favorite position m wins if she deviates to *Run*. (For the game with strategic voting, this conclusion follows from Corollary 3.1.) The entrant receives the payoff $b-c$, so in a one-candidate equilibrium the candidate's favorite position x satisfies $b-c \leq u(|m-x|)$, or equivalently $|m-x| \leq u^{-1}(b-c)$. In particular, if $b \geq c$ then no equilibrium with $x \neq m$ exists, because $u(z) < 0$ for every $z > 0$.

Under what conditions does a one-candidate equilibrium exist in which the candidate's position is m? The answers for the two games differ. Suppose that voting is sincere and another citizen's favorite position is m. If that citizen deviates to *Run* then she ties with the existing candidate and hence obtains the payoff $\frac{1}{2}b-c$ rather than 0. Consequently in this case a one-candidate equilibrium in which the candidate's position is m exists only if $b \leq 2c$. But if voting is strategic, the voting subgame following the entry of another citizen with favorite position m has an equilibrium in which the original candidate wins, because all citizens are indifferent between the her and the entrant. An entrant whose favorite position differs from m loses, so in this case a one-candidate equilibrium exists even if b is large; the only condition required is $u(|m-x_0|) \leq b-c$, so that the existing candidate does not prefer to deviate to *Out*.

Proposition 10.5: One-candidate equilibria of electoral competition game with citizen-candidates

Let $\langle n, (\hat{x}_1, \ldots, \hat{x}_n), u, b, c, x_0 \rangle$ be an electoral competition game with citizen-candidates and denote by m the median of the citizens' favorite positions.

a. Whether voting is sincere or strategic, in every equilibrium in which one citizen runs as a candidate, her position x is m if $b \geq c$ and satisfies $|m - x| \leq u^{-1}(b - c)$ if $b < c$.

b. If either voting is sincere and only one citizen has favorite position m or voting is strategic, the game has a one-candidate equilibrium in which the candidate's favorite position is m if and only if $u(|m - x_0|) \leq b - c$. If voting is sincere and more than one citizen has favorite position m, the game has such an equilibrium if and only if $u(|m - x_0|) \leq b - c$ and $b \leq 2c$.

Proof

Part a is proved in the text. To prove part b, first suppose that voting is sincere. Consider the action profile in which a citizen, say i, with favorite position m chooses *Run* and every other citizen chooses *Out*. Citizen i's payoff is $b-c$, and if she deviates to *Out* her payoff changes to $u(|m-x_0|)$. If a citizen whose favorite position differs from m deviates to *Run*, she loses, and hence is worse off. If another citizen with favorite position m deviates to *Run*, her payoff changes from 0 to $\frac{1}{2}b - c$. Thus if i is the only citizen with favorite position m, the action profile is an equilibrium if and only if $u(|m - x_0|) \leq b - c$, and if another citizen has favorite position m then it is an equilibrium if and only if $u(|m - x_0|) \leq b - c$ and $b \leq 2c$.

Now suppose that voting is strategic. Let s be a strategy profile in which a citizen, say i, with favorite position m runs as a candidate, every other citizen chooses *Out*, and the citizens vote as follows. In the subgame reached if the citizens adhere to s and in any subgame following the entry of i and another citizen with favorite position m, all citizens vote for i. In every other subgame, the citizens' vote profile is any Nash equilibrium of the subgame in which each citizen's action is weakly undominated, the existence of which is ensured by Proposition 4.2. (In particular, in any subgame in which two citizens run as candidates, by Corollary 3.1 each citizen who is not indifferent between the candidates votes for the candidate whose favorite position she prefers.) The outcome of s is a win for i, whose payoff is $b-c$.

The citizens' action profile in the subgame following the entry of i alone is the only one available to them. Their action profile in the subgame in which i and another citizen with favorite position m run as candidates is a Nash equilibrium in which each citizen's action is weakly undominated because no change in any citizen's vote affects the outcome. Their action profile in every other subgame is a Nash equilibrium by construction.

Now consider changes in the citizens' actions at the start of the game. If i deviates to *Out*, her payoff changes from $b-c$ to $u(|m-x_0|)$, so she is no better off if $u(|m-x_0|) \leq b-c$. If another citizen with favorite position m deviates to *Run*, she loses (because all the citizens continue to vote for i), and is thus worse off, and if a citizen with a favorite position different from m deviates to *Run* she also loses. Thus s is an equilibrium if $u(|m-x_0|) \leq b-c$.

Now let s be a subgame perfect equilibrium in which each citizen's action in every voting subgame is weakly undominated, exactly one citizen, say i, runs as a candidate, and i's favorite position is m. Then i is no better off deviating from s_i to *Out* at the start of the game, so that $u(|m-x_0|) \leq b-c$.

One-candidate equilibria of the game with strategic voting in which $b > 2c$ and the candidate's favorite position is m are vulnerable to uncertainty regarding the outcome of the election. In these equilibria, the deviation of another citizen to *Run* results in that candidate's losing. If the deviator's favorite position is m, then all citizens are indifferent between the candidates and in the equilibrium a majority of them continue to vote for the original candidate. If the deviator's favorite position x differs from m, then every citizen votes for the candidate whose position she prefers (given the assumption that citizens' voting strategies are weakly undominated), and hence the original candidate wins, because a majority of citizens prefer m to x. However, if x is close to m, the margin of victory of the original candidate is small. In fact, if only one citizen has favorite position m and x is the closest favorite position to m, this margin of victory is exactly one vote.

Now suppose that with positive probability each citizen fails to vote, independently of every other citizen, and this probability is the same for every citizen. Then a deviator with favorite position $x \neq m$ wins with positive probability. Suppose specifically that only one citizen has favorite position m and x is the closest favorite position to m. Then the deviator wins if the number of citizens who vote among the $\frac{1}{2}(n-1)$ who prefer x to m exceeds the number who do so among the $\frac{1}{2}(n+1)$ who prefer m to x (where n is the number of citizens). The probability of this event is less than $\frac{1}{2}$, but if n is large it is close to $\frac{1}{2}$ (Eguia 2007, Theorem 1). Thus in this case if $b > 2c$, one-candidate equilibria do not exist in the model with strategic voting when the number of citizens is large, just as by part b of the result they do not exist when voting is sincere.

Agglomerated equilibria

Do the games have multi-candidate equilibria in which all the candidates' positions are the same? Neither game has such equilibria in which the common position differs from the median of the citizens' favorite positions. The reason is that such configurations are vulnerable to entry: an entrant whose favorite position is the median, for example, wins whether voting is sincere or strategic, because optimal strategic voting between two distinct alternatives is sincere (Corollary 3.1). But the conclusions regarding equilibria in which the favorite position of every candidate is the median of the citizens' favorite positions differ between the games. Under sincere voting, if fewer than a third of the citizens' favorite positions are equal to the median then either an entrant whose favorite position is slightly less than the median or one whose favorite position is slightly greater than the median wins, because after her entry the votes of the citizens who prefer the median are split equally among the existing candidates. However, under strategic voting, the subgame following such entry has an equilibrium in which the votes of all the citizens who prefer the median go to one specific candidate among those whose favorite position is the median, so that the entrant loses. Thus if voting is strategic the game has an equilibrium in which two or more citizens run as candidates and all of their favorite positions are the median, but if voting is sincere and fewer than a third of the citizens' favorite positions are equal to the median then it has no such equilibrium.

Proposition 10.6: Agglomerated equilibria in electoral competition game with citizen-candidates

Let $\langle n, (\hat{x}_1, \ldots, \hat{x}_n), u, b, c, x_0 \rangle$ be an electoral competition game with citizen-candidates and denote by m the median of the citizens' favorite positions.

a. The game has no equilibrium in which two or more citizens run as candidates and all of their favorite positions are the same, different from m.

b. If voting is sincere and fewer than a third of the citizens have favorite positions of m then the game has no equilibrium in which two or more citizens run as candidates and all of their favorite positions are m.

c. For any $k \geq 2$, if voting is strategic, $b \geq kc$, and at least k citizens have favorite position m, then the game has an equilibrium in which k citizens run as candidates and all of their favorite positions are m.

Proof

Consider an action profile in which $k \geq 2$ citizens with favorite position z run as candidates and every other citizen chooses *Out*. If voting is sincere, the candidates tie. If voting is strategic, the action profile is consistent with equilibrium only if the candidates tie, because a candidate who loses is better off deviating to *Out*. In both cases each candidate's payoff is $b/k - c$. If a candidate deviates to *Out*, her payoff becomes 0, so for equilibrium we need $b \geq kc$.

a. If $z \neq m$ then whether voting is sincere or strategic, a citizen with favorite position m wins if she deviates to *Run*. (If voting is strategic, this conclusion follows from Corollary 3.1.) Her payoff is thus $b - c$ rather than $u(|m - z|) < 0$, so that for equilibrium we need $b < c$. This condition is inconsistent with $b \geq kc$, so no equilibrium exists in which $k \geq 2$ citizens with favorite position $z \neq m$ run as candidates and every other citizen chooses *Out*.

b. Now suppose that $z = m$ and voting is sincere. If fewer than a third of the citizens have favorite positions of m, then either more than a third of them have favorite positions less than m or more than a third have favorite positions greater than m (or both). The two cases are symmetric; suppose the former. Let j be a citizen whose favorite position is largest among the citizens with favorite positions less than m. If j deviates to *Run*, she obtains the votes of all citizens with favorite positions less than m, whereas each of the $k \geq 2$ original candidates gets an equal share of the votes of the remaining citizens. Thus j wins and obtains the payoff $b - c$ rather than $u(|m - \hat{x}_j|) < 0$. So for the original action profile with k candidates to be an equilibrium we need $b < c$, which is inconsistent with the requirement $b \geq kc$.

c. If $z = m$ and voting is strategic, the subgame following the deviation of any citizen j from *Out* to *Run* has an equilibrium in weakly undominated strategies in which all of the citizens who prefer m to j's position vote for the same candidate with position m, so that j loses. Thus if $b \geq kc$ and at least k citizens have favorite position m, the game has an equilibrium in which k citizens with favorite position m are candidates.

The equilibria of the game with strategic voting in part *c* are vulnerable to uncertainty for the same reason that the one-candidate equilibria for $b > 2c$ in the previous result are vulnerable. Suppose as before that each citizen independently fails to vote with the same probability. Then if a citizen with favorite position x close to m deviates to *Run*, she wins with positive probability, making her deviation worthwhile if b is large enough.

Dispersed equilibria

Equilibria in which the candidates' positions are dispersed are possible in both models. In such an equilibrium with two candidates, the outcome must be a tie, because otherwise the loser can deviate to *Out* without affecting the outcome, saving the cost c. In a two-alternative voting subgame of the game with strategic voting, the only weakly undominated action of a citizen who is not indifferent between the alternatives is to vote sincerely, for her preferred candidate, so whether voting is sincere or strategic the candidates' equilibrium positions must be symmetric about the median m of the citizens' favorite positions, say $m - \delta$ and $m + \delta$ for some $\delta > 0$. If one of the candidates deviates to *Out*, the other candidate wins, so for each value of the entry cost c there is a lower bound on δ for which an equilibrium of this type may exist.

To investigate the possibility of such an equilibrium further, consider deviations by citizens from *Out* to *Run*. First suppose that a citizen whose favorite position is outside $(m-\delta, m+\delta)$ deviates to *Run*. If voting is sincere, she loses. If voting is strategic, the resulting voting subgame has a Nash equilibrium in which every citizen votes as she would in the absence of the entrant, so that the entrant loses. In this equilibrium no citizen votes for her least preferred candidate, so if the number of citizens is at least five then by Proposition 4.1*b* no citizen's action is weakly dominated.

Now consider the deviation to *Run* for a citizen whose favorite position is in $(m - \delta, m+\delta)$. If voting is sincere, then if δ is large enough such an entrant surely wins, and if she wins she may be better off. Part *b* of the next result gives a condition for such an entrant not to win, so that the configuration is an equilibrium. If voting is strategic, then no matter how large is δ, as long as the number of citizens with favorite position m is not too large the voting subgame following the entry of a candidate between $m-\delta$ and $m+\delta$ has an equilibrium in which the entrant loses. In this equilibrium, the citizens whose favorite positions are m vote for the entrant and every other citizen votes for the remaining candidate whom she prefers; no citizen votes for her least preferred candidate, so no citizen's action is weakly dominated. Thus in this case the model with strategic voting, unlike the one with sincere voting, has equilibria in which the separation between the candidates' positions is arbitrarily large.

Proposition 10.7: Two-candidate equilibria of electoral competition game with citizen-candidates

Let $\langle n, (\hat{x}_1, \ldots, \hat{x}_n), u, b, c, x_0 \rangle$ be an electoral competition game with citizen-candidates and denote the median of the citizens' favorite positions by m.

a. Whether voting is sincere or strategic, in any two-candidate equilibrium in which the candidates' positions differ, these positions are $m - \delta$ and $m + \delta$ for some $\delta > 0$ and either (*i*) $b > 2c$ or (*ii*) $b \leq 2c$ and $\delta \geq \frac{1}{2}u^{-1}(b - 2c)$.

b. Suppose that voting is sincere and the citizens' favorite positions are equally-spaced: for some $\Delta > 0$ we have $\hat{x}_i - \hat{x}_{i-1} = \Delta$ for $i = 2, \ldots, n$. If p is a positive integer such that (*i*) the conditions in part *a* are satisfied for $\delta = p\Delta$ and (*ii*) fewer than a third of the citizens' favorite positions are in $[m - \frac{1}{2}p\Delta, m + \frac{1}{2}p\Delta]$, then the game has a two-candidate equilibrium in which the candidates' positions are $m - p\Delta$ and $m + p\Delta$.

c. If voting is strategic and fewer than $\frac{1}{3}(n-4)$ of the citizens' favorite positions are m (which requires $n \geq 7$), then for every value of δ satisfying the conditions in part *a* for which there exist citizens with favorite positions $m - \delta$ and $m + \delta$, the game has a two-candidate equilibrium in which the candidates' positions are $m - \delta$ and $m + \delta$.

Proof

a. The argument in the text shows that the candidates' positions in such an equilibrium are $m - \delta$ and $m + \delta$. (Note that while such positions are necessary for a tie, they are not sufficient if voting is sincere and m is the favorite position of more than one citizen.) The payoff of each candidate is $\frac{1}{2}b - c + \frac{1}{2}u(2\delta)$. If she deviates to *Out*, her payoff becomes $u(2\delta)$, so her entry is optimal if and only if $\frac{1}{2}b - c + \frac{1}{2}u(2\delta) \geq u(2\delta)$, or $b - 2c \geq u(2\delta)$, establishing part *a*.

b. Under condition (*i*), neither candidate is better off deviating to *Out*. Under condition (*ii*), any citizen who deviates from *Out* to *Run* loses and her entry either does not affect the outcome or changes it from a tie between the two existing candidates to a win for the one she likes less. Thus she is worse off.

c. Let s be a strategy profile in which two citizens, i with favorite position $m-\delta$ and j with favorite position $m+\delta$, run as candidates, every other citizen chooses *Out*, and the citizens vote as follows. In each case I argue that the profile of votes is a Nash equilibrium of the subgame in which no citizen's action is weakly dominated.

Subgame that results if citizens adhere to s

Citizens with favorite positions less than m vote for i, citizens with favorite positions greater than m vote for j, and the votes of citizens with favorite position m (who are indifferent between the candidates) are split between i and j in such a way that the candidates tie.

This action profile is a Nash equilibrium in which every citizen with favorite position different from m votes for the candidate she prefers. Thus by Corollary 3.1 every citizen's action is weakly undominated.

Subgame in which one citizen runs as a candidate

Every citizen votes for the candidate (she has no choice).

Subgame in which candidates are i, j, and a citizen l with favorite position in $(m-\delta, m+\delta)$

Every citizen whose favorite position is m votes for l and every other citizen votes for i if her favorite position is less than m and for j if her favorite position is greater than m.

Denote by n_m the number of citizens with favorite position m. Then l receives n_m votes, and i and j each receives $\frac{1}{2}(n-n_m)$ votes. Thus given that $n_m < \frac{1}{3}(n-4)$, i and j tie for first place and each receives at least two more votes than does l. Hence no change in any citizen's vote causes l to win; every change in a citizen's vote either does not affect the outcome or causes it to deteriorate for the citizen. No citizen is voting for her least preferred candidate, so by Proposition 4.1*b* every citizen's action is weakly undominated.

Subgame in which candidates are i, j, and a citizen l with favorite position outside $(m-\delta, m+\delta)$

Every citizen votes for i if her favorite position is less than m and for j

if her favorite position is greater than m. The votes of citizens whose favorite positions are m are split equally between i and j.

Any change in the vote of a citizen with favorite position m does not affect her payoff, while any change in the vote of any other citizen causes the winner to become her less preferred member of $\{i, j\}$. No citizen is voting for her least preferred candidate, so by Proposition 4.1*b* every citizen's action is weakly undominated.

Other subgames

Every citizen votes according to an arbitrary Nash equilibrium of the subgame in which every citizen's vote is weakly undominated. (Such a Nash equilibrium exists by Proposition 4.2.)

I now argue that no citizen can increase her payoff by deviating at the start of the game. By the argument for part *a*, neither candidate can increase her payoff by deviating to *Out*. Suppose that a citizen who is not a candidate deviates to *Run*. Given the equilibrium of the resulting voting subgame, such an entrant loses, so that the deviation makes her worse off.

Note that if δ is large the equilibrium in part c is vulnerable to deviations by groups of moderate citizens to vote for an entrant whose favorite position is between $m-\delta$ and $m+\delta$.

In the model with sincere voting, equilibria with many dispersed candidates are also possible.

Exercise 10.7: Multi-candidate dispersed equilibria of electoral competition game with citizen-candidates and sincere voting

Let $\langle n, (\hat{x}_1, \ldots, \hat{x}_n), u, b, c, x_0 \rangle$ be an electoral electoral competition game with citizen-candidates and sincere voting. Under the assumption in part b of Proposition 10.7 that the citizens' favorite positions are equally-spaced, show that for any integer k with $3 \le k \le n$, if $b/k - c$ is large enough the game has a Nash equilibrium in which k citizens run as candidates.

If voting is strategic, Proposition 4.3 implies that equilibria in which winning candidates occupy more than two distinct positions are not possible if the payoff function is strictly concave. However, if the cost c of running as a candidate is sufficiently small, equilibria with any number of candidates, two of whom win, exist. In these equilibria, no losing candidate is better off exiting because if she does so all citizens vote for the winner she likes least, causing that candidate to win outright. The citizens have no positive incentive to switch their votes in this way, but doing so is consistent with equilibrium and no citizen's action is weakly dominated. The next exercise invites you to establish this result.

Exercise 10.8: Multi-candidate dispersed equilibria of electoral competition game with citizen-candidates and strategic voting

Let $\langle n, (\hat{x}_1, \ldots, \hat{x}_n), u, b, c, x_0 \rangle$ be an electoral electoral competition game with citizen-candidates and strategic voting. Denote by m the median of the citizens' favorite positions. Show that for any integer $k \geq 4$, any number $\delta > 0$, and any positions x_1, x_2, $\ldots$, x_k with $x_1 = m - \delta$, $x_2 = m + \delta$, $x_i < x_1$ for some $i \geq 3$, $x_i > x_2$ for some $i \geq 3$, and $u_i(x_1) \neq u_i(x_2)$ for $i = 1, \ldots, k$, there exists $\underline{c}$ such that if $c \leq \underline{c}$ then the game has an equilibrium in which k citizens, with positions x_1, x_2, $\ldots$, x_k, run as candidates, candidates 1 and 2 tie for first place, and all other candidates lose.

Comments

The character of the equilibria appears to survive if the set of positions is multidimensional rather than one-dimensional. For example, if voting is sincere then a two-candidate equilibrium in which the candidates' positions differ exists if the candidates' positions differ enough that neither of them prefers to withdraw but not so much that an entrant can win. If voting is strategic, such an equilibrium exists if the first of these two conditions holds, with the entry of a candidate leading all the citizens to vote for the existing candidate the entrant likes least.

The citizen-candidate model has the merit of yielding tractable multi-candidate equilibria. The price is the assumption that each citizen is limited to running on her favorite position or not becoming a candidate—she cannot choose her position. One environment in which that restriction may be inappropriate is that in which candidates face a sequence of elections. Such a candidate may be able to credibly select a position different from her favorite position, knowing that voters can punish her in future elections if she deviates from that position while in office.

Notes

Section 10.1 is based on Osborne (1993). The model and results in Section 10.2 are due to Feddersen et al. (1990). The model in Section 10.4 was developed independently by Osborne and Slivinski (1996) (sincere voting) and Besley and Coate (1997) (strategic voting).

The much-discussed claim mentioned before Proposition 10.4 that plurality rule tends to lead to two-candidate electoral competitions appears to have been first stated explicitly in print by Droop (1881, 164) (see Riker 1982, 756). Riker writes (p. 754) “It is customary to call the law by Duverger's name, not because he had much to do with developing it but rather because he was the first to dare to claim it was a law.” It appears in Duverger (1951, 247) (and in translation in Duverger 1964, 217).

Exercise 10.1 is based on Haan and Volkerink (2001) and Brusco et al. (2012). The model and result in Exercise 10.5 are due to Jeffrey S. Rosenthal and Phillip Morenz (personal communication, 1992).

Solutions to exercises

Exercise 10.1

Denote the option of not running as a candidate by *Out*.

If all n candidates choose the position m, each of them wins with probability $1/n$. If one of them deviates to *Out* she wins with probability 0, and if she deviates to a position different from m then either her probability of getting to the second round is 0 or this probability is positive, and if it is positive then whenever she gets to the second round she loses (her opponent's position is m). Thus no deviation makes her better off, so that the strategy profile is a Nash equilibrium.

If k candidates choose the position m, with $k < n$, and the remainder choose *Out*, then any candidate who deviates from *Out* to the position m changes her probability of winning from 0 to $1/(k+1)$. Thus this strategy profile is not a Nash equilibrium.

Now suppose that k candidates choose the position $m-\delta$, k choose $m+\delta$, and the remainder choose *Out*. Then each candidate who chooses a position wins with probability $1/(2k)$.

If $k=1$ then either candidate can deviate to m and win with probability 1, so the strategy profile is not a Nash equilibrium.

Now suppose that $k \geq 2$.

- First consider deviations by one of the candidates who is choosing a position. If such a candidate deviates to *Out* then she wins with probability 0 and if she deviates to a position less than $m-\delta$ or greater than $m+\delta$

then she either does not make it to the second round or, if she does, loses in the second round. If she deviates to a position between $m-\delta$ and $m+\delta$ then if δ is sufficiently large she obtains the most votes on the first round and subsequently wins the second round. A sufficient condition for such a beneficial deviation not to exist is that δ is small enough that the vote share of a candidate at each position between $m-\delta$ and $m+\delta$ is at most $1/(2k)$.

- Now consider deviations by a candidate who is choosing *Out*. Under the assumption on δ in the previous point, no position for such a candidate generates a positive probability of her winning.

We conclude that if δ is small enough that the vote share of a candidate at each position between $m-\delta$ and $m+\delta$ is less than $1/(2k)$ then a strategy profile in which k candidates choose the position $m-\delta$ and k choose $m+\delta$, with $k \geq 2$, and the remaining candidates choose *Out*, is a Nash equilibrium.

Exercise 10.2

Denote by m the median of the citizens' favorite positions. If an office-seeker j has favorite position m and $U_j(m)+b-c \geq -D$ then the game has a subgame perfect equilibrium in which j is a candidate with position m, every other office-seeker chooses *Out*, all citizens vote for j in every subgame in which she is the only candidate or every candidate's position is m, each citizen votes for the candidate she prefers in every subgame in which there are two candidates, and in any other subgame the profile of the citizens' actions is any Nash equilibrium in which the citizens' actions are weakly undominated (the existence of which is ensured by Proposition 4.2). For this strategy profile, j's payoff is $U_j(m)+b-c$ and the payoff of every other office-seeker l is $U_l(m)$. The strategy profile is a subgame perfect equilibrium because if j deviates to another position she still wins, and is worse off; if she deviates to *Out* she gets the payoff $-D$; if another office-seeker enters at m, she loses (all citizens continue to vote for j); and if another office-seeker enters at a position other than m she loses.

If the favorite position of at least one office-seeker is less than m, the favorite position of at least one office-seeker is greater than m, and $b \geq c$, then the game has no subgame equilibrium with one candidate in which the candidate's position differs from m. To see why, let (x, v) be a strategy profile in which $x_j < m$ and $x_i = Out$ for all $i \neq j$. If an office-seeker l with favorite position greater than m deviates to enter at m, she wins (because the only undominated action of each citizen in the resulting subgame is to vote for the candidate she prefers) and obtains the payoff $U_l(m)+b-c$, which exceeds

$U_l(x_j)$.

The subgame perfect equilibria with two or more candidates who choose the position m are the same as the equilibria in the model with office-motivated candidates, and exist under the same conditions. In such an equilibrium with k candidates, the payoff of candidate j is $U_j(m)+b/k-c$; if she deviates to *Out*, her payoff becomes $U_j(m)$, so that for equilibrium we need $k \leq b/c$. The payoff of an office-seeker, say l, who is not a candidate is $U_l(m)$; if she deviates to become a candidate at m then the citizens all vote for one of the existing candidates and l loses, so that her payoff becomes $U_l(m)-c$.

Like the game with purely office-motivated candidates, the game appears to have no equilibrium with more than one position occupied. The logic in the proof of Proposition 10.1 rules out equilibria in which the candidates tie, and equilibria in which they do not tie are not possible because a citizen voting for a losing candidate benefits by shifting her vote to the winning candidate whom she likes best, so that losing candidates attract no votes and hence do not affect the outcome.

Exercise 10.3

For each player j, denote by $\hat{x}_j$ her favorite position.

In the subgame following player 1's choice of *Out*, if $x_0 \neq \hat{x}_2$ then the optimal action of player 2 is to choose $\hat{x}_2$ (and win), and if $x_0 = \hat{x}_2$ then $\hat{x}_2$ and *Out* are both optimal choices for her. In each case the outcome is $\hat{x}_2$.

Now consider the subgame following the entry of player 1 at m. If player 2 enters at m, the outcome is m. If player 2 enters at another position, the outcome is also m, because player 2 loses. Thus entering at m and *Out* are both optimal actions for player 2.

Now consider the subgame following the entry of player 1 at some position $x_1 < m$.

First suppose that player 1's favorite position, $\hat{x}_1$, is less than m. Then by assumption player 2's favorite position, $\hat{x}_2$, is greater than m.

- If $\hat{x}_2 - m < m - x_1$ then by entering at $\hat{x}_2$ player 2 wins and makes the policy outcome $\hat{x}_2$; she can do no better than that, so it is her optimal action.
- If $\hat{x}_2 - m \geq m - x_1$ then if player 2 enters at any position x_2 with $x_1 < x_2 < 2m - x_1$ she wins and induces the outcome x_2, which she likes better the closer x_2 is to $2m - x_1$. If she enters at $2m - x_1$ itself, she ties with player 1, which is worse for her than the outcome $2m - x_1$, and if she enters at a position greater than $2m - x_1$ she loses. Thus she has no

optimal action, but for any $\varepsilon > 0$ there exists $\delta > 0$ such that no action yields her a payoff greater by more than ε than her payoff for any action in $[2m - x_1 - \delta, 2m - x_1]$.

Thus for any value of $\hat{x}_2$, player 2's optimal or approximately optimal action in response to player 1's entry at x_1 generates an outcome $\min\{\hat{x}_2, 2m - x_1 - \delta\}$ for some small $\delta > 0$, which player 1 likes less than m.

Now suppose that $\hat{x}_1 > m$, so that $\hat{x}_2 < m$. If $\hat{x}_2 \leq x_1$ then every position $x_2 \leq x_1$ is optimal for player 2; in each case, the policy outcome is x_1, and no position for player 2 generates a policy outcome closer to $\hat{x}_2$, because any position less than x_1 leads player 1 to win. If $\hat{x}_2 > x_1$ then player 2's optimal position is $\hat{x}_2$, which leads her to win, so that the policy outcome is $\hat{x}_2$.

We conclude that the action *Out* and every position $x_1 < m$ for player 1 causes every optimal or approximately optimal action of player 2 to generate an outcome that is worse for player 1 than m. A symmetric argument applies if $x_1 > m$. Thus player 1's optimal action is m; player 2 responds optimally by either also choosing m or by choosing *Out*. Hence every approximate subgame perfect equilibrium generates either the outcome in which both players choose m or the outcome in which player 1 chooses m and player 2 chooses *Out*.

Exercise 10.4

Suppose that player 1 enters at a position $x_1 < \hat{x}_2$. If player 2 enters at the same position then player 3 wins outright if she enters at $\hat{x}_2$ (she gets more than a third of the vote and players 1 and 2 split the remainder equally), generating the best possible outcome for both her and player 2. (Other positions for player 2 may have the same implications.) Thus every subgame perfect equilibrium of the subgame following x_1 generates the policy outcome $\hat{x}_2$.

If player 1 enters at the position $\hat{x}_2$ then players 2 and 3 can do no better than stay out, and if player 1 enters at a position $x_1 > \hat{x}_2$ then by entering at $\hat{x}_2$ player 2 ensures that the policy outcome is $\hat{x}_2$ (player 3 optimally either stays out or, if entering at $\hat{x}_2$ causes her to tie for first place with player 2, enters at that position).

If player 1 stays out, then by entering at $\hat{x}_2$ player 2 ensures that the policy outcome is $\hat{x}_2$ (again player 3 either enters at the same position or stays out).

Thus the policy outcome of every subgame perfect equilibrium is $\hat{x}_2$. In an equilibrium, player 1 either stays out or enters. If she enters, player 3 enters at $\hat{x}_2$ and if player 3 would not win outright unless player 2 entered, then player 2 enters at a position that causes player 3 to win outright.

Exercise 10.5

I begin by establishing that a player does not enter if, in the event no subsequent player enters, she would lose.

Claim *For any nonterminal history h, in any subgame perfect equilibrium of the subgame following h, the player who moves first chooses a position (rather than Out) only if she wins in the event that every subsequent player (if any) chooses Out.*

Proof. I prove the claim by induction on the length of a history.

The claim is true for any history of length $n-1$, because after n moves the game ends, and n prefers *Out* to losing.

Now suppose that the claim is true for every history of length at least r, where $1 \leq r \leq n-2$. Consider a history h of length $r-1$. Suppose, contrary to the claim, that the subgame following h has a subgame perfect equilibrium in which player r chooses a position such that she loses if every subsequent player chooses *Out*. Then in this equilibrium, at least one subsequent player must enter (otherwise player r's entry is not optimal for her). Take the last player to do so, say player t. By the claim for histories of length $t-1$, player t wins. But then player r loses, so that the strategy profile is not a subgame perfect equilibrium. Thus the claim is true for every history of length $r-1$, and hence by induction for every history of any length. ◁

Now, if player 1 enters at m, no subsequent player can enter and win if no further players enter, because no position garners more votes than m. So by the claim, no subsequent player enters and hence player 1 wins. Thus the game has a subgame perfect equilibrium with the outcome in which player 1 enters at m and every other player chooses *Out*.

If player 1 enters at a position different from m then by entering at m player 2 ensures, by the claim, that no subsequent player enters, so that player 2 wins and player 1 loses.

Thus the game has a unique subgame perfect equilibrium outcome, in which player 1 enters at m and every other player chooses *Out*.

Exercise 10.6

I break the argument into steps.

Step 1 *In every subgame perfect equilibrium of the subgame following the entry of player* 1 *at m, players* 2 *and* 3 *enter at m.*

Proof. Suppose that player 1 enters at m. Then if player 2 enters at m, player 3's best action is to enter at m also, because her entry at another position either does not get her into a runoff or gets her into a runoff that she loses. Thus if player 2 enters at m then she wins with probability $\frac{1}{3}$, so that in every subgame perfect equilibrium of the subgame following player 1's entry at m, player 2 enters. Suppose that she enters at a position $x_2 < m$. If her vote share in the event that player 3 does not enter is positive (that is, her position is not extreme), then for some $\varepsilon > 0$ player 3 gets into the runoff by entering at $m + \varepsilon$ and wins outright. If player 2's vote share in the event that player 3 does not enter is zero, then player 3's best action is to enter at m, in which case players 1 and 3 each win with probability $\frac{1}{2}$ and player 2 does not win. Thus for every position at which player 2 enters other than m, she ultimately loses.

Hence player 2's best action in the subgame following player 1's entry at m is to enter at m, in which case player 3 optimally enters at m and each of the three players wins with probability $\frac{1}{3}$. ◁

Step 2 *In every subgame perfect equilibrium of the game, player* 1 *enters and wins with probability at least* $\frac{1}{3}$.

Proof. By Step 1, if player 1 enters at m she wins with probability $\frac{1}{3}$, so in every subgame perfect equilibrium of the game she wins with probability at least $\frac{1}{3}$, and hence enters. ◁

Step 3 *In every subgame perfect equilibrium of the game, player* 2 *enters.*

Proof. Let s^* be a subgame perfect equilibrium strategy profile in which player 2 does not enter in the subgame following player 1's action s_1^*. By Step 2, s_1^* is entry at some position, so player 3's optimal action in the subgame following the history in which players 1 and 2 follow their strategies in s^* (and hence player 2 does not enter) is to enter: if player 1 enters at m then player 3's optimal action is to enter at m (at any other position she loses), and if player 1 enters at a position different from m then all of player 3's optimal actions, among which is entry at m, cause her to win outright. By Step 2, in any subgame perfect equilibrium player 1 wins with positive probability, so in s^* she must enter at m. If she does so, player 2 optimally enters at m, because then player 3 optimally enters at m and player 2 wins with positive probability, whereas if player 2 enters at another position she loses given player 3's optimal action. ◁

Step 4 *In every subgame perfect equilibrium of the game, player* 3 *enters.*

Proof. By Step 3, in any subgame perfect equilibrium player 2 enters. By entering at the same position as player 2, player 3 ultimately wins with positive probability: she ties with player 2 in the first round and thus gets into the runoff with positive probability, and in the runoff she wins with the same probability as does player 2, which is positive because player 2 optimally enters. Thus in every subgame perfect equilibrium player 3 enters. ◁

Exercise 10.7

Let K be a set of k citizens whose vote shares are equal when the set of candidates is K. If any citizen j deviates from *Out* to *Run* then her vote share becomes less than $1/k$ because the fraction of citizens with favorite positions between any two candidates is less than $2/k$. At least one other candidate's vote share remains $1/k$, so j loses. Her entry causes the vote shares of the candidate closest to her on the left (if any) and the one closest on the right (if any) to fall, so that they lose. Thus j's entry makes her worse off.

Now suppose that a candidate, citizen i, deviates from *Run* to *Out*. Then the outcome changes to either a tie between the two candidates adjacent to her or an outright win for one of them. Thus i's payoff changes from $b/k + \sum_{j\in K} u_i(\hat{x}_j)/k - c$ to at most $\max_{j\in K\setminus\{i\}} u_i(\hat{x}_j)$. Hence a sufficient condition for no candidate to be better off deviating to *Out* is

$$b/k - c \geq \max_{i\in K}\left(\max_{j\in K\setminus\{i\}} u_i(\hat{x}_j) - \sum_{j\in K} u_i(\hat{x}_j)/k\right).$$

Exercise 10.8

Suppose that in the subgame following x, a citizen votes for candidate 1 if her favorite position is less than m and for candidate 2 if her favorite position is greater than m, and the votes of citizens with favorite position m are split between candidates 1 and 2 in such a way that these candidate tie. This action profile is a Nash equilibrium and every citizen votes for her favorite between candidates 1 and 2, so her action is weakly undominated. In a subgame following the entry of a new candidate, assume that the citizens' votes remain the same, so that the entrant loses and her deviation thus makes her worse off. Assume that in the subgame following the deviation to *Out* by any candidate with position less than m, all citizens vote for candidate 2, and in the subgame following the deviation to *Out* by any candidate with position greater than m, all citizens vote for candidate 1. Neither candidate 1 nor candidate 2 is any citizen's least favorite candidate, because candidates exist whose positions are less than x_1 and greater than x_2, so that in each case the action profile is a Nash equilibrium of the subgame in which no citizen's action is weakly dominated. None of the other subgames are reached by a deviation from x, so in any such subgame the action profile can be any Nash equilibrium in which the citizens' actions are weakly undominated, at least one of which exists by Proposition 4.2.

Given this voting behavior, if a citizen, say i, who is running as a candidate with a position less than m and losing deviates to *Out*, the outcome changes from a tie between candidates 1 and 2 to a win for candidate 2. Thus i's payoff changes from $\frac{1}{2}u_i(x_1) + \frac{1}{2}u_i(x_2) - c$ to $u_i(x_2)$, so she is no better off if $c \leq \frac{1}{2}(u_i(x_1) - u_i(x_2))$. Similarly, a citizen i who is running as a candidate with a position of at least m and losing is no better off deviating to *Out* if $c \leq \frac{1}{2}(u_i(x_2) - u_i(x_1))$. Finally, the payoff of the citizen, say i, with favorite position x_1 who runs as a candidate is $\frac{1}{2}b + \frac{1}{2}u_i(x_2) - c$. If she deviates to *Out*, this payoff becomes $u_i(x_2)$, so *Run* is optimal for her if $c \leq \frac{1}{2}(b - u_i(x_2))$. Similarly, *Run* is optimal for the citizen with favorite position x_2 who runs as a candidate if $c \leq \frac{1}{2}(b - u_i(x_1))$.

11 Distributive politics

11.1 Two-candidate competition with exogenous incomes under uncertainty 371
11.2 Voting over transfer systems when income is endogenous 378
11.3 Voting over linear transfer systems 386
11.4 Coalitional bargaining over redistribution 394

How does political power affect the distribution of material resources among the members of a society? This chapter analyzes models that address this question for the political system of majority rule.

Synopsis

One starting point is a collective choice problem in which the set of alternatives is the set of all distributions of a fixed amount of a material resource and each individual cares only about the amount she is assigned. This problem has no Condorcet winner. For every distribution, every member of the bare majority consisting of the individuals assigned the smallest amounts prefers the distribution in which her amount is increased by an equal share of the total amount assigned to the complementary minority, each member of which is assigned zero. For example, in a society consisting of three individuals, for any distribution (c_1, c_2, c_3) with $0 \leq c_1 \leq c_2 \leq c_3$, where c_i is the amount of the resource assigned to individual i, individuals 1 and 2, a majority, prefer the distribution $(c_1 + \frac{1}{2}c_3, c_2 + \frac{1}{2}c_3, 0)$.

An implication of this observation is that a two-candidate electoral competition game with majority rule in which the set of positions is the set of all distributions of the resource has no Nash equilibrium: for any distribution proposed by one candidate, the other candidate can propose a distribution that is preferred by a majority of individuals.

One way to escape this conclusion is to assume that the individuals care also about other (exogenous) features of the candidates' platforms, and the candidates are uncertain about these preferences. If the uncertainty is great enough, there may exist a distribution of the resource with the property that each candidate believes that no other distribution increases her probability of winning, because the amount by which every other distribution raises the payoffs of a majority of citizens is not likely to outweigh the preference of these citizens for the other features of the other candidate's platform. A model of this type is analyzed in Section 11.1. Proposition 11.1 characterizes a Nash equilibrium, if one exists, and the rest of the section explores properties of an equilibrium in some examples.

Another approach considers a collective choice problem in which the set of alternatives is the set of the individuals' favorite distributions rather than the set of all possible distributions. The model in Section 11.2 assumes that individuals differ in their earning power and that a tax-subsidy system specifies transfers as a function of income (individuals with the same income pay/receive the same tax/subsidy). The fact that an individual with high earning power can, by choosing her hours of work appropriately, obtain the same income and hence pay/receive the same tax/subsidy as one with low earning power, limits the taxes that the favorite system of an individual with low earning power can extract from one with high earning power. In a simple version of the model, an individual's favorite system imposes a 100% tax on individuals with lower earning power and equalizes the post-tax income of individuals with higher earning power. The favorite system of the individual with median earning power is not a Condorcet winner of the associated collective choice problem, but it comes close. Each individual is indifferent between the favorite systems of all individuals with higher earning power, because they all give her zero consumption. A richer version of the model generates, for some parameters, favorite systems that assign positive consumptions to individuals with lower earning power, with the amount assigned by the favorite system of an individual with earning power w to an individual with any given lower earning power decreasing in w. In this case, the favorite system of an individual with median earning power is a strict Condorcet winner of the associated collective choice problem.

Section 11.3 returns to a collective choice problem in which the set of alternatives is the set of all possible tax-subsidy systems, but restricts "possible" to mean linear in income. Proposition 11.2 gives conditions under which for any finite set of linear transfer systems the associated collective choice problem has single-crossing preferences with respect to the ordering of the individuals by their pre-tax income. As a result, the favorite transfer system of the individual with the median pre-tax income is a Condorcet winner, and hence, by Proposition 8.1, the outcome of a Nash equilibrium of an electoral competition game with two office-motivated candidates.

Section 11.4 takes a different approach, modeling the tax system as the outcome of society-wide bargaining. The distribution of income that emerges in this model is a compromise influenced by the possibility of any majority threatening to expropriate the remaining individuals, and these individuals, in response, threatening to destroy their endowments. In an equilibrium, these threats are not carried out, but the possibility that they could be carried out shapes the outcome. Two solution concepts based on different principles, the Shapley value and the core, yield the same conclusion: each individual's income is taxed at the rate of 50% and the revenue is divided equally among all individuals (Propositions 11.3 and 11.4).

11.1 Two-candidate competition with exogenous incomes under uncertainty

Consider a model of a society in which the citizens' incomes are given (they do not depend on the citizens' actions) and two political candidates propose tax-subsidy schemes. Each citizen cares about both her post-tax income and the candidates' policies on issues other than redistribution, which are fixed. The candidates know the citizens' incomes but may be uncertain about their preferences regarding other policies. Suppose, for example, that the society consists of three citizens with total income 1, and candidate 1 proposes a tax-subsidy scheme that generates the distribution of income $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. If candidate 2 proposes a scheme that generates the distribution $(\frac{1}{2}, \frac{1}{2}, 0)$ and each citizen is indifferent between the candidates' positions on other issues, then in the absence of uncertainty candidate 2 knows that she will obtain the votes of citizens 1 and 2, and hence win. But if the citizens are not neutral regarding the candidates' positions on other issues and candidate 2 is uncertain of their leanings, she may believe that the probability that the distribution $(\frac{1}{2}, \frac{1}{2}, 0)$ earns her the votes of citizens 1 and 2 is less than 1: with positive probability citizens 1 and 2 may like the non-distributional policies of candidate 1 enough to vote for candidate 1 even though candidate 2 offers them more post-tax income. In fact, candidate 2 may believe that the probability of her winning when she proposes the distribution $(\frac{1}{2}, \frac{1}{2}, 0)$ is less than her probability of winning when she proposes $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. For example, the former may result in only a slightly higher probability that citizens 1 and 2 vote for her and a sharply lower probability that citizen 3 does so. As a result, the action pair in which each candidate proposes the policy $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ may be a Nash equilibrium of the game.

I explore this idea using the two-candidate model of electoral competition in Section 8.5.2, in which the citizens care about both the policy chosen and the identity of the winning candidate, with the latter dependence reflecting characteristics of the candidate assumed to be fixed, like her competence or positions on issues other than distribution. Specifically, the model is an electoral competition game with two office-motivated candidates and uncertain partisanship in which the set of positions for each candidate is the set of distributions of consumption (post-transfer income) among the citizens and each citizen's payoff for a candidate's position depends only on the amount of consumption the distribution assigns to her.

In this game, two candidates simultaneously propose distributions of consumption among the citizens 1, ..., n, where, for convenience, n is odd. (You can equivalently think of the candidates proposing tax-subsidy schemes, given that the citizens' incomes are exogenous.) Each citizen i prefers the distribution $(c_1^1, \ldots, c_n^1)$ proposed by candidate 1 to the distribution $(c_1^2, \ldots, c_n^2)$ proposed by candidate 2 if and only if $g_i(c_i^1) > g_i(c_i^2) + \theta_i$, where g_i is an increasing function and θ_i is a number reflecting candidate 2's advantage (positive or negative) over candidate 1 on non-distributional policies. The candidates know how the citizens evaluate consumption but not how they evaluate other aspects of the candidates' policies: they know the functions g_i but are uncertain about the values of the numbers θ_i. Each candidate believes that $\theta_1, \ldots, \theta_n$ are independent draws from nonatomic distributions $F_1, \ldots, F_n$, so that each citizen i votes for candidate 1 if $\theta_i < g_i(c_i^1) - g_i(c_i^2)$, independently of the votes of the other citizens, an event with probability $F_i(g_i(c_i^1) - g_i(c_i^2))$. Under this assumption, each candidate believes that candidate 1's probability of winning, which is the probability that at least $(n+1)/2$ citizens vote for candidate 1 (given that n is odd), is

$$P\big(F_1(g_1(c_1^1) - g_1(c_1^2)), \ldots, F_n(g_n(c_n^1) - g_n(c_n^2))\big),$$

where P is the function defined in (8.5). Similarly, each candidate believes that candidate 2's probability of winning is

$$1 - P\big(F_1(g_1(c_1^1) - g_1(c_1^2)), \ldots, F_n(g_n(c_n^1) - g_n(c_n^2))\big).$$

Definition 11.1: Two-candidate electoral competition game of redistribution with uncertain partisanship

A *two-candidate electoral competition game of redistribution with uncertain partisanship* $\langle I, (g_i)_{i \in I}, (y_i)_{i \in I}, (F_i)_{i \in I}, \{1,2\} \rangle$, where

- $I = \{1, \ldots, n\}$ for an odd integer n (the set of citizens)

and for each $i \in I$

- $g_i : \mathbb{R}_+ \to \mathbb{R}$ is increasing (the component of i's payoff function that relates to consumption (post-tax income))

- $y_i \geq 0$ (i's pre-tax income)
- F_i is a nonatomic probability distribution over $\mathbb{R}$

is an electoral competition game with two office-motivated candidates and uncertain partisanship $\langle I, X, (v_i)_{i\in I}, (F_i)_{i\in I}, \{1,2\}\rangle$ in which

$$X = \{(c_1,\ldots,c_n) \in \mathbb{R}^n : \textstyle\sum_{i=1}^n c_i = \sum_{i=1}^n y_i \text{ and } c_i \geq 0 \text{ for } i = 1,\ldots,n\}$$

and

$$v_i(c_1,\ldots,c_n) = g_i(c_i) \text{ for all } (c_1,\ldots,c_n) \in X \text{ and } i = 1,\ldots,n.$$

Proposition 8.8 gives conditions under which the candidates' positions are the same in an interior Nash equilibrium of an electoral competition game with two office-motivated candidates and uncertain partisanship, if one exists. This result is not applicable to the game defined here because no member of $X \subset \mathbb{R}^n$ is interior and the functions v_i are not strictly concave (each such function is constant in c_j for $j \neq i$). However, the following closely related result holds. Its proof runs parallel to that of Proposition 8.8.

Proposition 11.1: Nash equilibrium of two-candidate electoral competition game of redistribution with uncertain partisanship

Let $\langle I, (g_i)_{i\in I}, (y_i)_{i\in I}, (F_i)_{i\in I}, \{1,2\}\rangle$ be a two-candidate electoral competition game of redistribution with uncertain partisanship. Assume that $I = \{1,\ldots,n\}$, the payoff function g_i of each citizen i for consumption is continuously differentiable and strictly concave, and each function F_i is continuously differentiable, with $F_i'(\theta) > 0$ for all $\theta \in \mathbb{R}$. If the game has a Nash equilibrium (c^{1*}, c^{2*}) with $c_i^{j*} > 0$ for $j = 1, 2$ and $i = 1,\ldots,n$ then this equilibrium has the following properties.

a. The candidates' plans are the same ($c^{1*} = c^{2*}$) and for some number λ

$$P_i'(F_1(0),\ldots,F_n(0))F_i'(0)g_i'(c_i^*) = \lambda \text{ for citizens } i = 1,\ldots,n, \qquad (11.1)$$

where $c^* = c^{1*} = c^{2*}$, P is given by (8.5), and P_i' is the derivative of P with respect to its ith argument.

b. If F_i is the same for all $i \in I$ then the common value of c^{1*} and c^{2*} is the solution of

$$\max_{c \in \mathbb{R}^n} \sum_{i=1}^n g_i(c_i) \text{ subject to } \sum_{i=1}^n c_i = \sum_{i=1}^n y_i \text{ and } c_i \geq 0 \text{ for } i = 1, \ldots, n.$$

Proof

a. By the definition of Nash equilibrium, for $j = 1, 2$ the action c^{j*} of candidate j maximizes j's probability of winning, given the other candidate's plan: $c^{1*} = (c_1^{1*}, \ldots, c_n^{1*})$ is a solution of

$$\begin{aligned} &\max_{(c_1^1, \ldots, c_n^1)} P\big(F_1(g_1(c_1^1) - g_1(c_1^{2*})), \ldots, F_n(g_n(c_n^1) - g_n(c_n^{2*}))\big) \\ &\qquad \text{subject to } \sum_{i=1}^n c_i^1 = \sum_{i=1}^n y_i \text{ and } c_i^1 \geq 0 \text{ for } i = 1, \ldots, n, \end{aligned} \tag{11.2}$$

and $c^{2*} = (c_1^{2*}, \ldots, c_n^{2*})$ is a solution of

$$\begin{aligned} &\max_{(c_1^2, \ldots, c_n^2)} \big(1 - P\big(F_1(g_1(c_1^{1*}) - g_1(c_1^2)), \ldots, F_n(g_n(c_n^{*1}) - g_n(c_n^2))\big)\big) \\ &\qquad \text{subject to } \sum_{i=1}^n c_i^2 = \sum_{i=1}^n y_i \text{ and } c_i^2 \geq 0 \text{ for } i = 1, \ldots, n. \end{aligned} \tag{11.3}$$

The derivatives of the equality constraint function in (11.2) with respect to the variables are all 1, and in particular are not all 0, so by Proposition 16.14 if c^{1*} is a solution of (11.2) with $c_i^{1*} > 0$ for $i = 1, \ldots, n$ then there is a unique number λ such that

$$P_i'(\pi(c^{1*}, c^{2*}))F_i'(g_i(c_i^{1*}) - g_i(c_i^{2*}))g_i'(c_i^{1*}) = \lambda \text{ for } i = 1, \ldots, n, \tag{11.4}$$

where the function $\pi : \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}_+^n$ is defined by

$$\pi(c^1, c^2) = \big(F_1(g_1(c_1^1) - g_1(c_1^2)), \ldots, F_n(g_n(c_n^1) - g_n(c_n^2))\big) \text{ for all } (c^1, c^2).$$

Now define the function $W : \mathbb{R}_+^n \to \mathbb{R}$ by

$$W(c) = \sum_{i=1}^n P_i'(\pi(c^{1*}, c^{2*}))F_i'(g_i(c_i^{1*}) - g_i(c_i^{2*}))g_i(c_i). \tag{11.5}$$

Each function g_i is strictly concave and all the coefficients of $g_i(c_i)$ in the definition of W are positive, so W is strictly concave. Consider the problem

$$\max_c W(c) \text{ subject to } \sum_{i=1}^n c_i = \sum_{i=1}^n y_i \text{ and } c_i \geq 0 \text{ for } i = 1, \ldots, n. \tag{11.6}$$

Given that W is strictly concave, this problem has a unique solution, say $\hat{c}$. By Proposition 16.14, if $\hat{c}_i > 0$ for $i = 1, \ldots, n$ then there is a unique number λ such that (11.4) is satisfied by $c^{1*} = \hat{c}$, and by Proposition 16.15, if there is a number λ and a vector c^{*1} with $c_i^{*1} > 0$ for $i = 1, \ldots, n$ that satisfies (11.4) then, given the concavity of W and the linearity of the constraint in (11.6), $c^{*1} = \hat{c}$. Thus (11.2) has a unique solution, which is the solution of (11.6).

Now, if $c_i^{2*} > 0$ for $i = 1, \ldots, n$ then the fact that c^{2*} is a solution of (11.3) means that there is a unique number λ such that c^{2*} satisfies the same condition, (11.4). Hence c^{2*} is also the unique solution of (11.6) and thus $c^{1*} = c^{2*}$, so that (11.4) reduces to (11.1).

b. Given $c^{1*} = c^{2*}$,

$$W(c) = \sum_{i=1}^{n} P_i'(F_1(0), \ldots, F_n(0)) F_i'(0) g_i(c_i) \text{ for all } c \in \mathbb{R}^n.$$

Thus if every distribution F_i is the same, then, given $P_i'(p, \ldots, p) = P_k'(p, \ldots, p)$ for all p and all i and k, c^* is a solution of (11.6) if and only if it is a maximizer of $\sum_{i=1}^{n} g_i(c_i)$ subject to the same constraints.

Like Proposition 8.8, this result does not assert that the game has an equilibrium; it only gives properties of an equilibrium if one exists. In the absence or near-absence of uncertainty, the game does not have an equilibrium, because for any distribution c proposed by one candidate, the other candidate can ensure that it wins with high probability by selecting the bare majority who are assigned the lowest total consumption in c and proposing to add to each of their allocations under c an equal share of the total consumption of the complementary minority. I know of no result that gives conditions under which the game necessarily has a Nash equilibrium, but some games do. Suppose, for example, that $n = 3$, $\sum_{i=1}^{3} y_i = 1$, and for $i = 1, 2, 3$ we have $g_i(z) = \sqrt{z}$ for all z and F_i is a normal distribution with mean 0 and standard deviation σ. My computations suggest that $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is a Nash equilibrium of this game if σ exceeds 0.24.

Character of equilibrium

Proposition 11.1b says if F_i is the same for every citizen then in any Nash equilibrium both candidates propose the distribution $(c_1^*, \ldots, c_n^*)$ that maximizes the sum of $g_i(c_i)$ across all citizens. Given the differentiability and concavity of each function g_i, $g_i'(c_i^*)$ is thus the same for all citizens. The derivative of g_i is a measure of the significance for i of consumption relative to a candidate's nondistributional policy. The smaller is the derivative, the larger is the increase in consumption needed to offset a given reduction in the appeal of the non-distributional policy. More loosely, the smaller is the derivative, the more the individual values non-distributional policy relative to consumption. Thus if, for every consumption level, citizen i attaches greater marginal value to non-distributional policy (in terms of consumption) than does citizen i', then in any equilibrium (if one exists), i is assigned less consumption than is i'.

Now suppose that g_i is the same for every citizen, equal to g, but F_i differs among them. First suppose that every F_i is symmetric about 0, with $F_i(x) = 1 - F_i(-x)$, and hence $F_i(0) = \frac{1}{2}$, and has a density f_i, so that $f_i(x) = f_i(-x)$. Then (11.1) is

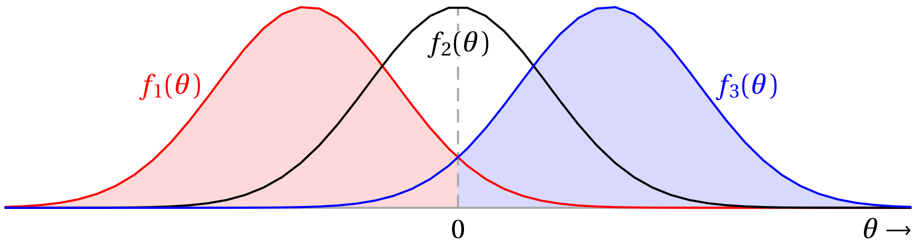
$$P_i'(\tfrac{1}{2}, \dots, \tfrac{1}{2}) f_i(0) g'(c_i^*) = \lambda \text{ for } i = 1, \dots, n.$$

The function P is symmetric in its arguments, so $P_i'(\frac{1}{2}, \dots, \frac{1}{2})$ is the same for all i and hence $f_i(0)g'(c_i^*)$ is the same for all i. Given the strict concavity of g, c_i^* is thus large when $f_i(0)$ is large. That is, an individual for whom a small difference between the candidates' policies results in a large difference in the probability of voting for each candidate receives a larger amount of consumption than does an individual who is unlikely to be swayed by a small difference in the candidates' policies. Roughly speaking, individuals who are influenced more by policy than by the candidates' characteristics receive larger amounts of consumption in an equilibrium (if one exists).

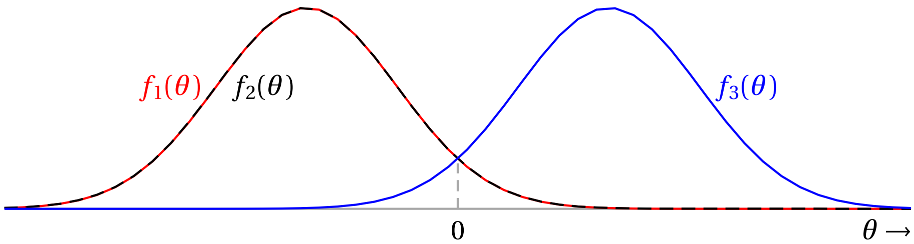
Now consider an example with three citizens in which the densities of the distributions F_1, F_2, and F_3 are the ones shown in Figure 11.1a. Citizen 1 is a partisan of candidate 1 in the sense that θ_1 is more likely to be negative than positive, so that if the amounts of consumption proposed for citizen 1 by the candidates are the same, citizen 1 is more likely to favor candidate 1. Similarly, citizen 3 is a partisan of candidate 2, and citizen 2 is neutral (she is equally likely to vote for each candidate if the candidates propose the same amount of consumption for her). To find the implications of (11.1) for this example, first note that for each citizen i we have $P_i'(p_1, p_2, p_3) = p_j(1-p_k) + p_k(1-p_j)$, where j and k are the citizens other than i. (The function P is defined in (8.5).) Now, $F_1(0)$ (the area shaded red in Figure 11.1a) is equal to $1 - F_3(0)$ (the area shaded blue), so denoting their common value by p we have

$$P_2'(F_1(0), F_2(0), F_3(0)) = p^2 + (1-p)^2$$
$$P_1'(F_1(0), F_2(0), F_3(0)) = P_3'(F_1(0), F_2(0), F_3(0)) = \tfrac{1}{2}p + \tfrac{1}{2}(1-p) = \tfrac{1}{2}.$$

By Proposition 11.1a, if (c_1^*, c_2^*, c_3^*) is an equilibrium then $P_i'(F_1(0), F_2(0), F_3(0)) f_i(0) g'(c_i^*)$ is the same for every citizen. Now, $f_1(0) = f_3(0) < f_2(0)$ and, given $p > \frac{1}{2}$, $p^2 + (1-p)^2 > \frac{1}{2}$, so $g'(c_1^*) = g'(c_3^*) > g'(c_2^*)$ and hence $c_1^* = c_3^* < c_2^*$. That is, citizens 1 and 3 are assigned the same amount of consumption, and this amount is less



(a) Citizen 1 is a partisan of candidate 1, citizen 2 is neutral, and citizen 3 is a partisan of candidate 2.



(b) Citizens 1 and 2 are partisans of candidate 1 and citizen 3 is a partisan of candidate 2.

Figure 11.1 Examples of densities for the distributions F_1, F_2, and F_3 in a two-candidate electoral competition game of redistribution with uncertain partisanship with three citizens.

than the amount assigned to citizen 2. Intuitively, the fact that citizen 2 is equally likely to vote for each candidate if they both offer her the same amount of consumption, whereas the other citizens have partisan leanings, means that a candidate's redistributing consumption from citizens 1 and 3 to citizen 2 increases the candidate's probability of winning.

Finally consider an example with three citizens in which the distributions F_1, F_2, and F_3 have the densities shown in Figure 11.1b. Denoting $F_1(0) = F_2(0) = 1 - F_3(0)$ by p, we have

$$P_1'(F_1(0), F_2(0), F_3(0)) = P_2'(F_1(0), F_2(0), F_3(0)) = p^2 + (1-p)^2$$
$$P_3'(F_1(0), F_2(0), F_3(0)) = 2p(1-p).$$

Given $p > \frac{1}{2}$, we have $p^2 + (1-p)^2 > 2p(1-p)$, so that since $f_1(0) = f_2(0) = f_3(0)$, $g'(c_1^*) = g'(c_2^*) > g'(c_3^*)$ and hence $c_1^* = c_2^* > c_3^*$. That is, citizens 1 and 2 are assigned the same amount of consumption, and this amount is greater than the amount assigned to citizen 3. A citizen's vote is pivotal only if the other two citizens vote for different candidates. Thus the votes of citizen 1 and citizen 2 are both likely to be pivotal, but that of citizen 3 is not. Hence the candidates find it advantageous to direct more consumption to citizens 1 and 2.

To summarize roughly, in the equilibria in these examples (if one exists) a citizen receives a larger amount of consumption the more sensitive is her payoff to her own consumption relative to the candidates' non-distributional policies, the less partisan she is, and the more likely her vote is to be pivotal.

11.2 Voting over transfer systems when income is endogenous

11.2.1 Main idea

Suppose that every individual's income is fixed independently of her actions. Then an individual who cares only about her own consumption and can impose arbitrary taxes on the other individuals optimally expropriates all the income of these individuals. If the tax she can impose on an individual may depend only on the individual's income, she is slightly constrained: she has to share the tax revenue with the individuals whose incomes are the same as hers. If the individuals' incomes are not fixed, but depend on the individuals' actions—like their choices of hours of work—then she is more constrained. In this case, the amount of tax she can extract from the other individuals is limited by the fact that some of these individuals may, by choosing appropriate hours of work, be able to earn the same income as she does, and hence pay the same tax or receive the same subsidy as she does. It is this case that I consider now.

The main model in this section assumes that each individual i has a given earning power w_i and chooses her hours (or intensity) of work $h_i \in [0,1]$, generating an income of $w_i h_i \in [0, w_i]$; she cares about both her hours of work and her consumption, which is equal to her post-transfer income. The number of individuals is finite, equal to n. I assume for convenience that n is odd and no two individuals have the same earning power; I index the individuals so that $w_1 < w_2 < \cdots < w_n$.

As a prelude, first assume that each individual i cares *only* about her consumption; assume also that if two different values of her hours of work h_i generate the same amount of consumption, she chooses the larger value. A transfer system $T : \mathbb{R}_+ \to \mathbb{R}$ is a function that assigns to each possible income y a transfer $T(y)$, so that the consumption of an individual with income y is $y - T(y)$. Thus if $T(y) > 0$ the transfer is a tax and if $T(y) < 0$ it is a subsidy. Under these assumptions, for any transfer system T and any number w, any individual with earning power at least w has the option of consuming $w - T(w)$ (by choosing her hours of work so that she earns the income w). Thus no transfer system can reduce the consumption of an individual j below the maximum of $w - T(w)$ for $w \le w_j$.

Among transfer systems that balance the budget, which one is best for individual i? She wants her consumption to be as large as possible, which, given the need to balance the budget, limits everyone else's consumption. The value she chooses for her own consumption can be achieved by any individual with higher

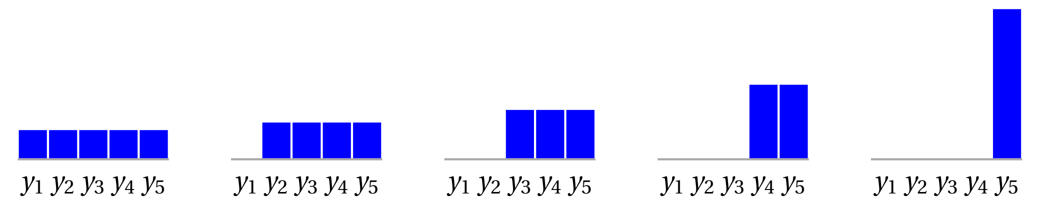


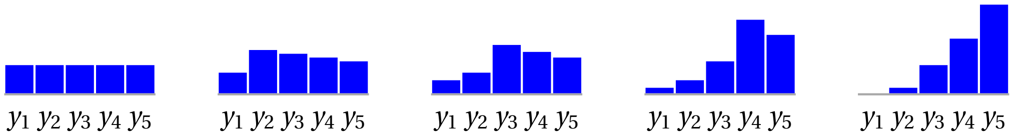
Figure 11.2 The levels of consumption assigned to the various income levels by the favorite transfer systems of the individuals in a model in which each individual can control her income. The society contains five individuals, who are ordered by their earning power, smallest to largest.

earning power, so her best plan reduces to zero the consumptions of the individuals with earning power smaller than hers and equalizes the consumptions of the individuals with higher earning power. An example with five individuals is shown in Figure 11.2.

Consider the collective choice problem for which the set of alternatives is the set of these favorite transfer systems (not the set of all feasible transfer systems). (I discuss an interpretation of this model at the end of the section.) For $n \geq 5$ this problem has no Condorcet winner: each individual's favorite transfer system is beaten by the favorite transfer system of another individual, as the following argument shows. Denote by T_i^* the favorite transfer system of each individual i and by i^* the individual with the median earning power. For any $i < i^*$, every individual j with $j \geq i^*$ prefers $T_{i^*}^*$ to T_i^*, so that T_i^* is beaten by $T_{i^*}^*$. For any $i > i^*$, every individual j with $j \leq i^*$ prefers T_1^* to T_i^*, so that T_i^* is beaten by T_1^*. And every individual j with $j > i^*$ prefers $T_{i^*+1}^*$ to $T_{i^*}^*$, i^* has the reverse preference, and every individual j with $j < i^*$ is indifferent between these systems, so that if $n \geq 5$ then $T_{i^*}^*$ is beaten by $T_{i^*+1}^*$. (If you have done Exercise 1.10, you will recognize the preference profile.)

However, a variant of the problem has a strict Condorcet winner. In this variant, each individual $i > i^*$ gives the individuals $j < i^*$ a small positive amount of consumption, with the amount declining in the value of i for each value of j, as in Figure 11.3. Each individual $j < i^*$ is then no longer indifferent between the favorite transfer systems of the individuals with high earning power. Instead, she prefers $T_{i^*}^*$, the favorite system of the individual with the median earning power, to T_i^* for $i > i^*$. As a consequence, $T_{i^*}^*$ is a strict Condorcet winner.

Such a variant of the problem is generated by a model in which each individual cares about her hours of work as well as her income. In this model, an individual with earning power w who chooses to work at an intensity $h < 1$ can be imitated by any individual with earning power at least wh. Thus her optimal transfer scheme may need to assign positive consumption to individuals with



(a) Individual 1 (b) Individual 2 (c) Individual 3 (d) Individual 4 (e) Individual 5

Figure 11.3 The payoffs assigned to the various income levels by the favorite transfer systems of the individuals in a model in which each individual can control her income.

earning power less than hers, otherwise they will choose a work intensity large enough that they earn the same income as she does, and hence be assigned the same transfer as she is.

The impact on the individuals' favorite systems of this alternative assumption is greatest when the dispersion of the individuals' earning powers is small: an individual can imitate only others whose incomes she can obtain by working at a sufficiently high intensity, which requires their earning power to be not too much greater than hers. No general result is available, but I now briefly present a model and give a complete example.

11.2.2 Model and example

A society consists of a finite number of individuals, each endowed with one unit of time, which she divides between work and leisure. Individuals may differ in their earning power; if individual i devotes the amount of time h to work, she obtains the income $w_i h$. Each individual uses her income to purchase a consumption good with price 1, and cares about the amount of her consumption and the amount of time she works. When an individual works for h units of time, her payoff is $u_i(w_i h, 1-h)$.

Definition 11.2: Society

A *society* $\langle N, (u_i)_{i\in N}, (w_i)_{i\in N}\rangle$ consists of

- a finite set N (of individuals, each endowed with one unit of time)

and for each $i \in N$

- a differentiable function $u_i : \mathbb{R}_+ \times [0,1] \to \mathbb{R}$ that is increasing in each of its arguments ($u_i(c,l)$ is i's payoff when her consumption is c and her amount of leisure is l)
- a number $w_i > 0$ (i's earning power).

A transfer system assigns to each possible income a number that is at most equal to the income. If the number is positive it is a tax, and if it is negative it is a subsidy. Note that the definition of a transfer system does not include a feasibility requirement: the net amount of income disbursed is not restricted by any budget. A feasibility condition is imposed later.

Definition 11.3: Transfer system

A *transfer system* is a function $T : \mathbb{R}_+ \to \mathbb{R}$ with $T(y) \leq y$ for all y. For pre-transfer income y, post-transfer income is $y - T(y)$.

Consider a society $\langle N, (u_i)_{i\in N}, (w_i)_{i\in N}\rangle$ in which each individual i chooses how to allocate her endowment of one unit of time between work and leisure, generating (pre-transfer) income $w_i h$ when she works for h units of time. For a transfer system T, this income results in consumption of $w_i h - T(w_i h)$, so i chooses h to solve the problem

$$\max_h u_i(w_i h - T(w_i h), 1-h) \text{ subject to } 0 \leq h \leq 1. \tag{11.7}$$

Reformulating this problem facilitates its analysis. For each individual $i \in N$, define the function $v_i : [0, w_i] \times \mathbb{R}_+ \to \mathbb{R}$ by

$$v_i(y,c) = u_i(c, 1-y/w_i) \text{ for all } (y,c), \tag{11.8}$$

so that $v_i(y,c)$ is i's payoff when she works enough hours to generate the (pre-transfer) income y and consumes c. Then individual i's optimization problem (11.7) may be formulated as

$$\max_{(y,c)} v_i(y,c) \text{ subject to } c = y - T(y) \text{ and } 0 \leq y \leq w_i. \tag{11.9}$$

Figure 11.4 illustrates the solution of an example of this problem. The orange curve gives the amount of consumption that each amount of pre-transfer income yields, given the transfer system. The transfer system assigns subsidies to values of y for which this curve is above the 45° line and taxes to values of y for which it is below the 45° line. The blue curve is a set of pairs (y,c) that yield i the same payoff—in economic jargon, an indifference set or indifference curve. The slope of the indifference curve of v_i through (y,c) at (y,c) is

$$-\frac{v'_{i,1}(y,c)}{v'_{i,2}(y,c)} = \frac{1}{w_i}\frac{u'_2(c,1-y/w_i)}{u'_1(c,1-y/w_i)} > 0,$$

where the subscripts 1 and 2 denote the index of the variable with respect to which the derivative is taken. The economic reason that the slope is positive is

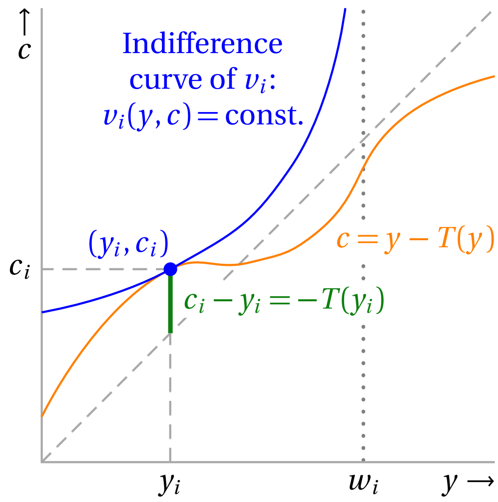


Figure 11.4 The pair (y_i, c_i) that solves individual i's optimization problem (11.9), given the transfer system T.

that an individual obtains more income only by working longer, so that if $y' > y$ then we need $c' > c$ for (y,c) and (y',c') to yield the same payoff.

The pair (y_i, c_i) in Figure 11.4 is the solution of (11.9)—the pair chosen by the individual, given the transfer system. The length of the green line segment is the subsidy she receives, $-T(y_i)$.

Individual's favorite transfer system

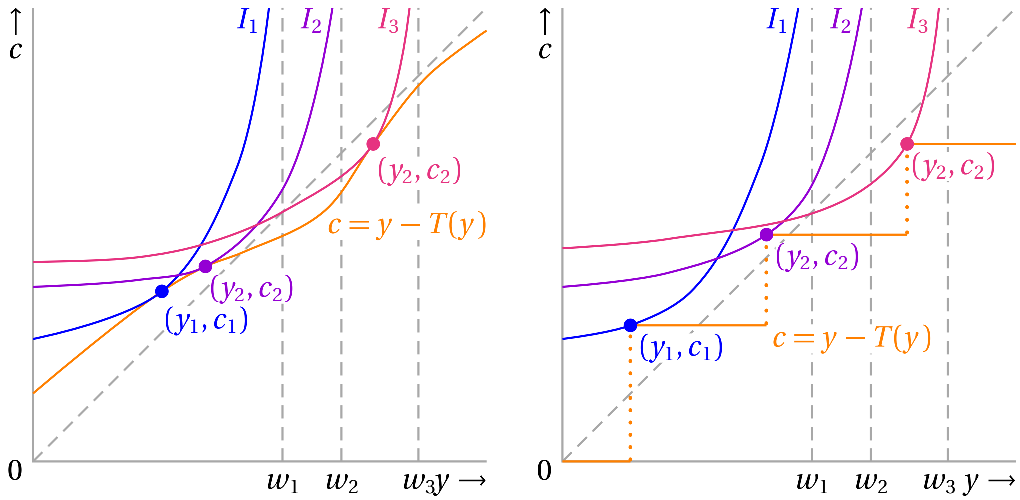
To be feasible, a transfer system must collect in taxes at least as much as it distributes in subsidies. Among feasible systems, the best one for any given individual is the one for which her payoff is highest, given that every individual (including her) chooses her hours of work optimally. I now present a convenient formulation of the problem of finding such a selfishly-optimal system.

Fix a transfer system T, and for each $i \in N$ let (y_i, c_i) be a solution of (11.9). Then in particular no individual i is better off choosing the pair (y_j, c_j) chosen by any other individual j for whom $y_j \le w_i$. She is also no better off choosing the pair $(0, -T(0))$, achieved if she does not work, and thus is no better off choosing $(0,0)$, given that $T(0) \le 0$. That is, for any individual $i \in N$ we have

$$\begin{aligned} &v_i(y_i, c_i) \ge v_i(y_j, c_j) \text{ for all } j \in N \text{ with } y_j \le w_i \\ &v_i(y_i, c_i) \ge v_i(0,0). \end{aligned} \tag{11.10}$$

Figure 11.5a illustrates these conditions for an example of a society containing three individuals.

Conversely, let $((y_i, c_i))_{i \in N}$ be a profile satisfying (11.10) with $y_1 < y_2 < \cdots < y_n$.



(a) The pairs (y_i, c_i) optimal for three individuals given the transfer system T. The curves labeled I_1, I_2, and I_3 are indifference curves of the individuals.

(b) A transfer system T for which for each individual i, each member of the profile $((y_i, c_i))_{i \in N}$, which satisfies (11.10), is optimal.

Figure 11.5

Define the (discontinuous) transfer system T by

$$T(y) = \begin{cases} y & \text{if } y < y_1 \\ y - c_i & \text{if } y_i \leq y < y_{i+1} \text{ for } i = 1, \ldots, n-1 \\ y - c_n & \text{if } y_n \leq y. \end{cases}$$

(An example for three individuals is given in Figure 11.5b.) I claim that for this transfer system, for each individual $i \in N$ the pair (y_i, c_i) is a solution of (11.9). The reason is that for any profile $((y_i, c_i))_{i \in N}$ satisfying (11.10) we have $c_j < c_k$ whenever $y_j < y_k$, otherwise k prefers (y_j, c_j) to (y_k, c_k).

In summary, $((y_i, c_i))_{i \in N}$ satisfies (11.10) if and only if for some transfer system T, for each $i \in N$ the pair (y_i, c_i) is a solution of (11.9) Thus rather than working with transfer systems we can work with transfer plans, defined as follows.

Definition 11.4: Transfer plan

A *transfer plan* is a profile $((y_i, c_i))_{i \in N}$ with $(y_i, c_i) \in \mathbb{R}_+ \times \mathbb{R}_+$ for all $i \in N$.

A selfishly-optimal transfer plan for any individual k maximizes k's payoff among the plans that satisfy (11.10) and raise at least as much in taxes as they distribute in subsidies.

Definition 11.5: Selfishly-optimal transfer plan for an individual

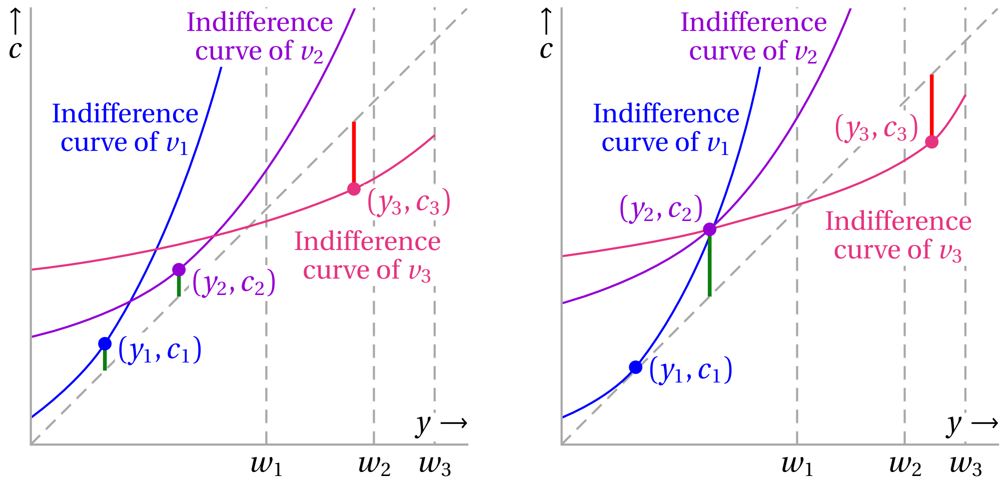
Let $\langle N,(u_i)_{i\in N},(w_i)_{i\in N}\rangle$ be a society, for each $i\in N$ let v_i be the function defined in (11.8), and let $k\in N$. A transfer plan $((y_i,c_i))_{i\in N}$ is *selfishly-optimal for individual* k if it is a solution of the problem

$$\max_{((y_i,c_i))_{i\in N}} v_k(y_k,c_k) \text{ subject to}$$
$$\begin{gathered} v_i(y_i,c_i)\ge v_i(y_j,c_j) \text{ for all } i\in N \text{ and all } j\in N \text{ for which } y_j\le w_i,\\ v_i(y_i,c_i)\ge v_i(0,0) \text{ for all } i\in N,\\ 0\le y_i\le w_i \text{ and } c_i\ge 0 \text{ for all } i\in N, \quad \text{and} \quad \textstyle\sum_{i\in N}(y_i-c_i)\ge 0. \end{gathered} \tag{11.11}$$

In the remainder of this section, I provide a diagrammatic analysis of the individuals' selfishly-optimal plans and an example of a society with three individuals in which the selfishly-optimal plan of the individual with the median earning power is a strict Condorcet winner of the collective choice problem in which the set of alternatives is the set of the individuals' selfishly-optimal plans.

My analysis is restricted to societies in which every individual's payoff function u_i is the same and has the property that the optimal amount of consumption for each individual in the absence of transfers is increasing in her earning power. That is, the value of c that maximizes $u_i(c,1-c/w)$ is increasing in w. This condition is equivalent to the slope $-v'_{i,1}(y,c)/v'_{i,2}(y,c)$ of the indifference curve of v_i through (y,c) at (y,c) being smaller for larger values of w_i, for each value of (y,c). (This equivalence is demonstrated in Mirrlees 1971, footnote 1, for example.) An interpretation of the condition is that the amount of additional consumption required to compensate for the extra work necessary to earn an additional unit of income is smaller for individuals with greater earning power.

Consider the transfer plan of individual 2 for the society with three individuals shown in Figure 11.6a. This plan satisfies the constraints in (11.11) for $k=2$: each individual i likes (y_i,c_i) better than (y_j,c_j) for $j\neq i$ and better than $(0,0)$, and the total tax paid (by individual 3), the length of the vertical red line segment, exceeds the total subsidy paid out (to individuals 1 and 2), the sum of the lengths of the green line segments. But this plan is not optimal for individual 2. First, the total tax exceeds the total subsidy, so that c_2 can be increased, making individual 2 better off, without violating the budget constraint or any other constraint. Second, individual 1 likes (y_1,c_1) better than (y_2,c_2), so that c_1 can be reduced while keeping (y_1,c_1) the best of the three pairs for her, which relaxes the budget constraint and allows c_2 to be increased. Similarly, c_3 can be reduced, increasing the tax on individual 3. Finally, the slope of individual 1's indifference curve through (y_1,c_1) and the slope of individual 3's indifference curve through (y_3,c_3)



(a) A suboptimal plan.

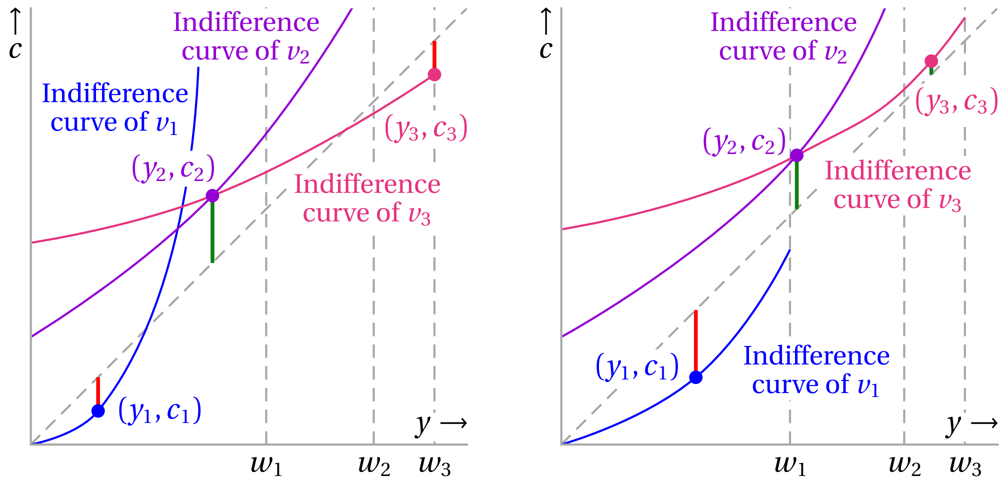
(b) A plan without suboptimal features of the plan in Figure 11.6a.

Figure 11.6 Transfer plans for individual 2 in a society with three individuals.

are both different from 1, so the pairs (y_1, c_1) and (y_3, c_3) can be moved along the indifference curves to increase the tax or reduce the subsidy for each individual without affecting her payoff. Figure 11.6b shows a plan that cannot be improved by any such changes (although it might be improved by other changes). Similar considerations apply to the selfishly-optimal plans of the other individuals.

Figure 11.7 shows two examples of transfer plans for individual 2 in a society with three individuals that illustrate other possibilities for an optimal plan. In Figure 11.7a, the plan pushes individual 1 down to her lowest possible payoff: she is indifferent between (y_1, c_1) and $(0,0)$. Although she is not indifferent between (y_1, c_1) and (y_2, c_2), the plan may still be optimal for individual 2. In Figure 11.7b, (y_2, c_2) is unattainable by individual 1 because $y_2 > w_1$. In this case, an optimal plan for individual 2 may not exist: plans in which y_2 is closer to w_1 may be better for her, but a plan in which $y_2 = w_1$ may be significantly worse, because (y_2, c_2) is then attainable for individual 1, who consequently has the option of receiving the same transfer as individual 2.

For a society with five individuals in which the values of w_i are large enough not to constrain individual 3's selfishly-optimal plan, Figure 11.8 shows a plan that satisfies the analogues of the necessary conditions for optimality for the plan of individual 2 in a society with three individuals illustrated in Figure 11.6b. The diagram suggests that, under the assumptions on preferences that I am making, an individual has more leverage in raising taxes from individuals whose earning powers are further from hers. As a consequence, under some conditions the payoffs in the individuals' selfishly-optimal plans plausibly take the form given



(a) Individual 1 is indifferent between (y_1, c_1) and $(0, 0)$.

(b) Individual 1's earning power is insufficient to allow her to obtain (y_2, c_2).

Figure 11.7 Transfer plans for individual 2 for a society with three individuals.

in Figure 11.3, so that the plan of the individual with the median earning power is a strict Condorcet winner of the collective choice problem in which the set of alternatives is the set of the individuals' selfishly-optimal plans.

A computed example for a society with three individuals is given in Figure 11.9. Each individual's preferences over the three selfishly-optimal plans are single-peaked with respect to the ordering $p_1^* < p_2^* < p_3^*$, where p_i^* denotes the selfishly-optimal plan of individual i, so that p_2^* is a strict Condorcet winner of the collective choice problem. No general result is available; see the Notes section at the end of the chapter for information about published results.

I close the section with an interpretation of the collective choice problem in which the set of alternatives is the set of the individuals' selfishly-optimal plans. Suppose that each individual chooses whether to become a candidate, and if she does so and is elected she implements her favorite plan (along the lines of the electoral competition game with citizen-candidates considered in Section 10.4). If one of the candidates is the individual with the median earning power, then that individual is elected by any voting method that elects the candidate whose favorite policy is the strict Condorcet winner of the collective choice problem. (Proposition 5.3 shows that every binary agenda has this property.)

11.3 Voting over linear transfer systems

The model I now discuss differs in two main respects from the one in the previous section. First, the transfer system is restricted to be linear (more properly,

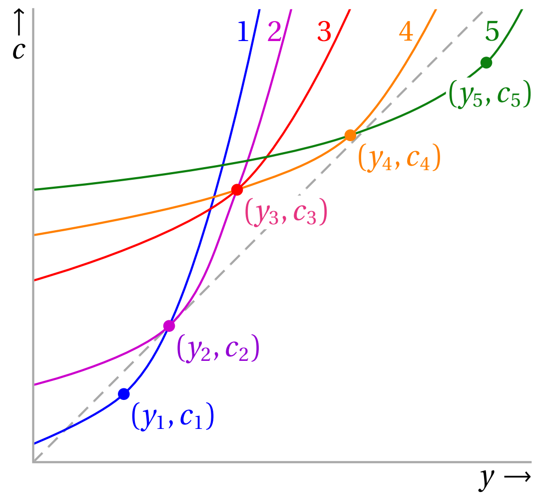


Figure 11.8 A transfer plan of individual 3 for a society with five individuals.

affine): income is taxed at the constant rate t and every individual receives a fixed subsidy r. Second, the set of alternatives in the collective choice problem is the set of all linear transfer systems, not only the ones that are selfishly-optimal for some individual.

Definition 11.6: Linear transfer system

For any $(t, r) \in [0,1] \times \mathbb{R}_+$ the *linear transfer system* (t, r) is the transfer system T for which $T(y) = ty - r$.

11.3.1 Exogenous incomes

First suppose that the individuals' incomes are given; no individual makes a choice that affects her income. For budget balance we need $t\overline{y} = r$, where $\overline{y}$ is the individuals' average income. Then the after-tax income of an individual with income y is $y - t(y - \overline{y})$. This payoff is increasing in t if $y < \overline{y}$ and decreasing in t if $y > \overline{y}$, so that for $y \neq \overline{y}$ the individual's preferences over t are single-peaked with respect to the ordering $\geq$ of t, with the favorite tax rate of an individual with income y equal to 1 if $y < \overline{y}$ and to 0 if $y > \overline{y}$. Suppose that the number of individuals is finite and none of them has an income of exactly $\overline{y}$. Consider a collective choice problem in which the set of alternatives is a finite set of feasible linear transfer systems that includes $(0,0)$ (no redistribution) and $(1, \overline{y})$ (complete equalization of incomes). Proposition 1.4 implies that the Condorcet winner of this problem is $(1, \overline{y})$ if the individuals' median income is less than the mean and $(0,0)$ if it is greater than the mean. Hence the Condorcet winner entails the full

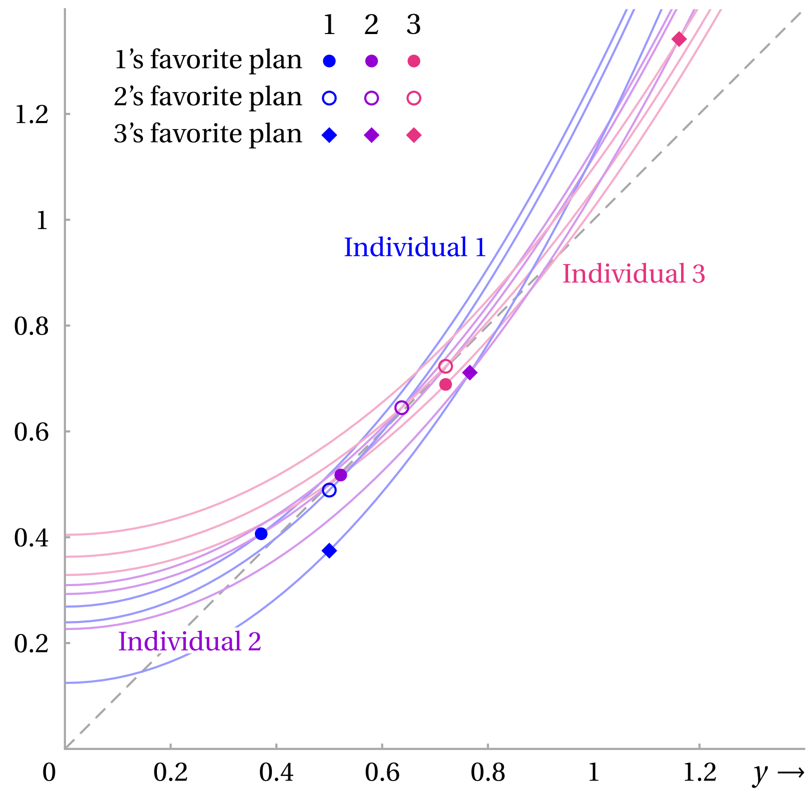


Figure 11.9 The selfishly-optimal transfer plans of the individuals in the society $\langle\{1,2,3\},(u,u,u),(w_1,w_2,w_3)\rangle$ where $w_1 = 2$, $w_2 = 2.2$, $w_3 = 2.4$, and $u(c,l) = c - 4(1-l)^2$ for all (c,l), so that $v_i(y,c) = c - 4(y/w_i)^2$ for $i = 1, 2, 3$.

equalization of income in a society in which the median income is less than the mean and no redistribution in a society in which the median income is greater than the mean. An example of a distribution for which the median is less than the mean is shown in Figure 11.10.

11.3.2 Endogenous incomes and incentive effects

If the income an individual earns depends on the amount of time she works, an increase in the tax rate may cause her to devote less time to work, reducing the revenue from the tax. Thus an increase in the tax rate may not be desirable. In particular, a tax rate of 1 may no longer be preferred by a majority of individuals to any other rate, even if the distribution of earning power is skewed to the left.

Consider a society $\langle N,(u_i)_{i\in N},(w_i)_{i\in N}\rangle$. Suppose that the individuals' earnings are subject to a linear transfer system (t,r). If individual i chooses to work for h units of time then her payoff is $u_i((1-t)w_i h + r, 1-h)$. She chooses h to maximize this payoff. Assume that the maximizer, denoted $h_i^*(t,r)$, is unique,

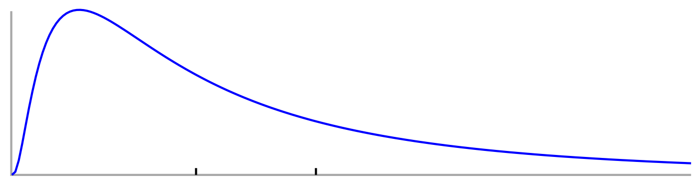


Figure 11.10 A distribution of income for which the median is less than the mean. (The distribution shown is a lognormal.)

and denote her maximal payoff by $V_i(t,r)$:

$$V_i(t,r) = \max_{h\in[0,1]} u_i((1-t)w_i h + r, 1-h) = u_i((1-t)w_i h_i^*(t,r) + r, 1-h_i^*(t,r)). \quad (11.12)$$

The result in this section applies to societies in which the individuals can be ordered by their pre-tax incomes independently of the transfer system: that is, for any two individuals i and j, either $w_i h_i^*(t,r) > w_j h_j^*(t,r)$ for every linear transfer system (t,r) or $w_i h_i^*(t,r) < w_j h_j^*(t,r)$ for every such system.

Definition 11.7: Society with individuals ordered by pre-tax income

The individuals in a society $\langle N, (u_i)_{i\in N}, (w_i)_{i\in N}\rangle$ *can be ordered by pre-tax income* if for every linear transfer system (t,r) and every individual $i \in N$ the problem

$$\max_{h\in[0,1]} u_i((1-t)w_i h + r, 1-h)$$

has a unique solution $h_i^*(t,r)$, and for some linear order $\geq$ on N

$$j < i \iff w_j h_j^*(t,r) > w_i h_i^*(t,r) \quad \text{for all } (t,r) \in [0,1]\times\mathbb{R}_+ \text{ with } h_i^*(t,r) > 0. \quad (11.13)$$

This condition is strong. Intuition suggests that in a diverse society the pre-tax incomes of some individuals are higher than those of other individuals under some transfer systems but lower under other transfer systems. Here are two examples.

Example 11.1: Cobb-Douglas payoff functions

Consider a society $\langle N, (u_i)_{i\in N}, (w_i)_{i\in N}\rangle$ for which for each $i \in N$ we have $u_i(y, 1-h) = y^{\beta_i}(1-h)^{1-\beta_i}$ for all y and h, where $\beta_i \in (0,1)$. Then

$$w_i h_i^*(t,r) = w_i \max\left\{0, \beta_i - \frac{(1-\beta_i)r}{(1-t)w_i}\right\}.$$

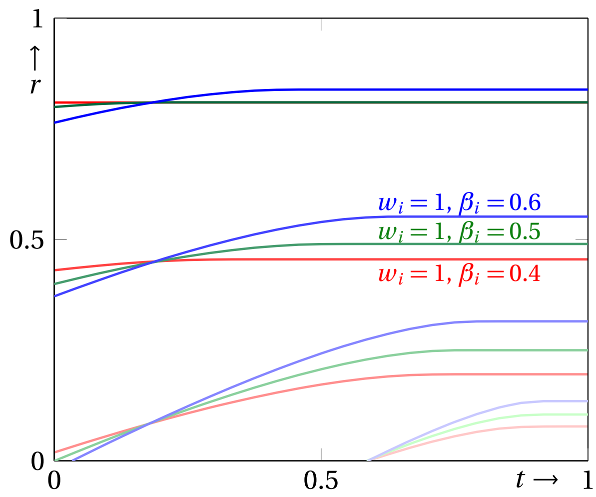


Figure 11.11 Indifference sets for the preferences over linear transfer systems of individuals with the payoff function specified in Example 11.1 for $w_i = 1$ and various values of β_i. The indifference sets for $\beta_i = 0.6$ are blue, those for $\beta_i = 0.5$ are green, and those for $\beta_i = 0.4$ are red; sets corresponding to higher payoffs are darker.

Thus if w_i is the same for all $i \in N$, then $w_i h_i^*(t, r)$ is increasing in β_i when it is positive, so that the ordering of the individuals defined by $j < i$ if and only if $\beta_j > \beta_i$ satisfies (11.13). Some indifference curves for $w_i = 1$ and various values of β_i are shown in Figure 11.11. The function $w_i h_i^*(t, r)$ is not differentiable at any (t, r) for which $\beta_i = (1 - \beta_i)r/((1 - t)w_i)$, so that the payoff function V_i is not differentiable at any such point. However, although the next result, Proposition 11.2, assume differentiability, it holds also if $w_i h_i^*(t, r)$ is piecewise differentiable, as it is in this example.

Alternatively, if β_i is the same for all $i \in N$, then $w_i h_i^*(t, r)$ is increasing in w_i when it is positive, so that again (11.13) is satisfied.

If the individuals differ in both w_i and β_i, (11.13) may be violated. For example, the pre-tax income of an individual with $w_i = 16$ and $\beta_i = 0.2$ is greater for the transfer system $(0.45, 0.25)$ than it is for the transfer system $(0.8, 0.8)$, but is less for an individual with $w_i = 2$, and $\beta_i = 0.9$.

Example 11.2: Quasilinear payoffs

Consider a society $\langle N, (u_i)_{i \in N}, (w_i)_{i \in N} \rangle$ for which for each $i \in N$ we have $u_i(y, 1 - h) = y + v(1 - h)$ for all y and h, where v is an increasing concave differentiable function. Then

$$w_i h_i^*(t,r) = w_i \max\{0, 1 - (v')^{-1}((1-t)w_i)\}.$$

Given the concavity of v, this expression is increasing in w_i when it is positive. Thus (11.13) is satisfied by the ordering $\geq$ of the individuals defined by $j < i$ if and only if $w_j > w_i$.

Now consider a collective choice problem in which the alternatives are finitely many linear transfer systems. If (t,r) and (t',r') are alternatives with $t' > t$ and $r' < r$, then every individual prefers (t,r) to (t',r'), so (t',r') can be eliminated from consideration. Thus we can assume that if (t,r) and (t',r') are alternatives with $t' > t$ then $r' \geq r$. For convenience, I make the stronger assumption that $t' > t$ if and only if $r' > r$.

The next result says that for a society in which the individuals can be ordered by pre-tax income, the collective choice problem has single-crossing preferences with respect to the ordering of the individuals by pre-tax income. Hence by Proposition 1.5, if each median individual according to the ordering has a unique favorite linear transfer system then the favorite linear transfer system of a median individual is a Condorcet winner of the problem. The key point in the argument is that the slope at any point (t,r) of i's indifference curve for her payoff function V_i through (t,r) is her pre-tax income $w_i h_i^*(t,r)$, so that the assumption that the individuals can be ordered by pre-tax income implies that the slopes of their indifference curves are ordered independently of the transfer system.

Proposition 11.2: Single-crossing preferences over linear transfer systems

Let $\langle N, (u_i)_{i\in N}, (w_i)_{i\in N}\rangle$ be a society in which the individuals can be ordered by pre-tax income. For each $i \in N$ denote by $\succsim_i$ the preference relation represented by the function V_i defined in (11.12): $(t,r) \succsim_i (t',r')$ if and only if $V_i(t,r) \geq V_i(t',r')$. Assume that V_i is continuously differentiable and $V'_{i,2}(t,r) \neq 0$ for all $(t,r) \in [0,1] \times \mathbb{R}_+$.

Let $\mathscr{T}$ be a finite set of linear transfer systems such that if $(t,r) \in \mathscr{T}$ and $(t',r') \in \mathscr{T}$ then $t' < t$ if and only if $r' < r$. Then the collective choice problem $\langle N, \mathscr{T}, \succsim\rangle$ has single-crossing preferences with respect to the ordering of the individuals by pre-tax income.

As a consequence, if each median individual according to this ordering has a unique favorite alternative in $\mathscr{T}$ then

- if the number of individuals is even then each of these favorite alternatives is a Condorcet winner of $\langle N, \mathscr{T}, \succsim \rangle$

- if the number of individuals is odd then the favorite alternative of the (unique) median individual is the strict Condorcet winner of $\langle N, \mathscr{T}, \succsim \rangle$.

Proof

Let (t^1, r^1) be a linear transfer system, let $i \in N$, and consider i's indifference set that contains (t^1, r^1):

$$\{(t, r) \in [0,1] \times \mathbb{R}_+ : V_i(t, r) = V_i(t^1, r^1)\}.$$

By the implicit function theorem, the function g on $(0,1)$ defined by $V_i(t, g(t)) = V_i(t^1, r^1)$ (i's indifference curve through (t^1, r^1)) is continuously differentiable and $g'(t^1) = -V'_{i,1}(t^1, r^1)/V'_{i,2}(t^1, r^1)$. By the envelope theorem

$$V'_{i,1}(t, r) = -w_i h_i^*(t, r) u'_{i1}((1-t) w_i h_i^*(t, r) + r, 1 - h_i^*(t, r))$$
$$V'_{i,2}(t, r) = u'_{i,1}((1-t) w_i h_i^*(t, r) + r, 1 - h_i^*(t, r)),$$

so that

$$-\frac{V'_{i,1}(t, r)}{V'_{i,2}(t, r)} = w_i h_i^*(t, r). \tag{11.14}$$

That is, for any linear transfer system (t, r), the slope at (t, r) of i's indifference curve through (t, r) is her pre-tax income when she chooses her hours of work optimally, given (t, r).

I argue that the collective choice problem $\langle N, \mathscr{T}, \succsim \rangle$ has single-crossing preferences with respect to the ordering $\geq$ of the individuals by pre-tax income. For any individual i and linear transfer system (t, r) with $h_i^*(t, r) > 0$, if $j < i$ then by (11.13) and (11.14) the slope of j's indifference curve through (t, r) at (t, r) is greater than the slope of i's indifference curve through (t, r) at (t, r), as for (t^1, r^1) in Figure 11.12. Thus j's indifference curve through (t, r) lies above i's for all tax rates greater than t and below it for all tax rates less than t.

Now suppose that for $(t^1, r^1) \in \mathscr{T}$ and $(t^2, r^2) \in \mathscr{T}$ we have $t^1 < t^2$, so that $r^1 < r^2$, and $(t^1, r^1) \succsim_i (t^2, r^2)$. Then given $r^2 > r^1$, the slope at (t^1, r^1) of i's indifference curve through (t^1, r^1) is positive, so $h_i^*(t^1, r^1) > 0$ by (11.14). Thus given the property of the indifference curves in the previous paragraph, $(t^1, r^1) \succ_j (t^2, r^2)$, as in Figure 11.12. Similarly, if $t^1 < t^2$, $(t^2, r^2) \succsim_i$

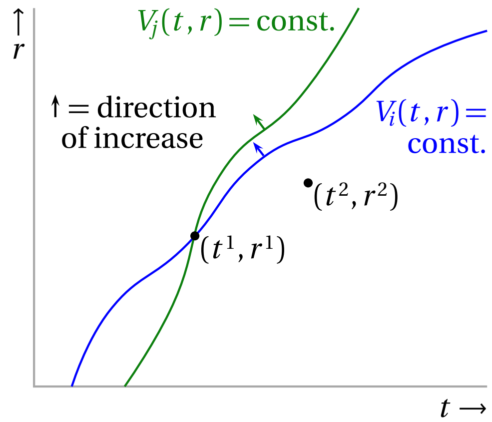


Figure 11.12 An illustration of the argument in the proof of Proposition 11.2. Individual i prefers (t^1, r^1) to (t^2, r^2), and so does any individual j whose indifference curves are more steeply sloped.

(t^1, r^1), and $j > i$ then $w_j h_j^*(t,r) < w_i h_i^*(t,r)$, so that $(t^2, r^2) \succ_j (t^1, r^1)$. Thus the conditions for single-crossing are satisfied.

The claims about the Condorcet winners follow from Proposition 1.5.

Suppose that (11.13) is satisfied by an ordering of the individuals by earning power (w_i) (as in Example 11.1 when the individuals' preferences over income and leisure are the same, and in Example 11.2), and that the political system generates a transfer system that is a Condorcet winner among the alternatives. Then by Proposition 11.2 the linear transfer system that the political system generates is the favorite, among the alternatives, of the individual with median earning power. The slope of an individual's indifference curve is her pre-tax income (see (11.14)), so if individuals with different pre-tax incomes disagree on the ordering of transfer systems, the individual with the lower pre-tax income prefers the system with a higher tax rate (and higher fixed component). Thus if, in a given society, the voting franchise is expanded among individuals with low earning power, so that the median individual's earning power decreases, the tax rate generated by the political system does not decrease, and may increase.

Recall that Proposition 8.1 says that in a Nash equilibrium of an electoral competition game with two office-motivated candidates, the policy chosen by each candidate is a Condorcet winner of the underlying collective choice problem. Thus Proposition 11.2 implies that in such a Nash equilibrium each candidate selects the favorite tax system of the individual with median pre-tax income if the number of individuals is odd, and the favorite tax system of an individual whose pre-tax income is one of the medians if the number of individuals is even.

11.4 Coalitional bargaining over redistribution

The ideas underlying the analysis in this section are that the size and wealth of each group in society determine its power, and the distribution of power shapes the factors that determine the transfer system, like the electoral system, the rules under which it operates (e.g. the rules on campaign spending), and the mechanisms by which a government makes decisions. However, the electoral system and the mechanisms of government decision-making are not modeled explicitly. Instead, the model aims to deduce directly from the distribution of power the distribution of payoffs that emerges.

The setting for the model is a society in which each individual is endowed with an amount of a consumption good; she does not have to work to obtain this good. Each individual's payoff is the amount of the good she ultimately obtains, after she pays the tax or receives the subsidy specified by the transfer system. For convenience, the number of individuals in the economy is assumed to be odd.

Definition 11.8: Endowed society

An *endowed society* $\langle N, (e_i)_{i \in S} \rangle$ consists of

- a finite set N (of individuals) with an odd number of members that is at least 3
- a number $e_i \geq 0$ for each $i \in N$ (the amount of a consumption good with which i is endowed).

A nonempty subset of N is a *coalition*. For any coalition S, $e(S) = \sum_{i \in S} e_i$, the total endowment of S. The payoff of each individual is the amount of the consumption good that she ultimately obtains.

The power of each coalition is delimited by two central assumptions:

- any majority has the option to expropriate any amount of the endowment of the complementary minority
- any minority has the option to destroy its endowment.

The following definition includes the assumption that such actions exist, and also the assumption that for any distribution of the total endowment $e(N)$ and any coalition S, actions for S and its complement $N \setminus S$ exist that achieve that distribution.

Definition 11.9: Coalitional redistribution game

A *coalitional redistribution game* $\langle N, (e_i)_{i\in N}, (A_S)_{S\subseteq N}, (h_S)_{S\subseteq N}\rangle$ consists of an endowed society $\langle N, (e_i)_{i\in N}\rangle$ and, for each coalition S, a set A_S of actions and a payoff function $h_S : A_S \times A_{N\setminus S} \to \mathbb{R}_+$ such that

a. for every distribution of the total endowment $e(N)$ of the society there are actions in A_S and $A_{N\setminus S}$ that achieve the distribution:

$$\{(h_S(\sigma_S, \sigma_{N\setminus S}), h_{N\setminus S}(\sigma_S, \sigma_{N\setminus S})) : (\sigma_S, \sigma_{N\setminus S}) \in A_S \times A_{N\setminus S}\} = \{(\pi_S, \pi_{N\setminus S}) \in \mathbb{R}^2_+ : \pi_S + \pi_{N\setminus S} \le e(N)\}$$

b. if S is a majority (has more than $\frac{1}{2}|N|$ members) then there exists $\widehat{\sigma}_S \in A_S$ such that

$$h_S(\widehat{\sigma}_S, \sigma_{N\setminus S}) \ge e(S) \text{ and } h_{N\setminus S}(\widehat{\sigma}_S, \sigma_{N\setminus S}) = 0 \text{ for all } \sigma_{N\setminus S} \in A_{N\setminus S} \quad (11.15)$$

and $\widehat{\sigma}_{N\setminus S} \in A_{N\setminus S}$ such that

$$h_S(\sigma_S, \widehat{\sigma}_{N\setminus S}) \le e(S) \text{ and } h_{N\setminus S}(\sigma_S, \widehat{\sigma}_{N\setminus S}) = 0 \text{ for all } \sigma_S \in A_S. \quad (11.16)$$

(The action $\widehat{\sigma}_S$ may be interpreted at the expropriation of any endowment of $N \setminus S$ that $N \setminus S$ does not destroy, and $\widehat{\sigma}_{N\setminus S}$ may be interpreted as the destruction of $N \setminus S$'s endowment.)

In the solution concept I use for a coalitional redistribution game, the *possibility* of each group's using its extreme actions (expropriation, destruction of endowment) determines the distribution of payoffs. No group takes those extreme actions; indeed, no coalition is singled out as the one that forms. But the compromise is shaped by the existence of these actions.

First we derive an index of the strength of each coalition S by analyzing two-player games in which the players are S and $N \setminus S$. For reasons that will become apparent, I refer to these games as "threat games". Then we derive a distribution of payoffs that balances these strengths—a compromise. I now describe each of these components in detail.

Threat games

Fix a coalition S and consider a two-player strategic game in which the players are S and its complement $N \setminus S$. The action chosen in this game by each player is interpreted as the action the player will take if negotiations break down—a

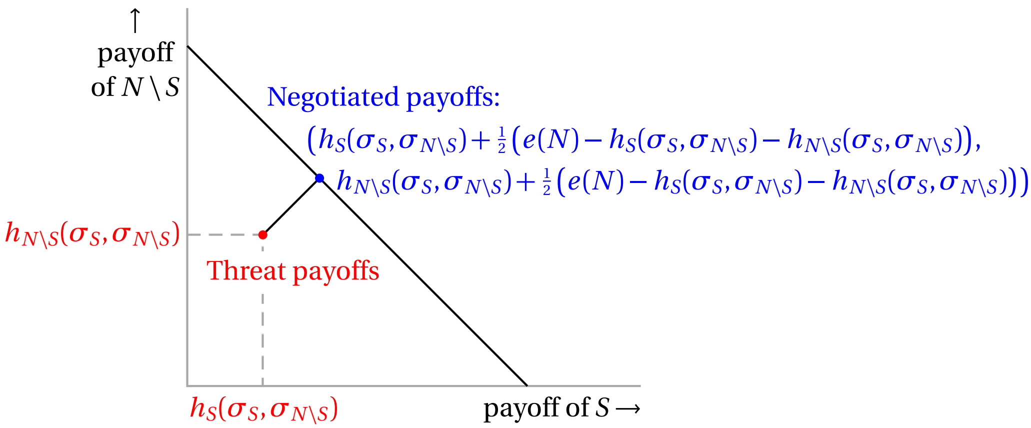


Figure 11.13

threat. The sum of the payoffs of S and $N \setminus S$ when these threats are carried out is typically less than the total payoff available, $e(N)$. The model assumes that bargaining results in S and $N \setminus S$ splitting equally the difference between $e(N)$ and this sum. Each player knows that her negotiated payoff is determined in this way, and chooses her threat to maximize her negotiated payoff, given the threat chosen by the other player.

More precisely, suppose that the players choose the actions (threats) σ_S and $\sigma_{N\setminus S}$. If they carry out these threats, their payoffs are $h_S(\sigma_S,\sigma_{N\setminus S})$ and $h_{N\setminus S}(\sigma_S,\sigma_{N\setminus S})$, and hence the surplus they forego is $e(N) - (h_S(\sigma_S,\sigma_{N\setminus S}) + h_{N\setminus S}(\sigma_S,\sigma_{N\setminus S}))$. In the negotiated outcome, this surplus is split equally between them, so that their negotiated payoffs are

$$u_S(\sigma_S,\sigma_{N\setminus S}) = h_S(\sigma_S,\sigma_{N\setminus S}) + \tfrac{1}{2}(e(N) - h_S(\sigma_S,\sigma_{N\setminus S}) - h_{N\setminus S}(\sigma_S,\sigma_{N\setminus S}))$$
$$u_{N\setminus S}(\sigma_S,\sigma_{N\setminus S}) = h_{N\setminus S}(\sigma_S,\sigma_{N\setminus S}) + \tfrac{1}{2}(e(N) - h_S(\sigma_S,\sigma_{N\setminus S}) - h_{N\setminus S}(\sigma_S,\sigma_{N\setminus S})).$$

These payoffs are illustrated in Figure 11.13.

Definition 11.10: Threat game between S and $N \setminus S$

Given a coalitional redistribution game $\langle N, (e_i)_{i\in N}, (A_S)_{S\subseteq N}, (h_S)_{S\subseteq N}\rangle$ and a coalition $S \subset N$ with more than $\frac{1}{2}|N|$ members, the *threat game* $\langle N, (e_i)_{i\in N}, A_S, A_{N\setminus S}, h_S, h_{N\setminus S}\rangle$ *between* S *and* $N\setminus S$ is the following two-player strategic game.

Players

S and $N \setminus S$.

Actions

The sets of actions of S and $N \setminus S$ are A_S and $A_{N\setminus S}$.

Payoffs

The payoff functions $u_S : A_S \times A_{N\setminus S} \to \mathbb{R}$ of S and $u_{N\setminus S} : A_S \times A_{N\setminus S} \to \mathbb{R}$ of $N \setminus S$ are defined by

$$u_S(\sigma_S, \sigma_{N\setminus S}) = \tfrac{1}{2}\big(e(N) + h_S(\sigma_S, \sigma_{N\setminus S}) - h_{N\setminus S}(\sigma_S, \sigma_{N\setminus S})\big)$$
$$u_{N\setminus S}(\sigma_S, \sigma_{N\setminus S}) = \tfrac{1}{2}\big(e(N) - h_S(\sigma_S, \sigma_{N\setminus S}) + h_{N\setminus S}(\sigma_S, \sigma_{N\setminus S})\big)$$

for all $(\sigma_S, \sigma_{N\setminus S})$.

I assume that each player chooses her action (threat) in this game to maximize her payoff, given the other player's action. That is, the pair of actions is a Nash equilibrium; we may aptly characterize it as a pair of optimal threats. For every pair $(\sigma_S, \sigma_{N\setminus S})$ of actions in the game, the sum of the players' payoffs is the same, equal to $e(N)$, so that the game is strictly competitive and hence every Nash equilibrium yields the same pair of payoffs (Proposition 16.5). The next result shows that the pair $(\widehat{\sigma}_S, \widehat{\sigma}_{N\setminus S})$ of extreme actions given in Definition 11.9b is a pair of optimal threats, and calculates the resulting negotiated payoffs.

Lemma 11.1: Nash equilibrium of threat game

Let $\langle N, (e_i)_{i\in N}, (A_S)_{S\subseteq N}, (h_S)_{S\subseteq N}\rangle$ be a coalitional redistribution game and let $S \subset N$ be a coalition with more than $\frac{1}{2}|N|$ members. The pair $(\widehat{\sigma}_S, \widehat{\sigma}_{N\setminus S})$ of actions given in Definition 11.9b is a Nash equilibrium of the threat game $\langle N, (e_i)_{i\in N}, A_S, A_{N\setminus S}, u_S, u_{N\setminus S}\rangle$, and in every Nash equilibrium the payoff of S is $\frac{1}{2}\big(e(N) + e(S)\big)$ and that of $N \setminus S$ is $\frac{1}{2}\big(e(N) - e(S)\big) = \frac{1}{2}e(N \setminus S)$.

Proof

We have $h_S(\widehat{\sigma}_S, \widehat{\sigma}_{N\setminus S}) = e(S)$ and $h_{N\setminus S}(\widehat{\sigma}_S, \widehat{\sigma}_{N\setminus S}) = 0$, so

$$u_S(\widehat{\sigma}_S, \widehat{\sigma}_{N\setminus S}) = \tfrac{1}{2}\big(e(N) + e(S)\big)$$
$$u_{N\setminus S}(\widehat{\sigma}_S, \widehat{\sigma}_{N\setminus S}) = \tfrac{1}{2}\big(e(N) - e(S)\big).$$

Now, by (11.16), for any $\sigma_S \in A_S$ we have

$$u_S(\sigma_S, \widehat{\sigma}_{N\setminus S}) = \tfrac{1}{2}\big(e(N) + h_S(\sigma_S, \widehat{\sigma}_{N\setminus S}) - h_{N\setminus S}(\sigma_S, \widehat{\sigma}_{N\setminus S})\big) \le \tfrac{1}{2}\big(e(N) + e(S)\big),$$

and by (11.15), for any $\sigma_{N\setminus S} \in A_{N\setminus S}$ we have

$$u_S(\widehat{\sigma}_S, \sigma_{N\setminus S}) = \tfrac{1}{2}\big(e(N) - h_S(\widehat{\sigma}_S, \sigma_{N\setminus S}) + h_{N\setminus S}(\widehat{\sigma}_S, \sigma_{N\setminus S})\big) \le \tfrac{1}{2}\big(e(N) - e(S)\big).$$

Thus $(\widehat{\sigma}_S, \widehat{\sigma}_{N\setminus S})$ is a Nash equilibrium of the game.

For every $(\sigma_S, \sigma_{N\setminus S}) \in A_S \times A_{N\setminus S}$ we have $u_S(\sigma_S, \sigma_{N\setminus S}) + u_{N\setminus S}(\sigma_S, \sigma_{N\setminus S}) = e(N)$, so the threat game is strictly competitive. Hence it has a unique Nash equilibrium payoff pair (Proposition 16.5).

Compromise

The model uses the Nash equilibrium payoffs in the threat game between S and $N \setminus S$ as measures of the strengths of S and $N \setminus S$, for each coalition S. The coalitional game in which the worth of each coalition is its equilibrium payoff in the threat game is called the Harsanyi coalitional form of the redistribution game (after John C. Harsanyi, 1920–2000).

Definition 11.11: Harsanyi coalitional form of coalitional redistribution game

The *Harsanyi coalitional form* of the coalitional redistribution game $\langle N, (e_i)_{i\in N}, (A_S)_{S\subseteq N}, (h_S)_{S\subseteq N}\rangle$ is the coalitional game with transferable payoff $\langle N, v\rangle$ in which the worth $v(S)$ of each coalition S is the payoff of S in a Nash equilibrium of the threat game $\langle N, (e_i)_{i\in N}, A_S, A_{N\setminus S}, h_S, h_{N\setminus S}\rangle$ between S and $N \setminus S$.

The Harsanyi coalitional form of a coalitional redistribution game is given in the following result, which follows immediately from Lemma 11.1.

Lemma 11.2: Harsanyi coalitional form of coalitional redistribution game

Let $\langle N, (e_i)_{i\in N}, (A_S)_{S\subseteq N}, (h_S)_{S\subseteq N}\rangle$ be a coalitional redistribution game. The Harsanyi coalitional form of this game is the coalitional game $\langle N, v\rangle$ for which

$$v(S) = \begin{cases} \frac{1}{2}\big(e(N) + e(S)\big) & \text{if } |S| > \frac{1}{2}|N| \\ \frac{1}{2}e(S) & \text{if } |S| < \frac{1}{2}|N| \end{cases} \tag{11.17}$$

for each $S \subseteq N$.

Giving every coalition S its worth $v(S)$ in the Harsanyi coalitional form is not feasible: no distribution of the total endowment yields these payoffs. To see why, consider a distribution of the total endowment in which each individual $i \in N$ receives x_i, so that $\sum_{i\in N} x_i = e(N)$. For the total payoff of every coalition S to be its worth $v(S)$ in the Harsanyi coalitional form, we need $\sum_{i\in S} x_i = v(S)$ for every coalition S. Denote by $\mathscr{S}$ the set of coalitions that are bare majorities, with

$k = \frac{1}{2}(n+1)$ members, and by l the number of such coalitions. For each coalition $S \in \mathscr{S}$ we have $v(S) = \frac{1}{2}(e(N) + e(S))$ by (11.17), so for the equality $\sum_{i \in S} x_i = v(S)$ to hold for all $S \in \mathscr{S}$ we need

$$\sum_{S \in \mathscr{S}} \sum_{i \in S} x_i = \tfrac{1}{2} \sum_{S \in \mathscr{S}} (e(N) + e(S)) = \tfrac{1}{2} l\, e(N) + \tfrac{1}{2} \sum_{S \in \mathscr{S}} \sum_{i \in S} e_i.$$

Reversing the order of each of the double summations and using the fact that each individual belongs to kl/n of the coalitions in $\mathscr{S}$, the left-hand side is

$$\sum_{i \in N} \sum_{\{S \in \mathscr{S} : i \in S\}} x_i = (kl/n) \sum_{i \in N} x_i = (kl/n) e(N)$$

and the right-hand side is

$$\tfrac{1}{2} l\, e(N) + \tfrac{1}{2} \sum_{i \in N} \sum_{\{S \in \mathscr{S} : i \in S\}} e_i = \tfrac{1}{2} l\, e(N) + \tfrac{1}{2}(kl/n) \sum_{i \in N} e_i = \tfrac{1}{2} l\, e(N) + \tfrac{1}{2}(kl/n) e(N).$$

Thus the two sides are equal if and only if $k = n$, which is not satisfied for any value of $n \geq 3$. Hence the distribution on which the individuals agree cannot give each coalition S its worth $v(S)$; it must entail compromise.

I present two models of compromise. One is the Shapley value, which assigns a payoff to each individual based on the impact her membership of a coalition has on the coalition's worth. Suppose that the individuals arrive in a given order. Let i be an individual, and let S be the set of individuals who arrive before i. Then i's arrival increases the worth of the set of individuals who have arrived by $v(S \cup \{i\}) - v(S)$. We can think of this amount as i's contribution for this order of arrival. The Shapley value assigns to each individual the average of her contributions over all orders. A property that imparts to it the flavor of a compromise is that the amount by which the payoff it assigns to any individual j decreases when any another individual i is excluded from the game is the same for all i and j (Proposition 16.11); no other solution concept has this property.

The next result says that the Shapley value of the Harsanyi coalitional form of a coalitional redistribution game involves a fixed subsidy equal to half of the total endowment and a 50% tax rate.

Proposition 11.3: Shapley value of coalitional redistribution game

Let $\langle N, (e_i)_{i \in N}, (A_S)_{S \subseteq N}, (h_S)_{S \subseteq N} \rangle$ be a coalitional redistribution game and let $\langle N, v \rangle$ be its Harsanyi coalitional form. The Shapley value of $\langle N, v \rangle$ assigns the payoff

$$\tfrac{1}{2}(\overline{e} + e_i)$$

to each individual $i \in N$, where $\overline{e} = e(N)/|N|$, the average endowment.

Proof

Let $i \in N$ and $n = |N|$.

Step 1 *The Shapley value of $\langle N, v\rangle$ is the same as the Shapley value of the coalitional game $\langle N, q\rangle$ for which*

$$q(S) = \begin{cases} e(S) & \text{if } |S| > \frac{1}{2}n \\ 0 & \text{if } |S| < \frac{1}{2}n. \end{cases}$$

Proof. By Lemma 16.1 the Shapley value of $\langle N, q\rangle$ is equal to the Shapley value of the game $\langle N, q^{\#}\rangle$ where $q^{\#}(S) = q(N) - q(N \setminus S)$ for each coalition S, or

$$q^{\#}(S) = \begin{cases} e(N) & \text{if } |S| > \frac{1}{2}n \\ e(S) & \text{if } |S| < \frac{1}{2}n. \end{cases}$$

By Lemma 11.2 we have $v(S) = \frac{1}{2}q(S) + \frac{1}{2}q^{\#}(S)$ for each coalition S, so the additivity of the Shapley value implies the result. ◁

Step 2 *The Shapley value of $\langle N, q\rangle$ assigns to each individual $i \in N$ the payoff $\frac{1}{2}(\bar{e} + e_i)$.*

Proof. The payoff of individual i in the Shapley value of $\langle N, q\rangle$ is the average, over all orderings R of N, of i's marginal contribution $q(S_i^R \cup \{i\}) - q(S_i^R)$ in the ordering R, where S_i^R is the set of individuals who precede i in R (see (16.3)).

- If i's position in R is $\frac{1}{2}(n-1)$ or less, her marginal contribution is 0, because $q(S) = 0$ if $|S| < \frac{1}{2}n$.

- If i's position in R is $\frac{1}{2}(n+1)$ (the middle position), her marginal contribution is $e(S_i^R \cup \{i\})$, because $q(S_i^R) = 0$ and $q(S_i^R \cup \{i\}) = e(S_i^R \cup \{i\})$. There are $(n-1)!$ orderings in which i is in this position, and each other individual comes before i in half of these orderings and after i in the other half, so the sum of i's marginal contributions over all the orderings is
$$(n-1)!\left[\tfrac{1}{2}e(N \setminus \{i\}) + e_i\right].$$

- If i's position in R is $\frac{1}{2}(n+1)+1$ or greater, her marginal contribution is e_i, because $q(S_i^R) = e(S_i^R)$ and $q(S_N^R \cup \{i\}) = e(S_i^R \cup \{i\}) = e(S_i^R) + e_i$. Thus the sum of i's marginal contributions over all the orderings in which she has a given position of $\frac{1}{2}(n+1)+1$ or greater is $(n-1)!e_i$.

There are $\frac{1}{2}(n-1)$ such positions for her, so the sum of her marginal contributions over all orderings in which her position is $\frac{1}{2}(n+1)+1$ or greater is

$$\tfrac{1}{2}(n-1)(n-1)!\,e_i.$$

The average of these marginal contributions is

$$\frac{1}{n!}\Big[\tfrac{1}{2}(n-1)!e(N\setminus\{i\})+(1+\tfrac{1}{2}(n-1))(n-1)!\,e_i\Big]$$
$$=\frac{1}{2n}e(N\setminus\{i\})+\frac{n+1}{2n}e_i=\tfrac{1}{2}\overline{e}+\tfrac{1}{2}e_i,$$

where $\overline{e}$ is the average endowment, $e(N)/n$. ◁

Another approach to modeling compromise is related to the solution concept of the core. The core of a coalitional game with transferable payoff $\langle N,v\rangle$ is the set of distributions of payoff among the players with the property that the total payoff of every coalition S is at least $v(S)$. I have argued that for the Harsanyi coalitional form $\langle N,v\rangle$ of a coalitional redistribution game no payoff distribution satisfies this property, so the core of such a game is empty. For any payoff distribution y, we can view $v(S)-y(S)$, where $y(S)=\sum_{i\in S}y_i$, as the extent of S's dissatisfaction with y. Suppose that, in the absence of a distribution in which every coalition is satisfied, we look for a distribution that minimizes the dissatisfaction. Precisely, we look for a distribution for which the maximal dissatisfaction across all coalitions is minimal. It turns out that the only distribution with this property is the one for which $y_i=\frac{1}{2}(\overline{e}+e_i)$ for all $i\in N$, the same as the Shapley value.

Proposition 11.4: Dissatisfaction-minimizing payoff distribution of coalitional redistribution game

Let $\langle N,(e_i)_{i\in N},(A_S)_{S\subseteq N},(h_S)_{S\subseteq N}\rangle$ be a coalitional redistribution game and let $\langle N,v\rangle$ be its Harsanyi coalitional form. Exactly one payoff distribution $(y_i)_{i\in N}$ minimizes $\max_{S\subseteq N}(v(S)-y(S))$, namely the one that assigns the payoff

$$\tfrac{1}{2}(\overline{e}+e_i)$$

to each individual $i\in N$, where $\overline{e}=e(N)/|N|$, the average endowment.

Proof

Let $n = |N|$ and $z_i = \frac{1}{2}(\overline{e} + e_i)$ for each $i \in N$. I first argue that the payoff distribution $(y_i)_{i\in N} = (z_i)_{i\in N}$ minimizes $\max_{S\subseteq N}(v(S) - y(S))$.

Using Lemma 11.2 we have

$$v(S) - z(S) = \begin{cases} \frac{1}{2}\big(e(N) + e(S) - e(N)|S|/n - e(S)\big) & \text{if } |S| > \frac{1}{2}n \\ \frac{1}{2}\big(e(S) - e(N)|S|/n - e(S)\big) & \text{if } |S| < \frac{1}{2}n \end{cases}$$
$$= \begin{cases} \frac{1}{2}e(N)(1 - |S|/n) & \text{if } |S| > \frac{1}{2}n \\ -\frac{1}{2}e(N)|S|/n & \text{if } |S| < \frac{1}{2}n, \end{cases}$$

so that the solutions of $\max_{S\subseteq N}(v(S) - z(S))$ are the coalitions of size $\frac{1}{2}(n+1)$ (a bare majority), and the maximum is $\frac{1}{2}e(N)(1 - \frac{1}{2}(n+1)/n) = \frac{1}{4}(n-1)e(N)/n$.

Now suppose, contrary to the claim, that there is a payoff distribution $(y_i)_{i\in N}$ for which $v(S) - y(S) < \frac{1}{4}(n-1)e(N)/n$ for every coalition S. For a coalition with $\frac{1}{2}(n+1)$ members, this inequality is $\frac{1}{2}e(N) + \frac{1}{2}e(S) - y(S) < \frac{1}{4}(n-1)e(N)/n$, so that we need

$$y(S) > \tfrac{1}{2}e(S) + \tfrac{1}{4}(n+1)e(N)/n \text{ for every coalition } S \text{ with } |S| = \tfrac{1}{2}(n+1).$$

Denote by k the number of coalitions with $\frac{1}{2}(n+1)$ members. Each individual is a member of $\frac{1}{2}(n+1)k/n$ of these coalitions, so that adding the inequalities over all coalitions with $\frac{1}{2}(n+1)$ members we get

$$\tfrac{1}{2}(n+1)ky(N)/n > \tfrac{1}{4}(n+1)ke(N)/n + \tfrac{1}{4}(n+1)ke(N)/n$$
$$= \tfrac{1}{2}(n+1)ke(N)/n,$$

which violates $y(N) = e(N)$, as required by the feasibility of $(y_i)_{i\in N}$.

Thus the payoff distribution $(y_i)_{i\in N} = (z_i)_{i\in N}$ is a minimizer of $\max_{S\subseteq N}(v(S) - y(S))$. It is the only minimizer because if $(x_i)_{i\in N}$ is a payoff distribution that differs from $(z_i)_{i\in N}$ then $z(S) > x(S)$ for some bare majority S, so that $v(S) - x(S) > v(S) - z(S) = \frac{1}{4}(n-1)e(N)/n$, the value of $\max_{S\subseteq N}(v(S) - z(S))$.

Comments

- The analysis is limited to individuals with linear payoff functions, so that transferring endowment among individuals is equivalent to transferring payoff, which allows us to use the solution concepts of the Shapley value and the dissatisfaction-minimizing payoff distribution. If payoff is not transferable, designing appealing solution concepts that model compromise is challenging.

- The solution concept of the dissatisfaction-minimizing payoff distribution is closely related to that of the nucleolus, a standard solution concept for coalitional games with transferable payoff. The nucleolus consists of the payoff distributions $(y_i)_{i \in N}$ that minimize the largest dissatisfaction $v(S) - y(S)$ and, subject to doing so, minimize the second largest dissatisfaction, and so forth. (See, for example, Moulin 1988, Section 5.4 and Osborne and Rubinstein 1994, Section 14.3.3.)

- Every possible coalition is treated in the same way by both solution concepts. However, coalitions may differ in the likelihood that they exercise their bargaining power. For example, a coalition of poor (low endowment) individuals may be more likely than one consisting of a mix of rich and poor to do so because of a shared identify or a more easily defined common purpose. In such cases, a different analysis may be appropriate.

- If we modify a coalitional redistribution game so that a minority cannot destroy its endowment, then every majority coalition can obtain the entire endowment of society, and an individual with a large endowment is no longer any more powerful than one with a small endowment. In this case, both solution concepts assign every individual the same payoff, equal to the average endowment.

- In an endowed society, incentive effects are absent: no individual chooses how much to work, so that taxation does not affect the total amount of payoff available.

Exercise 11.1: Variant of coalitional redistribution game in which wealth conveys power

Consider a variant of a coalitional redistribution game in which a coalition can expropriate its complement if it has a majority of the wealth rather than a majority of the votes. That is, the condition in part *b* of Definition 11.9 that S has more than $\frac{1}{2}|N|$ members is replaced by the condition that $e(S) > \frac{1}{2}e(N)$ (or equivalently $e(S) > e(N \setminus S)$). Assume that no coalition has exactly half of the total endowment $e(N)$ (to avoid having to specify the actions available to such a coalition). Arguments parallel to those in Step 1 of Proposition 11.3 to the conclusion that the Shapley value of the

Harsanyi coalitional form $\langle N, v\rangle$ of this game is the Shapley value of the variant of the coalitional game $\langle N, q\rangle$ in that step in which the conditions $|S| > \frac{1}{2}n$ and $|S| < \frac{1}{2}n$ are replaced by $e(S) > \frac{1}{2}e(N)$ and $e(S) < \frac{1}{2}e(N)$. For an endowed society $\langle N, (e_i)_{i\in N}\rangle$ in which one individual has more than half of the total endowment, what is the Shapley value of this game? Which payoff distributions $(y_i)_{i\in N}$ minimize $\max_{S\subseteq N}(v(S) - y(S))$?

Notes

The model in Section 11.1 is due Coughlin (1986) and Lindbeck and Weibull (1987). (Much of the analysis in these papers concerns a model in which each candidate's objective is to maximize her expected vote share rather than her probability of winning. For the reasons discussed in the comment on page 234, I do not consider this model.)

The model and analysis in Section 11.2 are based on Röell (2012) (a revised version of a paper from 1996); my exposition draws also on Brett and Weymark (2017, 2020).[1]

The model in Section 11.3.2 was first studied by Itsumi (1974) and Romer (1975), who consider whether individuals' preferences over linear tax systems are single-peaked. The single-crossing condition was developed by Rothstein (1990, 1991) (under the name order restricted preferences) and Gans and Smart (1996). Proposition 11.2 is based on Roberts (1977, Theorem 2) and Gans and Smart (1996, Proposition 1).

The model in Section 11.4 is a variant with finitely may individuals and transferable payoff of the model in Aumann and Kurz (1977); Proposition 11.3 is a version of their main result.

[1]These papers state results that may appear to be stronger than the ones that I discuss. Note, however, the following points. 1. Röell's results rest on her Theorem 6, the proof of which is incomplete because it does not show that the variable Γ defined in equation (31) of the paper is nonzero. 2. The model in Brett and Weymark (2020) does not impose an upper bound on hours of work, so that some of the issues I discuss do not arise. 3. Both Röell (2012) and Brett and Weymark (2020) show only that the individuals' preferences satisfy a variant of single-peakedness in which the preference inequalities in (1.3) are weak (Theorem 7 in Röell 2012, Theorem 4 in Brett and Weymark 2020). This property, which is satisfied by the payoffs in Figure 11.2 for the model in which the individuals care only about their consumption, is not sufficient for the existence of a Condorcet winner (see Exercise 1.10), so that Theorem 5 in Brett and Weymark (2020) does not follow from Theorem 4. In personal correspondence, Weymark argues that the proof of Theorem 4 may be modified to show that each individual's preferences have a single plateau, with strict preferences on each side of the plateau, in which case at least one of the favorite alternatives of an individual with median earning power is a Condorcet winner (see the text preceding Exercise 1.9).

Solutions to exercises

Exercise 11.1

First consider the Shapley value. The variant of the coalitional game $\langle N, q\rangle$ in Step 1 of Proposition 11.3 is the coalitional game $\langle N, q'\rangle$ where

$$q'(S) = \begin{cases} e(S) & \text{if } e(S) > \frac{1}{2}e(N) \\ 0 & \text{if } e(S) < \frac{1}{2}e(N). \end{cases}$$

Let individual 1 be the one who has more than half of the total endowment $e(N)$. The marginal contribution $q'(S_i^R \cup \{i\}) - q'(S_i^R)$ of any individual $i \neq 1$ in an ordering R is 0 if she precedes individual 1 ($q'(S_i^R \cup \{i\}) = q'(S_i^R) = 0$) and her endowment e_i if she follows individual 1 ($q'(S_i^R \cup \{i\}) = e(S_i^R) + e_i$ and $q'(S_i^R) = e(S_i^R)$). Every individual $i \neq 1$ precedes individual 1 in half of the orderings and follows her in the other half, so the average of her marginal contributions, and hence the payoff she is assigned by the Shapley value, is $\frac{1}{2}e_i$. Thus the payoff assigned to individual 1 by the Shapley value is the remaining endowment, $e(N) - \frac{1}{2}\sum_{i\in N\setminus\{1\}} e_i$. That is, every individual $i \neq 1$ is taxed at the rate of 50% and the proceeds go to individual 1.

Now consider the dissatisfaction-minimizing payoff distributions. The Harsanyi coalitional form of the game is given by

$$v(S) = \begin{cases} \frac{1}{2}\big(e(N) + e(S)\big) & \text{if } 1 \in S \\ \frac{1}{2}e(S) & \text{if } 1 \notin S. \end{cases}$$

Denote by $(z_i)_{i\in N}$ the Shapley value payoff distribution: for each $i \in N$ let $z_i = \frac{1}{2}e_i$ for $i \neq 1$ and $z_1 = e(N) - \frac{1}{2}\sum_{i\in N\setminus\{1\}} e_i$. Notice that

$$v(S) - z(S) = 0 \text{ for all } S \subseteq N.$$

Thus $(z_i)_{i\in N}$ minimizes $\max_{S\subseteq N}(v(S) - y(S))$ and is the only payoff distribution that does so.

12 Money in electoral competition

12.1 Mobilizing citizens to vote 409
12.2 Informing citizens of candidates' qualities 421

The models of electoral competition in the previous chapters have a glaring omission: money. In a mass election, a candidate may spend significant resources informing potential voters of her position and persuading them to cast their votes for her. She may try to make voting easier for her supporters and more difficult for those of her opponents; she may trumpet her accomplishments and impugn her opponents. Interested outside organizations may spend resources to engage in similar activities. Everyone may have a vote, but how they cast that vote may be affected by the campaign efforts of the candidates and outside organizations. A mass election may be less about the aggregation of the citizens' preferences and more about the manipulation of their votes by the wealthy members of society. And the policies adopted by the elected representatives may have less to do with the ones on which they campaigned and more to do with the preferences of wealthy lobbyists. In short, omitting money from the analysis of elections may be a critical flaw.

In fact, one perspective is that studying elections is the wrong place to start an investigation of the determinants of the policies societies adopt. In this view, these policies are determined by the distributions of wealth and power; the existence of elections and interest groups should not be treated as exogenous, but as an implication of the distributions of wealth and power. In some societies, elections act as one medium through which power is exercised, while in others the wealthy wield power more directly. Such a perspective underlies the model in Section 11.4, but the models I present in this chapter treat elections and interest groups as given. Even under this assumption, analyzing the issues is challenging. Much work in the field responds to the challenge by studying models in which the payoff functions and distributions involved have specific functional forms, making the generality of the conclusions hard to assess. I take a different approach, presenting some simple but relatively general models.

Synopsis

In the models I present there are two candidates and a single interest group. In Section 12.1 the set of positions is an interval of real numbers and there is a continuum of citizens, some of whom (*informed*) know the candidates' positions, and some of whom (*uninformed*) do not. Uninformed citizens vote only if prodded to do so by the interest group. The group's options, given its budget and technology, are characterized by a collection of sets of uninformed citizens; it can induce all the citizens in any one of these sets to vote. If its technology allows it to target citizens according to their preferences, for example, it may be able to mobilize all citizens with favorite positions in some interval. First the interest group selects a set of uninformed citizens to mobilize to vote, then the two candidates, who are office-motivated, simultaneously select positions, and finally each citizen who is either (*i*) informed or (*ii*) uninformed and mobilized votes for the candidate whose position she prefers. By Proposition 8.4, the subgame following the interest group's move has a unique subgame perfect equilibrium, in which the position of each candidate is the median of the voters' favorite positions. By judiciously selecting the set of citizens to mobilize, the interest group can move this median to a position it favors. Parts *c* and *d* of Proposition 12.1 specify the limits of this manipulation for an interest group that is unable to target its mobilization efforts and for one that can perfectly target the efforts according to the citizens' preferences.

Section 12.2 analyzes a model in which an interest group may provide citizens with verifiable information about the candidates' qualities. For the sake of tractability, there are only two possible policies, 0 and 1, two candidates, and a single citizen. Candidate 2's quality is known, whereas candidate 1 has two possible qualities, one better than candidate 2's and one worse. For any given candidate quality, the citizen prefers policy 0 to policy 1, but she prefers policy 1 implemented by a high-quality candidate 1 to policy 0 implemented by candidate 2. The interest group, by contrast, prefers policy 1 to policy 0. It observes candidate 1's quality and decides whether to offer to reveal this quality in exchange for candidate 1's choosing policy 1. If it makes this offer, candidate 1 decides whether to accept it. Candidate 2 is assumed to choose policy 0, because there is no advantage to her choosing policy 1, given the citizen's preferences. The citizen observes the policies chosen by the candidates, but not candidate 1's quality unless the interest group reveals it. Proposition 12.3 shows that this game has a weak sequential equilibrium in which if candidate 1's quality is high, the interest group offers to reveal it in exchange for candidate 1's choosing policy 1, and candidate 1 accepts this offer, but if candidate 1's quality is low, the interest group does not offer to reveal it. The outcome is that if candidate 1's quality is high she chooses policy 1 and is elected and if her quality is low candidate 2, who chooses policy 0, is elected. This outcome is better for the citizen than the best outcome in the absence of the interest group if and only if candidate 1's quality is sufficiently unlikely to be high. The interest group identifies the quality of candidate 1, but the cost of its doing so is that candidate 1 chooses policy 1, so if candidate 1's quality is likely to be high, the presence of the interest group makes the citizen worse off.

12.1 Mobilizing citizens to vote

Model

In the model in this section, some citizens vote only if mobilized by an interest group. I refer to these citizens as uninformed, although the reason they need to be prodded to vote may not be lack of information. The setting for the model is a society defined as follows.

Definition 12.1: Society with informed and uninformed citizens

A *society with informed and uninformed citizens* $\langle X, I, N, U, F, G^N \rangle$ consists of

- X, a closed interval of real numbers (the set of possible positions)
- I, a compact interval of real numbers (the set of citizens), $N \subset I$ (the set of informed citizens), and $U \subset I$ (the set of uninformed citizens), where $N \cup U = I$ and $N \cap U = \varnothing$ (each point in I is a citizen's name)
- $F : X \to [0,1]$, a continuous and increasing function with $F(\underline{x}) = 0$ and $F(\overline{x}) = 1$ (for each $x \in X$, $F(x)$ is the fraction of citizens with favorite positions at most x)
- $G^N : X \to [0,1]$, a continuous and increasing function with $G^N(\underline{x}) = 0$ and $G^N(x) < F(x)$ for all $x \in X$ (for each $x \in X$, $G^N(x)$ is the fraction of citizens who are informed and have favorite positions at most x).

Figure 12.1 illustrates an example of such a society, with $X = [\underline{x}, \overline{x}]$. The density of the favorite positions of the informed citizens is shown in green and that of the uninformed citizens in orange. The sum of the two, represented by the upper boundary of the orange area, is the density of F, the distribution of all citizens' favorite positions. The median of all citizens favorite positions, which I denote by m, is the median of F. The fraction of all citizens who are informed is $G^N(\overline{x})$,

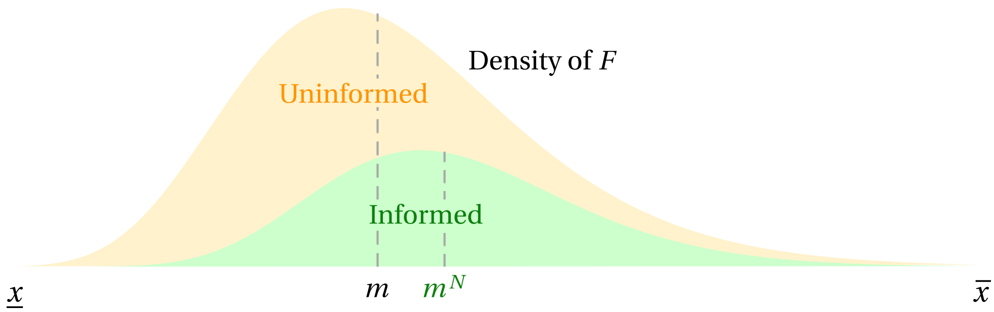


Figure 12.1 An illustration of a society with informed and uninformed citizens. The distribution of the favorite positions of informed citizens is shown in green and that of the uninformed citizens is shown in orange. The median of the informed citizens' favorite positions is m^N and that of all citizens' favorite positions is m.

so that the median of the informed citizens' favorite positions, which I denote m^N, is defined by the condition $G^N(m^N) = \frac{1}{2}G^N(\overline{x})$.

If only informed citizens vote, then by Proposition 8.4 the two-candidate electoral competition game with a continuum of citizens and office-motivated candidates in which two office-motivated candidates choose positions simultaneously has a unique Nash equilibrium, in which both candidates choose m^N.

Now suppose that before the candidates commit to positions, an interest group with preferences over positions can mobilize some of the uninformed citizens to vote. (Perhaps it does so by helping them to register to vote or by persuading them that the issues at stake are important enough to make their voting worthwhile.) Assume that the interest group's action affects only whether a citizen votes, not how she casts her vote; a citizen who is mobilized votes for the candidate she prefers. The group may be able to direct its efforts precisely to citizens with certain preferences, or may be able only to increase participation by uninformed citizens across the board. The model captures the limits the group faces by specifying a collection of sets of uninformed citizens, with the interpretation that, given the group's technology, its budget, and the citizens' characteristics, it is capable of mobilizing all the members of any one of the sets. For example, if the group knows the citizens' preferences and can precisely target its mobilization efforts, the collection may consist of all sets of uninformed citizens of at most a certain size with certain preferences.

After the group selects the set of uninformed citizens to mobilize, two candidates simultaneously choose positions and each citizen who is either informed or mobilized votes for her favorite candidate. The interest group cares about the position of the winner of the election, while each candidate is office-motivated, caring only about winning the election, not about the position of the winner.

Denote the collection of all the sets of uninformed citizens (subsets of U)

that the interest group is capable of mobilizing by $\mathscr{S}$, and the favorite position of each citizen $i \in I$ by $\hat{x}_i$. If the group mobilizes $S \in \mathscr{S}$, for any position $x \in X$ the fraction of citizens who vote and have favorite positions at most x is

$$G_S(x) = G^N(x) + \phi(\{i \in S : \hat{x}_i \leq x\}),$$

where for any set $Z \subseteq I$ of citizens $\phi(Z)$ is the fraction of citizens in Z. Thus the distribution function of the voters' favorite positions is the function F_S defined by $F_S(x) = G_S(x)/G_S(\bar{x})$. If the candidates choose the positions x_1 and x_2, the outcome $O(S,(x_1,x_2))$ of the game is the winner of the election for the electorate $N \cup S$:

$$O(S,(x_1,x_2)) = O_{F_S}(x_1,x_2), \tag{12.1}$$

where O_{F_S} is the function defined in (8.1). Each candidate prefers to win than to tie than to lose, as in the game without the interest group.

Definition 12.2: Two-candidate electoral competition game with vote mobilization by an interest group

A *two-candidate electoral competition game with vote mobilization by an interest group* $\langle\langle X, I, N, U, F, G^N\rangle, \mathscr{S}, u_g\rangle$, where $\langle X, I, N, U, F, G^N\rangle$ is a society with informed and uninformed citizens and

- $\mathscr{S}$ is a collection of subsets of U that includes $\varnothing$ ($\mathscr{S}$ is the collection of all sets of citizens the interest group is able to mobilize, taking into account both its technological and budgetary constraints; one such set is the empty set)
- $u_g : X \to \mathbb{R}$ is a single-peaked function (which represents the interest group's preferences)

is an extensive game with perfect information and simultaneous moves with the following components.

Players
The set of players is $\{1, 2, g\}$ (1 and 2 are candidates and g is an interest group).

Terminal histories
A terminal history is a sequence $(S,(x_1,x_2))$ where $S \in \mathscr{S}$ and $(x_1,x_2) \in X \times X$.

Player function
The player function P is defined by

- $P(\varnothing) = g$ (the interest group moves at the start of the game)

- $P(S) = \{1,2\}$ for every $S \in \mathscr{S}$ (the candidates move simultaneously after the interest group).

Actions

The set $A_i(h)$ of actions of each player i after the history h is given by $A_g(\varnothing) = \mathscr{S}$ and $A_1(S) = A_2(S) = X$.

Preferences

Each candidate prefers a terminal history $(S,(x_1,x_2))$ for which $O(S,(x_1,x_2))$, given in (12.1), is a win for her to one in which it is a tie to one in which it is a win for the other candidate.

The interest group's preferences are represented by the payoff function defined by

$$\begin{cases} u_g(x_j) & \text{if } O(S,(x_1,x_2)) = \textit{win for } j \ (=1,2) \\ \frac{1}{2}u_g(x_1) + \frac{1}{2}u_g(x_2) & \text{if } O(S,(x_1,x_2)) = \textit{tie}. \end{cases}$$

How much can the interest group affect the equilibrium?

How does the outcome of a subgame perfect equilibrium of this game depend on the collection $\mathscr{S}$ of sets of uninformed citizens that the interest group is capable of mobilizing? If the interest group mobilizes the set S then the set of voters is $N \cup S$, so that by Proposition 8.4 the subgame following the interest group's choice of S has a unique Nash equilibrium, in which each candidate chooses the median of the voters' favorite positions. As S varies, how does this median change? What are its smallest and largest possible values? That is, how far left and right is it possible for the interest group to move the equilibrium outcome? The next proposition, 12.1, answers these questions.

First suppose that the interest group is unable to target its efforts to citizens with specific preferences: every set it can mobilize is a random sample of uninformed citizens. If it can mobilize only small sets, the winning position it can induce is close to m^N, the median of the favorite positions of the informed individuals. As the size of the sets it can mobilize increases, the winning position it can induce moves towards m, the median of all the citizens' favorite positions. An example in which it can mobilize 30% of the uninformed citizens is shown in Figure 12.2.

Now suppose that it can target its mobilization efforts to citizens with specific preferences, and is in fact capable of mobilizing any set of uninformed citizens. Suppose that it starts mobilizing citizens with favorite positions on the far right, close to $\overline{x}$, and then gradually expands its efforts to citizens on the right with less

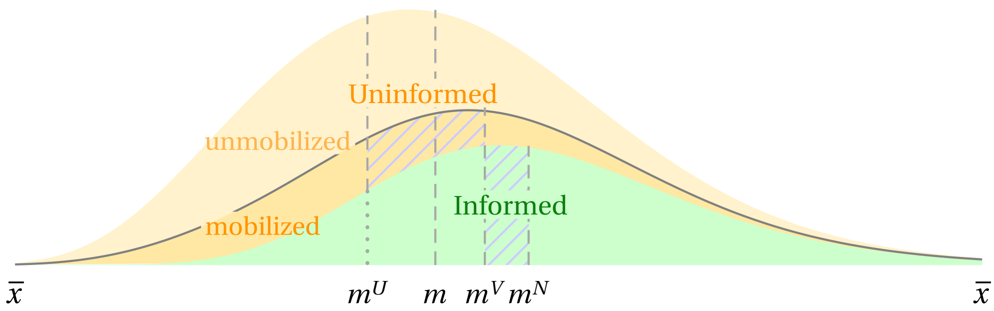


Figure 12.2 An example of distributions of informed and uninformed citizens in a two–candidate electoral competition game with vote mobilization by an interest group. In this example, the interest group can mobilize 30% of the uninformed citizens, generating the density of voters indicated by the black line, with median m^V. (The areas with blue hatching are equal.)

extreme preferences. Then the median of the voters' favorite positions gradually moves to the right from m^N. It continues to do so until it reaches the value $\overline{m}$ for which the uninformed citizens who are mobilized are exactly those with favorite positions at least $\overline{m}$. At this point the density of the distribution of the voters' favorite positions is indicated by the black line in Figure 12.3 and $G^N(\overline{m}) = 1 - F(\overline{m})$. As more citizens, with favorite positions less than $\overline{m}$, are mobilized, the median of the voters' favorite positions decreases. Thus the largest possible value for the median of the voters' favorite positions is $\overline{m}$.

A symmetric argument shows that the smallest possible value of the median of the voters' favorite positions is achieved when the interest group mobilizes all uninformed citizens with favorite positions at most $\underline{m}$, where $F(\underline{m}) = G^N(\overline{x}) - G^N(\underline{m})$.

The next result states these conclusions precisely and shows that for every position $z \in [\underline{m}, \overline{m}]$ there is a set of uninformed citizens that, if mobilized, causes the median of the voters' favorite positions to be z.

Lemma 12.1: Properties of median of favorite positions of set of informed and mobilized citizens

Let $\langle\langle X, I, N, U, F, G^N\rangle, \mathscr{S}, u_g\rangle$ be a two-candidate electoral competition game with vote mobilization by an interest group, with $X = [\underline{x}, \overline{x}]$. Denote by m the median of F (the distribution of all citizens' favorite positions) and by m^N the position for which $G^N(m^N) = \frac{1}{2}G^N(\overline{x})$ (the median of the distribution of the informed citizens' favorite positions).

a. Suppose that mobilization cannot be targeted: let $\lambda \in [0,1]$ and suppose that $\mathscr{S}$ consists of every subset S of the set U of uninformed cit-

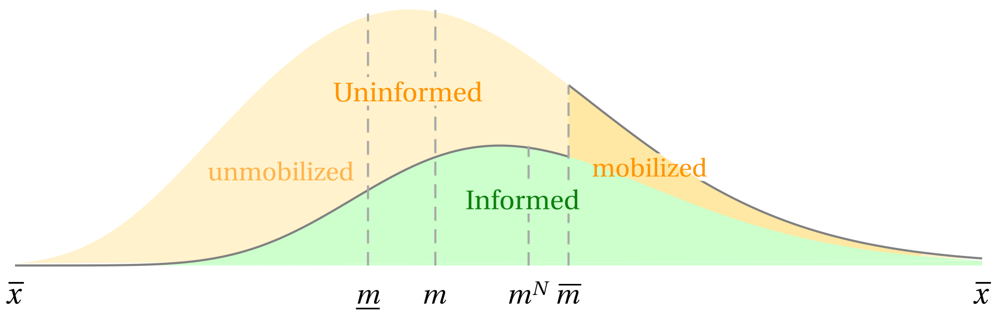


Figure 12.3 An example of distributions of informed and uninformed citizens in a two–candidate electoral competition game with vote mobilization by an interest group. If the interest group can mobilize arbitrary sets of uninformed citizens, the furthest to the right that it can move the median of the voters' favorite positions is $\overline{m}$ defined by $G^N(\overline{m}) = 1 - F(\overline{m})$, and the furthest to the left it can move it is $\underline{m}$ defined by $F(\underline{m}) = G^N(\overline{x}) - G^N(\underline{m})$.

izens that contains the fraction λ of U and for which the median of the favorite positions of its members is the median of the favorite positions of the members of U. Then for every $S \in \mathscr{S}$ the median of the favorite positions of the citizens in $N \cup S$ is the same; denote it $m^V(\lambda)$. The function m^V is continuous, with $m^V(0) = m^N$ and $m^V(1) = m$. If $m < m^N$ it is decreasing, and if $m > m^N$ it is increasing.

b. There is a unique position $\overline{m}$ such that $G^N(\overline{m}) = 1 - F(\overline{m})$ and a unique position $\underline{m}$ such that $F(\underline{m}) = G^N(\overline{x}) - G^N(\underline{m})$, and $\underline{m} < m < \overline{m}$ and $\underline{m} < m^N < \overline{m}$.

c. For any set $S \subset U$, the median of the distribution of the favorite positions of the citizens in $N \cup S$ is in $[\underline{m}, \overline{m}]$.

d. For any position z, let $\bar{S}(z)$ be the set of citizens in U with favorite positions at least z. For every position $x \in [m^N, \overline{m}]$ there exists a position $\bar{z}(x) \in [\overline{m}, \overline{x}]$ such that the median of the favorite positions of the citizens in $N \cup \bar{S}(\bar{z}(x))$ is x. We have $\bar{z}(m^N) = \overline{x}$ and $\bar{z}(\overline{m}) = \overline{m}$.

e. For any position z, let $\underline{S}(z)$ be the set of citizens in U with favorite positions at most z. For every position $x \in [\underline{m}, m^N]$ there exists a position $\underline{z}(x) \in [\underline{x}, \underline{m}]$ such that the median of the favorite positions of the citizens in $N \cup \underline{S}(\underline{z}(x))$ is x. We have $\underline{z}(\underline{m}) = \underline{m}$ and $\underline{z}(m^N) = \underline{x}$.

Proof

a. For $S \in \mathscr{S}$, the median $m^V(\lambda)$ of the favorite positions of the citizens in $N \cup S$ satisfies $G^N(m^V(\lambda)) + \lambda G^U(m^V(\lambda)) = \frac{1}{2}(G^N(\bar{x}) + \lambda G^U(\bar{x}))$, where G^U is the distribution function of the uninformed citizens, defined by $G^U(x) = F(x) - G^N(x)$ for all x. Thus $m^V(\lambda)$ is independent of S and m^V is continuous, given that F and G^N are continuous. We have $\frac{1}{2}(G^N(\bar{x}) + \lambda G^U(\bar{x})) = G^N(m^N) + \lambda G^U(m^U)$, so $G^N(m^N) - G^N(m^V(\lambda)) = \lambda(G^U(m^V(\lambda)) - G^U(m^U))$. (That is, the areas of the regions with blue hatching in Figure 12.2 are equal.) Thus $m^V(0) = m^N$ and $m^V(1) = m$; if $m < m^N$ then $m^V(\lambda) \leq m^N$ for all λ and m^V is decreasing, and if $m > m^N$ then $m^V(\lambda) \geq m^N$ for all λ and m^V is increasing.

b. The functions G^N and F are both continuous, with $G^N(\underline{x}) = F(\underline{x}) = 0$, $F(\bar{x}) = 1$, and $0 < G^N(\bar{x}) < 1$, so the equations $F(x) + G^N(x) = 1$ and $F(x) + G^N(x) = G^N(\bar{x})$ have solutions by the Intermediate Value Theorem. The solutions are unique because G^N and F are increasing, and $\underline{m} < \overline{m}$ because $G^N(\bar{x}) < 1$.

If $z \leq m$ then $1 - F(z) \geq \frac{1}{2}$ and $G^N(z) < F(z) \leq \frac{1}{2}$, and if $z \geq m$ then $F(z) \geq \frac{1}{2}$ and $G^N(\bar{x}) - G^N(z) < 1 - F(z) \leq \frac{1}{2}$, so $\underline{m} < m < \overline{m}$.

If $z \leq m^N$ then $1 - F(z) > G^N(\bar{x}) - G^N(z) \geq \frac{1}{2}G^N(\bar{x})$ and $G^N(z) \leq \frac{1}{2}G^N(\bar{x})$, and if $z \geq m^N$ then $F(z) > \frac{1}{2}G^N(\bar{x})$ and $G^N(\bar{x}) - G^N(z) \leq \frac{1}{2}G^N(\bar{x})$, so $\underline{m} < m^N < \overline{m}$.

c. Let x^V be the median of the favorite positions of the citizens in $N \cup S$. If S does not include all the citizens with favorite positions greater than x^V then adding such citizens to S increases the median, and if S includes citizens with favorite positions less than x^V then removing such citizens from S also increases the median. Thus a subset S of U for which the median x^V of the favorite positions of the citizens in $N \cup S$ is maximal consists of all members of U whose favorite positions are at least x^V. That is, $G^N(x^V) = 1 - F(x^V)$, and hence $x^V = \overline{m}$.

A symmetric argument shows that any set S for which the median of the favorite positions of the citizens in $N \cup S$ is minimal consists of all members of U whose favorite positions are at most $\underline{m}$.

d. For any position $z \in [\overline{m}, \bar{x}]$, let $\mu(z)$ be the median of the favorite positions of the citizens in $N \cup \bar{S}(z)$. We have $\mu(\overline{m}) = \overline{m}$ by the definition of $\overline{m}$ and $\mu(\bar{x}) = m^N$ because no citizen has a favorite position larger than $\bar{x}$. The result follows from the continuity of μ, which is a consequence of the continuity of F and G^N.

e. The argument is analogous to the argument for part *d*.

This result implies that in any subgame perfect equilibrium of the vote-mobilization game, the candidates' common position lies between $\underline{m}$ and $\overline{m}$. If the interest group is unable to target its mobilization efforts to citizens with specific preferences—if it can mobilize only random samples of uninformed citizens—then the winning positions it can induce lie between m and m^N and depend on the size of the set of citizens it can mobilize. If it has enough resources to mobilize all uninformed citizens then it can induce the position m, whereas if its resources allow the mobilization of only a small (random) subset of uninformed citizens then it can induce only positions close to m^N. If it can mobilize any set of citizens and can target its mobilization efforts precisely, then it can induce any position in between $\underline{m}$ and $\overline{m}$.

Proposition 12.1: Subgame perfect equilibrium of electoral competition game with vote mobilization by an interest group

Let $\langle\langle X, I, N, U, F, G^N\rangle, \mathscr{S}, u_g\rangle$ be a two-candidate electoral competition game with vote mobilization by an interest group, with $X = [\underline{x}, \overline{x}]$. Denote by m the median of F (the distribution of all citizens' favorite positions), by m^N the position for which $G^N(m^N) = \frac{1}{2}G^N(\overline{x})$ (the median of the distribution of the informed citizens' favorite positions), and by $\hat{x}_g$ the interest group's favorite position. Let $\underline{m}$ be the unique (by Lemma 12.1*b*) position for which $F(\underline{m}) = G^N(\overline{x}) - G^N(\underline{m})$ and $\overline{m}$ the unique position for which $G^N(\overline{m}) = 1 - F(\overline{m})$.

a. In every subgame perfect equilibrium of the game the candidates choose the same position, the median of the favorite positions of the citizens in $N \cup S$, where S is the set chosen by the interest group at the start of the game. This position lies in $[\underline{m}, \overline{m}]$.

b. (No mobilization possible) If $\mathscr{S} = \{\varnothing\}$, the game has a unique subgame perfect equilibrium, in which each candidate's position is m^N.

c. (Mobilization cannot be targeted) Let $\lambda \in [0,1]$ and suppose that $\mathscr{S}$ consists of every subset S of the set U of uninformed citizens that contains at most the fraction λ of U and for which the median of the favorite positions of its members is the median of the favorite positions of the members of U. By Lemma 12.1*a* the median of the favorite positions of the members of $N \cup S$ for every $S \in \mathscr{S}$ that contains exactly the fraction λ of U is independent of S; denote it $m^V(\lambda)$. If $m < m^N$ then in every subgame perfect equilibrium each candidate's position is

$$\begin{cases} m^V(\lambda) & \text{if } \hat{x}_g \le m^V(\lambda) \\ \hat{x}_g & \text{if } m^V(\lambda) < \hat{x}_g < m^N \\ m^N & \text{if } \hat{x}_g \ge m^N. \end{cases}$$

If $m^N < m$ then the common position satisfies conditions symmetric with these ones.

d. (Mobilization can be perfectly targeted) Suppose that for every $x \in X$ the subsets of U consisting of the citizens with favorite positions at least x and the citizens with favorite positions at most x are both in $\mathscr{S}$. Then in every subgame perfect equilibrium each candidate's position is

$$\begin{cases} \underline{m} & \text{if } \hat{x}_g \le \underline{m} \\ \hat{x}_g & \text{if } \underline{m} < \hat{x}_g < \overline{m} \\ \overline{m} & \text{if } \hat{x}_g \ge \underline{m}. \end{cases}$$

Proof

a. Let Γ be a two-candidate electoral competition game with a continuum of citizens and office-motivated candidates in which the set of citizens is $N \cup S$. The subgame of the game here following any action S of the interest group is equivalent to a variant of Γ in which the interest group is a third player, with no actions, and its analysis parallels that of Γ. In particular, by the arguments in the proof of Proposition 8.4, Γ has a unique Nash equilibrium, in which each candidate's position is the median of the favorite positions of the citizens in $N \cup S$.

By Lemma 12.1*c*, for every $S \subset U$ the median of the favorite positions of the citizens in $N \cup S$ is in $[\underline{m}, \overline{m}]$, so in every outcome of a subgame perfect equilibrium the candidates' common position is in this interval.

b. The result follows from part *a* for $S = \varnothing$.

c. The result follows from part *a* and Lemma 12.1*a*.

d. First suppose that $\hat{x}_g \le \underline{m}$. By Lemma 12.1*c*, the median of $N \cup S$ is at least $\underline{m}$ for every set S, so by part *a* the interest group can achieve no outcome better than the one in which both candidates choose $\underline{m}$. By Lemma 12.1*e*, the interest group can achieve this outcome (by choosing S to be the set of all uninformed citizens with favorite positions at most $\underline{m}$).

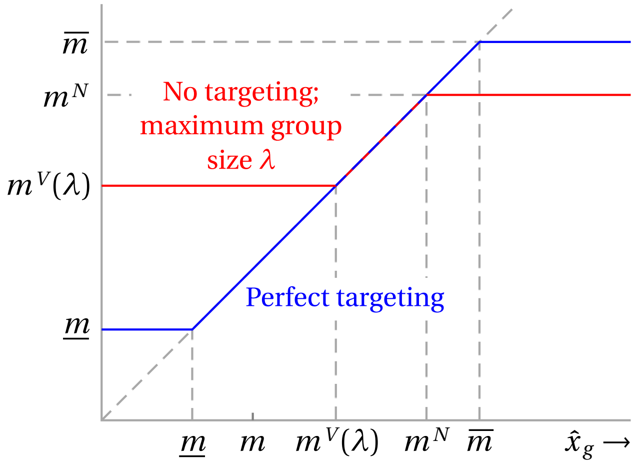


Figure 12.4 The electoral outcomes in a two-candidate electoral competition game with vote mobilization by an interest group as a function of the favorite position $\hat{x}_g$ of the interest group and the group's ability to target citizens with specific preferences, for an example in which $m < m^N$. (See Proposition 12.1*c* and *d*.)

> Now suppose that $\underline{m} < \hat{x}_g \leq m^N$. By Lemma 12.1*e* there is a position $z \in [\underline{x}, \underline{m}]$ such that when the interest group mobilizes the set S of uninformed citizens with favorite positions at most z, the median of the voter's favorite positions is $\hat{x}_g$. Thus by part *a*, when the interest group mobilizes S, both candidates choose the position $\hat{x}_g$. No outcome is better for the interest group, so it is the outcome of every subgame perfect equilibrium.
>
> Similar analyses, using Lemma 12.1*d*, apply when $m^N \leq \hat{x}_g \leq \overline{m}$ and $\hat{x}_g \geq \overline{m}$.

Parts *c* and *d* of the result are illustrated in Figure 12.4. The result is a formalization of the idea that an interest group with resources available to mobilize citizens to vote can bend the outcome of an election in its favor. If it is unable to target citizens with specific preferences then it can move the electoral outcome only closer to the median m of all the citizens' favorite positions, while if it is able to target citizens with specific preferences and its targeting ability is sufficiently precise it can move the electoral outcome away from m.

Comments

Interest group cares about size of mobilized set In the game I have defined, the interest group cares only about the position of the winner of the election, not about the size of the set of citizens it mobilizes. If it cares about the size of the set, because, for example, it incurs a cost that increases in this size, it optimally balances the benefit of mobilizing more citizens (to change the candidates' common equilibrium position) with the cost of doing so.

Two interest groups Suppose that there are two interest groups rather than one, with favorite positions on opposite sides of m, and for some position p between their favorite positions each group is able to mobilize all of the uninformed citizens with favorite positions on the same side of p as theirs (and is insensitive to the cost of doing so). Then the game in which the interest groups choose their mobilization sets simultaneously before the candidates choose positions has a subgame perfect equilibrium in which each group mobilizes all the citizens it can, so that every uninformed citizen is mobilized, and the outcome is that both candidates choose m.

If the groups care about the sizes of the sets they mobilize, preferring small sets to large (because of the cost of mobilization), and differ sufficiently in these preferences, the game has an equilibrium in which the winning position differs from m. (By Lemma 12.1c it is between $\underline{m}$ and $\overline{m}$.)

Interest group and candidates move simultaneously Consider the strategic game in which the interest group and the candidates move simultaneously: no player can commit to an action before the others move.

If the interest group cares only about the winning position, not about the size of the set it mobilizes, then every action profile (S, x_1, x_2) in which x_1 and x_2 are both equal to the median favorite position of the citizens in $N \cup S$ is a Nash equilibrium of this game. Given that the candidates' positions are the same, the set the interest group mobilizes has no effect on the winning position, and given this set, the only pair of mutually optimal positions for the candidates is the one in which both candidates choose the median favorite position of the citizens in $N \cup S$, by the arguments for the game without an interest group.

If the interest group cares about the size of the set of citizens it mobilizes and, for any given outcome, prefers to mobilize a small set than a large one, then the simultaneous move game has a single Nash equilibrium, in which the interest group mobilizes no one and the candidates' positions are both the median favorite position of the informed citizens. That is, the interest group has no effect.

Thus if no player can commit to an action before the others move, the model has no interesting equilibrium.

Interest group moves after candidates The subgame perfect outcome of the game survives in the variant in which the interest group moves after the candidates rather than before them, as you are asked to show in the next exercise.

Exercise 12.1: Mobilization game in which interest group moves after candidates

Let $(S^*,(x_1^*,x_2^*))$ be the outcome of a subgame perfect equilibrium of the game in which the interest group moves first, as described in Proposition 12.1. Consider the extensive game in which the candidates move (simultaneously) first, before the interest group. Show that if the interest group has an optimal action for every pair of the candidates' positions then this game has a subgame perfect equilibrium in which the candidates' positions are x_1^* and x_2^* and the interest group mobilizes S^* after this history.

Unlike the equilibria described in Proposition 12.1, however, such an equilibrium is sensitive to the assumption that the interest group cares only about the winning position, not about the size of the set of citizens it mobilizes. Suppose that the interest group incurs a cost $c(S)$ that increases with the size of the set S of citizens it mobilizes, and that its preferences are represented by $u_g(x_i)-c(S)$, where x_i is the winning position. (The cost of mobilizing citizens to vote may also reasonably depend on the candidates' positions—mobilizing citizens to vote when the positions are similar may require more effort than when they are far apart. Such a dependence reinforces the following argument.) Consider an equilibrium of the game in which the interest group moves first that has the form described in Proposition 12.1*d* with $m^N < x_1^* = x_2^* \leq \hat{x}_g$. If in the game in which the candidates move first they choose the positions x_1^* and x_2^* then the only optimal action of the interest group in the subsequent subgame is to mobilize no one, so that the median of the voters' favorite positions is m^N. Suppose that candidate 1 deviates to x_1 slightly less than x_1^*. Then if the interest group continues to mobilize no one, the outcome changes from x_1^* to x_1, which is worse for the interest group. To deter this deviation, in the subgame following (x_1,x_2^*) the interest group needs to mobilize enough citizens to move the median of the voters' favorite positions from m^N to a position greater than $\frac{1}{2}(x_1+x_2^*)$. The cost of doing so may be significant, so that depending on the size of $u_g(x_1)$ relative to $u_g(x_1^*) = u_g(x_2^*)$ the interest group might optimally continue to mobilize no one. In this case, in no subgame perfect equilibrium is the pair of the candidates' positions (x_1^*,x_2^*).

The significance of the interest group's moving first is that it commits to a budget for mobilization before the candidates commit to positions. If it is better off in the equilibrium of the game in which it does so than in the one in which it does not commit to a budget before the candidates commit to positions, it has an interest in committing and may be able to do so by raising money early in the election campaign. That is, the interest group may be able to take actions that make it a first-mover and may have an interest in so doing.

Interest group commits to position-contingent mobilization Another possible assumption about the timing of the players' actions is that the interest group first selects a candidate and a function that specifies the set of citizens the group will mobilize for each position the candidate chooses, and then the candidates choose positions. Let $(S^*, (x_1^*, x_2^*))$ be the outcome of a subgame perfect equilibrium of the game in which the interest group moves first, as described in Proposition 12.1. Then the game with position-contingent mobilization has a subgame perfect equilibrium in which the interest group proposes to candidate 1 that it will mobilize S^* regardless of the position the candidate chooses, and the candidates subsequently choose x_1^* and x_2^*. The game has no subgame perfect equilibrium with a better outcome for the interest group by the argument for the original game that S^* is optimal. (It has other subgame perfect equilibria with the same outcome, in which it proposes to mobilize S^* if candidate 1 chooses x_1^* and other sets if candidate 1 chooses a different position.)

Policy-motivated candidates If the candidates are policy-motivated and the distribution of the citizens' favorite positions is uncertain, as in a two-candidate electoral competition game with policy-motivated candidates and uncertain median, then in the absence of the interest group the candidates' equilibrium positions differ (Proposition 9.4). I conjecture that the addition of an interest group with mobilization options that moves either before or after the candidates shifts the candidates' equilibrium positions towards the interest group's favorite position, but I know of no formal results for this model.

12.2 Informing citizens of candidates' qualities

Suppose that interest groups can provide verifiable information about candidates that the candidates themselves are unable to provide. If the interest groups' preferences differ from those of the majority of citizens and they provide information in exchange for the candidates committing to positions that the interest groups favor, are the citizens better off or worse off than in their absence? The model in this section is intended to examine this question. The analysis of a suitable model rapidly increases in complexity and opacity as the number of players and the number of their possible actions increase. The model I present is one of the simplest that can address the issues.

There are two possible policies, 0 and 1, two candidates, and a single citizen. Each candidate chooses a policy and the citizen votes for one of the candidates. Candidate 1 has two possible qualities, l and h; in line with standard terminology, I refer to l and h as the two possible *types* of candidate 1. Candidate 2's quality is known to be 0, which is between l and h. Candidate 1 knows her quality, but the citizen does not; the citizen believes that it is l with probability p and h with probability $1-p$. Each candidate is office-motivated: she prefers to win (that is, to obtain the citizen's vote), in which case her payoff is 1, than to lose, in which case her payoff is 0.

The citizen's payoff from policy x carried out by a candidate with quality q is $u(x,q)$. For each candidate quality, the citizen prefers policy 0 to policy 1:

$$u(0,q) > u(1,q) \text{ for } q \in \{l,0,h\}.$$

For each policy, the citizen prefers a candidate of quality h to one of quality 0 to one of quality l:

$$u(x,h) > u(x,0) > u(x,l) \text{ for } x \in \{0,1\}.$$

Finally, to make it possible for candidate 1 to attract the citizen's vote if she chooses policy 1 and the interest group informs the citizen that her quality is h, the citizen prefers policy 1 carried out by a candidate of quality h to policy 0 carried out by a candidate of quality 0:

$$u(1,h) > u(0,0).$$

Given that candidate 2's quality is known and the citizen prefers policy 0 to policy 1 conditional on the quality of the candidate offering the policy, there is no advantage to candidate 2's selecting policy 1. Thus I assume that she selects policy 0. She has no other decisions to make, so she does not appear as a player in the games I analyze, although her existence affects the citizen's options.

In the main model, an interest group that prefers policy 1 to policy 0, in contrast to the citizen, can reveal candidate 1's quality in exchange for her committing to policy 1. To assess the interest group's impact, I first find the equilibrium outcomes in its absence and then analyze a model in which it is present.

12.2.1 Model without interest group

Consider the extensive game in Figure 12.5, in which no interest group is present. The game begins in the center, with a move of chance (c) that determines the quality of candidate 1. This quality is l with probability p and h with probability $1-p$. Candidate 1 observes this move—that is, she knows her quality—and then chooses a policy, 0 or 1. The citizen observes the policy chosen by the candidate but not the candidate's quality, and then selects (votes for) one of the candidates.

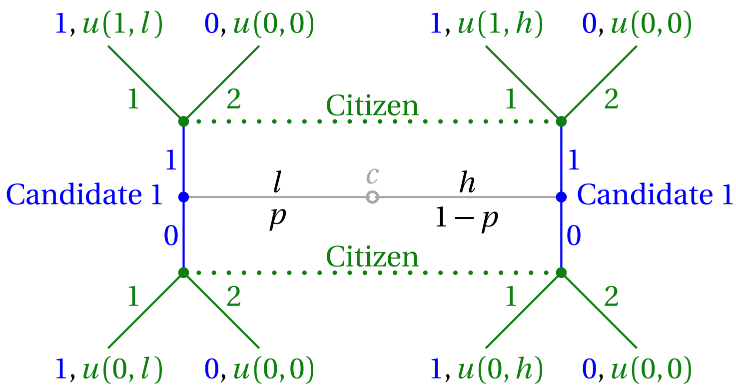


Figure 12.5 A policy game with a candidate of uncertain quality. The players are a citizen and candidate 1. The game begins with a move of chance (in the center of the figure), which determines the quality of candidate 1, either l or h. Candidate 1 observes this move and then chooses policy 0 or policy 1. The citizen does not observe candidate 1's quality. At each information set, the citizen selects (votes for) candidate 1 or candidate 2. For each terminal history, candidate 1's payoff is listed first and the citizen's second.

Definition 12.3: Policy game with a candidate of uncertain quality

A *policy game with a candidate of uncertain quality* $\langle l, h, p, u\rangle$, in which the players are a citizen and candidate 1, where

- l and h are possible qualities for candidate 1
- $p \in (0,1)$ (the probability that candidate 1's quality is l)
- $u : \{0,1\} \times \{l,h\} \to \mathbb{R}$ with $u(0,q) > u(1,q)$ for $q = l$ and $q = h$, $u(x,h) > u(x,0) > u(x,l)$ for $x = 0$ and $x = 1$, and $u(1,h) > u(0,0)$ (the citizen's payoff function),

is the extensive game with imperfect information shown in Figure 12.5, where for each terminal history candidate 1's payoff is listed first and the citizen's second.

I claim that in every weak sequential equilibrium of any such game $\langle l, h, p, u\rangle$ both types of candidate 1 (quality l and quality h) choose the same policy. Consider an assessment in which they choose different policies. Suppose that type h chooses policy 0 and type l chooses policy 1. Then if the citizen's beliefs are consistent with the candidate's strategy, they assign probability 1 to the history $(h, 0)$ at her information set following the policy 0 (the lower of the two sets in Figure 12.5) and probability 1 to the history $(l, 1)$ in her information set following the policy 1. Consequently the citizen optimally votes for candidate 1 if candidate 1 chooses policy 0 and for candidate 2 if candidate 1 chooses policy 1. But then if type l of candidate 1 deviates and chooses policy 0, pretending to be type h, the citizen votes for her and she is better off, so the assessment is not a weak sequential equilibrium. A similar argument applies to an assessment in which type h chooses policy 1 and type l chooses policy 0.

Every game $\langle l, h, p, u\rangle$ has weak sequential equilibria in which both types of candidate 1 choose policy 0, and also ones in which both types of candidate 1 choose policy 1. For some games, candidate 1 loses in some of these equilibria. In these cases the citizen's belief leads her to vote for candidate 2 not only if candidate 1 adheres to her strategy but also if she deviates from it.

Proposition 12.2: Weak sequential equilibria of policy game with a candidate of uncertain quality

Let $\langle l, h, p, u\rangle$ be a policy game with a candidate of uncertain quality.

a. In every weak sequential equilibrium of the game both types of candidate 1 choose the same policy.

b. The game has weak sequential equilibria in which both types of candidate 1 choose policy 0 and ones in which both types choose policy 1.

c. Let $x \in \{0,1\}$. In the weak sequential equilibria in which both types of candidate 1 choose policy x, the citizen votes for candidate 1 if $pu(x,l)+(1-p)u(x,h) > u(0,0)$ and for candidate 2 if $pu(x,l)+(1-p)u(x,h) < u(0,0)$. If $pu(x,l)+(1-p)u(x,h) = u(0,0)$ then equilibria exist in which the citizen votes for candidate 1 and in which she votes for candidate 2. In each of these equilibria the citizen's payoff is $\max\{u(0,0), pu(x,l)+(1-p)u(x,h)\}$.

Proof

A proof of part a is given in the text preceding the proposition.

To prove parts b and c, suppose that both types of candidate 1 choose policy 0. For the citizen's beliefs to be consistent with this strategy, at her lower information set the citizen assigns probability p to the history $(l,0)$ and probability $1-p$ to the history $(h,0)$. Thus at this information set she optimally votes for candidate 1 if $pu(0,l)+(1-p)u(0,h) > u(0,0)$, for candidate 2 if the reverse inequality holds, and for either candidate in the case of equality. Her payoff is thus $\max\{u(0,0), pu(0,l)+(1-p)u(0,h)\}$. If she votes for candidate 1 then candidate 1's strategy of choosing policy 0 is optimal for her regardless of the beliefs of the citizen at her upper information set and her consequent optimal action. If she votes for candidate 2 then for each type of candidate 1 to be no better off deviating to policy 1, the citizen's belief at her upper information set must assign sufficient probability to the history $(l,1)$ to make a vote for candidate 2 optimal for her. Given that $u(1,l) < u(1,0) < u(0,0)$, such a belief exists, and given that this information set is not reached if the players follow their strategies, such a belief is consistent with equilibrium.

A similar argument applies if both types of candidate 1 choose policy 1.

12.2.2 Model with interest group

Now add an interest group to the model. The resulting game is shown in Figure 12.6, where the pair (x,q) attached to each terminal history consists of the position x of the chosen candidate and her quality q, not the players' payoffs as in Figure 12.5. As before, chance (c) first determines the quality of candidate 1. The interest group observes this quality and can offer to (convincingly) reveal it (action A, for "advertise") in exchange for candidate 1's selecting policy 1, which the interest group prefers to policy 0. Candidate 1 can accept this offer, in which case she selects policy 1 and the citizen is informed of her quality, or reject it, in which case she selects policy 0 and the citizen is not informed of her quality. If the interest group does not make an offer (action 0) then candidate 1 chooses policy 0 or policy 1 as before. The citizen observes the policy chosen by candidate 1 but does not observe her quality unless the interest group offers to reveal it and candidate 1 accepts this offer. Also, the citizen does not observe whether the interest group offered to reveal candidate 1's quality unless the interest group makes an offer and candidate 1 accepts it. That is, the four histories $(l,A,0)$, $(h,A,0)$, $(l,0,0)$, and $(h,0,0)$ are in the same information set, indicated by the dotted rectangle in Figure 12.6.

The preferences of the citizen and the candidates are the same as in the game without the interest group. In particular, for any given candidate quality, the citizen prefers policy 0 to policy 1. The interest group, by contrast, prefers policy 1 to policy 0, and does not care about the quality of the candidate. (Perhaps the quality represents the candidate's policies on issues about which the interest group is not interested.) Specifically, if the policy of the winner of the election is x then the interest group's payoff is x if it does not advertise candidate 1's quality and $x - c$ if it does, where $0 \leq c \leq 1$.

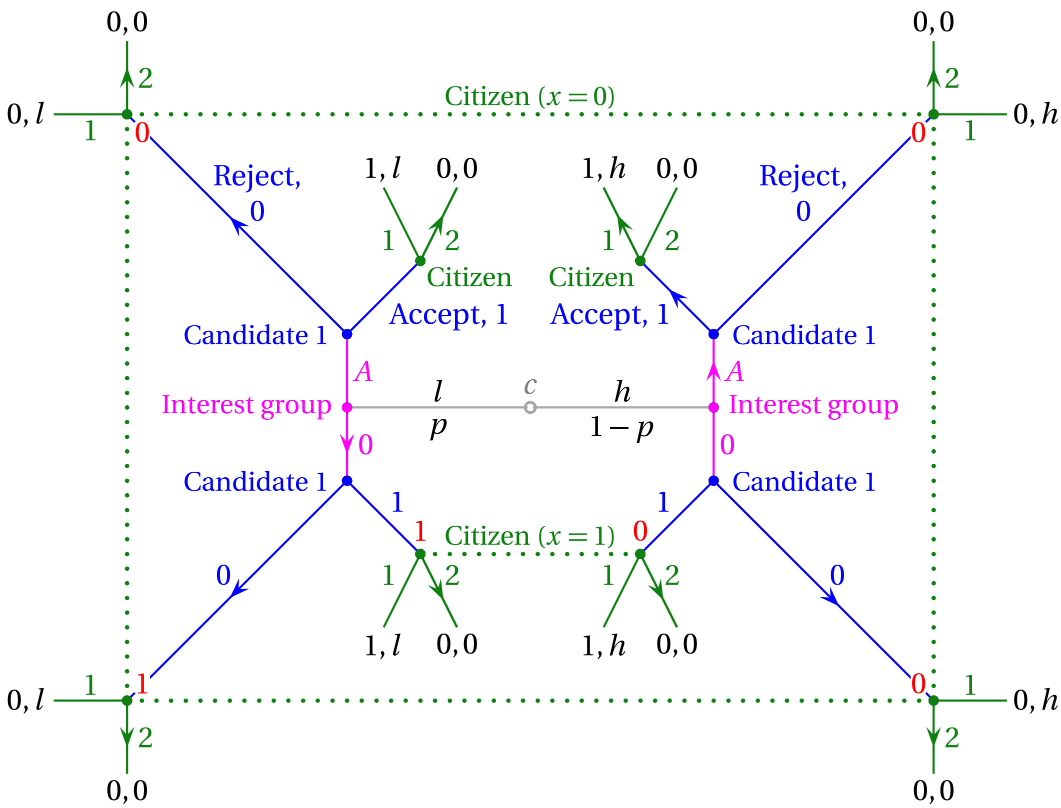


Figure 12.6 A policy game with a candidate of uncertain quality and interest group. The pair (x, q) attached to each terminal history consists of the position x of the chosen candidate and her quality q (*not the players' payoffs*). The arrows indicate the strategy profile and the numbers in red the belief system in one weak sequential equilibrium.

Definition 12.4: Policy game with a candidate of uncertain quality and interest group

A *policy game with a candidate of uncertain quality and an interest group* $\langle l, h, p, u, A, c\rangle$, in which the players are a citizen, candidate 1, and an interest group, where

- l and h are possible qualities for candidate 1
- $p \in (0,1)$ (the probability that candidate 1's quality is l)
- $u : \{0,1\} \times \{l, h\} \to \mathbb{R}$ with $u(0,q) > u(1,q)$ for $q = l$ and $q = h$, $u(x,h) > u(x,0) > u(x,l)$ for $x = 0$ and $x = 1$, and $u(1,h) > u(0,0)$ (the citizen's payoff function),
- A is the action of the interest group to advertise candidate 1's quality

- $c \in [0,1]$ (the interest group's advertising cost),

is an extensive game with imperfect information with the terminal histories, player function, chance probabilities, and information partitions shown in Figure 12.6. The players' preferences are represented by payoff functions that assign the following payoffs to each terminal history (q, z, x, v), where $q \in \{l, h\}$ (the quality of candidate 1), $z \in \{A, 0\}$ (the action of the interest group, to offer to reveal candidate 1's quality or not), $x \in \{0, 1\}$ (the policy chosen by candidate 1, where 0 entails the rejection of the interest group's offer, if it made one, and 1 entails the acceptance of such an offer), and $v \in \{1, 2\}$ (the candidate elected):

$$\text{Citizen} \quad \begin{cases} u(x,q) & \text{if } v = 1 \\ u(0,0) & \text{if } v = 2 \end{cases}$$

$$\text{Candidate 1} \quad \begin{cases} 1 & \text{if } v = 1 \\ 0 & \text{if } v = 2 \end{cases}$$

$$\text{Interest group} \quad \begin{cases} x - c & \text{if } z = A \text{ and } x = 1 \\ x & \text{otherwise.} \end{cases}$$

If the citizen's expected payoff from policy 0 carried out by candidate 1, evaluated according to the prior probabilities p and $1-p$, is at least her payoff $u(0,0)$ from choosing candidate 2, then this game has weak sequential equilibria that correspond to the equilibria of the game in which the interest group is absent: the interest group does not offer to reveal the quality of either type of candidate 1. Both types of candidate 1 choose policy 0 whether or not the interest group offers to reveal their types; at the information set reached when candidate 1 chooses policy 0, the belief assigns probability p to the history $(l, 0, 0)$ and probability $1-p$ to the history $(h, 0, 0)$ (as required by the consistency condition) and the citizen votes for candidate 1. The citizen votes for candidate 2 after the history $(l, A, 1)$ and for candidate 1 after the history $(h, A, 1)$. The belief at the information set $\{(l, 0, 1), (h, 0, 1)\}$ is arbitrary; the citizen's action at this set is the one that is optimal given her belief.

More interestingly, every policy game with a candidate of uncertain quality and an interest group has a weak sequential equilibrium in which if candidate 1's quality is h, the interest group offers to reveal this quality in exchange for candidate 1's selecting policy 1, and candidate 1 accepts this offer, but if candidate 1's quality is l then the interest group does not offer to reveal it. In this equilibrium the outcome is policy 0 implemented by candidate 2 (with quality 0) if candidate 1's quality is l and policy 1 implemented by candidate 1 if her quality if h.

The arrows and beliefs (the red numbers) in Figure 12.6 indicate this equilibrium.

Proposition 12.3: Weak sequential equilibrium of policy game with candidate of uncertain quality and interest group

A policy game with a candidate of uncertain quality and an interest group $\langle l, h, p, u, A, c\rangle$ has a weak sequential equilibrium in which

- the interest group offers to reveal candidate 1's quality if it is h but not if it is l
- if the interest group offers to reveal candidate 1's quality, the candidate accepts the offer and chooses policy 1 if her quality is h, and rejects the offer and chooses policy 0 if her quality is l; if the interest group does not offer to reveal her quality, she chooses policy 0 regardless of her quality
- the citizen votes for candidate 1 if the interest group reveals that the candidate's quality is h and the candidate chooses policy 1, and otherwise votes for candidate 2
- at the information set reached after the interest group does not offer to reveal candidate 1's type and candidate 1 chooses policy 1, the citizen believes that the candidate's quality is l
- at the information set reached after candidate 1 chooses policy 0, the citizen believes that the candidate's quality is l and the interest group did not offer to reveal the candidate's quality.

The outcome is that (*a*) if candidate 1's quality is h, the interest group advertises this quality, candidate 1 chooses policy 1, and the citizen votes for candidate 1, and (*b*) if candidate 1's quality is l, the interest group does not offer to advertise this quality, candidate 1 chooses policy 0, and the citizen votes for candidate 2. The citizen's expected payoff in the equilibrium is $pu(0,l)+(1-p)u(1,h)$.

Proof

Strategy of interest group

After the history l, actions A and 0 both lead to the election of candidate 2 and hence policy 0, so that in particular the action 0 is optimal.

After the history h, the action A leads to the election of candidate 1, who chooses policy 1, and the action 0 leads to the election of candidate 2, who chooses policy 0. Given $c \leq 1$, the latter outcome is no better for the interest group than the former, so that the former is optimal.

Strategy of candidate 1

After each of the histories (l, A), $(l, 0)$, and $(h, 0)$, the choice of either policy 0 or policy 1 leads to the election of candidate 2, so choosing policy 0 after each of these histories is optimal.

After the history (h, A), the choice of policy 1 leads to the election of candidate 1 and the choice of policy 0 leads to the election of candidate 2, so the former action is optimal.

Strategy of citizen

After the history $(l, A, 1)$, electing candidate 1 yields the payoff $u(1, l)$ whereas electing candidate 2 yields the payoff $u(0,0)$, so given $u(0,0) > u(0, l) > u(1, l)$, the citizen's electing candidate 2 is optimal.

After the history $(h, A, 1)$, electing candidate 1 yields the payoff $u(1, h)$ whereas electing candidate 2 yields the payoff $u(0,0)$, so given $u(1, h) > u(0,0)$, the citizen's electing candidate 1 is optimal.

At the information set $\{(l,0,1),(h,0,1)\}$, the citizen's belief assigns probability 1 to the history $(l,0,1)$, so that she believes that electing candidate 1 will yield the payoff $u(1, l)$ whereas electing candidate 2 will yield the payoff $u(0,0)$, and hence given $u(0,0) > u(0, l) > u(1, l)$, her electing candidate 2 is optimal.

At the information set $\{(l,A,0),(l,0,0),(h,A,0),(h,0,0)\}$ (the rectangle in Figure 12.6), the citizen's belief assigns probability 1 to the history $(l,0,0)$, so that she believes that electing candidate 1 will yield the payoff $u(0, l)$ whereas electing candidate 2 will yield the payoff $u(0,0)$, and hence given $u(0,0) > u(0, l)$, her electing candidate 2 is optimal.

Beliefs

Given the strategy profile, the probability of reaching the information set $\{(l,0,1),(h,0,1)\}$ is 0, so any belief at this information set is consistent with the strategy profile. The probability of the history $(l,0,0)$ conditional on reaching the information set $\{(l,A,0),(l,0,0),(h,A,0),(h,0,0)\}$ is 1, so the belief at this information set is consistent with the strategy profile.

Figure 12.7 shows the citizen's expected payoffs in the equilibria of the game

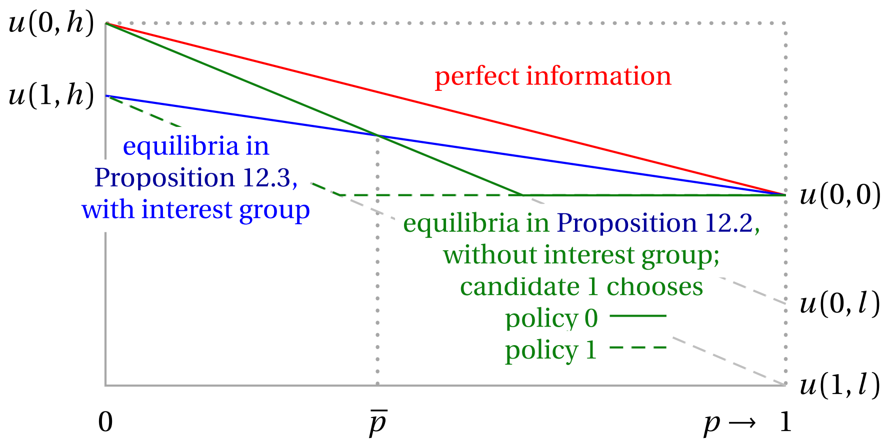


Figure 12.7 The citizen's expected payoffs in equilibria of policy games with a candidate of uncertain quality with and without an interest group, as a function of p, the probability that candidate 1's quality is l.

without the interest group given in Proposition 12.2 and in the equilibrium of the game with the interest group given in Proposition 12.3. In the latter equilibrium, the interest group identifies the type of candidate 1, but the price of its doing so is that if candidate 1's type is h then she chooses policy 1, which is worse for the citizen than policy 0. As a consequence, if the probability that candidate 1's quality is h is sufficiently large—if the probability that her type is l is less than $\overline{p}$ in the figure—the citizen is better off in the equilibrium without the interest group in which both types of candidate 1 choose policy 0.

This analysis of the citizen's payoff ignores the cost of advertising. If this cost is ultimately borne by the citizen then the changes in her payoff need to be adjusted appropriately.

Notes

Section 12.1 is not closely based on any published model, but draws elements from the models of Baron (1994), Grossman and Helpman (1996), and Herrera et al. (2008).

Section 12.2 is based on Prat (2006), which takes elements from Coate (2004) and Ashworth (2006).

Solutions to exercises

Exercise 12.1

By Proposition 12.1a, x_1^* and x_2^* are both equal to the median of the favorite positions of the citizens in $N \cup S^*$. Given that $x_1^* = x_2^*$, the interest group is indifferent among all the sets of citizens it can mobilize, so that in particular S^* is optimal for it after the history (x_1^*, x_2^*). Now consider a history (x_1, x_2^*) with $x_1 \neq x_1^*$. If the interest group continues to choose S^*, then candidate 1 loses and the outcome remains $x_1^* = x_2^*$. The interest group benefits from changing the set of citizens it mobilizes only if doing so causes candidate 1 to win and x_1 is between $\hat{x}_g$ and $x_1^* = x_2^*$. To cause candidate 1 to win, the interest group must mobilize a set S for which the median favorite position m^V of the set of voters is closer to x_1 than to x_2^*. However, in that case in the game in which it moves first it could achieve the outcome m^V, which it prefers to $x_1^* = x_2^*$, by mobilizing S, contradicting the fact that $(S^*, (x_1^*, x_2^*))$ is the outcome of a subgame perfect equilibrium of the game in which it moves first. So for no value of x_1 can the interest group profitably deviate from S^*. The same argument applies to histories (x_1^*, x_2) with $x_2 \neq x_2^*$.

If the interest group chooses S^* after all such histories, as well as after (x_1^*, x_2^*), then each candidate i faces the same electorate whether she chooses x_i^* or deviates from it, so that by the arguments in the proof of Proposition 8.4 for the strategic game in which the interest group is absent, neither candidate i can increase her payoff by deviating from x_i^*, given the position of the other candidate and the interest group's strategy.

We conclude that if the interest group has an optimal action for every pair (x_1, x_2) with $x_1 \neq x_1^*$ and $x_2 \neq x_2^*$ then the game in which the candidates move first has a subgame perfect equilibrium with the outcome $((x_1^*, x_2^*), S^*)$.

13 Two-period electoral competition with imperfect information

13.1 Selecting and controlling incumbents with unknown preferences 435
13.2 Inducing an incumbent to exert effort 441
13.3 Candidates who pander to voters 443

Suppose that after serving a term in office, an incumbent politician stands for reelection. How does the possibility that she is reelected affect the policy she chooses in her first term? Does it induce her to choose a policy well aligned with the citizens' preferences? Or does it lead different types of incumbent to choose different policies, allowing citizens to reelect only ones who will implement policies favorable to them in the future? Could it lead a well-informed incumbent to choose a policy that the imperfectly-informed citizens believe is good for them even though it is not, as the incumbent knows?

Synopsis

The models I present are among the simplest that capture the main ideas that have been studied. They have two periods, a single citizen, and two possible policies, x and y. In the first period, the incumbent chooses a policy. The citizen has limited information: she may not know the incumbent's preferences; she may observe only the outcome of the policy chosen by the incumbent, not the policy itself; and she may not know the policy that is best for her until the second period. She either reelects the incumbent for a second term or elects a challenger. The second period is the last, so the officeholder in that period chooses the policy that maximizes her payoff in that period.

In the model in Section 13.1, the citizen prefers x to y. The incumbent, together with the second-period challenger, is drawn from a pool of candidates in which the proportion π have the same preferences as the citizen (their preferences are *consonant* with citizen's) and the proportion $1-\pi$ prefer y to x (their preferences are *dissonant* from the citizen's). The citizen does not know the incumbent's type. In each period in which a candidate is in office, she receives the payoff v' if the policy is the one she favors and v if it is the other policy. She receives an additional payoff of $b > 0$ if she is the incumbent in the first period and is reelected in the second period. Her total payoff is the sum of her payoffs in the two periods. The model, like the ones in the subsequent sections, is an extensive game with imperfect information. I apply to these models the notion of weak sequential equilibrium. (All of the equilibria I discuss satisfy the additional conditions imposed by the notion of (full) sequential equilibrium.)

If the benefit to the incumbent of holding office in the second period is large, the prospect of not being reelected induces behavior favorable to the citizen in the first period: the game has an equilibrium in which all incumbents choose x, the policy preferred by the citizen, in the first period. The cost of achieving this good outcome in the first period is that all incumbents, consonant and dissonant, are reelected, and dissonant ones choose y in the second period. In an extension of the model to many periods in which the future is accorded sufficient weight, the disadvantage of reelecting a dissonant candidate is absent, because they face the same incentive to choose x in the future as they do in the present. In this case, the game has an equilibrium in which the outcome is the best one possible for the citizen: x in every period.

If the benefit to the incumbent of holding office in the second period is small, the game has an equilibrium in which a consonant incumbent chooses x but a dissonant one chooses y. The first-period outcome is not so good for the citizen, but the incumbent's behavior allows her to reelect only consonant incumbents, improving the outcome in the second period.

In the models in Section 13.2, all incumbents are dissonant. (Perhaps they are drawn from a social class different from that of the citizen.) Generating a good outcome for the citizen requires the incumbent to expend costly effort. The citizen observes only the outcome, which depends both on the effort and random factors out of the incumbent's control. A higher bar for reelection may induce an incumbent to exert more effort, improving the first-period outcome for the citizen. But it may alternatively make it not worthwhile for the incumbent to try to reach the bar, so that she reduces her effort.

In the model in Section 13.3, the citizen is uncertain not only about the incumbent's preferences but also about the policy that is best for her, which depends on a state known to the incumbent but not to her. Proposition 13.3 shows that a simple model has a unique equilibrium, in which some incumbents choose the policy that the citizen believes is most likely to be best for her, even if the incumbent knows that another policy is in fact best. As a consequence, under some conditions the mechanism in which a policy is chosen by a randomly selected citizen rather than an elected representative is better for the citizen.

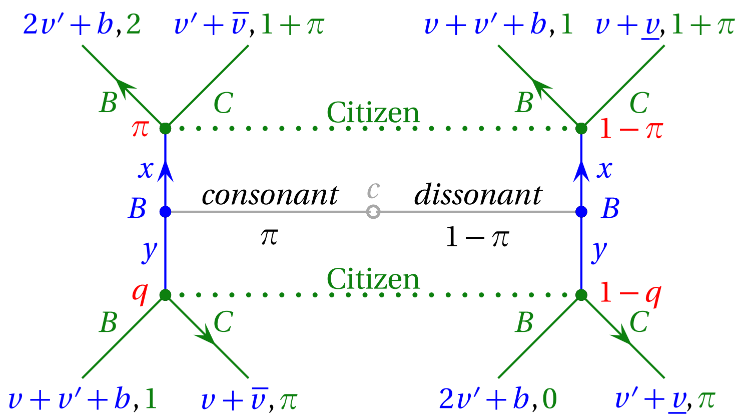


Figure 13.1 A two-period electoral competition game with unobserved incumbent preferences. The game begins at the small circle in the center of the figure, with a move of chance. The incumbent's payoff is listed first, and $\overline{v} = \pi v' + (1-\pi)v$ and $\underline{v} = (1-\pi)v' + \pi v$. If $b \geq (1-\pi)(v'-v)$ then for any $q \leq \pi$ the assessment for which the strategy profile is indicated by the arrows and the belief system is given by the numbers in red is a weak sequential equilibrium.

13.1 Selecting and controlling incumbents with unknown preferences

The citizen prefers x to y. The incumbent, together with the second-period challenger if the incumbent is not reelected, is drawn from a pool of candidates in which the proportion π have the same preferences as the citizen (the candidate's type is *consonant*) and the proportion $1-\pi$ prefer y to x (the candidate's type is *dissonant*). The citizen observes the policy chosen by the incumbent, but not her preferences. In each period in which a candidate is in office, she receives the payoff v' if the policy is the one she favors and v if it is the other policy. She receives an additional payoff of $b > 0$ if she is the incumbent in the first period and is reelected in the second period. A candidate's total payoff is the sum of her payoffs in the two periods.

This situation may be modeled as the extensive game with imperfect information in Figure 13.1, in which $\overline{v} = \pi v' + (1-\pi)v$ and $\underline{v} = \pi v + (1-\pi)v'$. (Ignore for the moment the arrows and numbers in red.) The players are the citizen and the incumbent (B). The game begins with a move of chance, which determines whether the incumbent's type is consonant or dissonant. The incumbent then chooses a policy, x or y. The citizen observes this policy but not the incumbent's type, and either reelects the incumbent (B) or elects the challenger (C). The second period is the last, so whoever is elected chooses her favorite policy; this choice is not included as an action in the game, but is reflected in the payoffs.

Definition 13.1: Two-period electoral competition game with unobserved incumbent preferences

A *two-period electoral competition game with unobserved incumbent preferences* $\langle \pi, v, v', b \rangle$, where $\pi \in (0,1)$, $0 < v < v'$, and $b \geq 0$ is the extensive game with imperfect information shown in Figure 13.1, where $\underline{v} = \pi v + (1-\pi)v'$ and $\overline{v} = \pi v' + (1-\pi)v$.

Note that an incumbent who is induced by the carrot of reelection to select x in the first period when, in the absence of the incentive, she would not do so, chooses y in the second period, which the citizen does not like. Thus if the threat to not reelect an incumbent who chooses y is effective in inducing all incumbents to choose x, making the first-period outcome better for the citizen, the second-period outcome is poor for the citizen if the incumbent is dissonant. Conversely, if the citizen's strategy is successful in selecting only consonant incumbents for reelection, then consonant and dissonant incumbents choose different policies in the first period, generating a poor outcome for the citizen in the first period if the incumbent is dissonant. Specifically, if dissonant incumbents are induced to choose x in the first period, like consonant incumbents, and are thus reelected, the citizen's expected payoff is v' in the first period and $\pi v' + (1-\pi)v$ in the second period, whereas if consonant incumbents choose x and dissonant ones choose y in the first period, the citizen's payoffs are $\pi v' + (1-\pi)v < v'$ in the first period and $\pi v' + (1-\pi)(\pi v' + (1-\pi)v) > \pi v' + (1-\pi)v$ in the second period.

If the benefit b of holding office is sufficiently large, the game has a weak sequential equilibrium in which both types of incumbent choose x and the citizen reelects an incumbent if and only if she chooses x. In such an equilibrium, the belief system assigns probability π to the history (*consonant*, x) at the upper information set and some probability $q \leq \pi$ to the history (*consonant*, y) at the lower information set, as illustrated in Figure 13.2. This belief system is weakly consistent with the strategy profile: given the strategy profile, the history (*consonant*, x) occurs with probability π and the history (*dissonant*, x) occurs with probability $1-\pi$, and the bottom information set is not reached, so that weak consistency imposes no constraint on the beliefs there. At the top information set the citizen is indifferent between B and C, because observing the policy x conveys no information about the incumbent's type, and at the bottom information set she is no better off choosing B than C, given that $q \leq \pi$. A consonant incumbent decreases her payoff if she deviates to y and a dissonant one does not benefit from deviating to y if $b \geq (1-\pi)(v'-v)$.

If $b \leq (1-\pi)(v'-v)$ then the game has an equilibrium in which a consonant

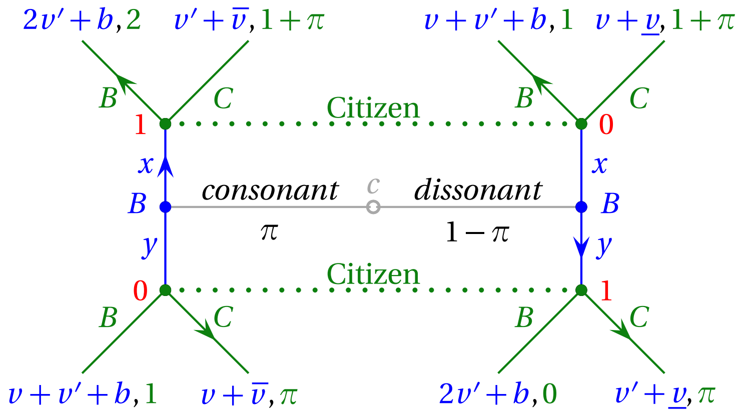


Figure 13.2 A weak sequential equilibrium of a two-period electoral competition game with unobserved incumbent preferences in which both types of incumbent choose the policy x.

incumbent chooses x, a dissonant incumbent chooses y, and the citizen reelects an incumbent if and only if she chooses x. In this equilibrium, the outcome is the same as it would be if the citizen knew the incumbent's type. An assessment with this strategy profile is shown in Figure 13.1. Given the belief system, the citizen's strategy is optimal: she prefers B to C at her top information set and C to B at her bottom information set. Also, given the citizen's strategy, the action of a consonant incumbent is optimal because $2v' + b > v + \overline{v}$, and the action of a dissonant incumbent is optimal if $v' + \underline{v} \geq v + v' + b$, or $b \leq (1 - \pi)(v' - v)$. Thus the assessment is a weak sequential equilibrium if this inequality is satisfied: if the gain from holding office in the second period is small relative to the difference between the incumbent's evaluations of the two policies. The assessment is not a weak sequential equilibrium if b is greater than $(1 - \pi)(v' - v)$ because then a dissonant incumbent gains by deviating to x in the first period, and thus getting reelected in the second period. This maneuver hurts her in the first period, when she implements x rather than her favored policy y, but helps her in the second period, when she implements y rather than enduring the favored policy of a randomly-determined challenger.

The game has no weak sequential equilibrium in which a consonant incumbent chooses y and a dissonant one chooses x. In any assessment with these strategies the weak consistency of the belief system with the strategy profile requires the belief at the top information set to assign probability 1 to the history $(dissonant, x)$ and the belief at the bottom information set to assign probability 1 to the history $(consonant, y)$, so that the citizen optimally reelects an incumbent if and only if she chooses y. But then a dissonant incumbent gains by deviating to y (and getting reelected).

When b is large, the game does, however, have a weak sequential equilibrium in which both types of incumbent choose y, the citizen reelects an incumbent if and only if she chooses y, and the belief system assigns probability π to the history (*consonant*, y) at the bottom information set (as required by the weak consistency of the beliefs with the strategies) and at most π to the history (*consonant*, x) at the top information set. The citizen is indifferent between B and C at the bottom information set and does not benefit from deviating to B at the top information set. A consonant incumbent does not benefit from deviating to x if $v + v' + b \geq v' + \overline{v}$, or $b \geq \pi(v' - v)$, and a dissonant incumbent is worse off deviating to x.

In this equilibrium, if the citizen observes the policy x then she believes that the incumbent is more likely to be dissonant than she was initially. A history in which the policy x is chosen does not happen if the incumbent adheres to her strategy, and hence the citizen has no basis on which to form a belief about the incumbent's type if she observes x. Thus the belief that incumbent is more likely to be dissonant than she was originally is weakly consistent with the strategy profile. Is it reasonable? Suppose that we interpret the equilibrium as a steady state: in each of a long sequence of periods, a citizen and an incumbent have been randomly selected from large populations, and in every period the selected incumbent has chosen y and has been reelected. Now, in the current period, the selected citizen observes the policy x. The citizen might conclude that incumbent has made a mistake, or that something has changed—perhaps the incumbent's options, or her payoffs. Depending on the cause of the mistake or the nature of the change in the game, the citizen's belief that the probability that the incumbent is consonant is $q \leq \pi$ could be reasonable.

A different approach, initiated by Cho and Kreps (1987), considers the implications for the citizen's belief of the choice of the policy x being deliberate. In this case, the citizen might reason that the incumbent must be consonant: deviating to x holds no possible advantage for a dissonant incumbent, since y gets her reelected and thus leads to the best possible outcome for her, but *is* advantageous for a consonant incumbent if it results in the citizen reelecting her. And if a deviation to x leads the citizen to believe that the incumbent is consonant, then she should reelect her, making a deviation by a consonant incumbent worthwhile and thus destabilizing the equilibrium. A large body of work following Cho and Kreps (1987) explores the implications of arguments of this type for the stability of weak sequential equilibria.

My arguments about the weak sequential equilibria of the game are summarized in the next result.

Proposition 13.1: Weak sequential equilibria of two-period electoral competition game with unobserved incumbent preferences

Let $\langle \pi, v, v', b \rangle$ be a two-period electoral competition game with unobserved incumbent preferences. Each weak sequential equilibrium of this game has one of the following three forms.

a. (Pooling on x) If $b \geq (1-\pi)(v'-v)$ then both types of incumbent choose x, the citizen reelects an incumbent who chooses x and not one who chooses y, and the belief system assigns probability π to the history $(consonant, x)$ at the information set reached when the incumbent chooses x and, for some $q \leq \pi$, probability q to the history $(consonant, y)$ at the information set reached when the incumbent chooses y.

b. (Separating) If $b \leq (1-\pi)(v'-v)$ then a consonant incumbent chooses x and a dissonant one chooses y, the citizen reelects an incumbent who chooses x and not one who chooses y, and the belief system assigns probability 1 to the history $(consonant, x)$ at the information set reached when the incumbent chooses x and probability 1 to the history $(dissonant, y)$ at the information set reached when the incumbent chooses y.

c. (Pooling on y) If $b \geq \pi(v'-v)$ then both types of incumbent choose y, the citizen reelects an incumbent who chooses y and not one who chooses x, and for some $q \leq \pi$ the belief system assigns probability q to the history $(consonant, x)$ at the information set reached when the incumbent chooses x and probability π to the history $(consonant, y)$ at the information set reached when the incumbent chooses y.

The citizen's expected payoff in the equilibria in parts a and b of this result, as a function of π, are shown in Figure 13.5. This payoff is increasing in π except at the point at which the pooling equilibrium replaces the separating equilibrium.

A long horizon

In the two-period model, if the citizen induces a dissonant incumbent to choose x in the first period then the outcome in the second period is undesirable: the dissonant incumbent, with no prospect of further reelection, chooses y. If the game lasts for more than two periods, the citizen may be able to generate a bet-

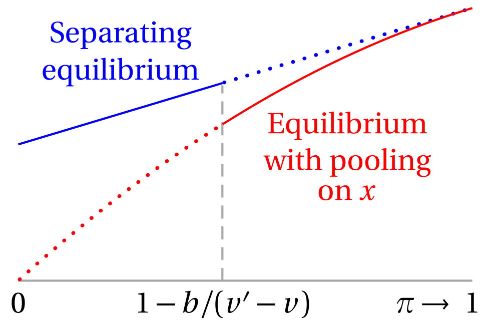


Figure 13.3 The expected payoff of the citizen in a separating equilibrium of a two-period electoral competition game with unobserved incumbent preferences, which exists if $\pi \le 1-b/(v'-v)$, and an equilibrium with pooling on x, which exists if $\pi \ge 1-b/(v'-v)$.

ter outcome. Suppose, at the other extreme, that the horizon is infinite. In each period, the incumbent from the previous period competes with a challenger randomly chosen from a large set of candidates, the fraction π of which are consonant, and the remainder dissonant. The set of candidates is large enough that the probability that an incumbent who is not reelected will subsequently be the challenger is negligible; I take it to be zero. A candidate obtains the payoff v' in every period in which the outcome is the one she favors (x if she is consonant, y if she is dissonant) and v in every period in which it is the other outcome, plus $b > 0$ in each period (including the first) in which she is the incumbent. The citizen obtains the payoff 1 in every period in which the outcome is x and 0 in every period in which it is y. Every player has the same discount factor, $\delta \in (0,1)$, and her total payoff is the discounted average ($1-\delta$ times the discounted sum) of her payoffs in each period.

Suppose that in each period the citizen reelects an incumbent if and only if she chose the policy x in the previous period, and every incumbent, consonant or dissonant, chooses policy x in every period regardless of history. A dissonant incumbent then obtains $v+b$ in every period, and hence a discounted average payoff of $v+b$. If she deviates to y in any period, she is better off in that period, obtaining the payoff $v'+b$, and worse off in every subsequent that period, obtaining v. Her discounted average payoff is $(1-\delta)(v'+b)+\delta v$. This payoff is less than $v+b$ if $v'-v$ is small, b is large, and/or δ is close to 1. Thus in these cases, the game has a subgame perfect equilibrium in which the citizen obtains the best possible outcome: x in every period. Inducing an incumbent to choose x by not reelecting her if she chooses y does not doom the citizen to y in the future if the incumbent is dissonant, as it does in the two-period model, because in the future the incumbent will face the same incentives that she faces today. (A term limit would weaken this incentive.)

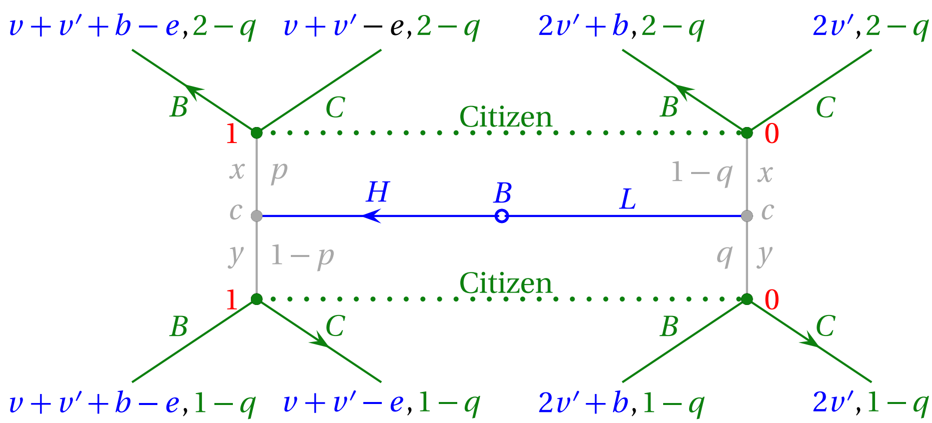


Figure 13.4 A two-period electoral competition game with unobserved incumbent actions. The game begins at the small circle in the center of the figure, with the choice of x or y by the incumbent. If $(p-(1-q))b \geq v'-v+e$ then the assessment for which the strategy profile is indicated by the arrows and the belief system is given by the numbers in red is a weak sequential equilibrium.

13.2 Inducing an incumbent to exert effort

Uncertainty about the incumbent's motivation is only one of the possible limitations of the citizen's information. The citizen may not observe the incumbent's action, but rather the outcome of the action, which is related probabilistically to the action. In this section I briefly discuss models that focus on this limitation in the citizen's information. In these models, the citizen knows the candidates' motivations.

In the first model, an incumbent interacts with a citizen as shown in Figure 13.4. The incumbent takes one of two actions, H and L, which may be interpreted as the effort she expends. Then chance determines the outcome, which is either x or y; it is x with probability $p > \frac{1}{2}$ if the action is H and with probability $1-q < \frac{1}{2}$ if the action is L. The citizen observes the outcome, but not the action, and either reelects the incumbent or elects a challenger. In both cases, the officeholder chooses L in the second period and the outcome is x with probability $1-q$ and y with probability q.

The citizen's payoff depends on the outcomes in the two periods. She prefers x to y; her payoff is 1 in each period in which the outcome is x, and 0 in each period in which it is y, and her total payoff is the sum of her payoffs in the two periods. The incumbent's payoff depends on both her action and the outcome. She prefers L to H and, unlike the citizen, y to x. Specifically, her payoff in each period is v' if the outcome is y and $v < v'$ if it is x, minus $e > 0$ if her action is H and plus $b \geq 0$ if she is reelected. Given that she prefers y to x and L is more likely to yield y, she chooses L if she is reelected. For the same reason, a challenger also chooses L if elected.

Definition 13.2: Two-period electoral competition game with unobserved incumbent actions

A *two-period electoral competition game with unobserved incumbent actions* $\langle p, q, v, v', b, e\rangle$, where $p \in (\frac{1}{2}, 1)$, $q \in (\frac{1}{2}, 1)$, $0 < v < v'$, $b \geq 0$, and $e \geq 0$ is the extensive game with imperfect information shown in Figure 13.4.

By reelecting the incumbent only if the outcome is x, can the citizen induce the incumbent to choose H? The game has a weak sequential equilibrium with such a result if the benefit b to holding office is high enough relative to $v' - v + e$. This equilibrium is indicated in Figure 13.4. Note that at each of her information sets, the citizen is indifferent between reelecting the incumbent and electing the challenger. Thus both actions are optimal for her at each information set, regardless of her belief. Hence to check whether an assessment is a weak sequential equilibrium we need to check only the optimality of the incumbent's action. If the citizen reelects the incumbent only if the outcome is x then the incumbent's payoff to H is at least her payoff to L if and only if $(p - (1-q))b \geq v' - v + e$. Otherwise the incumbent's payoff to L is greater than her payoff to H. Thus the weak sequential equilibria of the game are given as follows.

Proposition 13.2: Weak sequential equilibria of a two-period electoral competition game with unobserved incumbent actions

Let $\langle p, q, v, v', b\rangle$ be a two-period electoral competition game with unobserved incumbent actions. Each weak sequential equilibrium of this game has one of the following forms.

a. If $(p - (1-q))b \geq v' - v + e$ then the incumbent chooses H, the citizen reelects an incumbent if the outcome is x and not if it is y, and the belief system assigns probability 1 to the history (H, x) at the information set reached when the outcome is x and probability 1 to the history (H, y) at the information set reached when the outcome is y.

b. The incumbent chooses L, the citizen (*i*) reelects the incumbent regardless of the outcome, or (*ii*) elects the challenger regardless of the outcome, or (*iii*) reelects the incumbent only if the outcome is y, or, if $(p-(1-q))b \leq v'-v+e$, (*iv*) reelects the incumbent only if the outcome is x, and the belief system assigns probability 1 to the history (L, x) at the information set reached when the outcome is x and probability 1 to the history (L, y) at the information set reached when the outcome is y.

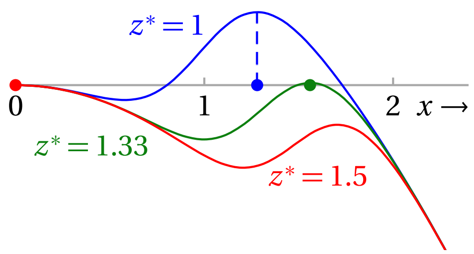
If the citizen raises the bar for reelection from y (that is, no bar at all) to x, then the incumbent's action either remains L, if the cost of meeting the bar outweighs the benefit, or changes from L to H. In a model with a richer set of actions, the tradeoff the citizen faces when choosing the reelection bar is smoother.

Suppose that the incumbent chooses an action (effort level) $x \in \mathbb{R}$. The outcome $z \in \mathbb{R}$ depends probabilistically on x; denote the probability that the outcome is at most z when the incumbent's effort is x by $F(z, x)$. For each value of z, $F(z, x)$ is decreasing in x: more effort makes higher outcomes more likely. The citizen observes z and either reelects the incumbent or elects the challenger. The incumbent likes to be in office and dislikes effort. For simplicity, assume that her payoff does not depend on the policy chosen by the officeholder (unlike the payoff of the incumbent in the previous models in this chapter). If she is reelected her payoff is $b - c(x)$, and if she is not reelected it is $-c(x)$, where c is an increasing function.

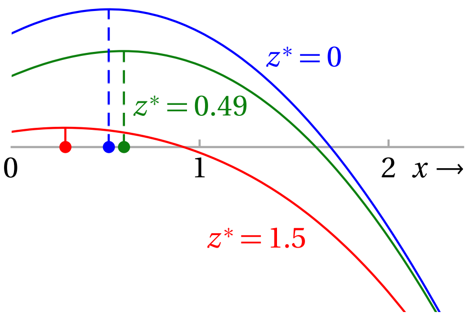
Assume that for some number z^*, the citizen reelects the incumbent if the outcome is at least z^*. Then the expected payoff of an incumbent as a function of her effort x is $(1 - F(z^*, x))b - c(x)$. Examples are shown in Figure 13.5 for the case in which $F(\cdot, x)$ is a normal distribution with mean x, $b = 3$, and $c(x) = x^2$ for all x. Increasing the reelection standard z^* from 0 initially induces the incumbent to exert more effort, but at some point the cost of exerting enough effort that the outcome is likely to meet the standard overwhelms the benefit of reelection, and the incumbent optimally reduces her effort. In the examples, she reduces it to zero if the variance of the distribution of outcomes is small, and reduces it gradually if this variance is large. In these examples the citizen is better off when the outcome is a more precise signal of the incumbent's effort.

13.3 Candidates who pander to voters

A citizen may be uncertain not only of the candidates' preferences and actions, but also of the policy that is best for her. We can model this latter uncertainty by assuming that the policy best for the citizen depends on a state that is known to the candidates but not the citizen. The candidates' preferences may differ from those of the citizen, as in the model in Section 13.1, but the citizen, unlike the one in the model in Section 13.3, observes the policy chosen by the officeholder. In this environment, electing a representative to choose a policy may not



(a) Variance of distribution of outcomes given effort: 0.0625. Reelection cutoff that induces most effort: $z^* = 1.33$.



(b) Variance of distribution of outcomes given effort: 1. Reelection cutoff that induces most effort: $z^* = 0.49$.

Figure 13.5 The payoff of an incumbent as a function of her effort x for various cutoffs z^* for reelection in two examples of a two-period electoral competition game with continua of possible actions and outcomes. In each example the distribution of outcomes when the incumbent's effort is x is a normal distribution with mean x, $b = 3$, and $c(x) = x^2$ for all x. The payoff-maximizing effort levels are indicated by small disks.

be optimal for the citizen. On the one hand, given that the candidates are better informed than the citizen, having an elected representative choose a policy makes it possible that the chosen policy is better than the one the citizen would choose. On the other hand, the competition to be elected may induce a candidate to select a policy that she knows the citizen believes is most likely to be optimal, even though, as the candidates know, this policy is not best for the citizen. Depending on the balance of these effects, the citizen may be better off choosing a policy herself or selecting a candidate randomly in each period than electing a representative on the basis the policies she chose as an incumbent.

Model

As previously, there are two periods, the participants are an incumbent, a challenger, and a citizen, and there are two possible policies, x and y. In the first period, the incumbent chooses a policy. The citizen observes this policy and either reelects the incumbent or elects the challenger. In the second period, the officeholder chooses a policy.

There are two possible states, called x and y like the policies. The policy best for the citizen depends on the state: policy x is best in state x and policy y is best in state y. The incumbent knows the state, and hence the policy that is best for the citizen, but the citizen does not; she believes that the state is x with probability p and y with probability $1 - p$. With probability π, a candidate (the incumbent or the challenger) has preferences that are consonant with the

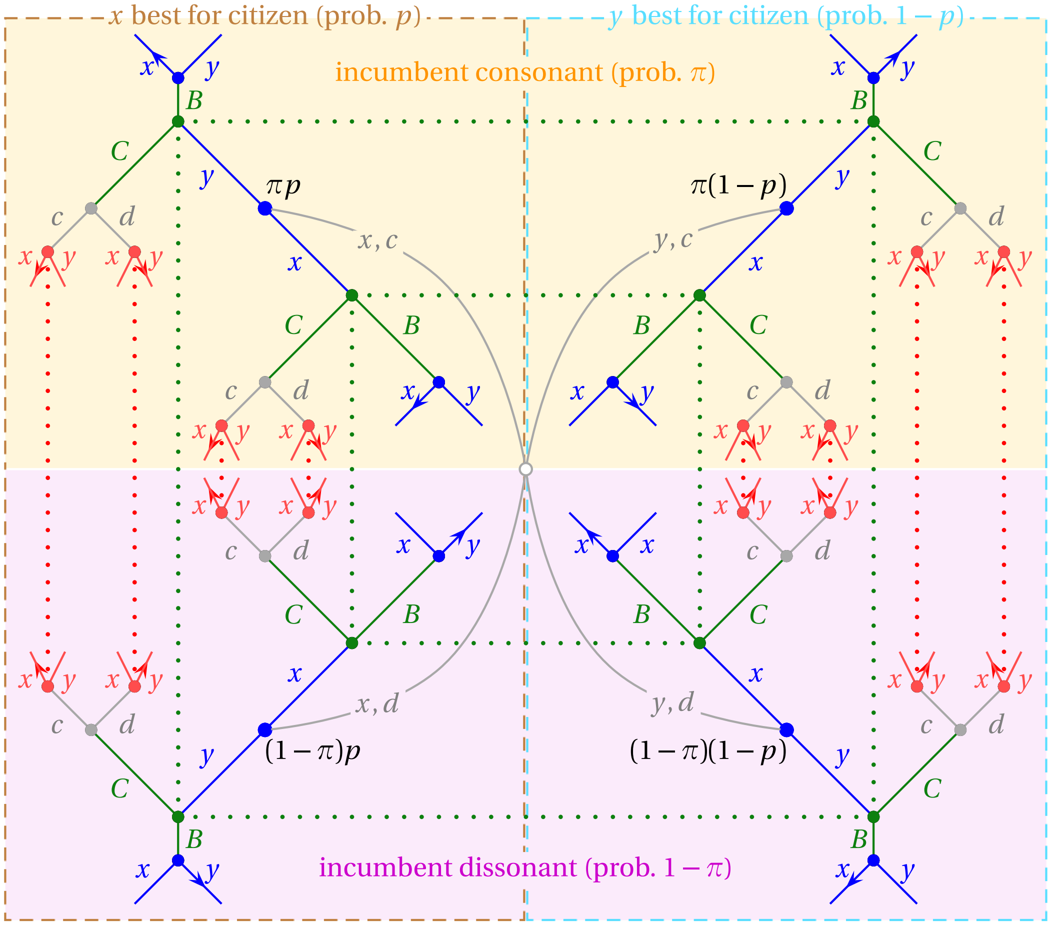


Figure 13.6 The structure of a two-period game of electoral competition in which the candidates, but not the citizen, know the best policy for the citizen. The game begins with the move of chance indicated by the small circle in the center. Blue elements belong to the incumbent, red elements to the challenger, green elements to the citizen, and gray elements to chance.

citizen's—she prefers x in state x and y in state y—and with probability $1-\pi$ she has preferences that are dissonant from the citizen's—she prefers y in state x and x in state y.

The structure of an extensive game with imperfect information that reflects these assumptions is shown in Figure 13.6. Play begins with the move of chance indicated by the small circle in the center of the figure. Chance independently determines the state (x with probability p, y with probability $1-p$) and the incumbent's preferences (consonant with the citizen's with probability π, dissonant from the citizen's with probability $1-\pi$). The branch labeled x, c, for example, corresponds to state x and preferences for the incumbent that are consonant with the citizen's, an event with probability πp. The incumbent observes the move of chance and chooses a policy (x or y) for period 1. The citizen observes the policy, but not the incumbent's type or the state, and thus has two information sets, indicated by the dotted squares. The inner information set is reached if the incumbent chooses policy x, and the outer one is reached if the incumbent chooses policy y. At each information set, the citizen selects the incumbent (B) or the challenger (C). If she selects the challenger, chance determines the challenger's type (c (consonant) with probability π and d (dissonant) with probability $1-\pi$). Finally, the candidate the citizen selected chooses a policy (x or y) for period 2. (The state remains the same in period 2 as it was in period 1.)

(An alternative and perhaps more natural specification assumes that the challenger's type, like the incumbent's, is chosen by chance initially, rather than after the citizen selects the challenger. This specification is more complicated to depict, but its analysis is the same. The important point is that the citizen does not know the challenger's type when she selects a candidate, which is true in both formulations.)

The citizen's payoff is the number of periods (0, 1, or 2) in which the policy matches the state. She does not receive this payoff until the game ends, and in particular cannot use the part due to the policy in period 1 to infer the state. (If the state were chosen independently in each period, we could allow the citizen to observe her first-period payoff before period 2. The game would then be more complicated, but the results would be the same.)

A candidate (incumbent or challenger) receives the payoff v' in each period in which the policy is the one she prefers and v in each period in which it is the other policy, with $v' > v > 0$. In addition, if she is in office in the second period then she receives an additional payoff $b \geq 0$, which reflects her degree of office-motivation. At the moment I assume that all candidates have the same degree of office-motivation; later I assume that they may differ in this degree. Thus, for example, if the state is x, the incumbent's type is c, the incumbent chooses policy x in the first period, is reelected, and chooses policy x again in the second period, then her payoff is $2v' + b$. For a history that differs only in that the challenger is elected in the second period and chooses the policy y, the incumbent's payoff is $v' + v$.

In any weak sequential equilibrium of the game, the officeholder in period 2 chooses her favorite action, as indicated by the arrows in Figure 13.6. Thus we may find the weak sequential equilibria of the game by finding the weak sequential equilibria of the game in Figure 13.7, in which the players are the incumbent and the citizen (in addition to chance) and the part of each terminal history in period 2 in Figure 13.6 is replaced by the payoffs of the incumbent and the citizen when the second-period policy-maker chooses her favorite action. Subsequently

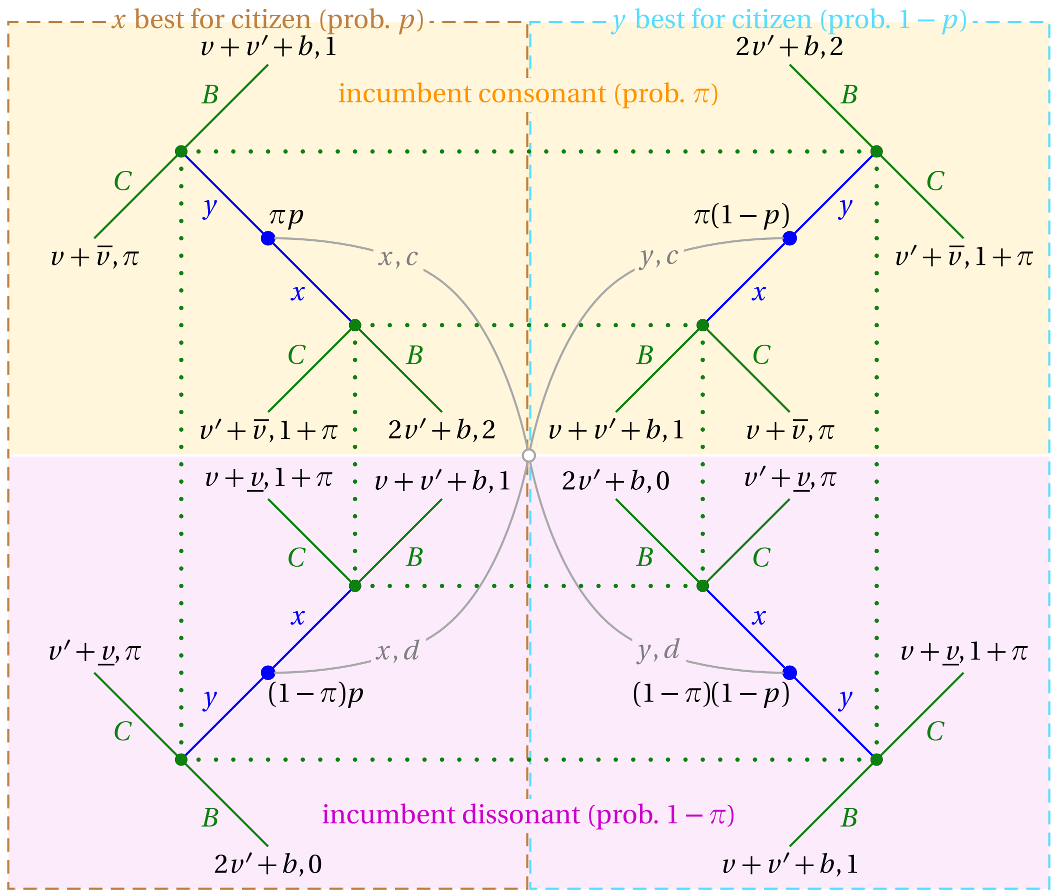


Figure 13.7 The game in Figure 13.6 with the second period replaced by the payoffs to the players' unique optimal actions in that period. The game begins with the move of chance indicated by the small circle in the center. The players are an incumbent (blue) and a citizen (green). The incumbent's payoff is listed first, and $\overline{v} = \pi v' + (1-\pi)v$ and $\underline{v} = (1-\pi)v' + \pi v$.

I work with a variant of this reduced game.

In the variant, the incumbent has many possible degrees of office-motivation. For an incumbent to be willing to implement her less-preferred policy to get reelected, her office-motivation has to be sufficient that she prefers to endure her less-preferred policy in the first period and be reelected than to choose her favored policy in the first period and put up with the policy chosen by a random challenger in the second period. But if the office-motivation of every incumbent has such a magnitude, then in an equilibrium all incumbents choose the same policy, which has the unfortunate implication that the citizen's information set corresponding to the other policy is not reached. Consequently any belief regarding the history that led to this information set is consistent with the equilibrium strategy profile, leading to a multiplicity of equilibria. The model I specify avoids this problem by allowing the incumbent's degree of office-motivation to take any value in a finite set $\mathscr{B}$. Thus in this model, the incumbent's type is a pair (h, b), where h is the concordance of her preferences with those of the citizen (consonant or dissonant) and b is her degree of office-motivation. The subsequent result (Proposition 13.3) assumes that $\mathscr{B}$ contains two values, one high and one low. The optimal behavior of the incumbents with low office-motivation imposes discipline on the citizen's belief at the information set following the policy that is not chosen by any high-office-motivation incumbent.

In the model, chance starts by determining a state (x or y), the concordance of the incumbent's preferences with those of the citizens (*consonant* or *dissonant*), and the incumbent's degree of office-motivation (a member of $\mathscr{B}$). The incumbent observes these values and chooses a policy (x or y). Then the citizen, who observes the incumbent's policy but not the move of chance, selects either the incumbent or the challenger, who chooses her favorite policy in the second period. An example in which $\mathscr{B} = \{0, b\}$ is given in Figure 13.8.

Definition 13.3: Two-period electoral competition game with well-informed candidates with uncertain motivations

A *two-period electoral competition game with well-informed candidates with uncertain motivations* $\langle \{B, V, C\}, \{x, y\}, p, \pi, v', v, \mathscr{B}, \rho \rangle$, where

- B, V, and C are labels (B and V are the names of an incumbent and a citizen, and C represents the citizen's action of selecting a challenger rather than the incumbent)
- x and y are labels (the names both of two policies and two states)
- $p \in (0, 1)$ (the probability the state is x)
- $\pi \in (0, 1)$ (the probability that a candidate's preferences are the same as (consonant with) the citizen's preferences)
- $v' \geq 0$ (the component of the incumbent's payoff attributable to the policy in a period in which the policy is her favorite policy)
- $v \geq 0$ with $v < v'$ (the component of the incumbent's payoff attributable to the policy in a period in which the policy is not her favorite policy)

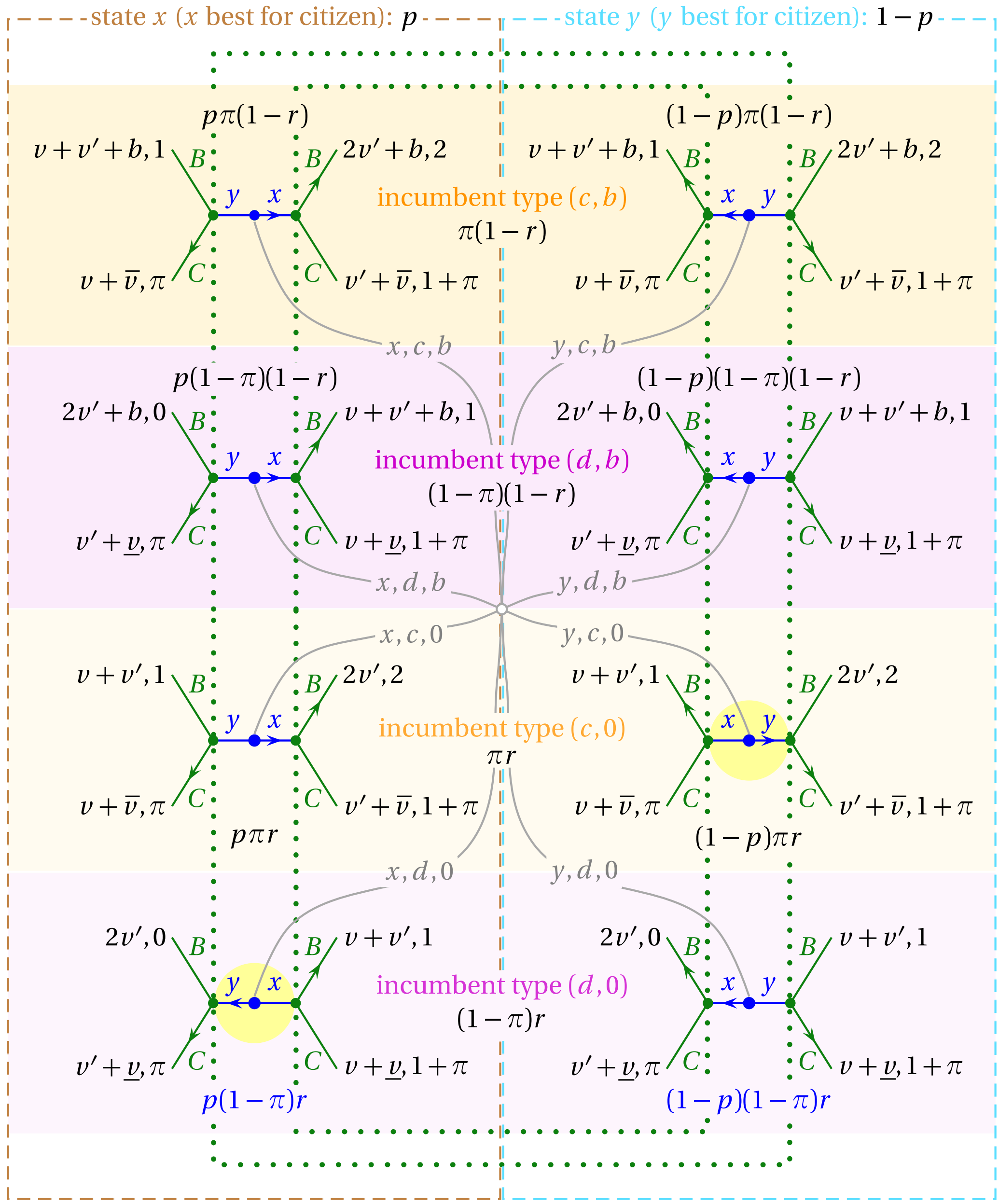


Figure 13.8 A two-period electoral competition game with well-informed candidates with uncertain motivations in which incumbents have two possible degrees of office–motivation, 0 (lower half of figure) and b (upper half of figure). Blue elements belong to the incumbent, green elements to the citizen, and gray elements to chance; probabilities are indicated in black. The incumbent's payoff is listed first; $\overline{v} = \pi v' + (1-\pi)v$ and $\underline{v} = (1-\pi)v' + \pi v$. The arrows indicate the actions chosen in any weak sequential equilibrium when $p > \frac{1}{2}$ and $v + b > \max\{\underline{v}, \overline{v}\}$, as given in Proposition 13.3. The cases in which the incumbent chooses y are highlighted.

- $\mathscr{B}$ is a finite set of nonnegative numbers (the possible values for the component of the incumbent's payoff attributable to her holding office)
- ρ is a probability distribution over $\mathscr{B}$ (the probabilities of the various degrees of office-motivation)

is an extensive game with imperfect information with the following components.

Players
B (an *incumbent*) and V (a *citizen*).

Terminal histories
The terminal histories are the sequences $((s,h,\beta),z,D)$ for $(s,h,\beta) \in \{x,y\} \times \{c,d\} \times \mathscr{B}$, $z \in \{x,y\}$, and $D \in \{B,C\}$

Player function
The player function P is defined by

- $P(\varnothing) = chance$ (the game begins with a move of chance)
- $P(s,h,\beta) = B$ for each $(s,h,\beta) \in \{x,y\} \times \{c,d\} \times \mathscr{B}$ (the incumbent moves after chance determines the state and the incumbent's type)
- $P((s,h,\beta),z) = V$ for every $(s,h,\beta,z) \in \{x,y\} \times \{c,d\} \times \mathscr{B} \times \{x,y\}$ (the citizen moves after the incumbent chooses a policy).

Chance probabilities
At the initial history $\varnothing$ chance selects the state $s \in \{x,y\}$ (the best policy for the citizen), the concordance of the incumbent's preferences with the citizen's preferences (consonant (c) or dissonant (d)), and the incumbent's degree of office-motivation $\beta \in \mathscr{B}$, independently. The state s is x with probability p and y with probability $1-p$, the incumbent's preferences are consonant (c) with the citizen's with probability π and dissonant (d) from them with probability $1-\pi$, and the incumbent's degree of office-motivation is β with probability $\rho(\beta)$.

Information partitions
Player B's information partition consists of one set for each move of chance. Player V's information partition consists of the following two sets:

$$\{(s,h,\beta,x) : s \in \{x,y\}, h \in \{c,d\}, \beta \in \mathscr{B}\}$$
$$\{(s,h,\beta,y) : s \in \{x,y\}, h \in \{c,d\}, \beta \in \mathscr{B}\}.$$

Preferences

The preferences of each player over the set of lotteries over terminal histories are represented by the expected value of the following payoffs. For the terminal history $((s,h,\beta),z,D)$, the payoff of player B is the sum of two components: an amount attributable to the first-period policy,

$$\begin{cases} v' & \text{if } h=c \text{ and } s=z, \text{ or } h=d \text{ and } s\neq z \\ v & \text{if } h=c \text{ and } s\neq z, \text{ or } h=d \text{ and } s=z, \end{cases}$$

and an amount attributable to the second-period policy,

$$\begin{cases} v'+\beta & \text{if } D=B \text{ (citizen selects incumbent)} \\ \pi v'+(1-\pi)v & \text{if } D=C \text{ and } h=c \text{ (citizen selects challenger,} \\ & \qquad\qquad\qquad \text{incumbent is consonant)} \\ (1-\pi)v'+\pi v & \text{if } D=C \text{ and } h=d \text{ (citizen selects challenger,} \\ & \qquad\qquad\qquad \text{incumbent is dissonant).} \end{cases}$$

If player V, the citizen, selects the incumbent ($D=B$) her payoff is

$$\begin{cases} 0 & \text{if } z\neq s \text{ and } h=d \\ 1 & \text{if } z=s \text{ and } h=d, \text{ or } z\neq s \text{ and } h=c \\ 2 & \text{if } z=s \text{ and } h=c \end{cases}$$

and if she selects the challenger ($D=C$) it is

$$\begin{cases} \pi & \text{if } z\neq s \\ 1+\pi & \text{if } z=s. \end{cases}$$

A strategy for the incumbent is a policy for each state and each of her types (h,β) (that is, for each move of chance). A strategy for the citizen is a function that assigns to each policy (x and y) either B (reelect incumbent) or C (elect challenger). A belief system assigns to each of the citizen's information sets a probability distribution over the histories in the set.

I now study the weak sequential equilibria of the game in which the incumbent has two possible degrees of office-motivation, 0 and b, with $v+b>\max\{\underline{v},\overline{v}\}$, where

$$\overline{v}=\pi v'+(1-\pi)v \quad \text{and} \quad \underline{v}=(1-\pi)v'+\pi v. \tag{13.1}$$

First consider an incumbent whose degree of office-motivation is 0 (types $(c,0)$ and $(d,0)$); her actions appear in the bottom two panels of Figure 13.8. For such an incumbent, choosing her favorite policy is unambiguously optimal. If she chooses the other policy, she sacrifices the payoff $v' - v$ in the first period and gains at most $v' - \max\{\underline{v}, \overline{v}\} < v' - v$ in the second period. She gains nothing from holding office per se, so there is no advantage in her choosing her less-favored policy in the first period in order to be reelected and thus choose her favorite policy in the second period, because if she chooses her favorite policy in the first period and is not reelected then with positive probability the challenger will choose her favorite policy in the second period. Thus in any weak sequential equilibrium, every incumbent whose degree of office-motivation is 0 chooses her favorite policy.

Now consider incumbents whose degree of office-motivation is b. Given that $v + b > \max\{\underline{v}, \overline{v}\}$, such an incumbent optimally chooses her less-favored policy in the first period if doing so gets her reelected, because the amount she thereby loses in the first period, $v' - v$, is less than the amount she gains in the second period, which is $v' + b - \overline{v}$ if her type is (c, b) and $v' + b - \underline{v}$ if her type is (d, b). Thus if, for example, the citizen reelects only incumbents who choose x, then an incumbent whose preferences are the same as the citizen's optimally chooses x even if she knows that the state is y.

The next result shows that in every weak sequential equilibrium of the game all office-motivated incumbents choose the policy more likely to be best for the citizen, given the citizen's information that the state is x with probability p and y with probability $1 - p$, rather than the policy they know is best for the citizen given their knowledge of the state. Incumbents who behave in this way are sometimes said to "pander" to the citizen. The equilibrium is shown in Figure 13.8 for the case in which $p > \frac{1}{2}$, so that the policy more likely to be best for the citizen, given the citizen's information, is x. The policy x is chosen by both incumbents with positive office-motivation and ones with no office-motivation who prefer x, given the state (that is, type $(c, 0)$ in state x and type $(d, 0)$ in state y). If only the former chose x, then the citizen would be indifferent between reelecting the incumbent and electing the challenger following the policy x, because the probability that the incumbent shares her preferences is the same as the probability that the challenger does so. The fact that also the latter choose x tips the balance in favor of reelecting the incumbent, given that x is the more likely state.

Proposition 13.3: Equilibrium of two-period electoral competition game with well-informed candidates with uncertain motivations

Let $\langle \{B, V, C\}, \{x, y\}, p, \pi, v', v, \mathscr{B}, \rho \rangle$ be a two-period electoral competition game with well-informed candidates with uncertain motivations in which $\mathscr{B} = \{0, b\}$ for $b > 0$, and $\rho(0) = r$ and $\rho(b) = 1 - r$, with $r \in (0, 1]$. If $p > \frac{1}{2}$ (x

is more likely than y to be best for the citizen) and $v + b > \max\{\underline{v}, \overline{v}\}$ (the incumbent's office-motivation is sufficiently high), where $\underline{v}$ and $\overline{v}$ are given in (13.1), then the game has a unique weak sequential equilibrium. In this equilibrium, each office-motivated incumbent (types (c, b) and (d, b)) chooses policy x in both state x and state y, and every incumbent with no office-motivation chooses her favorite policy.

Proof

Step 1 *In every weak sequential equilibrium the incumbent chooses x after the moves of chance $(x, c, 0)$ and $(y, d, 0)$, and y after the moves of chance $(y, c, 0)$ and $(x, d, 0)$.*

Proof. After the moves of chance $(x, c, 0)$ and $(y, d, 0)$ the incumbent's worst payoff if she chooses x, the policy she prefers, exceeds her best payoff if she chooses y, and after the moves of chance $(y, c, 0)$ and $(x, d, 0)$ her worse payoff if she chooses y, the policy she prefers, exceeds her best payoff if she chooses x. ◁

Step 2 *The game has no weak sequential equilibrium in which the citizen chooses the same action at each of her information sets.*

Proof. Suppose that the citizen chooses the same action at both of her information sets. Then the incumbent's first-period policy does not affect whether she is reelected, so she optimally chooses x after the moves of chance (x, c, b) and (y, d, b) and y after the moves of chance (y, c, b) and (x, d, b). Given Step 1, the only belief that is consistent with the incumbent's strategy is given as follows. The probability that chance chooses $(x, c, 0)$ or (x, c, b) is $p\pi$, so at the citizen's inner information set (following the policy x) the belief assigns probability $p\pi/\Pi$ to the incumbent's being consonant and probability $(1 - p)(1 - \pi)/\Pi$ to her being dissonant, where Π is the sum of the numerators of these expressions. At her outer information set (following the policy y) it assigns probability $(1-p)\pi/(1-\Pi)$ to the incumbent's being consonant and probability $p(1-\pi)/(1-\Pi)$ to her being dissonant.

Given this belief and the incumbent's strategy, the citizen's payoffs to her actions at her inner information set are

$$B : 2p\pi/\Pi$$
$$C : [p\pi(1 + \pi) + (1 - p)(1 - \pi)\pi]/\Pi.$$

The difference between these payoffs is $(2p-1)\pi(1-\pi)/\Pi$, which is positive given $p > \frac{1}{2}$ and $\pi \in (0,1)$. Thus the citizen's unique optimal action at her inner information set is B. The citizen's payoffs to her actions at her outer information set are

$$B : 2(1-p)\pi/(1-\Pi)$$
$$C : [(1-p)\pi(1+\pi)+p(1-\pi)\pi]/(1-\Pi).$$

The difference between these payoffs is $(1-2p)\pi(1-\pi)/(1-\Pi)$, which is negative. Thus the citizen's unique optimal action at her outer information set is C.

We conclude that the game has no weak sequential equilibrium in which the citizen chooses the same action at each of her information sets. ◁

Step 3 *In every weak sequential equilibrium the incumbent chooses the same policy after every move of chance* (s,h,β) *with* $\beta = b$ *(the ones in the top half of Figure 13.8).*

Proof. By Step 2, in every weak sequential equilibrium the citizen chooses B at one of her information sets and C at the other one. If she chooses B at her outer information set (following policy y) and C at her inner information set (following policy x), then given $v + b > \max\{\underline{v}, \overline{v}\}$, the incumbent optimally chooses y after every move of chance (s,h,β) with $\beta = b$. If the citizen chooses C at her outer information set and B at her inner information set, then the incumbent optimally chooses x after every such move of chance. ◁

Step 4 *In every weak sequential equilibrium the incumbent chooses* x *after every move of chance* (s,h,β) *with* $\beta = b$.

Proof. By Step 3 the only other possibility is that the incumbent chooses y after every move of chance (s,h,β) with $\beta = b$. Then if the citizen's inner information set is reached, by Step 1 the citizen believes that the move of chance was $(x,c,0)$ with probability $p\pi r/\Lambda$ and $(y,d,0)$ with probability $(1-p)(1-\pi)r/\Lambda$, where Λ is the sum of the numerators. The difference between the citizen's expected payoffs to B and C at her inner information set is thus proportional to $p\pi r(2-(1+\pi))+(1-p)(1-\pi)r(-\pi) = r\pi(1-\pi)(2p-1) > 0$, so that the citizen optimally chooses B at the information set. But then after the move of chance (x,c,b) the incumbent is better off deviating from y to x, regardless of whether the citizen chooses B or C at her outer information set. ◁

Step 5 *The game has a unique weak sequential equilibrium, and in this equilibrium types (c,b) and (d,b) of the incumbent choose x in both states, type $(c,0)$ chooses x in state x and y in state y, and type $(d,0)$ chooses y in state x and x in state y.*

Proof. By Steps 1 and 4 the incumbent's strategy has this form in every weak sequential equilibrium. Consider the assessment in which the incumbent's strategy takes this form, the citizen chooses B (reelect the incumbent) if the incumbent's policy is x and C if the incumbent's policy is y, and the belief system assigns to each history in each of the citizen's information sets the probability of the history occurring given the incumbent's strategy, conditional on the information set's being reached. This strategy profile is indicated by the arrows in Figure 13.8.

I argue that the assessment is a weak sequential equilibrium, the belief system is the only one that is consistent with the incumbent's strategy, and the citizen's strategy is the only one that is optimal given the belief system.

We have $\underline{v} \in (v, v')$, $\overline{v} \in (v, v')$, and $v + b > \max\{\underline{v}, \overline{v}\}$, so that the strategy of each type of incumbent is optimal given the citizen's strategy.

At the citizen's inner information set, following the incumbent's choice of policy x, the citizen's expected payoff from choosing B, given the incumbent's strategy, is $[2p\pi(1-r)+(1-p)\pi(1-r)+p(1-\pi)(1-r)+2p\pi r]/\Delta$ and her expected payoff from choosing C is $[(1-r)(\pi+p)+p\pi r(1+\pi)+(1-p)(1-\pi)r\pi]/\Delta$, where Δ is the sum of the probability that chance selects (s,h,b) for some pair (s,h), $(x,c,0)$, or $(y,d,0)$. The difference between the numerators of these expressions is $(1-\pi)r\pi(2p-1)$, which is positive given $p > \frac{1}{2}$. Thus selecting B is the only optimal action for the citizen at her inner information set.

At the citizen's outer information set, following the incumbent's choice of policy y, the citizen's expected payoff from choosing B, given the incumbent's strategy, is $2(1-p)\pi r/(1-\Delta)$ and her expected payoff from choosing C is $[(1-p)\pi r(1+\pi)+p(1-\pi)r\pi]/(1-\Delta)$. The difference between the numerators of these expressions is $(1-\pi)r\pi(1-2p)$, which is negative given $p > \frac{1}{2}$. Thus selecting C is the only optimal action for the citizen at her outer information set.

Both information sets are reached with positive probability, and the beliefs at each information set are derived from the strategy profile using Bayes' rule, so they are the only ones weakly consistent with the strategy profile. ◁

The citizen's expected payoff in the equilibrium in this result is

$$(1-r)(p+\pi)+r\pi[1+\pi+2p(1-\pi)]. \tag{13.2}$$

If, instead of selecting a representative to choose a policy for her, the citizen chooses one herself, she optimally selects x, the policy more likely to be best for her, and consequently obtains the expected payoff $2p$. By doing so she does not benefit from the candidates' superior information, but at the same time is not subject to the risk that a candidate's preferences differ from hers. The citizen's payoff when she chooses a policy herself exceeds her payoff (13.2) if the probability p that the best policy is x exceeds $\pi(1+\pi r)/[1+r-2\pi r(1-\pi)]$, which is less than 1 if $\pi < 1$. Thus however likely it is that a candidate's preferences are consonant with the citizen's preferences, for p sufficiently high the citizen is better off choosing the policy herself than electing a representative to do it for her under the rules of the game we are considering. When p is less than this bound, so that the best policy is relatively uncertain, handing the decision to a representative is better even though the representative's preferences may differ from the citizen's and she may choose the policy the citizen believes is best, rather than the policy that is in fact best, in order to get reelected.

Another option for the citizen is to select a candidate randomly in each period, without the possibility of reelection. The citizen's expected payoff in this case is 2π. This payoff exceeds her payoff (13.2) if the probability p that the best policy is x is less than $\pi(1-\pi r)/[1-r+2\pi r(1-\pi)]$. This number is greater than $\frac{1}{2}$, as p is assumed to be in Proposition 13.3, if $\pi < \frac{1}{2}$. Thus when a candidate's preferences are more likely to be discordant from than consonant with the citizen's preferences and the identity of the best policy is sufficiently uncertain, the outcome for the citizen is better when the incumbent has no possibility of reelection than it is in the equilibrium of the game we are considering. In this case, the reelection incentive leads the incumbent to choose policies that are worse for the citizen, on average, than the favorite policy of a randomly chosen candidate.

Notes

The models in Sections 13.1 and 13.2 express in a two-alternative model the main ideas in the example studied by Fearon (1999), whose work draws in part on the infinite-horizon example studied by Ferejohn (1986). Section 3.3 of Duggan and Martinelli (2015) (of which Duggan and Martinelli 2017 is a shortened version) analyze much more general models. Section 13.3 is based on Maskin and Tirole (2004). Duggan and Martinelli (2020) study a general model.

IV Exercising power

14 Bargaining 459
15 Rulers threatened by rebellion 493

14 Bargaining

14.1 Bargaining game with voting 461
14.2 Recurrent distributive bargaining game with voting 486

In one class of collective decision-making processes, individuals sequentially make and vote on proposals. Processes in this class are typically categorized as bargaining.

Synopsis

I present two models of bargaining processes. In Section 14.1, an outcome is an alternative. Negotiations may last many periods, but once agreement is reached, the game ends. In Section 14.2, an outcome is a sequence of alternatives, one in each period. The alternative in each period serves as the default in the next period: if the proposal in period t is not accepted, the alternative implemented in that period is the one from period $t-1$.

In both models, the first event is that an individual is selected randomly to make a proposal, which each individual then votes either for or against. In the model in Section 14.1, a bargaining game with voting, if a majority of individuals vote for the proposal, it is implemented and the game ends. Otherwise another individual is selected randomly to make a proposal, and a vote is again held. The pattern repeats until a proposal is accepted, when the game ends. In the model in Section 14.2, the acceptance of a proposal means that the proposal is implemented and its rejection means that the default alternative, from the previous period, is implemented. In both cases, negotiations remain open: in every future period a proposal may be made and voted upon.

Both models are extensive games. For a bargaining game with voting, the solution notion of subgame perfect equilibrium does not restrict the outcome: for every alternative x, the strategy profile in which every individual always proposes x and votes in favor of x and against any other alternative is a subgame perfect equilibrium (Proposition 14.1). In this equilibrium, no individual's vote makes a difference because all votes are unanimous; thus no individual has an incentive to change her behavior. However, an individual who is (unexpectedly) confronted with an alternative she prefers to x has nothing to lose by voting for it rather than against it as her equilibrium strategy mandates, and restricting individuals' strategies to not be weakly dominated in this sense holds the promise that it might eliminate some alternatives as equilibrium outcomes. If the individuals' votes are observable, however, it has little effect: for the types of sets of alternatives and preferences usually assumed, for most alternatives x the game has a subgame perfect equilibrium with undominated voting in which agreement on x is reached immediately (Proposition 14.2).

If only the outcome of each vote, not the vote cast by each individual, is observable, then the outcomes of subgame perfect equilibria in which each individual's voting strategy is not weakly dominated depend on the nature of the set of alternatives and the individuals' preferences. If the alternatives are distributions among the individuals of a fixed amount of a good, there are at least five individuals, and each individual cares only about the amount she receives and is sufficiently patient, then for every alternative x the game has a subgame perfect equilibrium with undominated voting in which agreement on x is reached immediately (Proposition 14.3). If, however, the set of alternatives is an interval of numbers and each individual's preferences over this set are single-peaked, then if the individuals are sufficiently patient, the outcome of any subgame perfect equilibrium with undominated voting is necessarily close to the median of the individuals' favorite positions (Proposition 14.4).

A requirement much stronger than subgame perfect equilibrium with undominated voting is that each player's strategy is stationary, always making the same proposal and casting the same vote regarding any given proposal, independent of the history. The appeal of this requirement is questionable, but if the alternatives are distributions among the individuals of a fixed amount of a good, it leads to a unique equilibrium outcome (Proposition 14.5). In this outcome the player first selected proposes a particular distribution of the available good among a minimal majority, the members of this majority vote in favor, and the game ends.

Section 14.2 studies a game with recurrent bargaining: negotiations are always open. The alternatives are distributions among the individuals of a fixed amount of a good. In each period, one of these distributions is the default alternative. In the first period, this default is the distribution in which no individual receives any of the good, and in each subsequent period it is the alternative implemented in the previous period. In each period, an individual is selected randomly. She can either pass, in which case the default alternative is implemented in the period, or she can propose another alternative. If a majority of individuals vote in favor of her proposal, it is implemented in the period, and if not, the default alternative from the previous period is implemented.

The section aims to convey the main features of the subgame perfect equilibria of this game by means of an example. Say that a player's strategy is stationary if (*i*) for any given default alternative the strategy always makes the same proposal and (*ii*) for any given default alternative and proposal it always votes in the same way. In sharp contrast to a distributive bargaining game with voting, in which the game ends once agreement is reached, a game with recurrent bargaining has many stationary equilibrium outcomes, and these outcomes may involve waste and assign a positive amount to more than a bare majority of the players. In the example I present, for almost any distribution x, including ones in which some of the good is wasted, the game has an equilibrium in which every player's strategy is stationary and x is implemented in every period. The players whose shares of the good are relatively high in x constitute a minimal majority; none of them wants to deviate because doing so would reopen negotiations and expose her to the risk that the outcome in every future period is worse for her than x.

14.1 Bargaining game with voting

A finite set N of individuals faces a collective choice problem in which the set X of alternatives is a compact convex subset of a Euclidean space. The process by which they may reach agreement takes place over a sequence of time periods, $t = 1, 2, \ldots$, so that we need to specify each individual's preferences over pairs (x, t) consisting of an alternative $x \in X$ and a time t at which agreement is reached, together with the outcome in which agreement is never reached. I assume that the preference relation of each individual i is represented by a function with values of the form $\delta_i^{t-1} u_i(x)$ for pairs (x, t), where $\delta_i \in (0, 1)$ and $u_i : X \to \mathbb{R}_+$, and the value 0 for the outcome in which agreement is never reached.

The model of the bargaining procedure is an extensive game with perfect information, simultaneous moves, and chance moves. The set of players is the set N of individuals; to avoid dealing with ties in votes, I assume that their number, n, is odd. First, chance selects one of the players. Each player i is chosen independently with probability $\rho_i \in (0, 1)$. The selected player proposes a member x of X, and then all players simultaneously cast votes for or against x; every player observes the votes cast. If a majority of the votes ($\frac{1}{2}(n+1)$ or more) are cast in favor of x, the game ends with the outcome x. Otherwise, play moves to the next period, in which again a player chosen by chance makes a proposal and a vote is held. Play continues in the same manner until a proposal is accepted, or goes on forever if all proposals are rejected. The first two periods of play are illustrated in Figure 14.1.

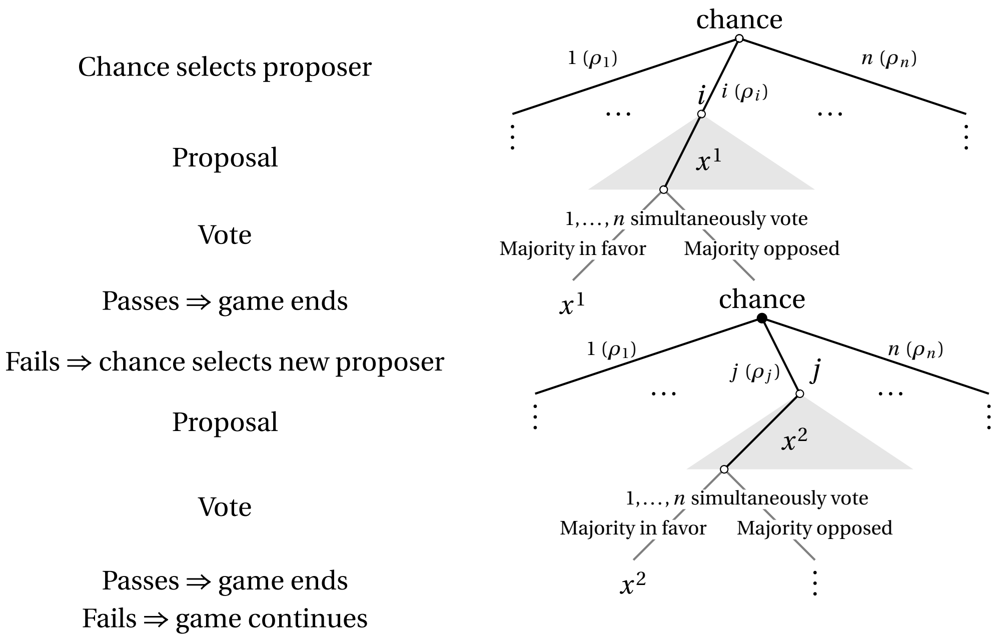


Figure 14.1 An illustration of the first two periods of a bargaining game with voting. Only one of the actions of only one of the players in each period is shown.

Definition 14.1: Bargaining game with voting

A *bargaining game with voting* $\langle N, X, (\rho_i)_{i \in N}, (\delta_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where

- $N = \{1, \ldots, n\}$, where $n \geq 3$ is an odd integer (the set of individuals)
- X is a compact convex subset of a Euclidean space (the set of alternatives)
- $\rho_i \in (0,1)$ for each $i \in N$, with $\sum_{i \in N} \rho_i = 1$ (the probability that player i is chosen to make a proposal in any given period, her *recognition probability*)
- $\delta_i \in (0,1)$ for each $i \in N$ (player i's discount factor)
- $u_i : X \to \mathbb{R}_+$ is continuous, with $u_i(x) > 0$ for all $x \in \text{int}\, X$ (u_i is player i's payoff function over alternatives)

is the extensive game with perfect information, simultaneous moves, and chance moves with the following components, where

$$V = \{(v_1, \ldots, v_n) : v_i \in \{\textit{for}, \textit{against}\} \text{ for } i = 1, \ldots, n\} \text{ (vote profiles)}$$
$$R = \{(v_1, \ldots, v_n) \in V : \textstyle\sum_{i=1}^{n} v_i < \frac{1}{2}(n+1)\} \text{ (vote profiles opposed)}$$
$$A = \{(v_1, \ldots, v_n) \in V : \textstyle\sum_{i=1}^{n} v_i \geq \frac{1}{2}(n+1)\} \text{ (vote profiles in favor).}$$

Players

The set N (of individuals).

Terminal histories

The set of terminal histories consists of

- $(i^1, x^1, v^1, i^2, x^2, v^2, \ldots, i^t, x^t, v^t)$ for any $t \geq 1$, $i^s \in N$ and $x^s \in X$ for $s = 1, \ldots, t$, $v^s \in R$ for $s = 1, \ldots, t-1$, and $v^t \in A$ (proposals through period $t-1$ are rejected and the proposal in period t is accepted)
- $(i^1, x^1, v^1, i^2, x^2, v^2, \ldots, i^t, x^t, v^t, \ldots)$ where $i^s \in N$, $x^s \in X$, and $v^s \in R$ for $s = 1, 2, \ldots$ (all proposals are rejected).

Player function

The player function P is defined as follows.

- $P(\varnothing) =$ chance and $P(i^1, x^1, v^1, i^2, x^2, v^2, \ldots, i^t, x^t, v^t) =$ chance for every $t \geq 1$ if $v^t \in R$ (chance moves at the start of the game and after a proposal is rejected).
- $P(i^1, x^1, v^1, i^2, x^2, v^2, \ldots, i^t) = i^t$ for every $t \geq 1$ (the player who moves (proposes an alternative) after chance is the one selected by chance)
- $P(i^1, x^1, v^1, i^2, x^2, v^2, \ldots, i^t, x^t) = N$ for every $t \geq 1$ (all players move (vote) after a player makes a proposal).

Actions

For any history h, the set $A_i(h)$ of actions of each player $i \in N$ after h is X if h ends with i's selection as the proposer and $\{\textit{for}, \textit{against}\}$ (a vote) if h ends with a proposal. The set of actions of chance at the beginning of the game and after any proposal is rejected is N.

Chance probabilities

For each player $i \in N$, chance selects i with probability ρ_i whenever it moves.

Preferences

The preference relation of each player $i \in N$ over lotteries over the set Z

of terminal histories is represented by the expected value of the payoff function $U^i : Z \to \mathbb{R}_+$ given by

$$U^i(h^t) = \delta_i^{t-1} u_i(x^t)$$

for all $t \geq 1$ and all terminal histories h^t ending in period t with a proposal x^t followed by a vote $v^t \in A$ to accept x^t, and $U^i(h^\infty) = 0$ for all terminal histories h^∞ in which all proposals are rejected.

A strategy for any player i in this game is a function that associates an alternative with each history ending in a move of chance that selects i to be the proposer and either *for* or *against* with each history ending in a proposal. An outcome is a terminal history, which is either a history ending in a vote in which a majority of the players vote *for* the latest proposal or a history in which no proposal is approved by a majority of the players.

A variant of the model assumes that while negotiations are taking place, each player receives a payoff from the status quo, which may or may not be a member of X. Denoting the status quo by q, in this variant the (discounted average) payoff of individual i to an agreement on $x \in X$ in period t is $(1 - \delta_i^{t-1})u_i(q) + \delta_i^{t-1} u_i(x)$ (assuming that u_i is defined for q as well as for every member of X). A bargaining game with voting is thus equivalent to this variant if $u_i(q) = 0$ for every player i (that is, if each player's payoff from the status quo is the same as her payoff from perpetual disagreement).

14.1.1 Subgame perfect equilibrium

A bargaining game with voting is an extension to many players of a (two-player) bargaining game of alternating offers (as defined, for example, in Section 7.2 of Osborne and Rubinstein 1994). The two-player game has a unique subgame perfect equilibrium. By contrast, a bargaining game with voting has many subgame perfect equilibria. For example, for every alternative $x \in X$ it has a subgame perfect equilibrium in which agreement is reached on x immediately. The argument for this result is simple. Suppose that every player proposes x whenever she is selected and votes for x and against every other alternative regardless of the history. Then if a player deviates by proposing an alternative different from x, her proposal is voted down and the outcome is x in a future period, which she likes less than x immediately. If she deviates by voting against x or for another alternative after some history, the outcome is unaffected, given the other players' strategies.

Proposition 14.1: Subgame perfect equilibrium of bargaining game with voting

Let $\langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a bargaining game with voting and let $x \in X$. For each player $i \in N$ let s_i^* be the strategy of player i in which she proposes x after every history that ends with her selection as the proposer and votes in favor of x and against every other proposal after every history that ends with a vote. Then the strategy profile s^* is a subgame perfect equilibrium. The outcome of s^* is that x is proposed and accepted immediately.

Proof

The strategy profile s^* satisfies the one-deviation property, which requires that no player can increase her payoff in any subgame by changing only her action at the start of the subgame, given the other players' strategies. Consider player i. If, after any history that ends with the selection of i as the proposer, she adheres to s_i^* and proposes x, then the outcome is x, whereas if she deviates from s_i^* and proposes $y \neq x$, then y is rejected and the outcome is x in the following period. If, after any history that ends with a proposal, i changes her vote, the outcome remains the same because every other player accepts the proposal if it is x and rejects it otherwise. The game satisfies the condition in Proposition 16.9, so the fact that s^* satisfies the one-deviation property means that it is a subgame perfect equilibrium.

The strategy profile s^* in this result has an unattractive feature. Consider a history ending in a proposed alternative y that player j (and possibly other players) prefers to x. Every player's strategy in s^* calls for her to vote against y, and if the players follow their strategies the outcome is x with one period of delay. Casting such a vote is optimal for every player, including player j, because no change in the vote of a single player affects the outcome, given the other players' strategies. For the same reason, voting *for* y also is optimal, given the other players' strategies. In fact, voting for y and voting against it are both optimal not only if the other players vote according to their strategies, but also if they cast another other votes, *unless* their votes are equally split between *for* and *against*, in which case j's voting *for* yields the outcome y and her voting *against* yields x with one period of delay (assuming that all players adhere to s^* in the following periods). Table 14.1 shows the outcome as a function of the players' votes. Player j prefers y to x, and hence to x with one period of delay, so her voting against y, as s_i^* specifies, is weakly dominated by her voting for y, given the be-

	number of other players voting against y						
	0	$\cdots$	$\frac{1}{2}(n-3)$	$\frac{1}{2}(n-1)$	$\frac{1}{2}(n+1)$	$\cdots$	$n-1$
j votes against y	y	$\cdots$	y	x^{+1}	x^{+1}	$\cdots$	x^{+1}
j votes for y	y	$\cdots$	y	y	x^{+1}	$\cdots$	x^{+1}

Table 14.1 The outcome in a bargaining game with voting following a history ending with the proposal y when all players adhere to the strategy s^* in the proof of Proposition 14.1 following the vote on y. The superscript $+1$ means that the outcome occurs with a period of delay.

havior specified by s^* in other periods. I refer to a subgame perfect equilibrium in which no player's vote in any period is weakly dominated in this sense as a subgame perfect equilibrium with undominated voting.

Definition 14.2: Subgame perfect equilibrium with undominated voting in bargaining game with voting

Let $\Gamma = \langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a bargaining game with voting. For each strategy s and history h that ends with a proposal, define the strategic game $G(h,s)$ as follows.

Players

The set of players is N.

Actions

The set of actions of each player is $\{$*for*, *against*$\}$ (the possible votes).

Preferences

The preferences of each player $i \in N$ over action profiles are represented by the function that assigns to the action profile v the payoff of player i in Γ for the history consisting of h followed by v followed by the sequence of action profiles generated by s after the history (h, v).

A strategy profile s in Γ is a *subgame perfect equilibrium with undominated voting* if it is a subgame perfect equilibrium of Γ and for every player $i \in N$ and every history h ending with a proposal, the action $s_i(h)$ is not weakly dominated in the game $G(h,s)$.

Does the restriction to undominated voting reduce the set of equilibrium outcomes? Roughly speaking, not much. The next result shows that every alternative in a subset of X that, for some games, consists of almost all members of X, is the outcome of a subgame perfect equilibrium with undominated voting. The idea behind the equilibrium is that if, when the players' strategies call for them to propose x, y is proposed and some player j votes in favor of it while every

	number of other players voting against y							
	0	$\cdots$	$\frac{1}{2}(n-3)$	$\frac{1}{2}(n-1)$	$\frac{1}{2}(n+1)$	$\cdots$	$n-2$	$n-1$
j votes against y	y	$\cdots$	y	x^{+1}	x^{+1}	$\cdots$	x^{+1}	x^{+1}
j votes for y	y	$\cdots$	y	y	x^{+1}	$\cdots$	x^{+1}	$\gamma^j(x)^{+1}$

Table 14.2 The outcome in a bargaining game with voting following a history ending with the proposal y when all players' strategies specify the proposal $x \neq y$, a vote for x and against every other alternative, and a switch to the proposal $\gamma^j(x)$ if a single player j deviates and votes if favor of an alternative different from x. The superscript $+1$ means that the outcome occurs with a period of delay.

other player adheres to her strategy and votes against it, the other players react by switching from proposing x to proposing an alternative $\gamma^j(x)$ that is worse for j than x and subsequently voting in favor of this alternative. Then even if j prefers y to x, voting for y does not weakly dominate voting against it, because if all the other players vote against it then voting for y leads to the outcome $\gamma^j(x)$ in the next period whereas voting against it, which causes no change in the other players' behavior, leads to the outcome x in the next period. The outcomes as a function of the players' votes are given in Table 14.2.

To construct a subgame perfect equilibrium along these lines, we need to ensure that the switch to proposing $\gamma^j(x)$ is optimal for every player, given the strategies of the remaining players. We do so by specifying that, just as a player's vote in favor of an alternative that she is supposed to vote against induces a switch to a regime in which she is punished, so a single player's deviation from the new regime by voting for a proposal that she is supposed to vote against induces a switch to a regime in which *she* is punished.

These strategies may conveniently be modeled as automata. We specify a set of states, the actions taken by each strategy in each state, and a rule for transitions between states. The states in this case are members of X. For any $\sigma \in X$, in state σ a player's strategy proposes σ and votes for σ and against every other proposal. The state is initially x and remains x unless an alternative different from x is proposed and exactly one player, j, votes in favor of this proposal, in which case the state changes to $\gamma^j(x)$, an alternative that is worse for j than x.

The states that can be reached are x, $\gamma^j(x)$ for each $j \in N$, $\gamma^k(\gamma^j(x))$ for each $k \in N$ and each $j \in N$, and so on. For each of these states σ, we need $\gamma^j(\sigma)$ to be worse than σ for j. A sufficient condition for the existence of such alternatives is that x is a member of a subset X^* of X with the property

$$\text{for all } i \in N \text{ and all } y \in X^* \text{ there exists } \gamma^i(y) \in X^* \text{ with } u_i(\gamma^i(y)) < u_i(y). \quad (14.1)$$

Suppose, for example, that X is the one-dimensional interval $[\underline{x}, \overline{x}]$ and each player's payoff function is single-peaked. For each player j, let $\hat{x}_j$ be j's favorite alternative. Then $X^* = (\underline{x}, \overline{x})$ satisfies (14.1), with

$$\gamma^j(y) = \begin{cases} \frac{1}{2}(y+\underline{x}) & \text{if } y \le \hat{x}_j \\ \frac{1}{2}(y+\overline{x}) & \text{if } y > \hat{x}_j. \end{cases}$$

Proposition 14.2: Subgame perfect equilibrium with undominated voting in bargaining game

Let $\langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a bargaining game with voting and let X^* be a subset of X that satisfies (14.1). Then for any $x \in X^*$ the game has a subgame perfect equilibrium with undominated voting in which at the start of the game each player proposes x and all players vote in favor of x.

Proof

For each player $i \in N$, let s_i^* be the strategy defined by the automaton in Figure 14.2. Note that given that $x \in X^*$, every state that can be reached is in X^*. According to the strategy profile s^*, every player proposes x if she is selected to make a proposal in the first period and votes in favor of x if it is proposed in this period, so the outcome of s^* is immediate agreement on x.

I claim that s^* is a subgame perfect equilibrium with undominated voting . The game satisfies the condition in Proposition 16.9, so to show that s^* is a subgame perfect equilibrium it suffices to show that s^* satisfies the one-deviation property.

Proposal in any state σ

If player i adheres to s_i^*, she proposes σ, which is accepted (all players vote in favor). If she deviates to propose $y \ne \sigma$, then all players vote against y and the state remains σ, so that the outcome is σ with one period of delay, which is no better for i than immediate agreement on σ.

Vote in any state σ regarding proposal y

First suppose that $y = \sigma$. Then all players vote in favor of σ, and the outcome is σ. If player i deviates and votes against σ, the outcome remains σ, because the state remains σ and the remaining players constitute a majority. Thus i's voting for σ is optimal.

I now argue that her voting for σ is undominated. If $\frac{1}{2}(n-1)$ of the other players vote for σ, her voting for σ leads to σ's being accepted, whereas

	State x (initial)	State $\gamma^j(y)$ ($j \in N$, $y \in X^*$)
i proposes	x	$\gamma^j(y)$
i votes	for x against any $y \neq x$	for $\gamma^j(y)$ against any $z \neq \gamma^j(y)$
Transitions	If, in any state σ, $y \neq \sigma$ is proposed and exactly one player, k, votes for it, go to state $\gamma^k(\sigma)$.	

Figure 14.2 The strategy of player i in the proof of Proposition 14.2. For each player j and each alternative $\sigma \in X^*$, $\gamma^j(\sigma)$ is the alternative in X^* for which $u_j(\gamma^j(\sigma)) < u_j(\sigma)$ given in (14.1). The state remains the same unless the condition in the transition cell is satisfied.

her voting against σ leads to its being rejected, in which case the state remains σ and hence the outcome is σ with one period of delay. Given that $\sigma \in \text{int}\, X$ we have $u_i(\sigma) > 0$, so that $\delta_i u_i(\sigma) < u_i(\sigma)$. Thus for these votes of the other players, player i's payoff from voting against σ is less than her payoff from voting for σ. Thus her voting for σ is not weakly dominated.

Now suppose that $y \neq \sigma$. Then all players vote against y, the state remains σ, and the outcome is thus σ with one period of delay. If player i deviates from s_i^* and votes for y while the other players adhere to s^* and vote against y, the state changes to $\gamma^i(\sigma)$, so that the outcome is $\gamma^i(\sigma)$ with one period of delay, which is worse for i than σ with one period of delay. Thus i's voting against y is optimal and undominated.

The equilibrium strategies in the proof of this result depend on each player's vote being observable, so that if a single player deviates from her strategy and votes in favor of a proposal different from the one the equilibrium specifies, given the history, she can be identified and punished. If only the numbers of votes for and against a proposal are observable, not each individual's vote, such punishments cannot be implemented. Under this assumption, the set of equilibrium outcomes depends on the form of the set of alternatives and the individuals' preferences over this set. I present two cases.

Bargaining over distribution when individuals' votes are unobserved

Suppose that the individuals have one unit of a good to distribute among themselves, some of which may be wasted, and the payoff of each individual is the amount of the good that she receives. The model of a bargaining game with voting was first explored under this assumption, with the interpretation that the individuals are legislators and the alternatives are bills that distribute benefits among the legislators' districts.

Definition 14.3: Distributive bargaining game with voting

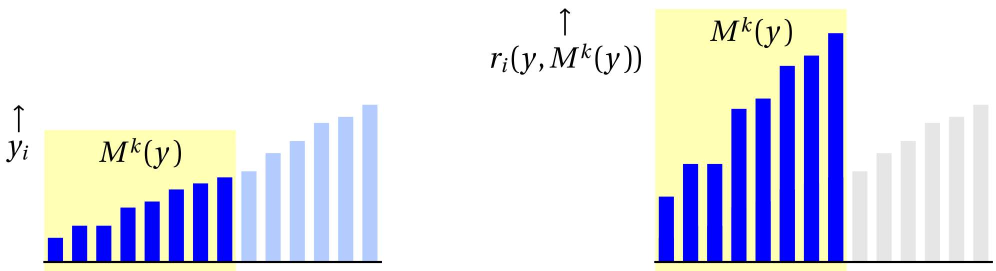
A bargaining game with voting $\langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ is *distributive* if

$$X = \{(x_1,\ldots,x_n) \in \mathbb{R}^n_+ : \textstyle\sum_{i=1}^n x_i \le 1\},$$

where $n = |N|$, and $u_i(x) = x_i$ for all $i \in N$ and all $x \in X$.

If in such a game the players number at least five and each player's discount factor is close enough to 1, then for any alternative $x \in X$, immediate agreement on x is the outcome of a subgame perfect equilibrium with undominated voting in which the strategies do not rely on the observability of the individuals' votes. That is, the conclusion that any, or almost any, alternative is the outcome of an equilibrium does not depend on each individual's vote being observable, as Proposition 14.2 assumes.

The reason that in such an equilibrium proposing x is optimal for every player—even those whose shares x_i are small—is that any other proposal, say y, is rejected, and in the subgame that is reached, the player who proposed y, say j, obtains the payoff 0. A subset of the set of all players excluding j that (*i*) has $\frac{1}{2}(n+1)$ members (a bare majority) and (*ii*) has the lowest total payoff under y among all such subsets, gangs up on j. Every member of this set votes against y, proposes a distribution z in which j's payoff is 0 and all of their payoffs are at least as high as they are under y (accounting for the fact that z is received with a period of delay), and votes for z and against every other proposal. (The fact that z is received with a period of delay is the reason the players' discount factors need to be sufficiently close to 1.) Why do these players engage in this behavior? For each of them, voting against y is not weakly dominated by voting in favor of y because all of them are better off when the outcome is z with a period of delay than when it is y immediately. Also, if any of them makes a proposal different from z, the remaining players gang up on *her* in the same way, reducing her payoff to 0.



(a) The subset $M^k(y)$ of $N \setminus \{k\}$ with $\frac{1}{2}(n+1)$ members for whom the sum of the payoffs in y is smallest.

(b) The payoffs of the players in the proposal $r(y, M^k(y))$.

Figure 14.3 The payoffs of the players in $N \setminus \{k\}$ for some $k \in N$, ordered by y_i.

Proposition 14.3: Subgame perfect equilibrium with undominated voting in distributive bargaining game with unobserved individual votes

Let $\langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a distributive bargaining game with voting and let $n = |N|$. If $n \geq 5$ and $\delta_i \in \left(\frac{1}{2}(n+1)/(n-1), 1\right)$ for all $i \in N$ then for every $x \in X$ the game has a subgame perfect equilibrium with undominated voting in which every player's action after every history depends only on the sequence of proposals and outcomes of the votes (majority in favor, majority opposed) and every player proposes x at the start of the game and votes in favor of this proposal.

Proof

I define a strategy profile s^* and argue that it is a subgame perfect equilibrium with undominated voting. In s^*, every player proposes x at the start of the game and all players vote in favor. If a player deviates to a different proposal and this proposal is rejected, the other players penalize her. To specify the penalization, for any proposal $y \in X$ and nonempty subset S of N, let $r(y, S)$ be the proposal that allocates all the payoff to the players in S, proportionally to their payoffs in y: for each $i \in N$,

$$r_i(y,S) = \begin{cases} y_i / \sum_{j\in S} y_j & \text{if } i \in S \text{ and } \sum_{j\in S} y_j > 0 \\ 1/|S| & \text{if } i \in S \text{ and } \sum_{j\in S} y_j = 0 \\ 0 & \text{if } i \in N \setminus S. \end{cases} \tag{14.2}$$

Now, for each player $k \in N$ and each proposal $y \in X$, let $M^k(y)$ be a subset of $N \setminus \{k\}$ with $\frac{1}{2}(n+1)$ members for which the total payoff $\sum_{i \in M^k(y)} y_i$ is smallest (refer to Figure 14.3a). If player k proposes $y \neq x$, then the members of $M^k(y)$, a majority, vote against y, and the next player to make a proposal chooses $r(y, M^k(y))$ (refer to Figure 14.3b). This behavior penalizes player k because $r_k(y, M^k(y)) = 0$ given that $k \notin M^k(y)$.

The strategy s_i^* may conveniently be described more precisely as the automaton shown in Figure 14.4. The states are x and $p^j(y)$ for each $j \in N$ and each $y \in X$. The initial state is x. In this state, s_i^* proposes x and votes in favor of x. If, in this state, player j proposes $y \neq x$, then s_i^* votes against the proposal if $i \in M^j(y)$ and in favor otherwise, and hence y is rejected. The state remains x if x is proposed, and changes to $p^j(y)$ if player j proposes $y \neq x$ and this proposal is rejected. In state $p^j(y)$, j is penalized: s_i^* proposes $r(y, M^j(y))$ and every member of $M^j(y)$ votes in favor of it. If a different alternative, say z, is proposed by player l, then i votes in favor of it if and only if $i \in M^l(z)$. Thus in state $p^j(y)$, $r(y, M^j(y))$ is accepted and every other proposal is rejected. If, in this state, some player l proposes $z \neq r(y, M^j(y))$ and z is rejected, the state changes to $p^l(z)$.

The outcome of the strategy profile s^* is that the player chosen to move first proposes x, and all players vote in favor.

I now argue that s^* has undominated voting and satisfies the one-deviation property. The game satisfies the condition in Proposition 16.9, so the fact that s^* satisfies the one-deviation property means that it is a subgame perfect equilibrium.

Proposal in state x

If player i adheres to her strategy and proposes x then her payoff is x_i. If she proposes $y \neq x$, then every player in $M^i(y)$, a majority, votes against the proposal and the state becomes $p^i(y)$, in which $r(y, M^i(y))$ is proposed and accepted (because all members of $M^i(y)$ vote in favor), resulting in a payoff for i of 0. Thus i's payoff if she proposes x is at least her payoff if she makes any other proposal.

Vote in state x regarding proposal y

First suppose that $y = x$. If every player follows her strategy then x is accepted and i's payoff is x_i.

A player's vote regarding x affects the outcome generated by s^* only if the other players' votes are split equally between *for* and *against*. In this case, if i votes *against*, then x is rejected, the state remains x, and x is accepted with one period of delay, so that i's payoff is $\delta_i x_i$. If i votes *for*, then x is accepted immediately and her payoff is x_i. The latter payoff is at least the former, so i's voting for x is optimal and undominated. (If $x_i = 0$, i's voting *for* and *against* yield her the same payoff regardless of the other players' votes.)

Now suppose that $y \neq x$, and let j be the player who proposes y. If every player adheres to her strategy, y is rejected, the state changes to $p^j(y)$, and each player i obtains $r_i(y, M^j(y))$ with one period of delay, which is worth $\delta_i r_i(y, M^j(y))$ to her.

As in the previous case, a player's vote regarding y affects the outcome generated by s^* only if the other players' votes are split equally between *for* and *against*.

Consider a player $i \notin M^j(y)$. If she deviates from her strategy and votes against y, then if the other players' votes are split equally between *for* and *against*, y is rejected and her payoff changes from y_i to $r_i(y, M^j(y)) = 0$. Thus i's voting for y is optimal and undominated.

Now consider a player $i \in M^j(y)$. If she deviates from her strategy and votes for y, then if the other players' votes are split equally between *for* and *against*, her payoff changes from $\delta_i r_i(y, M^j(y))$ to y_i. If $y_i = 0$ then certainly $y_i < \delta_i r_i(y, M^j(y))$, so assume that $y_i > 0$. Then, given (14.2),

$$\frac{y_i}{r_i(y, M^j(y))} = \sum_{l \in M^j(y)} y_l.$$

This sum, the total payoff of a subset of $N \setminus \{j\}$ with $\frac{1}{2}(n+1)$ players that has the smallest total payoff under y, is at most $\frac{1}{2}(n+1)/(n-1)$. (The maximum is achieved when $y_i = 1/(n-1)$ for all $i \in N \setminus \{j\}$.) By assumption, this number is less than δ_i. Thus $y_i < \delta_i r_i(y, M^j(y))$, and hence i's voting against y is optimal and undominated.

Proposal in state $p^j(y)$

If i adheres to her strategy and proposes $r(y, M^j(y))$ then the outcome is $r(y, M^j(y))$ and her payoff is $r_i(y, M^j(y))$.

If she proposes $z \neq r(y, M^j(y))$ then this proposal is rejected, the state changes to $p^i(z)$, and her payoff is $r_i(z, M^i(z)) = 0$. So i is not better off deviating from her strategy than adhering to it.

Vote in state $p^j(y)$ regarding proposal z

First suppose that $z = r(y, M^j(y))$. If every player follows her strategy then z is accepted and i's payoff is $r_i(y, M^j(y))$.

	State x (initial)	State $p^j(y)$ ($j \in N$, $y \in X$)
i proposes	x	$r(y, M^j(y))$
i votes	for x against $y \neq x$ proposed by j $\Leftrightarrow i \in M^j(y)$	for $r(y, M^j(y)) \Leftrightarrow i \in M^j(y)$ against $z \neq r(y, M^j(y))$ proposed by l $\Leftrightarrow i \in M^l(z)$
Transitions	if j proposes $y \neq x$ and y is rejected, go to $p^j(y)$	if l proposes $z \neq r(y, M^j(y))$ and z is rejected, go to $p^l(z)$

Figure 14.4 The strategy of player i in the proof of Proposition 14.3. The state remains the same unless the condition in the transition cell is satisfied. The function r is defined in (14.2) and $M^j(y)$ is a subset S of $N \setminus \{j\}$ with $\frac{1}{2}(n+1)$ members for which $\sum_{i \in S} y_i$ is smallest.

As in state x, a player's vote affects the outcome only if the other players' votes are split equally between *for* and *against*. In this case, if i votes *against*, then z is rejected, the state remains $p^j(y)$, and z is accepted with one period of delay, so that i's payoff is $\delta_i r_i(y, M^j(y))$. If i votes *for*, then z is accepted immediately and her payoff is $r_i(y, M^j(y))$. If $i \in M^j(y)$ then the latter payoff is at least the former and if $i \notin M^j(y)$ then both payoffs are zero. So in both cases the action specified by i's strategy is optimal and undominated.

Now suppose that $z \neq r(y, M^j(y))$. If every player follows her strategy then z is rejected and the state changes to $p^l(z)$, where l is the player who proposed z, with the outcome $r(z, M^l(z))$.

Again, a player's vote affects the outcome only if the other players' votes are split equally between *for* and *against*. In this case, if i votes *against*, then z is rejected, the state changes to $p^l(z)$, and $r(z, M^l(z))$ is accepted with one period of delay, so that i's payoff is $\delta_i r_i(z, M^l(z))$. If i votes *for*, then z is accepted immediately and her payoff is z_i. If $i \notin M^l(z)$ then $r_i(z, M^l(z)) = 0$, so the latter payoff is at least the former. If $i \in M^l(z)$ then $z_i < \delta_i r_i(z, M^l(z))$ by the argument for the vote of a player in $M^j(y)$ regarding y in state x. So in both cases the action specified by i's strategy is optimal and undominated.

	number of other players voting against x^*						
	0	$\cdots$	$\frac{1}{2}(n-3)$	$\frac{1}{2}(n-1)$	$\frac{1}{2}(n+1)$	$\cdots$	$n-1$
j votes against x^*	x^*	$\cdots$	x^*	α	α	$\cdots$	α
j votes for x^*	x^*	$\cdots$	x^*	x^*	α	$\cdots$	α

Table 14.3 The outcome in a bargaining game with voting following a history ending with the proposal x^*, where α is the lottery (over alternatives in future periods) that results if x^* is voted down.

Exercise 14.1: Distributive bargaining game with three players

Why is the strategy profile in the proof of the result not a subgame perfect equilibrium for $n = 3$?

Bargaining over one-dimensional policy when individuals' votes are unobserved

Now suppose that X is an interval of numbers and each individual's preferences over X are single-peaked. In this case, the preferences have a degree of commonality absent in a distributive bargaining game, where reducing one individual's payoff to 0 allows the payoffs of the remaining individuals to be raised. If X is an interval of numbers, an alternative that is bad for one individual is bad also for individuals with similar favorite positions, so that punishing one individual for a deviation means imposing a low payoff on other individuals, making it difficult to provide them with an incentive to take part in the punishment.

In fact, when X is an interval of numbers and the individuals' discount factors are close to 1, only alternatives close to the median x^* of the individuals' favorite alternatives are outcomes of subgame perfect equilibria with undominated voting in which every individual's action after every history depends only on the sequences of proposals and outcomes of the votes (pass, fail).

A key step in the argument is that if any player proposes x^* then a majority of individuals vote in favor of it. If x^* is voted down, the outcome is a lottery over alternatives in future periods that by assumption does not depend on the margin by which it loses or the identity of the individuals who voted for or against it. Denoting this lottery α, each individual j faces the situation in Table 14.3 when x^* is proposed. No alternative is preferred to x^* by a majority of individuals, so for a majority of individuals, voting for x^* weakly dominates voting against it—and hence x^* passes.

Now let $\underline{x}$ and $\overline{x}$ be the smallest and largest alternatives that are outcomes of subgame perfect equilibria of the type we are considering, and suppose that a majority of individuals prefer $\underline{x}$ to $\overline{x}$. I argue that $\overline{x}$ is close to x^*, and hence so too is $\underline{x}$. To pass, $\overline{x}$ needs the vote of at least one individual whose favorite position is at most x^*. The vote of such an individual i makes a difference to the outcome only if the votes of the remaining individuals are split equally between *for* and *against*. If in this case i votes in favor of $\overline{x}$ then the outcome is $\overline{x}$, whereas if she votes against $\overline{x}$ then (*a*) with positive probability she is selected to be the proposer and can, by proposing x^*, obtain the outcome x^* with one period of delay, given the previous argument, and (*b*) with the remaining probability the outcome is at worst $\overline{x}$ with one period of delay. Thus if her discount factor is close to 1, her voting against $\overline{x}$ weakly dominates her voting in favor of $\overline{x}$ unless $\overline{x}$ is close enough to x^*.

Proposition 14.4: Subgame perfect equilibrium with undominated voting in bargaining game over policy with unobserved individual votes

Let $\langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a bargaining game with voting for which X is a compact interval of numbers. Denote by x^* the median of the player's favorite alternatives. For any $\varepsilon > 0$ there exists $\overline{\delta} < 1$ such that if $\delta_i \geq \overline{\delta}$ for all $i \in N$ then a proposal x passes in a subgame perfect equilibrium with undominated voting of the game in which every player's action after every history depends only on the sequences of proposals and outcomes of the votes (pass, fail) only if $x \in [x^* - \varepsilon, x^* + \varepsilon]$.

Proof

Let s^* be a subgame perfect equilibrium with undominated voting in which every player's action after every history depends only on the sequences of proposals and outcomes of the votes.

Step 1 *If, after any history, a player's strategy in s^* proposes x^*, a majority of players vote in favor.*

Proof. If a majority votes against x^*, the outcome is a lottery over agreements in later periods (and no agreement at all), independent of the margin by which x^* loses and the identity of the players who vote for and against it, by the assumption on the form of the strategies. Thus each player j faces the situation in Table 14.3, where α is the lottery that results if x^* is voted down. Given that x^* is the median of the individuals' favorite alternatives, it is preferred to α by a majority of players. Thus the requirement that each player's vote in each period be undominated means that a majority of players vote for x^*. ◁

Denote by $\underline{x}$ the smallest alternative that passes in a subgame perfect equilibrium of the type described in the result and by $\overline{x}$ the largest such alternative. (I assume that such alternatives exist; if they do not, choose alternatives close to the infimum and supremum of the set of alternatives that pass.) Given Step 1, we have $\underline{x} \leq x^* \leq \overline{x}$. Suppose, without loss of generality, that a majority of players prefer $\underline{x}$ to $\overline{x}$, so that, in particular, $x^* < \overline{x}$.

Step 2 *For any $\varepsilon > 0$ there exists $\overline{\delta} < 1$ such that if $\delta_i \geq \overline{\delta}$ for all $i \in N$ then $\overline{x} < x^* + \varepsilon$ and $\underline{x} > x^* - \varepsilon$.*

Proof. Consider a subgame perfect equilibrium in which $\overline{x}$ passes. Given that the set of players who vote for it is a majority, this set includes at least one player, say i, whose favorite position is at most x^*, so that $u_i(x^*) > u_i(\overline{x})$. Consider this player's choice to vote for or against $\overline{x}$. Her vote makes a difference to the outcome only if the other players' votes are split equally between *for* and *against*, in which case her voting *for* leads to the outcome $\overline{x}$ and her voting *against* leads to a period of delay followed by an outcome that is at least as good for her as x^* if she is selected to propose, given Step 1, and is no worse than $\overline{x}$ otherwise (because no equilibrium outcome is worse for her than that). Thus if her discount factor is close enough to 1 then unless $\overline{x}$ is close to x^* her action of voting against $\overline{x}$ weakly dominates her action of voting for it, so that she optimally votes against it and hence it is rejected.

Finally, the fact that a majority of players prefer $\underline{x}$ to $\overline{x}$ means that a player with favorite position x^* does so, and hence given the continuity of each payoff function, if $\overline{x}$ is close to x^* then so too is $\underline{x}$. $\lhd$

14.1.2 Equilibrium in stationary strategies

A restriction much stronger than the ones I have discussed so far is stationarity, which requires that each player's strategy be history-independent. That is, each player makes the same proposal whenever she is selected to be the proposer and casts the same vote regarding any given proposal regardless of the history that preceded the proposal.

Definition 14.4: Stationary strategy in bargaining game with voting

A strategy σ_i of a player in a bargaining game with voting $\langle N, X, (\rho_i)_{i \in N}, (\delta_i)_{i \in N}, (u_i)_{i \in N} \rangle$ is *stationary* if the actions it prescribes after any two histories h and h' are the same if the subgames following h and h' are the same: for each $i \in N$, $\sigma_i(h) = \sigma_i(h') \in X$ if h and h' are histories ending with chance selecting i to be the proposer, and, for each proposal $x \in X$, $\sigma_i(h) = \sigma_i(h') \in \{\mathit{for}, \mathit{against}\}$ if h and h' are histories ending with the proposal x.

The results I have presented so far rely on strategies that are not stationary: their essence is that they react to the players' past actions. No such reaction is possible for a player using a stationary strategy. As Austen-Smith and Banks (2005, 249) write,

> stationary strategies rule out much of interest in many political interactions: reciprocity, both positive and negative, is impossible without strategies being sensitive to the realized decision history.

If we interpret the actions specified by a player's strategy after a deviation to be the other players' beliefs about her future actions, then the stationarity of a player's strategy means that the other players' beliefs do not change when a deviation occurs. As Rubinstein (1991, 912) writes,

> assuming passivity of beliefs eliminates a great deal of what sequential games are intended to model: namely, the changing pattern in players' behavior and beliefs, as they accumulate experience.

Despite these assessments, significant effort has been devoted to the analysis of equilibria in stationary strategies. I close the chapter with a brief discussion of this analysis.

The result I present characterizes the subgame perfect equilibria with undominated voting of distributive bargaining games with voting in which every player is identical (same recognition probability and same discount factor) and every player's strategy is stationary. The result shows that in every such equilibrium the payoff profile is the same and agreement is reached immediately, so that each player's expected payoff is $1/n$ (given that the players are identical). Thus if a proposal is rejected and then the players adhere to an equilibrium, each player receives the payoff $1/n$ with one period of delay, which is worth δ/n to her, where δ is the common discount factor. Hence for every player i, voting in favor of a proposal x with $x_i > \delta/n$ weakly dominates voting against x, and voting against a proposal x with $x_i < \delta/n$ weakly dominates voting in favor of x. So to induce a majority to vote in favor of a proposal, it is enough for a player to offer δ/n to each member of a set containing $\frac{1}{2}(n-1)$ of the other players, assembling a minimal majority (including herself). In a stationary equilibrium, every player i assembles the same minimal majority S^i whenever she is selected to make a proposal. As far as i is concerned, the identity of the members of S^i does not matter, but for all the players' payoffs to be equal, as is the case in every equilibrium, every player j has to be a member of exactly $\frac{1}{2}(n+1)$ of these sets. I call a collection of minimal majorities with this property balanced.

Definition 14.5: Balanced collection of minimal majorities

Let $\langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a bargaining game with voting and let $n = |N|$. A *minimal majority* is a subset of N with $\frac{1}{2}(n+1)$ members. A collection $\{S^j\}_{j\in N}$ of n minimal majorities is *balanced* if each player is a member of $\frac{1}{2}(n+1)$ of the sets S^j.

For every (odd) value of n, a balanced collection of minimal majorities exists. One such collection is given by

$$S^j = \{j, (j+1)(\text{mod } n), \ldots, (j+\tfrac{1}{2}(n-1))(\text{mod } n)\} \quad \text{for all } j \in N,$$

so that $S^1 = \{1, 2, \ldots, \frac{1}{2}(n+1)\}$, $S^2 = \{2, 3, \ldots, \frac{1}{2}(n+3)\}$, and so forth.

Proposition 14.5: Stationary subgame perfect equilibria of distributive bargaining game

Let $\langle N, X, (\rho_i)_{i\in N}, (\delta_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a distributive bargaining game with voting and let $n = |N|$. Suppose that $\rho_i = 1/n$ for all $i \in N$ and there exists $\delta \in (0,1)$ such that $\delta_i = \delta$ for all $i \in N$. For every subgame perfect equilibrium with undominated voting in which every player's strategy is stationary there is a balanced collection $\{S^j\}_{j\in N}$ of minimal majorities with $j \in S^j$ for all $j \in N$ such that the strategy of each player i always makes the proposal x for which

$$x_j = \begin{cases} 1 - \frac{1}{2}\delta(n-1)/n & \text{if } j = i \\ \delta/n & \text{if } j \in S^i \setminus \{i\} \\ 0 & \text{otherwise} \end{cases}$$

and votes for a proposal y if and only if $y_i \geq \delta/n$. Each player's expected payoff in every equilibrium is $1/n$.

Proof

I first argue that a strategy profile satisfying the conditions in the result is a subgame perfect equilibrium with undominated voting. Let s^* satisfy the conditions in the result. The game satisfies the condition in Proposition 16.9, so that s^* is a subgame perfect equilibrium if and only if it satisfies the one-deviation property. To show that it satisfies this property, I consider each type of subgame in turn.

Subgame starting with proposal by player i

If i adheres to s_i^*, her proposal is accepted and her payoff is $1 - \frac{1}{2}\delta(n-1)/n$.

If she makes a proposal y with $y_i > 1 - \frac{1}{2}\delta(n-1)/n$ then it is rejected, because the total payoff to the other players is less than $\frac{1}{2}\delta(n-1)/n$ and hence fewer than $\frac{1}{2}(n-1)$ of the other players receive at least δ/n. With probability $1/n$ she is selected to be the proposer again in the next period, in which case her payoff is $1 - \frac{1}{2}\delta(n-1)/n$ (with one period of delay), and with probability $1/n$ each of the other players is selected to be the proposer, in which case her payoff is either δ/n or 0 (with one period of delay). Given that $\delta < 1$, we have $\delta/n < 1 - \frac{1}{2}\delta(n-1)/n$, so that i's payoff is less than $1 - \frac{1}{2}\delta(n-1)/n$.

If she makes a proposal y with $y_i < 1 - \frac{1}{2}\delta(n-1)/n$ then either it is accepted, in which case her payoff is y_i, or it is rejected, in which case her payoff is less than $1 - \frac{1}{2}\delta(n-1)/n$ by the argument for the previous case.

Subgame starting with vote regarding a proposal y

If y is rejected, the outcome, after one period of delay, is given in the following table for any player $i \in N$.

proposer	i's payoff	probability
i	$1 - \frac{1}{2}\delta(n-1)/n$	$1/n$
j with $i \in S^j$	δ/n	$\frac{1}{2}(n-1)/n$
j with $i \notin S^j$	0	$\frac{1}{2}(n-1)/n$

Thus every player's expected payoff if y is rejected is

$$\delta\left[(1/n)\left(1 - \tfrac{1}{2}\delta(n-1)/n\right) + \tfrac{1}{2}((n-1)/n)(\delta/n)\right] = \delta/n.$$

Hence voting for y if and only if $y_i \geq \delta/n$, as s_i^* prescribes, is consistent with s_i^* being undominated and optimal.

Now let s be a subgame perfect equilibrium with undominated voting in which each player's strategy is stationary. I argue that s takes the form given in the result.

Step 1 *For the strategy profile s and each $i \in N$, the expected payoffs of player i at the start of the game and at the start of each subgame following the rejection of a proposal are all the same.*

Proof. The conclusion follows from the definition of a stationary strategy and the fact that the game and every such subgame are identical. ◁

Denote by $V_i(s)$ the expected payoff of player i at the start of the game (and hence by Step 1 at the start of each subgame beginning with the selection of a proposer by chance) for the strategy profile s.

Step 2 *The proposal that s specifies for each player at the start of any subgame following her selection as the proposer is accepted and hence $\sum_{i \in N} V_i(s) = 1$.*

Proof. Suppose that the proposal that s_j specifies for player j is rejected. Then play moves to the next period, at the start of which the expected payoff of each player i is $V_i(s)$. Now suppose that j deviates from s_j and proposes that the share of each player i be $V_i(s)/\sum_{k=1}^{n} V_k(s)$. Given that $\sum_{k=1}^{n} V_k(s) \leq 1$, this share is at least $V_i(s)$, and hence is greater than $\delta V_i(s)$. Thus every player votes for the proposal, yielding j a payoff of at least $V_j(s)$, which is more than the payoff $\delta V_j(s)$ she gets if she makes a proposal that is rejected.

If $\sum_{i \in N} V_i(s) < 1$ then any player can increase the amount assigned to herself in her proposal without changing the amounts assigned to the other players. ◁

Step 3 *If x is a proposal for which $x_j > \delta V_j(s)$ then j's strategy s_j votes in favor of x, and if $x_j < \delta V_j(s)$ then s_j votes against x.*

Proof. If x fails then j's payoff is $\delta V_j(s)$, so her voting in favor of x weakly dominates her voting against it if $x_j > \delta V_j(s)$ and her voting against it weakly dominates her voting for it if $x_j < \delta V_j(s)$. ◁

Step 4 *For every player $i \in N$, $V_i(s) > 0$.*

Proof. Suppose that i is selected to be the proposer. For some set S containing half of the remaining players, $\sum_{j\in S} V_j(s) \le \frac{1}{2}$, so that there is a proposal x^i with $x^i_i > 0$ and $x^i_j > V_j(s)$ for all $j \in S$, and by Step 3 the strategy of every player $j \in S$ votes in favor of x^i. ◁

Step 5 *Let x^i be the proposal made by s^i whenever i is selected to be the proposer. If for some player $j \ne i$ we have $x^i_j = \delta V_j(s)$ then j's strategy s_j votes in favor of x^i.*

Proof. Suppose that $x^i_j = \delta V_j(s)$. By Step 4, $x^i_j > 0$ and by Step 2, x^i passes. Suppose that j's strategy s_j votes against x^i. If $x^i_k < \delta V_k(s)$ for some player k then her voting against x^i weakly dominates her voting for it, so $x^i_k \ge \delta V_k(s)$ for every player k who votes in favor of x^i. Then i can increase her payoff by deviating to a proposal y^i in which $y^i_j = 0$ and $y^i_l > x^i_l$ for every player $l \ne j$, which passes because, by Step 3, every player who voted for x^i optimally votes for y^i. So s_j votes in favor of x^i. ◁

Step 6 *$V_i(s)$ is the same for every player $i \in N$.*

Proof. An optimal proposal of any player minimizes the amount allocated to the other players among the proposals that a majority of players vote in favor of. By Step 3 and Step 5, every player j votes in favor of a proposal that gives her at least $\delta V_j(s)$ and against one that gives her less than $\delta V_j(s)$, so every optimal proposal of player i gives $\delta V_j(s)$ to each member j of a minimal majority S^i for which $\sum_{j\in S^i\setminus\{i\}} V_i(s)$ is smallest, and nothing to the players outside S^i.

Suppose that the values of $V_j(s)$ are not the same for all $j \in N$.

If the number of players tied for the smallest value of $V_j(s)$ is at least $\frac{1}{2}(n+1)$, then none of the remaining players is included in any set S^i, so that every such player obtains a positive payoff only if she is the proposer. Thus the payoff of every such player k is $V_k(s) = (1/n)\big(1-\delta\sum_{l\ne k} V_l(s)\big) < 1/n$. Hence $V_j(s) < 1/n$ for every player, contradicting $\sum_{j\in N} V_j(s) = 1$ (from Step 2).

If the number of players tied for the smallest value of $V_j(s)$ is at most $\frac{1}{2}(n-1)$, then all of these players are included in every set S^i. Thus for every such player k, we have $V_k(s) = (1/n)\big(1-\delta\sum_{l\ne k} V_l(s)\big) + ((n-1)/n)\delta V_k(s)$, so that, given $\sum_{l\ne k} V_l(s) = 1 - V_k(s)$, we have $V_k(s) = 1/n$. That is, the smallest value of $V_j(s)$ is $1/n$, and hence $V_j(s) = 1/n$ for all $j \in N$, contradicting the assumption that the value of $V_j(s)$ are not the same for all $j \in N$. ◁

Step 7 $V_i(s) = 1/n$ *for all* $i \in N$.

Proof. By Step 2, $\sum_{j\in N} V_j(s) = 1$, so the result follows from Step 6. ◁

Step 7 and the optimality of s_i implies that the proposal made by s_i gives δ/n to each member of a set of $\frac{1}{2}(n-1)$ players and hence $1-\frac{1}{2}(n-1)\delta/n$ to i. Thus the expected payoff of each player j under s is

$$(1/n)\big(1-\tfrac{1}{2}(n-1)\delta/n\big) + (K/n)\delta/n,$$

where K is the number of players who include j in the set to whose members they give δ/n. For this expected payoff to equal $1/n$, as it must, $K = \frac{1}{2}(n-1)$. Hence the collection of minimal majorities to which the players' proposals assign positive payoffs is balanced, so that s takes the form in the result.

Comments

Equilibrium outcome The outcome of an equilibrium is that the player selected to make the first proposal offers δ/n to each member of a set containing $\frac{1}{2}(n-1)$ of the other players, who, together with her, make up a minimal majority (also called a minimal winning coalition); she takes the remainder of the pie. All members of the minimal majority vote in favor of this proposal, and hence the game ends; the payoff of each player outside the minimal majority is zero.

Payoffs A player's equilibrium payoff in a subgame that starts with her making a proposal, $1-\frac{1}{2}\delta(n-1)/n$, exceeds her equilibrium payoff in a subgame that starts with another player making a proposal, δ/n. If δ is close to 1 then the former payoff is close to $(n+1)/2n$, which is close to $\frac{1}{2}$ when n is large: the proposer gets almost half of the pie when the number of players is large.

Each player's payoff function u_i is assumed to be linear a distributive bargaining game with voting. If the payoff functions are not linear, the stationary equilibrium payoffs may not be unique (Banks and Duggan 2000, Example 4).

Discount factors Differences among the players' discount factors have two effects on an equilibrium. On the one hand, a player with a high discount factor finds it less costly to wait for a chance to be the proposer, which puts her in an advantageous position. On the other hand, she unattractive as a coalition partner, because her expected payoff if a proposal is rejected is high, so that she needs a high payoff to be induced to vote for a proposal. In the next exercise you are asked to show that in an example of a three-player distributive bargaining game with voting in which players 1 and 2 have the same discount factor and player 3 has a higher discount factor, the common equilibrium payoff of players 1 and 2 is greater than the equilibrium payoff of player 3.

Exercise 14.2: More patient players may obtain lower payoffs

Consider a distributive bargaining game with voting with three players in which the discount factors of players 1 and 2 are both δ and that of player 3 is $\delta' > \delta$, and every player's recognition probability is $\frac{1}{3}$. In every subgame perfect equilibrium with undominated voting in which each player's strategy is stationary, agreement is reached immediately and the payoffs of players 1 and 2 are the same. When player 3 is selected to be the proposer, she needs one of the two remaining (identical) players to vote in favor of her proposal for the proposal to be accepted. In a stationary equilibrium, she selects each of them with probability $\frac{1}{2}$. (This assumption extends the definition of a distributive bargaining game with voting, which restricts voting to be deterministic.) Show that the equilibrium payoff of player 3 is less than the common equilibrium payoffs of players 1 and 2.

Recognition probabilities Suppose that the players' discount factors are the same but their recognition probabilities differ. If i's recognition probability is at least as high as j's, then i's stationary equilibrium payoff is as least as high as j's (Eraslan 2002, Corollary 1). Figure 14.5 shows the limit, as the common discount factor approaches 1, of the stationary equilibrium payoffs of player 1 in a three-player distributive bargaining game with voting as a function of the recognition probabilities. Note that player 1's equilibrium payoff is $\frac{1}{3}$ whenever all the recognition probabilities are at most $\frac{1}{2}$, and hence by symmetry the other players' equilibrium payoffs are $\frac{1}{3}$ in these cases too. Among these cases are ones in which one player's recognition probability is close to zero and the other players' probabilities are both close to $\frac{1}{2}$.

Voting rules The characterization in Proposition 14.5 can be generalized to a game in which the votes of any number q of the players are needed for a proposal to be accepted, rather than $\frac{1}{2}(n+1)$. If $q = n$, so that every player's vote is needed for acceptance, and every player has the same discount factor and same recognition probability, then each player's expected payoff in a stationary subgame equilibrium with undominated voting is her recognition probability. Here is a rough argument for this result. Assume that the game has unique stationary equilibrium payoffs and that agreement is reached immediately in any stationary

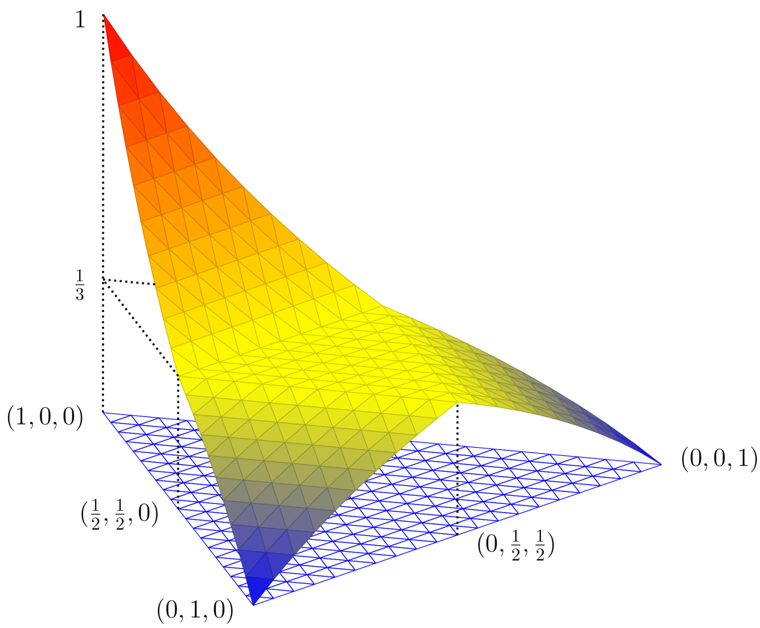


Figure 14.5 The limit of the payoffs of player 1 in a stationary subgame perfect equilibrium with undominated voting of a three-player distributive bargaining game with voting as the players' common discount factor approaches 1, as a function of the vector (ρ_1, ρ_2, ρ_3) of recognition probabilities. (Constructed from Table 1 (p. 12) of Imai and Salonen 2012.)

equilibrium. Denote the equilibrium payoff of each player i by v_i. If i is selected to be the proposer, she offers δv_j to every other player j to induce her to vote in favor of her proposal, leaving $1 - \delta \sum_{j \in N \setminus \{i\}} v_j = 1 - \delta(1 - v_i)$ for herself, and if another player is selected to be the proposer, that player offers δv_i to player i. Thus player i's expected payoff at the start of the game is

$$v_i = p_i(1 - \delta(1 - v_i)) + (1 - p_i)\delta v_i.$$

Solving this equation for v_i we obtain $v_i = p_i$. Note that the result implies a stark difference between a player's equilibrium payoff under majority rule and her payoff under unanimity rule for some profiles of recognition probabilities. For example, if $n = 3$ and the recognition probabilities are $(\varepsilon, \frac{1}{2}(1-\varepsilon), \frac{1}{2}(1-\varepsilon))$ for some small $\varepsilon > 0$, then for δ close to 1, player 1's payoff is ε under the unanimity rule and close to $\frac{1}{3}$ under majority rule.

Exercise 14.3: Game with deterministic rotating recognition and voting

Consider a variant of a distributive bargaining game with voting with three players in which, following a proposal by any player i, the remaining players vote sequentially. First player $(i+1)(\text{mod } 3)$ votes. If she votes in favor, the proposal is accepted. Otherwise, player $(i+2)(\text{mod } 3)$ votes. If she votes in favor, the proposal is accepted, and otherwise player $(i+1)(\text{mod } 3)$ makes a proposal. Play continues until a proposal is accepted. Show that the game has a stationary subgame perfect equilibrium in which each player's proposal gives her all the pie. (This equilibrium is the only stationary subgame perfect equilibrium of the game.)

14.2 Recurrent distributive bargaining game with voting

A bargaining game with voting ends when the individuals reach agreement on an alternative; no player has an opportunity to reopen the negotiations. In this section I briefly consider a model in which negotiations are always open.

Time is discrete, starting with period 1. In each period, one unit of a good is available to a finite set N of individuals; it may be distributed in any way among the individuals, and some of it may be wasted. An outcome is a sequence $(x^1, x^2, \ldots)$ of distributions of the good among the individuals, where for each $t = 1, 2, \ldots$ we have $x_i^t \in [0,1]$ for all $i \in N$ and $\sum_{i \in N} x_i^t \leq 1$.

In each period, there is a default distribution of the good. In period 1 this distribution is $x^0 = (0, \ldots, 0)$ and in every subsequent period t it is the distribution x^{t-1} from the previous period. At the start of each period t, an individual is selected by chance. She either chooses *pass*, in which case $x^t = x^{t-1}$ and the period ends, or proposes a distribution different from x^{t-1}, in which case all individuals vote simultaneously for or against the proposal. If a majority of the individuals vote for the proposal, it becomes the distribution x^t in period t, and otherwise the distribution in period t is x^{t-1}.

For each individual $i \in N$, let $\delta_i \in (0,1)$ and let $u_i : [0,1] \to \mathbb{R}$ be an increasing function. The preferences of individual i regarding lotteries over outcomes (sequences of distributions) are represented by the expected value of the discounted average of the sequence $(u_i(x_i^1), u_i(x_i^2), \ldots)$ for the discount factor δ_i, namely

$$(1-\delta_i)\sum_{t=1}^{\infty} \delta_i^{t-1} u_i(x_i^t).$$

Note that if $x_i^t = x_i$ for all t then this discounted average is $u_i(x_i)$, because $\sum_{t=1}^{\infty} \delta_i^{t-1} = 1/(1-\delta_i)$. I refer to a game defined in this way as a *recurrent distributive bargaining game with voting*.

When an individual makes a proposal in such a game, the default distribution for the period is relevant, because it will be the distribution in the current period if the individual's proposal is voted down; when an individual casts a vote, both the default distribution and the proposal are relevant for the same reason. Thus for this game I define a stationary strategy of any individual i to be a pair (s_i^p, s_i^v) of functions, with $s_i^p : X \to \{pass\} \cup X$ specifying i's proposal, as a function of the default distribution, whenever she is selected to be the proposer, and $s_i^v : X \times X \to$ $\{for, against\}$ specifying her vote, as a function of the default distribution and the proposal, whenever a ballot is held.

Unlike a distributive bargaining game with voting, a recurrent distributive bargaining game with voting has many subgame perfect equilibria with undominated voting in which every player's strategy is stationary, and in some of these equilibria some of the available good is wasted. These features are demonstrated well by the following three-player example.

Let $N = \{1,2,3\}$ and suppose that each player has the same recognition probability ($\rho_i = \frac{1}{3}$ for all $i \in N$) and discount factor ($\delta_1 = \delta_2 = \delta_3 = \delta$), and $u_i(x_i) = x_i$ for all $i \in N$ and all $x_i \in [0,1]$. The strategy profile I define is based on three distributions,

$$y^1 = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{6}), y^2 = (\tfrac{1}{3}, \tfrac{1}{6}, \tfrac{1}{3}), \text{ and } y^3 = (\tfrac{1}{6}, \tfrac{1}{3}, \tfrac{1}{3}).$$

Let $Y = \{y^1, y^2, y^3\}$. Here is the strategy of player i.

Proposal of player i

$$s_i^p(x) = \begin{cases} pass & \text{if } x \in Y \\ y^i & \text{if } x \notin Y. \end{cases}$$

Vote of player i

$$s_i^v(x,z) = \begin{cases} for & \text{if } x \in Y \text{ and } x_i = \frac{1}{6} \\ & \quad \text{or } x \notin Y,\, z \in Y, \text{ and } z_i \geq (1-\delta)x_i + \frac{5}{18}\delta \\ & \quad \text{or } x \notin Y,\, z \notin Y, \text{ and } z_i \geq x_i \\ against & \text{otherwise.} \end{cases}$$

Notice that some of the good is wasted in the distributions y^1, y^2, and y^3: in each case, only $\frac{5}{6}$ of the unit is assigned to the individuals. Notice also that if the default distribution is y^i then a proposal z that gives each player an amount of the good that is larger than the amount she gets in y^i is rejected because the two players whose share in y^i is $\frac{1}{3}$, say i and j, vote against it. Why do they do that? Because the acceptance of z would reopen negotiations. The player k chosen to be the proposer in the next period would propose y^k, and, if δ is close enough to 1, this proposal would be accepted, which means that rather than being assured of the share $\frac{1}{3}$ in every future period, as they are when z is rejected, i and j would face a future in which each of their shares is $\frac{1}{3}$ in every future period with probability $\frac{2}{3}$ but only $\frac{1}{6}$ in every future period with probability $\frac{1}{3}$. Given that a proposal that improves upon y^i for every player is rejected, a player who makes such a proposal reaps no benefit from doing so.

In the outcome of this strategy profile, the player chosen in period 1 to be the proposer, say i, selects y^i. This alternative gives her and one of the other players, say j, the fraction $\frac{1}{3}$ of the good. The default amount x_k^0 for each player k in period 1 is 0 and $\frac{1}{3} > \frac{5}{18}$, so players i and j vote for y^i and thus $x^1 = y^i$. In every subsequent period the player chosen to be the proposer passes, so that the distribution remains y^i. Thus the outcome of the strategy profile is the lottery in which the distribution is y^1 in every period, or y^2 in every period, or y^3 in every period, each with probability $\frac{1}{3}$. I denote this lottery by ξ. The payoff of each player for ξ is

$$\tfrac{2}{3} \cdot \tfrac{1}{3} + \tfrac{1}{3} \cdot \tfrac{1}{6} = \tfrac{5}{18}.$$

I now argue that the strategy profile is a subgame perfect equilibrium with undominated voting if $\delta \geq \frac{12}{13}$. The game satisfies the condition in Proposition 16.9, so that the strategy profile is a subgame perfect equilibrium if and only if it satisfies the one-deviation property. Here is an argument that it satisfies this property and that no player's vote after any history is weakly dominated.

Subgame following selection of player i when default distribution is x

Suppose that $x \in Y$. If i follows her strategy, the distribution is x in every subsequent period, so that i's payoff in the remainder of the game is x_i. If i deviates from her strategy and makes a proposal different from x, then only the single player for whom $x_i = \frac{1}{6}$ votes *for*, so x remains the distribution in every subsequent period, as it does if she follows her strategy.

Suppose that $x \notin Y$. If i follows her strategy, she proposes y^i, which assigns $\frac{1}{3}$ to her and to one of the other players. For at least two players j, we have $x_j \leq \frac{1}{2}$, so that $(1-\delta)x_j + \frac{5}{18}\delta \leq \frac{1}{2} - \frac{4}{18}\delta \leq \frac{1}{2} - \frac{4}{18}\frac{12}{13} = \frac{23}{78} < \frac{1}{3}$. Thus at least two players vote for y^i, so that it becomes the default distribution and hence, if all players follow their strategies subsequently, the distribution in every subsequent period. If i deviates from her strategy and proposes a member of Y different from y^i, at least two players vote for it by the same argument, so that it becomes the distribution in every subsequent period. Thus i does not benefit from the deviation. Finally, if i deviates from her strategy and proposes a distribution z not in Y, whether or not it is accepted the outcome subsequently is the lottery ξ. If she adheres to her strategy she receives $\frac{1}{3}$ in every future period, so regardless of the value of z, she is worse off if she deviates.

Subgame starting with vote on z when default distribution is x

Suppose that $x \in Y$ and $x_i = \frac{1}{6}$. If i follows her strategy, she votes for z. The other two players vote against z, so the distribution in the period is x, and that distribution persists in every future period. If i deviates from her strategy and votes against z and the other players' votes are such that her vote affects the outcome, i's deviation changes the distribution in the period from z to x, which means that the outcome in subsequent periods changes from the lottery ξ to the certain distribution in which she receives $\frac{1}{6}$ in every period. Thus i's voting for z is optimal and undominated.

Suppose that $x \in Y$ and $x_i = \frac{1}{3}$. If i follows her strategy, she votes against z, as does one of the other players, so the distribution in the period is x, which persists in every future period. An argument analogous to the one for the case in which $x_i = \frac{1}{6}$ shows that i is made worse off by a deviation that affects the outcome, so that her voting against z is optimal and undominated.

Suppose that $x \notin Y$ and $z \in Y$. If z is accepted, the distribution is z in every subsequent period, so that i receives the amount z_i in every period and hence the payoff z_i. If it is rejected, then i receives x_i in the current period and then the lottery ξ, and hence the payoff $(1-\delta)x_i + \frac{5}{18}\delta$. Thus i's voting for z is optimal and undominated if $z_i \geq (1-\delta)x_i + \frac{5}{18}\delta$ and her voting against it is optimal and undominated if the reverse inequality holds.

Suppose that $x \notin Y$ and $z \notin Y$. If z is accepted, the distribution is z in the current period and then the lottery ξ. If it is rejected, then i receives x_i in the current period and then the lottery ξ. Thus i's voting for z is optimal and undominated if $z_i \geq x_i$ and her voting against it is optimal and undominated if $z_i \leq x_i$.

This equilibrium is a member of a large class of equilibria: the strategy profile obtained by replacing y^1, y^2, and y^3 with $(\alpha_1, \alpha_2, \beta_3)$, $(\alpha_1, \beta_2, \alpha_3)$, and $(\beta_1, \alpha_2, \alpha_3)$ for any $(\alpha_1, \alpha_2, \alpha_3)$ and $(\beta_1, \beta_2, \beta_3)$ with $\alpha_i \in [0,1]$, $\beta_i \in [0,1]$, $\alpha_i > \beta_i$ for $i = 1$, 2, 3, and $\alpha_1 + \alpha_2 + \beta_3 \leq 1$, $\alpha_1 + \beta_2 + \alpha_3 \leq 1$, and $\beta_1 + \alpha_2 + \alpha_3 \leq 1$ is an equilibrium for values of δ sufficiently close to 1. These equilibria differ qualitatively from the equilibrium of a distributive bargaining game with voting, as characterized in Proposition 14.5.

- The set of players among whom the pie is shared may be larger than a minimal majority. If $\beta_i > 0$ for $i = 1, 2, 3$, the player chosen to be the first proposer offers a positive amount to every player, and this proposal is accepted.
- Some of the pie may be wasted. In the equilibria in which $\alpha_i = \frac{1}{6}$ and $\beta_i = \frac{1}{3}$ for $i = 1$, 2, 3, for example, the player chosen to be the first proposer distributes only $\frac{5}{6}$ of the pie.

- The game has multiple stationary equilibrium outcomes. In fact, for any distribution (x_1, x_2, x_3) with $x_i > 0$ for $i = 1, 2, 3$, there is a number $\delta^* \in (0, 1)$ such that if $\delta \geq \delta^*$ then the game has a stationary equilibrium in which the first player selected by chance proposes (x_1, x_2, x_3), which remains the alternative in every future period.

Anesi and Seidmann (2015) show that the example may be extended to games with more than three players and to voting rules that require more than a bare majority of the players to vote in favor for a proposal to be accepted, as long as the number of votes required is less than the number of players. (They show also that under unanimity rule, by contrast, the game has a unique stationary equilibrium payoff, which coincides with the stationary equilibrium payoff of a distributive bargaining game with voting.)

Notes

The model of a bargaining game with voting is due to Baron and Ferejohn (1989) and Banks and Duggan (2000). Proposition 14.2 is based on Cho and Duggan (2015), Propositions 14.3 and 14.5 are based on Baron and Ferejohn (1989), and Proposition 14.4 is based on Cho and Duggan (2009). My exposition draws also on Austen-Smith and Banks (2005, Section 6.2) and Eraslan (2002)[1]. Section 14.2 is based on Anesi and Seidmann (2015).

The example in Exercise 14.2 is due to Colin Stewart. Exercise 14.3 is based on Ali et al. (2019, Example 3).

Solutions to exercises

Exercise 14.1

If $x \neq (1, 0, 0)$ and player 3 proposes $y = (1, 0, 0)$ in state x, then $M^3(y) = \{1, 2\}$, and player 1's strategy calls for her to vote against y. But her doing so is not optimal: if she votes for y then her payoff is 1 and if she votes against y then her payoff is at most $\delta_1 < 1$. (If there are at least five players, $x \neq (1, 0, \ldots, 0)$, and player $k \neq 1$ proposes $y = (1, 0, \ldots, 0)$, then every set $M^k(y)$ contains only players whose payoffs in y are 0.)

Exercise 14.2

Denote the common equilibrium payoff of players 1 and 2 by v and that of player 3 by v'. Suppose that $v \leq v'$, so that $\delta v < \delta' v'$. Then when a player is

[1] Note that Corollary 2 in Eraslan (2002), which is stated also on p. 217 of Austen-Smith and Banks (2005), is incorrect, as Exercise 14.2 shows.

the proposer, she offers δv to either player 1 or player 2 and keeps the rest, so that her payoff is $1-\delta v$. Player 3 is never offered a positive payoff by another proposer, so her equilibrium payoff is $v' = \frac{1}{3}(1-\delta v)$. Player 1 needs to give δv to player 2 if she is the proposer, keeping $1-\delta v$ for herself, gets δv if player 2 is the proposer, and gets δv with probability $\frac{1}{2}$ if player 3 is the proposer. Thus her equilibrium expected payoff is $v = \frac{1}{3}(1-\delta v) + \frac{1}{3}\delta v + \frac{1}{6}\delta v = \frac{1}{3} + \frac{1}{6}\delta v > \frac{1}{3}$, which is inconsistent with $v \leq v'$. Thus $v > v'$.

Exercise 14.3

Consider the strategy profile in which every player, whenever she is chosen to be the proposer, proposes that she get the entire pie, and whenever she has to vote on a proposal, she votes in favor, regardless of the proposal. By the following argument, this strategy profile satisfies the one-deviation property, so that by Proposition 16.9 it is a subgame perfect equilibrium. Consider player 1.

Subgame starting with proposal

Player 1 cannot propose that she gets more; a proposal to get less is accepted, and player 1 is worse off.

Subgame following a proposal x by player 3

If player 1 accepts x then she gets x_1. If she rejects x then player 2 accepts it, and player 1 still gets x_1. So player 1 optimally accepts x.

Subgame following rejection by player 3 of proposal x by player 2

If player 1 accepts x then she gets x_1. If she rejects x then player 3 proposes $(0,0,1)$, which is accepted, so that player 1 gets 0. Thus player 1 optimally accepts x.

The arguments for players 2 and 3 are the same.

15 Rulers threatened by rebellion

15.1 Threat of revolt 494
15.2 Coordinating rebellion 501

Some societies are ruled by a dictator or an unchanging group of leaders. This chapter discusses two models intended to study the effect on a dictator's behavior of the threat of rebellion by the masses and the circumstances under which the masses can coordinate their actions and mount a successful rebellion.

Synopsis

Section 15.1 considers the choice of a dictator between carrying out policies that benefit the masses and handing over control to them. In the model, a wealthy dictator interacts with a mass of poor individuals over an infinite sequence of periods. In each period, the dictator decides how much of the wealth available in that period to give to the poor. In some periods, the environment is conducive to revolt. In those periods, the dictator has the option to hand over control to the mass of poor individuals, and if she does not do so those individuals have the option to revolt. Proposition 15.1 shows that if the probability that the environment is conducive to revolt is below some threshold then the game has a unique subgame perfect equilibrium. In this equilibrium, the dictator democratizes when the environment is conducive to revolt and gives no wealth to the poor when the environment is not conducive to revolt. If, when the environment is conducive to revolt she were not to democratize, the poor would revolt. When the probability that the environment is conducive to revolt is low, the dictator's expected payoff if she follows this strategy is high and is independent of the strategy of the poor, putting a lower bound on the dictator's payoff in any subgame perfect equilibrium. So when the environment is conducive to revolt, the poor optimally take that opportunity, knowing that if they do not then even if the dictator gives them all the wealth in that period, their best payoff in every equilibrium of the remainder of the game is low.

Section 15.2 studies the coordination problem the masses face when deciding whether to revolt. In the model, in each period in an infinite sequence a dictator chooses an action and then each member of a large set of citizens decides whether to rebel. The probability that a rebellion is successful is an increasing function of the fraction of citizens who rebel. If the citizens observe the dictator's action, they can coordinate their actions. In this case, Proposition 15.2 describes some equilibria and specifies the highest payoff the citizens can achieve in any equilibrium. If the citizens observe only a noisy signal of the dictator's action, then perfect coordination is not possible. Proposition 15.3 specifies the best equilibria for the citizens in this case, which are worse than the best equilibria when the dictator's action is observed. The remainder of the section analyzes informally a variant of the model, in which a dictator has the option to call an election. Her decision to exercise this option may be used by the citizens to coordinate their actions and in doing so achieve a payoff as high as in the best equilibrium for them in the model in which they observe the dictator's action.

15.1 Threat of revolt

Consider a society consisting of a small rich elite and a mass of poor individuals. Initially, the rich elite rules. Suppose that the poor individuals have the option of staging a revolution, after which they will have access to the entire wealth of the society minus an amount that is destroyed during the revolution. Two options of the elite that may forestall a revolution are to transfer wealth to the poor individuals and to transfer control of the distribution of wealth to these individuals, an action I refer to as democratization. Under what circumstances will the elite follow each of these strategies? The model I present studies the idea that if environments conducive to revolt are rare, the poor have an incentive to take such opportunities when they arise, inducing rulers to forestall revolt by preemptively democratizing.

15.1.1 Model

A society consists of two players, *Rich*, representing a small group of wealthy individuals, and *Poor*, representing a large number of impecunious individuals. One unit of wealth is available; *Rich* initially controls the distribution of this unit between *Rich* and *Poor*. *Rich* can give some wealth to *Poor*, in which case *Poor* may accept the offer or rebel, or can hand over control of the distribution of wealth to *Poor*. If *Poor* rebels, the fraction v of wealth is lost and all of the remainder is taken by *Poor*. If *Rich* hands over control to *Poor*, *Rich* gets the fraction d of wealth and *Poor* gets the remainder. (Perhaps *Rich* owns all the capital, and will

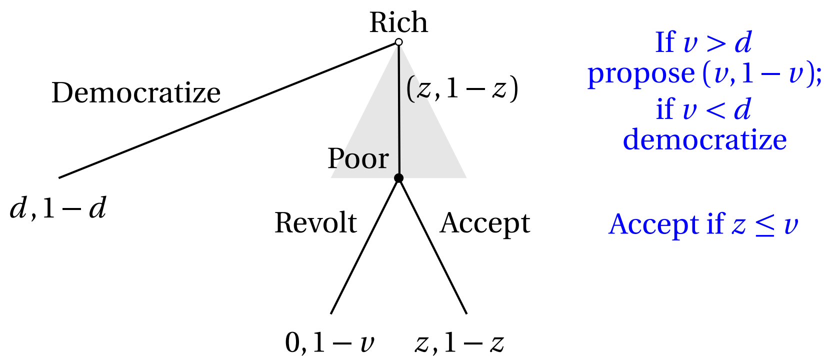


Figure 15.1 An extensive game with perfect information in which *Rich* initially controls the distribution of wealth. At the start of the game, its options are to democratize, in which case it receives the amount d, and to offer some amount to *Poor*, which may accept the offer or revolt, in which case some wealth is destroyed and *Rich* obtains nothing. (The diagram shows only one of the possible amounts *Rich* can offer to *Poor*.) In each payoff pair, the first number is the payoff of *Rich*.

not cooperate unless it receives at least the fraction d of wealth, although this strategic consideration is not part of the model.)

An extensive game with perfect information that models this society is given in Figure 15.1. If *Poor* revolts, it obtains the payoff $1 - v$, so in any subgame perfect equilibrium it accepts a proposed distribution $(z, 1-z)$ only if $z \leq v$. Thus if $v > d$ then *Rich* proposes $(v, 1-v)$, which *Poor* accepts, and if $v < d$ then *Rich* democratizes, obtaining the payoff d, which exceeds the best payoff for *Rich* that *Poor* accepts. If $v = d$, both equilibria exist. Hence in a subgame perfect equilibrium *Rich* democratizes only if the payoff of *Poor* under democracy is at most its payoff under revolution. However, at least in the case that this inequality between the payoffs is strict, the model seems to artificially restrict the actions of *Poor*. If it is better off under revolution than under democracy, why does it not have the option to revolt under democracy? If it has this option, it will take it, and in the only subgame perfect equilibrium of the resulting game, *Rich* offers *Poor* $1 - v$, which it accepts. That is, only in the singular case in which $v = d$ does the game have an equilibrium in which *Rich* democratizes; the game fails to capture the idea that *Rich* might democratize to stave off revolt.

An extension of the game addresses these shortcomings. In the new model, play takes place over an infinite sequence of periods, rather than in a single period, and sometimes the environment is conducive to revolt while sometimes it is not. The payoff of *Poor* after a revolt is assumed to be less than its payoff under democracy, so it has no incentive to revolt once democracy is established.

The game is illustrated in Figure 15.2. In each period, with probability $1 - q$

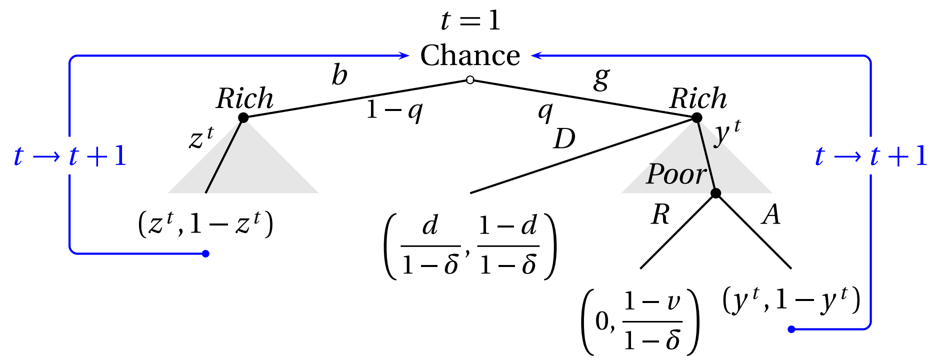


Figure 15.2 A dynamic game of revolt (see Definition 15.1). In each payoff pair, the first number is the payoff of *Rich*.

the environment is not conducive to revolt (bad, b). In this case, *Rich* chooses a distribution of wealth in the period and play moves to the next period. With probability q, the environment *is* conducive to revolt (good, g), in which case *Rich* has two types of action. One option is to democratize (D), resulting in the payoff pair $(d, 1-d)$ in every subsequent period and no further strategic options for either player. The other option is to propose some distribution $(z, 1-z)$ of wealth in the period; *Poor* can either accept this proposal or revolt. If *Poor* accepts the proposal then the payoff pair in the period is $(z, 1-z)$ and play continues to the next period. If *Poor* revolts, then the payoff pair is $(0, 1-v)$ in every subsequent period and neither player has any further strategic options; we assume $v > d$. Each player has the same discount factor, $\delta \in (0,1)$. Here is a precise definition of the game.

Definition 15.1: Dynamic game of revolt

A *dynamic game of revolt* $\langle\{Rich, Poor\}, q, \delta, v, d\rangle$, where

- $q \in (0,1)$ (the probability that the environment is conducive to revolt)
- $\delta \in (0,1)$ (the discount factor)
- $v \in [0,1)$ (the amount of per-period wealth destroyed by a revolt)
- $d \in [0, v)$ (the amount of wealth obtained by *Rich* each period under democracy)

is an extensive game with perfect information and chance moves with the following components.

Players
Rich and *Poor*

Terminal histories

The set of terminal histories is the set of all sequences

- (g, D) and $(e^1, e^2, \ldots, e^t, g, D)$ for all $t \geq 1$
- (g, y, R) for all $y \in [0,1]$ and $(e^1, e^2, \ldots, e^t, g, y^{t+1}, R)$ for all $t \geq 1$ and all $y^{t+1} \in [0,1]$
- $(e^1, e^2, \ldots)$

where $e^\tau = (b, z^\tau)$ or (g, y^τ, A) for some $z^\tau \in [0,1]$ and $y^\tau \in [0,1]$, for every $\tau \geq 1$.

Player function

Chance is assigned to the initial history and to every history ending with (b, z^t) for some $z^t \in [0,1]$ or (g, y^t, A) for some $y^t \in [0,1]$. *Rich* is assigned to every history ending with b or g, and *Poor* is assigned to every history ending with (g, y^t) for some $y^t \in [0,1]$.

Chance probabilities

Chance selects g with probability q and b with probability $1-q$, independent of history, whenever it moves.

Preferences

The preferences of *Rich* are represented by the payoff function

$$\begin{cases} \sum_{\tau=1}^{t} \delta^{\tau-1} x^\tau + \delta^t \dfrac{d}{1-\delta} & \text{if } h = (e^1, e^2, \ldots, e^t, g, D) \text{ for some } t \\ \sum_{\tau=1}^{t} \delta^{\tau-1} x^\tau & \text{if } h = (e^1, e^2, \ldots, e^t, g, y^{t+1}, R) \text{ for some } t \\ \sum_{\tau=1}^{\infty} \delta^{\tau-1} x^\tau & \text{if } h = (e^1, e^2, \ldots) \end{cases}$$

and those of *Poor* are represented by the payoff function

$$\begin{cases} \sum_{\tau=1}^{t} \delta^{\tau-1}(1-x^\tau) + \delta^t \dfrac{1-d}{1-\delta} & \text{if } h = (e^1, e^2, \ldots, e^t, g, D) \text{ for some } t \\ \sum_{\tau=1}^{t} \delta^{\tau-1}(1-x^\tau) + \delta^t \dfrac{1-v}{1-\delta} & \text{if } h = (e^1, e^2, \ldots, e^t, g, y^{t+1}, R) \text{ for some } t \\ \sum_{\tau=1}^{\infty} \delta^{\tau-1}(1-x^\tau) & \text{if } h = (e^1, e^2, \ldots) \end{cases}$$

where $x^\tau = z^\tau$ if $e^\tau = (b, z^t)$ and $x^\tau = y^\tau$ if $e^\tau = (g, y^\tau, A)$.

15.1.2 Subgame perfect equilibrium

When environments conducive to revolt are relatively rare (q is small), the game has a unique subgame perfect equilibrium, in which *Rich* democratizes whenever the environment is good (for revolt) and takes all the wealth when the environment is bad, and *Poor* revolts whenever *Rich* does not democratize. The outcome is that *Rich* democratizes on the first occurrence of a good environment.

In this equilibrium, unlike in the equilibria of the static model, *Rich* democratizes to stave off revolt. Given that *Rich* gives *Poor* no wealth when the environment is bad, even if *Rich* offers *Poor* all the wealth when the environment is good, *Poor* optimally revolts, because the expected time before the environment is good again is large. As a consequence, whenever the environment is good for revolt, *Rich* optimally democratizes to avoid being impoverished by revolution.

Here is the argument that the game has no other subgame perfect equilibrium. If *Rich* takes all the wealth when the environment is bad and democratizes when it is good its payoff is independent of the strategy of *Poor*, so this payoff is a lower bound for its payoff in a subgame perfect equilibrium. This payoff is relatively large when the probability q of a good environment is small. Now, in each period the total wealth available is at most 1, so the lower bound for the equilibrium payoff of *Rich* implies an upper bound for the equilibrium payoff of *Poor*, which is relatively small when q is small. As a consequence, if q is small, *Poor* revolts regardless of how much wealth *Rich* offers it in a period in which the environment is good, because even if *Rich* offers it all the wealth in such a period its expected payoff from accepting the offer and then getting the discounted value of its highest payoff in the game is less than its payoff if it revolts. Finally, given that *Poor* revolts, *Rich* optimally democratizes and, because its payoff is thus independent of the strategy of *Poor*, takes all of the wealth when the environment is bad.

Proposition 15.1: Subgame perfect equilibrium of dynamic game of revolt with democratization

Let $\langle\{Rich, Poor\}, q, \delta, v, d\rangle$ be a dynamic game of revolt with $0 < d < v < \delta$ and let $q^* = (1-\delta)(\delta - v)/(\delta(1-\delta+v-d))$. If $q \leq q^*$ then the following strategy pair is a subgame perfect equilibrium and if $q < q^*$ then it is the only subgame perfect equilibrium.

- The strategy of *Rich* assigns 1 to every history ending in b and D to every history ending in g. (*Rich* takes all the wealth whenever the environment is bad for revolt and democratizes whenever it is good for revolt.)

- For every number $y \in [0,1]$, the strategy of *Poor* assigns R to every history ending in (g, y). (If, when the environment is good for revolt, *Rich* does not democratize, *Poor* revolts regardless of how much wealth *Rich* offers it.)

Proof

The payoffs satisfy the condition in Proposition 16.9, so a strategy pair is a subgame perfect equilibrium if and only if it satisfies the one-deviation property.

I first argue that the strategy pair in the result satisfies the one-deviation property.

Action of *Rich* after history ending in b

If *Rich* deviates from 1 to any other number then it obtains less in the period of its deviation and the same in every subsequent period, so that it is worse off.

Action of *Rich* after history ending in g

If *Rich* deviates from D to any number y then its payoff in the resulting subgame is 0 rather than $d/(1-\delta)$, so it is worse off.

Action of *Poor* after history ending in (g, y)

If *Poor* follows its strategy and chooses R, its payoff in the resulting subgame is $(1-v)/(1-\delta)$. If it deviates and chooses A, its payoff is $1 - y + \delta V^P$, where V^P is its payoff from the strategy pair at the start of the game, so that

$$V^P = (1-q)(0+\delta V^P) + q\frac{1-d}{1-\delta}$$

and hence

$$V^P = \frac{q(1-d)}{(1-\delta)(1-\delta(1-q))}.$$

Thus *Poor* is no better off deviating than choosing R if and only if

$$1 - y + \frac{\delta q(1-d)}{(1-\delta)(1-\delta(1-q))} \le \frac{1-v}{1-\delta}.$$

This condition is satisfied for all values of y if and only if $q \le q^*$.

I now argue that the game has no other subgame perfect equilibrium.

Step 1 *The expected payoff of Rich in every subgame perfect equilibrium is at least*

$$\frac{(1-q)(1-\delta)+qd}{(1-\delta)(1-\delta(1-q))}.$$

Proof. If *Rich* uses the strategy that chooses $z = 1$ following every history ending in b and D after every history ending in g, its expected payoff V^R satisfies

$$V^R = (1-q)(1+\delta V^R) + q\frac{d}{1-\delta},$$

regardless of the strategy of *Poor*, so that V^R is the expression given in the result. ◁

Step 2 *The expected payoff of Poor in every subgame perfect equilibrium is at most*

$$\frac{q(1-d)}{(1-\delta)(1-\delta(1-q))}.$$

Proof. The sum of the players' payoffs in each period is at most 1, so that the sum of their total payoffs in the game is at most $1/(1-\delta)$. The result follows from Step 1. ◁

Step 3 *If $q < q^*$ then in every subgame perfect equilibrium the strategy of Poor chooses R after every history ending in (g, y) for any $y \in [0,1]$.*

Proof. If *Poor* chooses A after such a history, its payoff in the subgame following the history is at most $1 + \delta V^P$, where V^P is its maximal payoff in the game. If it chooses R, its payoff is $(1-v)/(1-\delta)$. Thus by Step 2 its choosing A is not optimal if

$$\frac{1-v}{1-\delta} > 1 + \delta\frac{q(1-d)}{(1-\delta)(1-\delta(1-q))},$$

which is equivalent to $q < q^*$. ◁

Step 4 *If $q < q^*$ then in every subgame perfect equilibrium the strategy of Rich chooses D after every history ending in g.*

Proof. By Step 3, if *Rich* chooses any value of y after a history ending in g its payoff is 0, whereas if it chooses D after such a history its payoff is $d/(1-\delta) > 0$. ◁

Step 5 *If $q < q^*$ then in every subgame perfect equilibrium the strategy of Rich chooses $z = 1$ after every history ending in b.*

Proof. Suppose that *Rich* chooses a value of z less than 1 after a history ending in b. If it deviates to 1, its payoff in the period increases and by Step 4 its payoff in every future period is unaffected. Thus the deviation increases its payoff. ◁

If good environments are not rare, the game has a subgame perfect equilibrium in which democratization does not occur. In this equilibrium, *Rich* can stave off a revolution by redistributing in good environments, even though it does not do so in bad environments.

Exercise 15.1: Subgame perfect equilibrium of dynamic game of revolt without democratization

Show that the following strategy pair is a subgame perfect equilibrium of a dynamic game of revolt $\langle\{Rich, Poor\}, q, \delta, v, d\rangle$ if $0 < d < v < \delta < 1$ and $q \geq 1 - v/\delta$, where $y^* = (v - \delta(1-q))/(1 - \delta(1-q))$.

- The strategy of *Rich* assigns 1 to every history ending in b and y^* to every history ending in g. (*Rich* takes all the wealth whenever the environment is bad for revolt and takes y^* whenever it is good for revolt.)

- The strategy of *Poor* assigns R to every history ending in (g, y) with $y \geq y^*$ and A to every history ending in (g, y) with $y < y^*$. (If, when the environment is good for revolt, *Rich* does not democratize, *Poor* revolts if *Rich* takes y^* or more and otherwise accepts the amount *Rich* offers it.)

15.2 Coordinating rebellion

A dynamic game of revolt models the poor as a unitary actor, who decides whether to accept a proposal or rebel. In this section I present a model with a large number of citizens, who face a coordination problem: rebellion is worthwhile for a citizen only if it is successful, which requires sufficiently many of her compatriots to rebel. In the model, the dictator's choice of how much wealth to assign to the masses is the trigger for a rebellion. If the masses observe only a noisy signal of this choice then their ability to coordinate their actions is limited and

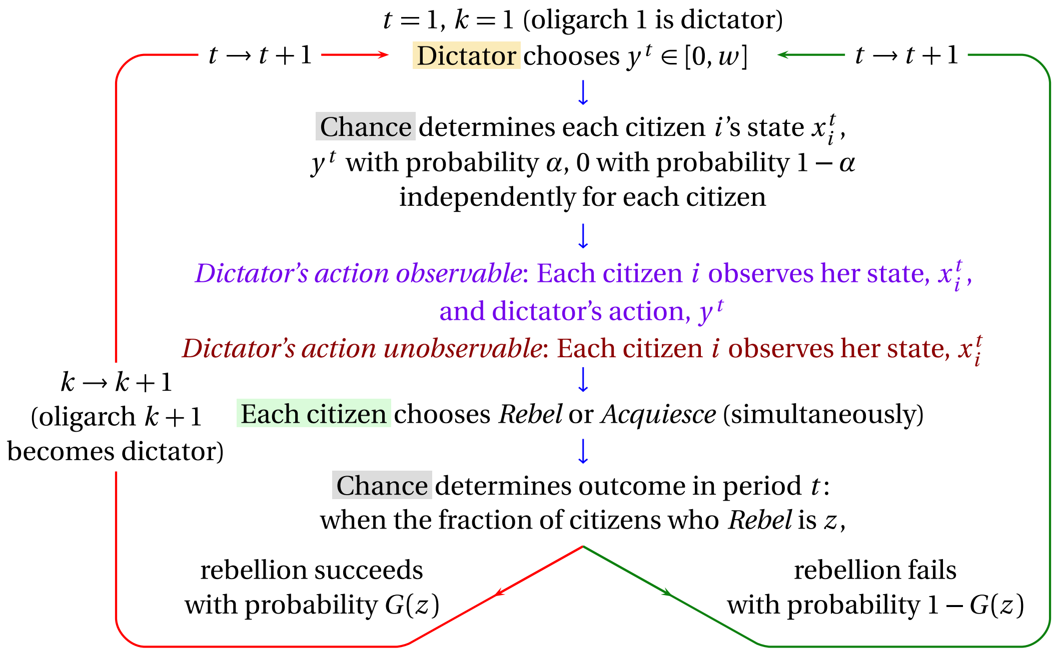


Figure 15.3 The structure of the interaction between the oligarchs and the citizens in an oligarchic society.

the most they receive in an equilibrium is less than the most they receive when the dictator's choice is perfectly observable (Proposition 15.2, Proposition 15.3). Giving the dictator the option of calling an election, the exercise of which (or lack thereof) is perfectly observed, facilitates coordination and restores an equilibrium in which the masses receive as much as they do when they perfectly observe the dictator's choice of how much wealth to assign them (Section 15.2.3).

The structure of the model is depicted in Figure 15.3. In each of an infinite sequence of periods, one of a countably infinite set of oligarchs interacts with a continuum of citizens. In each period, the ruling oligarch—the current dictator—chooses how much of a pie of size w to devote to public goods, and consumes the remainder. (One interpretation of the public good is that it represents the effort the dictator expends in governing the society.) In each period, each citizen's economic fortune depends on her *state*, which is random. When the dictator chooses y in some period, each citizen's state in that period is y with probability α and 0 with probability $1-\alpha$, independently of every other citizen's state. Each citizen observes her state in each period; her payoff depends on her state in a way that I specify subsequently.

In one version of the model, each citizen observes the dictator's action in each period, and in another version, she does not. In both cases, having observed her state, she chooses whether to rebel or acquiesce; the citizens make

Rebel			*Acquiesce*		
Rebellion succeeds	Rebellion fails		Rebellion succeeds	Rebellion fails	
	$z^t = 0$	$z^t > 0$		$z^t = 0$	$z^t > 0$
$x_i^t + b - L$	$x_i^t - c$	$x_i^t - c - L$	$x_i^t - L$	x_i^t	$x_i^t - L$

Table 15.1 The payoff of citizen $i \in I$ in any given period t in an oligarchic society $\langle I, K, w, b, c, L, \alpha, G, \delta \rangle$, where x_i^t is the citizen's state in period t and z^t is the fraction of citizens who choose *Rebel* in period t.

these choices simultaneously. The more citizens who rebel, the more likely the rebellion is to succeed. Specifically, when the fraction of citizens who rebel in period t is z^t, the probability that a rebellion succeeds in that period is $G(z^t)$, where G is an increasing function with $G(0) = 0$ and $G(1) = 1$ (rebellion surely succeeds if every citizen participates). If the rebellion succeeds, the dictator is ousted and replaced by the next oligarch in line. Otherwise the currently ruling oligarch continues as dictator. At the end of the period, all oligarchs and citizens observe the fraction of citizens who rebelled and the outcome of the rebellion (succeeded, failed).

In any given period t an oligarch obtains the payoff $w - y^t$ if she is the dictator in that period and chooses y^t, and the payoff 0 if she is not the dictator in that period. Her total payoff is the discounted sum of her payoffs in all periods, with discount factor $\delta \in (0, 1)$. If she is the dictator in every period and chooses y in every period, for example, her total payoff is $(w - y)/(1 - \delta)$.

Each citizen's payoff in any given period depends on her action in the period, the fraction of citizens who rebel in the period, and the outcome of the rebellion, as specified in Table 15.1. Suppose that citizen i rebels in period t. Then if the rebellion succeeds, her payoff in the period is $x_i^t + b - L$, where x_i^t is her state in the period, $b > 0$ represents the benefit, material or psychological, of participating in a successful rebellion, and $L > 0$ represents the loss incurred by all citizens due to the disruption of the rebellion. If the rebellion fails, her payoff is $x_i^t - c - L$ if a positive fraction of citizens rebel and $x_i^t - c$ if the fraction of citizens who rebel is zero (isolated rebellious actions are costly, but do not cause rebellion), where $c > 0$. If citizen i acquiesces in period t, her payoff in the period is $x_i^t - L$ if a positive fraction of citizens rebel (regardless of whether the rebellion succeeds or fails) and x_i^t if the fraction of citizens who rebel is zero.

The model in which each citizen observes only her own state in each period, not the dictator's action (the more interesting of the two cases), is captured by an extensive game with imperfect information with a set of players that is the union of a countably infinite set (the oligarchs) and a continuum (the citizens).

Defining this game, and an appropriate solution concept for it, entails technical challenges. Rather than doing so, I define an oligarchic society as a list of parameters, without specifying the actions available to the members of the society, each member's information when she takes an action, or the members' payoffs, and subsequently define a notion of equilibrium for such a society that is motivated by the structure of interaction shown in Figure 15.3 and the considerations in game-theoretic notions of equilibrium, but is not cast as a specific type of equilibrium of a specific game.

Definition 15.2: Oligarchic society

An *oligarchic society* $\langle I, K, w, b, c, L, \alpha, G, \delta \rangle$ consists of

- I, a nonempty interval (the set of citizens, a continuum)
- $K = \{1, 2, \ldots\}$ (oligarchs, each of whom is a potential dictator)
- $w > 0$ (the amount of wealth available in each period)
- $b > 0$ (the increment to the payoff of a citizen who participates in a successful rebellion)
- $c > 0$ (the decrement to the payoff of a citizen who participates in an unsuccessful rebellion)
- $L > 0$ (the decrement to the payoff of each citizen if a rebellion occurs)
- $\alpha \in [0, 1]$ (when the dictator chooses y, any given citizen's state is y with probability α and 0 with probability $1 - \alpha$, independently of the other citizens' states)
- $G : [0, 1] \to [0, 1]$, with $G(0) = 0$ and $G(1) = 1$, is increasing on $(0, 1)$ ($G(z)$ is the probability a rebellion succeeds if the fraction of citizens who participate in it is z)
- $\delta \in (0, 1)$ (the oligarchs' common discount factor).

The notions of equilibrium that I define are intended to capture the considerations in a stationary equilibrium of a game in which every oligarch uses the same strategy and every citizen uses the same strategy: the behavior of each oligarch and each citizen is optimal, given the other players' actions, is unvarying over time, and for any given period depends only on the events within that period.

15.2.1 Dictator's action observable

I start with the case in which the citizens observe the dictator's action before choosing whether to rebel. I define a (stationary) equilibrium as a pair (y^*, σ^*), where $y^* \in [0, w]$ is the action of each oligarch in every period in which she is the dictator and $\sigma^* : [0, w] \to [0, 1]$ defines the probability $\sigma^*(y)$ with which each citizen rebels in any given period when the ruling oligarch chooses y in that period. (I do not include the citizen's state as an argument of σ^* because, as I argue subsequently, a citizen's optimal action is independent of her state.)

First consider the optimality of y^*. Given σ^*, the fraction of the (continuum of) citizens who rebel in a period in which the dictator chooses y is $\sigma^*(y)$, so the probability that rebellion succeeds is $p^*(y) = G(\sigma^*(y))$. Thus the total payoff $V(y)$ of the oligarch who rules in period 1 and chooses y in every period, regardless of history, is equal to $w - y + (1 - p^*(y))\delta V(y)$, so that

$$V(y) = \frac{w - y}{1 - \delta(1 - p^*(y))}.$$

For the oligarchs' common action y^* to be optimal, we need $V(y^*) \geq V(y)$ for all $y \in [0, w]$.

Now consider the optimality of σ^*. A citizen reasonably bases her decision on whether to rebel in any given period on her estimate of the probability that a rebellion will succeed. Suppose that her estimate of this probability is p. Then if her state is x, her payoff is

$$\begin{cases} p(x + b - L) + (1 - p)(x - c - L) & \text{if she rebels} \\ x - L & \text{if she acquiesces,} \end{cases}$$

so that, independently of her state, she optimally rebels if $p > c/(b + c)$, optimally acquiesces if $p < c/(b + c)$, and is indifferent between the two actions if $p = c/(b + c)$. Let $\hat{p} = c/(b + c)$. Then the probability she assigns to rebelling is optimal if and only if it is 0 if $p < \hat{p}$ and 1 if $p > \hat{p}$. (Every probability is optimal if $p = \hat{p}$.) As for standard notions of equilibria in games, I assume that her belief about the probability p that a rebellion will succeed is correct. (Perhaps she has inferred this probability from her long experience playing the game or observing other people play the game or similar games.) That is, she believes that if the dictator chooses y then the fraction $\sigma(y)$ of her compatriots will rebel, so that her estimate of the probability that a rebellion will succeed is $G(\sigma(y))$. (Given that the set of citizens is a continuum, her participation in the rebellion has no effect on its success.) These considerations lead to the following definition of an observable-action equilibrium.

Definition 15.3: Observable-action equilibrium of oligarchic society

An *observable-action equilibrium* of a oligarchic society $\langle I, K, w, b, c, L, \alpha, G, \delta\rangle$ is a pair (y^*, σ^*), where $y^* \in [0, w]$ (each oligarch's action) and $\sigma^* : [0, w] \to [0, 1]$ ($\sigma^*(y)$ is each citizen's probability of rebelling when the ruling oligarch chooses y), such that

$$V(y^*) \geq V(y) \quad \text{for all } y \in [0, w] \tag{15.1}$$

and

$$\sigma^*(y) = \begin{cases} 0 & \text{if } p^*(y) < \hat{p} \\ 1 & \text{if } p^*(y) > \hat{p} \end{cases} \quad \text{for all } y \in [0, w] \tag{15.2}$$

where

$$V(y) = \frac{w - y}{1 - \delta(1 - p^*(y))} \quad \text{for all } y \in [0, w]$$

(the payoff of an oligarch when she is in power and chooses y),

$$p^*(y) = G(\sigma^*(y)) \quad \text{for all } y \in [0, w]$$

(the probability of a successful rebellion when the ruling oligarch chooses y), and $\hat{p} = c/(b + c)$.

An oligarchic society has an observable-action equilibrium in which the dictator takes all the wealth and the citizens always rebel and also an equilibrium in which the dictator takes all the wealth and the citizens never rebel. In addition, it has equilibria in which the dictator always takes some amount $y \leq (1 - \delta)w$ and all citizens rebel if she takes more. I describe these equilibria in more detail and then state a precise result.

Consider the pair $(0, \sigma^1)$ with $\sigma^1(y) = 1$ for all $y \in [0, w]$. That is, every oligarch takes all of w for herself whenever she is the dictator and every citizen rebels whatever the ruling oligarch does. Given that every citizen rebels, the rebellion succeeds; any citizen who deviates to acquiescence has no effect on the success of the rebellion and forgoes the payoff b that accrues to a citizen who participates in a successful rebellion. Thus the citizens' actions are optimal given the oligarchs' actions. Every oligarch rules for one period, and can do nothing to avoid being ousted from power. Hence $(0, \sigma^1)$ is an observable-action equilibrium.

Now consider the pair $(0, \sigma^0)$ with $\sigma^0(y) = 0$ for all $y \in [0, w]$. That is, every oligarch takes all of w for herself whenever she is the dictator and every citizen always acquiesces. The oligarch who is dictator in period 1 stays in power indefinitely and obtains w in every period, and no citizen's switching to rebel for any value of y has any effect on the outcome. Hence $(0, \sigma^0)$ is an observable-action equilibrium.

Observable-action equilibria in which each oligarch chooses a positive value of y also exist. For example, for any $y^* \in [0, \delta w]$ there is an equilibrium in which each citizen rebels if and only if $y < y^*$ and each oligarch chooses y^*. In these equilibria, the citizens' threat to rebel if the oligarch chooses $y < y^*$, which is credible because all citizens participate in such a rebellion, which thus succeeds, induces the oligarch to choose y^*. In no equilibrium does each oligarch choose $y > \delta w$ because the payoff of such an oligarch is less than $(w - \delta w)/(1 - \delta) = w$, and by choosing $y = 0$ an oligarch guarantees herself a payoff of w (she is in power for one period, in which she obtains w).

Proposition 15.2: Observable-action equilibria of oligarchic society

Let $\langle I, K, w, b, c, L, \alpha, G, \delta \rangle$ be an oligarchic society.

a. Define $\sigma^0 : [0, w] \to [0, 1]$ by $\sigma^0(y) = 0$ for all $y \in [0, w]$ (each citizen acquiesces for all values of y) and $\sigma^1 : [0, w] \to [0, 1]$ by $\sigma^1(y) = 1$ for all $y \in [0, w]$ (each citizen rebels for all values of y). Both $(0, \sigma^0)$ and $(0, \sigma^1)$ are observable-action equilibria of the society.

b. Let $y^* \in [0, \delta w]$ and define $\sigma^* : [0, w] \to [0, 1]$ by

$$\sigma^*(y) = \begin{cases} 0 & \text{if } y \geq y^* \\ 1 & \text{if } y < y^* \end{cases}$$

(each citizen rebels if and only if $y < y^*$). Then (y^*, σ^*) is an observable-action equilibrium of the society.

c. In every observable-action equilibrium (y, σ) of the society we have $y \leq \delta w$.

Proof

a. First consider $(0, \sigma^0)$. In the notation of Definition 15.3 we have $p^*(y) = 0$ for all y, so that $V(y) = (w - y)/(1 - \delta)$ for all y, and hence (15.1) is satisfied. Given $\hat{p} > 0$, (15.2) is also satisfied. Now consider $(0, \sigma^1)$. We have $p^*(y) = 1$ for all y, so that $V(y) = w - y \leq V(0)$ for all y, so that (15.1) is satisfied. Given $\hat{p} < 1$, (15.2) is also satisfied.

b. We have

$$p^*(y) = G(\sigma^*(y)) = \begin{cases} G(0) = 0 & \text{if } y \geq y^* \\ G(1) = 1 & \text{if } y < y^*. \end{cases}$$

Thus

$$V(y) = \begin{cases} (w - y)/(1 - \delta) & \text{if } y \geq y^* \\ w - y & \text{if } y < y^*. \end{cases}$$

Hence $V(y^*) \geq V(y)$ for all $y \in [0, w]$ if and only if $(w - y^*)/(1 - \delta) \geq w$, or $y^* \leq \delta w$. Thus (15.1) is satisfied.

Now, $\hat{p} \in (0, 1)$, so if $p^*(y) < \hat{p}$ then $y \geq y^*$, so that $\sigma^*(y) = 0$, and if $p^*(y) > \hat{p}$ then $y < y^*$, so that $\sigma^*(y) = 1$. Thus (15.2) is satisfied.

Hence (y^*, σ^*) is an observable-action equilibrium.

c. We have $V(0) = w/(1 - \delta(1 - p^*(0))) \geq w$, so by (15.1) for any observable-action equilibrium (y, σ) we have $V(y) \geq w$ and hence $y \leq \delta w(1 - p^*(y)) \leq \delta w$.

15.2.2 Dictator's action unobservable

Now consider the more interesting case in which the citizens do not observe the dictator's action. When the dictator's action is y, each citizen's state is y with probability α and 0 with probability $1 - \alpha$, and each citizen observes only her state, not the dictator's action. I now define a (stationary) equilibrium to be a pair (y^*, ρ^*), where $y^* \in [0, w]$ is the action of each oligarch in every period in which she is the dictator and $\rho^* : [0, w] \to [0, 1]$ defines the probability $\rho^*(x)$ with which each citizen rebels in any given period when her state is x.

The fraction of citizens who rebel when the dictator chooses y is now $\alpha\rho^*(y) + (1 - \alpha)\rho^*(0)$, which I denote $q^*(y)$. Thus the probability of a successful rebellion when the dictator chooses y is $G(q^*(y))$ and the total payoff $V(y)$ of the oligarch who rules in period 1 and chooses y in every period, regardless of history, is equal to $w - y + (1 - q^*(y))\delta V(y)$, so that

$$V(y) = \frac{w - y}{1 - \delta(1 - q^*(y))}.$$

As before, for the oligarchs' common action y^* to be optimal, we need $V(y^*) \geq V(y)$ for all $y \in [0, w]$.

Now consider the optimality of ρ^*. If a citizen's state is $x \in (0, w]$, she knows that the ruling oligarch has chosen x (and, in particular, if $x \neq y^*$ she knows that the ruling oligarch has deviated from y^*), and hence she knows that the state is x for the fraction α of citizens and 0 for the fraction $1 - \alpha$. Thus given the citizens' common strategy ρ^*, she knows that the fraction of citizens who will rebel is $q^*(x)$ and hence the probability of a successful rebellion is $G(q^*(x))$. If a citizen's state is 0, she has no information about the ruling oligarch's action, and in particular has no reason to believe that the ruling oligarch has deviated from y^*. In the spirit of the assumption of the consistency of beliefs with strategies in the standard solution concepts for extensive games with imperfect information, I assume that in this case she believes that the fraction of citizens who will rebel is $q(y^*)$ and hence the probability of a successful rebellion is $G(q(y^*))$. Given that a citizen optimally rebels if she believes that the probability of success exceeds $\hat{p}$ $(= c/(b+c))$, optimally acquiesces if she believes that this probability is less than $\hat{p}$, and is indifferent between the two actions if she believes that this probability is $\hat{p}$, we are led to the following definition of an equilibrium.

Definition 15.4: Unobservable-action equilibrium of oligarchic society

An *unobservable-action equilibrium* of a oligarchic society $\langle I, K, w, b, c, L, \alpha, G, \delta\rangle$ is a pair (y^*, ρ^*), where $y^* \in [0, w]$ (each oligarch's action) and $\rho^* : [0, w] \to [0, 1]$ ($\rho^*(x)$ is each citizen's probability of rebelling when her state is x), such that

$$V(y^*) \geq V(y) \quad \text{for all } y \in [0, w], \tag{15.3}$$

$$\rho^*(0) = \begin{cases} 0 & \text{if } q^*(y^*) < \hat{p} \\ 1 & \text{if } q^*(y^*) > \hat{p}, \end{cases} \tag{15.4}$$

and

$$\rho^*(x) = \begin{cases} 0 & \text{if } q^*(x) < \hat{p} \\ 1 & \text{if } q^*(x) > \hat{p} \end{cases} \quad \text{for all } x \in (0, w], \tag{15.5}$$

where

$$V(y) = \frac{w - y}{1 - \delta(1 - q^*(y))} \quad \text{for all } y \in [0, w] \tag{15.6}$$

(the payoff of an oligarch when she is in power and chooses y),

$$q^*(y) = G(\alpha\rho^*(y) + (1-\alpha)\rho^*(0)) \quad \text{for all } y \in [0, w]$$

(the probability of a successful rebellion when the ruling oligarch chooses y), and $\hat{p} = c/(b+c)$.

Analogues of the patterns of behavior in Proposition 15.2*a* are unobservable-action equilibria: if every citizen acquiesces regardless of her state, no rebellion every occurs and the ruling oligarch optimally chooses 0 and stays in power indefinitely; if every citizen rebels regardless of her state, a rebellion succeeds in every period, and the ruling oligarch optimally chooses 0 and remains in power for only one period.

The analogue of the pattern of behavior in Proposition 15.2*b*, however, is not generally an equilibrium. Suppose that for some number x^* every citizen rebels when her state is less than x^* and acquiesces when it is at least x^*. Then if the dictator chooses x^*, the fraction $1-\alpha$ of citizens, whose state is 0, rebel, so that rebellion succeeds with probability $q^*(x^*) = G(1-\alpha)$. Thus if $G(1-\alpha) < \hat{p}$ then rebelling is not optimal for the citizens whose state is 0 ((15.4) is violated) and if $G(1-\alpha) > \hat{p}$ then acquiescing is not optimal for the citizens whose state is x^* ((15.5) is violated). Only if $G(1-\alpha) = \hat{p}$ is the pattern of behavior an equilibrium.

The fact that citizens whose state is 0 do not know the value of y chosen by the dictator makes disciplining the dictator with rebellion more difficult. In an equilibrium in which the dictator chooses a positive value of y, a deviation by the dictator to $y = 0$ must increase the probability of rebellion. Such a deviation causes every citizen's state to become 0, so the probability of rebellion for an citizen with state 0 must be positive. As a consequence, rebellion occurs with positive probability also when the dictator does not deviate, because even then some citizens' states are (randomly) 0. Thus the impact of a deviation to 0 on the dictator's payoff is less than it is when her action is observable, and hence the maximum amount of wealth the dictator gives to the citizens in an equilibrium is also less than it is when her action is observable.

Further, there is a discrete difference between the case in which the dictator's action is observable and that in which it is not observable but α is close to 1. The reason is that for the probability of rebellion for a citizen with state 0 to be positive, the citizen must be indifferent between rebelling and acquiescing, which requires that if the dictator adheres to the equilibrium and chooses $y = y^*$, the probability of success of a rebellion must be equal to $\hat{p}$, independent of the value of α. When α is close to 1, that in turn requires that citizens with state 0 rebel with probability 1 and those with state y^* rebel with probability close to the number r for which $G(r) = \hat{p}$. By contrast, when the dictator's action is observable, no one rebels when the dictator adheres to her equilibrium strategy.

This argument leads to the conclusion that the largest value of y possible in an unobservable-action equilibrium is discretely less than δw, the value that Proposition 15.2*b* shows can be achieved in an observable-action equilibrium. Parts *c* and *d* of the next result specify the upper bound. In the equilibria defined in the result, all oligarchs choose some number y and if $G(1-\alpha) < \hat{p}$ then every citizen whose state is less than y rebels and some of those whose state is at least y do so, while if $G(1-\alpha) > \hat{p}$ then some of those whose state is 0 rebel, all of those whose state is between 0 and y do so, and none of those whose state is at least y do so.

Proposition 15.3: Unobservable-action equilibria of oligarchic society

Let $\langle I, K, w, b, c, L, \alpha, G, \delta\rangle$ be an oligarchic society and let $\hat{p} = c/(b+c)$.

a. Define $\rho^0 : [0,w] \to [0,1]$ by $\rho^0(x) = 0$ for all $x \in [0,w]$, and $\rho^1 : [0,w] \to [0,1]$ by $\rho^1(x) = 1$ for all $x \in [0,w]$. Both $(0,\rho^0)$ and $(0,\rho^1)$ are unobservable-action equilibria of the society.

b. In every unobservable-action equilibrium (y^*, ρ^*) with $y^* > 0$,

$$G(\alpha\rho^*(y^*) + (1-\alpha)\rho^*(0)) = \hat{p}$$

(the probability of a successful rebellion is equal to $\hat{p}$) and

$$y^* \le \frac{\delta w(G(\rho^*(0)) - \hat{p})}{1 - \delta(1 - G(\rho^*(0)))}. \tag{15.7}$$

c. Suppose that $G(1-\alpha) \le \hat{p}$. Let $y^* \in (0, \delta w(1-\hat{p})]$, let r be the unique number satisfying $G(\alpha r + 1 - \alpha) = \hat{p}$, and define $\rho^* : [0,w] \to [0,1]$ by

$$\rho^*(x) = \begin{cases} 1 & \text{if } 0 \le x < y^* \\ r & \text{if } y^* \le x \le w. \end{cases} \tag{15.8}$$

The pair (y^*, ρ^*) is an unobservable-action equilibrium of the society, and in every equilibrium (y, ρ) we have $y \le \delta w(1-\hat{p})$.

d. Suppose that $G(1-\alpha) > \hat{p}$. Let

$$0 < y^* \le \frac{\delta w(G(r) - \hat{p})}{1 - \delta(1 - G(r))}, \tag{15.9}$$

where r is the unique number satisfying $G((1-\alpha)r) = \hat{p}$, and define $\rho^* : [0,w] \to [0,1]$ by

$$\rho^*(x) = \begin{cases} r & \text{if } x = 0 \\ 1 & \text{if } 0 < x < y^* \\ 0 & \text{if } y^* \le x \le w. \end{cases} \tag{15.10}$$

The pair (y^*, ρ^*) is an unobservable-action equilibrium of the society, and in every equilibrium (y, ρ) the value of y is at most the right-hand side of (15.9).

Proof

a. First consider $(0, \rho^0)$. In the notation of Definition 15.4, we have $q^*(y) = G(0) = 0$ for all $y \in [0, w]$, so $V(y) = (w - y)/(1 - \delta)$ for all $y \in [0, w]$. Thus (15.3) is satisfied. Given that $q^*(x) = 0$ for all $x \in [0, w]$, (15.4) and (15.5) are also satisfied. Hence $(0, \rho^0)$ is an unobservable-action equilibrium of the society.

Now consider $(0, \rho^1)$. We have $q^*(y) = G(1) = 1$ for all $y \in [0, w]$, so $V(y) = w - y$ for all $y \in [0, w]$. Thus (15.3) is satisfied, and (15.4) and (15.5) are also satisfied. Hence $(0, \rho^1)$ is an unobservable-action equilibrium of the society.

b. Let $q^*(y) = G(\alpha\rho^*(y) + (1 - \alpha)\rho^*(0))$ for all $y \in [0, w]$, the probability of a successful rebellion, as given in Definition 15.4. If $q^*(y^*) < \hat{p}$ then $\rho^*(0) = \rho^*(y^*) = 0$ by (15.4) and (15.5), so that $q^*(0) = q^*(y^*) = 0$. If $q^*(y^*) > \hat{p}$ then $\rho^*(0) = \rho^*(y^*) = 1$ by (15.4) and (15.5), so that $q^*(0) = q^*(y^*) = 1$. Using (15.6), given $y^* > 0$, in both cases we thus have $V(0) > V(y^*)$, contradicting (15.3). Thus $q^*(y^*) = \hat{p}$.

Now by (15.6) we have

$$V(y^*) = \frac{w - y^*}{1 - \delta(1 - \hat{p})} \quad \text{and} \quad V(0) = \frac{w}{1 - \delta(1 - G(\rho^*(0)))}.$$

Thus the requirement of (15.3) that $V(y^*) \geq V(0)$ implies (15.7).

c. I first argue that (y^*, ρ^*) is an unobservable-action equilibrium. Given (15.8), we have

$$q^*(x) = \begin{cases} 1 & \text{if } 0 \leq x < y^* \\ G(\alpha r + 1 - \alpha) = \hat{p} & \text{if } x \geq y^*, \end{cases}$$

so that (15.4) and (15.5) are satisfied. Also

$$V(y) = \begin{cases} w - y & \text{if } 0 \leq y < y^* \\ \dfrac{w - y^*}{1 - \delta(1 - \hat{p})} & \text{if } y \geq y^*, \end{cases}$$

so that (15.3) is satisfied, given $y^* \leq \delta w(1 - \hat{p})$.

The right-hand side of (15.7) is increasing in $G(\rho^*(0))$, and hence attains its maximum when $G(\rho^*(0)) = 1$; this maximum is $\delta w(1 - \hat{p})$, so the last claim follows from part *b*.

d. I first argue that (y^*, ρ^*) is an unobservable-action equilibrium. We have

$$q^*(x) = \begin{cases} G(r) & \text{if } x = 0 \\ G(\alpha + (1-\alpha)r) & \text{if } 0 < x < y^* \\ G((1-\alpha)r) = \hat{p} & \text{if } x \geq y^*, \end{cases}$$

so that (15.4) and (15.5) are satisfied. Also

$$V(y) = \begin{cases} \dfrac{w}{1-\delta(1-G(r))} & \text{if } y = 0 \\ \dfrac{w-y}{1-\delta(1-G(\alpha+(1-\alpha)r))} & \text{if } 0 < y < y^* \\ \dfrac{w-y}{1-\delta(1-\hat{p})} & \text{if } y \geq y^*. \end{cases}$$

Now,

$$\frac{w}{1-\delta(1-G(r))} > \frac{w-y}{1-\delta(1-G(\alpha+(1-\alpha)r))}$$

for $y > 0$, so that (15.3) is satisfied if

$$\frac{w-y^*}{1-\delta(1-\hat{p})} \geq \frac{w}{1-\delta(1-G(r))},$$

which is true given (15.9).

The right-hand side of (15.7) is increasing in $G(\rho^*(0))$, and hence attains its maximum at the largest value of $\rho^*(0)$ that is consistent with $q^*(y^*) = G(\alpha\rho^*(y^*) + (1-\alpha)\rho^*(0)) = \hat{p}$. Given that $G(1-\alpha) > \hat{p}$, this value is the number r such that $G((1-\alpha)r) = \hat{p}$. Thus the right-hand side of (15.7) is equal to the right-hand side of (15.9).

If α is close to 0, then by part *d* the largest value of y in an equilibrium is close to 0, as one would expect: if almost no one observes the dictator's action, rebellion is a blunt tool. The largest value of y in an equilibrium is also close to 0 if the benefit b that citizens receive from participating in a successful rebellion is small, or the cost c they incur from participating in an unsuccessful rebellion is large, both of which cause $\hat{p}$ to be close to 1.

If α is close to, but less than, 1, then by part *c* the largest equilibrium value of y is at most $\delta w(1-\hat{p})$, which if $b > 0$ is discretely less than its largest equilibrium value of δw in an observable-action equilibrium. That is, the presence of a small amount of imperfect information significantly reduces the largest equilibrium value of y; the logic behind this result is discussed informally before the statement of the proposition.

In the equilibria in parts c and d of the result, rebellion occurs in every period, because a positive fraction of citizens rebel. Thus each citizen's payoff in each period is $\alpha y^* - L$. In the equilibrium $(0, \rho^0)$ in part a, in which rebellion never occurs, each citizen's payoff is 0 in each period. Thus each citizen is better off in an equilibrium (y^*, ρ^*) as specified in part c or d than in the equilibrium $(0, \rho^0)$ if and only if $\alpha y^* > L$.

15.2.3 Coordinating rebellion

To induce the oligarchs to choose values of y higher than the ones possible in unobservable-action equilibria, the citizens need to coordinate their rebellious impulses. One way for them to do so is simply by communicating: if the citizens who observe the value of y chosen by the dictator inform their compatriots, then an observable-action equilibrium can be implemented.

Another natural possibility is that the citizens condition their actions not only on events in the current period, but also on events in previous periods. That is, we could consider the possibility of an equilibrium in nonstationary strategies. To do so, we need to formulate a game and a solution concept precisely, which is challenging. Plausibly such a game has an equilibrium in which each oligarch chooses a number, say y^{**}, larger than the largest value of y in an unobservable-action equilibrium, with some citizens rebelling whenever their state is less than y^{**} and all citizens rebelling if the fraction of citizens who rebelled in the previous period was large enough to imply that the oligarch deviated and chose a value of y smaller than y^{**} in that period.

An additional possibility is that each citizen has an opportunity to take an action before choosing whether to rebel, and the number of citizens choosing the action is publicly observable. For example, each citizen might have the opportunity to demonstrate or to vote in a referendum. The model I describe adopts the latter interpretation.

Assume that in each period t, after each citizen i observes her state x_i^t, the ruling oligarch can give the citizens the opportunity to vote (simultaneously) to allow her to rule for another period. Suppose that the outcome of every vote is observed by all citizens. If the ruling oligarch does not give the citizens the opportunity to vote, or if she does so, a majority favors her resignation, and she does not resign, the citizens have the option to rebel. The structure of the interaction is shown in Figure 15.4.

I argue that this model has an equilibrium in which each oligarch chooses δw in every period, as in the observable-action equilibrium of an oligarchic society that is best for the citizens. That is, even though the citizens cannot observe

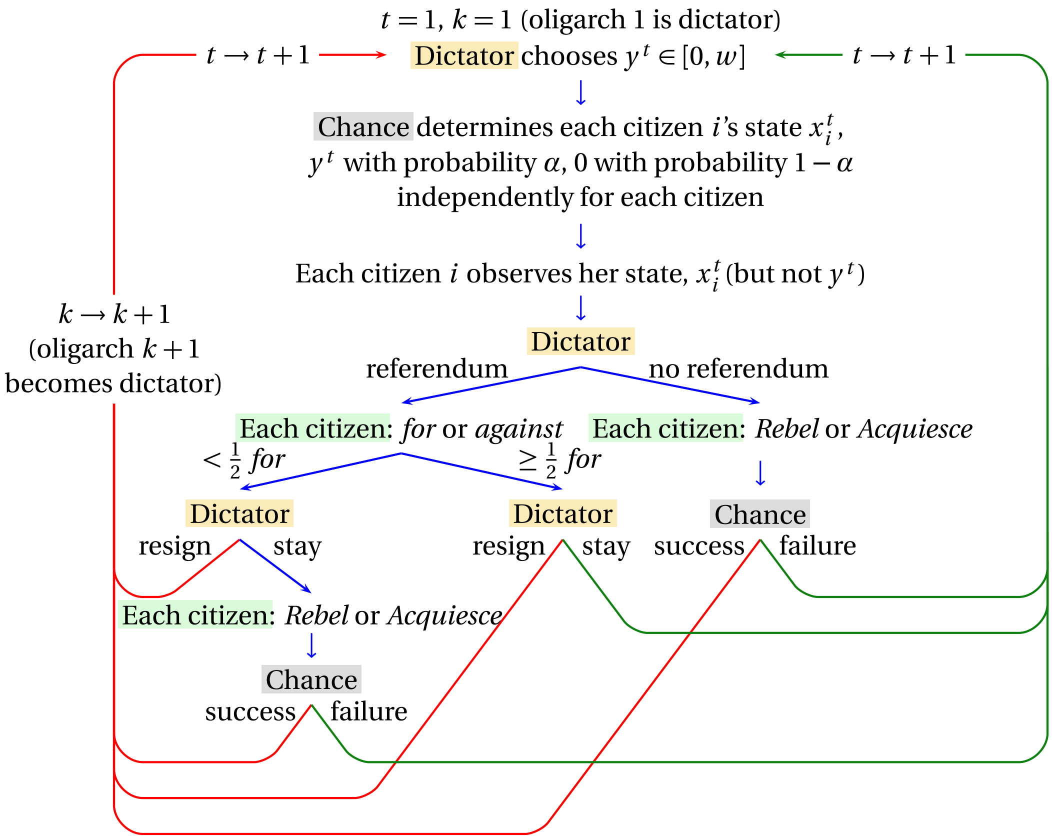


Figure 15.4 The structure of a dynamic games of rebellion with the option of a referendum. Whenever the citizens move, they do so simultaneously.

the actions of the ruling oligarch, the model has an equilibrium in which their payoffs are as high as possible in an equilibrium of the model in which they can observe these actions.

A formal model that captures this interaction, like one that captures the interaction in Figure 15.3 when the dictator's action is unobservable, is an extensive games with imperfect information. A citizen whose state is 0 does not know the value of y that the dictator chose, and when deciding the action to take has to form a belief about this value. A full precise specification of the game is complex, and as for the earlier model, I do not provide such a specification. In fact, in this case I do not specify a formal model at all, but merely sketch an informal analysis of equilibrium.

I argue that the (unformulated) game has an equilibrium in which after every history in which she is in power, every oligarch

- chooses δw
- holds a referendum

- resigns if and only if less than half of the citizens vote in her favor

and every citizen i

- rebels after any history that ends with the ruling oligarch's (*i*) not holding a referendum or (*ii*) holding one, losing the vote, and choosing to stay in power
- whenever her state in any period t is x_i^t,

$$\text{if } \alpha > \tfrac{1}{2} \quad \begin{cases} \text{votes } \textit{against} & \text{if } x_i^t < \delta w \\ \text{votes } \textit{for} & \text{if } x_i^t \geq \delta w \end{cases}$$

$$\text{if } \alpha \leq \tfrac{1}{2} \quad \begin{cases} \text{votes } \textit{against} \text{ with probability } \tfrac{1}{2} & \text{if } x_i^t = 0 \\ \text{votes } \textit{against} & \text{if } 0 < x_i^t < \delta w \\ \text{votes } \textit{for} & \text{if } x_i^t \geq \delta w. \end{cases}$$

Here is my argument.

Oligarchs

If a ruling oligarch adheres to her strategy, she chooses δw in every period, so that the state is δw for the fraction α of citizens and 0 for the fraction $1-\alpha$. The oligarch holds a referendum, and wins (if $\alpha > \frac{1}{2}$ she obtains the fraction α of the votes, and if $\alpha \leq \frac{1}{2}$ she obtains the fraction $\alpha + \frac{1}{2}(1-\alpha) = \frac{1}{2} + \frac{1}{2}\alpha$). The oligarch remains in power and obtains $w - \delta w$ in every future period, so her payoff in the game is $(w - \delta w)/(1-\delta) = w$.

If a ruling oligarch deviates from her strategy by choosing δw in some period but then not holding a referendum, all citizens rebel and the oligarch is thus removed from power. Her payoff in the rest of the game in this case is $w - \delta w$, which is less than w.

If a ruling oligarch deviates from her strategy by choosing a number greater than δw in some period, she is worse off in that period and, depending on her subsequent actions in the period, either remains in power or is removed from power, so that her payoff in the rest of the game is less than $w - \delta w + \delta(w - \delta w)/(1-\delta) = w$.

Suppose that a ruling oligarch deviates from her strategy by choosing $y^t < \delta w$ in some period t. If she does not hold a referendum, all the citizens rebel and she is removed from power. If she does hold a referendum, she loses: if $\alpha > \frac{1}{2}$ then all citizens vote against her and if $\alpha \leq \frac{1}{2}$ then the fraction $\alpha + \frac{1}{2}(1-\alpha) = \frac{1}{2} + \frac{1}{2}\alpha$ does so. Having lost, she either resigns or stays and is removed from power by a rebellion. Thus she obtains the payoff $w - y^t$ in the game. Among such deviations the best one for her is thus $y^t = 0$, which yields her the payoff w, equal to her payoff if she adheres to her strategy.

Citizens

If citizen i follows her strategy, she obtains x_i^t in every period t.

If citizen i deviates from her strategy by acquiescing rather than rebelling after a history that ends with the ruling oligarch's (*i*) not holding a referendum or (*ii*) holding one, losing the vote, and choosing to stay, her payoff in the period changes from $x_i^t + b - L$ to $x_i^t - L$ (see Table 15.1), so she is worse off.

If a citizen deviates from her strategy by changing her vote in some period, the outcome of the vote remains the same, given that there is a continuum of citizens.

Note that although the fact that no citizen's vote affects the outcome follows immediately from the assumption that there is a continuum of citizens, it remains true if the number of citizens is finite and large, because the margin of victory or loss is positive for all (positive) values of α.

Although the action that the dictator allows the citizens to take before deciding whether to rebel in this model is called voting in a referendum, any binary action plays the same role. The dictator could allow the citizens to participate in a demonstration, or prohibit them from doing so, and if she allows them to do so and more than half of them participate, they could all rebel if she does not resign. Or the dictator could allow them to take any other specific action, or prohibit them from doing so, and if she allows them to do so and more than half of them take the action, they could all rebel if she does not resign.

The action the citizens may be allowed to take acts as coordinating device: the fact that all citizens costlessly observe whether the action is allowed and the fraction of citizens who take it allows them to coordinate their actions. The original game, in which no such action is available, may have a nonstationary equilibrium in which a citizen's rebelling in one period is a costly signal of her state that is used by the citizens to coordinate rebellion in the following period, although the analysis of such equilibria seems complex.

Notes

Section 15.1 is based on Acemoglu and Robinson (2000) and Section 15.2 is based on Fearon (2011).

Solutions to exercises

Exercise 15.1

The payoffs satisfy the condition in Proposition 16.9, so a strategy pair is a subgame perfect equilibrium if and only if it satisfies the one-deviation property.

I now argue that the strategy pair in the result satisfies the one-deviation property. Note that $y^* \geq 0$ given that $q \geq 1 - v/\delta$.

Action of *Rich* after history ending in b

If *Rich* deviates from 1 to any other number then it obtains less in the period of its deviation and the same in every subsequent period, so that it is worse off.

Action of *Rich* after history ending in g

If *Rich* deviates from y^* to any number $y < y^*$ then it obtains less in the period of its deviation and the same in every subsequent period, so that it is worse off.

If *Rich* deviates to a number $y > y^*$ then its payoff in the resulting subgame is 0, so it is no better off (and is worse off if $y^* > 0$).

If *Rich* deviates to D then its payoff in the resulting subgame is $d/(1-\delta)$, so for the strategy pair to be a subgame perfect equilibrium we need

$$\frac{d}{1-\delta} \leq y^* + \delta V^R,$$

where V^R is its payoff from the strategy pair at the start of the game, so that

$$V^R = (1-q)(1+\delta V^R) + q(y^* + \delta V^R),$$

and hence

$$V^R = \frac{1-q+qy^*}{1-\delta}.$$

Thus the condition for equilibrium is

$$\frac{d}{1-\delta} \leq y^* + \frac{\delta(1-q+qy^*)}{1-\delta}$$

or

$$d \leq \delta(1-q) + (1-\delta(1-q))y^* = v,$$

which is satisfied, given the assumption on d and v.

Action of *Poor* after history ending in (g, y)

The payoff of *Poor* from accepting y^* is $1-y^*+\delta V^P$, where V^P is its payoff from the strategy pair at the start of the game, so that

$$V^P = (1-q)(0+\delta V^P) + q(1-y^* + \delta V^P),$$

and hence

$$V^P = \frac{q(1-y^*)}{1-\delta}.$$

Thus its payoff from accepting y^* is

$$1 - y^* + \frac{\delta q(1-y^*)}{1-\delta} = \frac{1-v}{1-\delta},$$

which is its payoff from choosing R. Thus R is an optimal response to any $y \geq y^*$ and A is an optimal response to any $y \leq y^*$.

V Appendix

16 Preferences, profiles, and games 523

16 Preferences, profiles, and games

16.1 Preferences and payoffs 523
16.2 Sets of individuals and profiles 528
16.3 Brouwer's fixed-point theorem 529
16.4 Strategic games 529
16.5 Bayesian games 533
16.6 Extensive games 535
16.7 Coalitional games 546
16.8 Optimization 550

This appendix provides definitions of most of the formal concepts used in the body of the book. It includes only brief explanations and discussions, and is intended only to remind you of the definitions. For detailed explanations and discussions using the same terminology and notation as this book see Osborne and Rubinstein (2023), Osborne (2004), and Osborne and Rubinstein (1994). (The first and last of these books, and several chapters of the second, are freely available in electronic form.)

16.1 Preferences and payoffs

A binary relation on a set X indicates, for each ordered pair (x,y) with $x \in X$ and $y \in X$, whether x and y are related in a certain way. For example, $\geq$ is a binary relation on the set of real numbers: for every pair (x,y) of real numbers, $x \geq y$ means that x is at least y. Two other binary relations on the set of real numbers are $>$ and $=$. Formally, a binary relation B on X may be defined as a subset of $X \times X$, the set of all ordered pairs (x,y) with $x \in X$ and $y \in X$; if $(x,y) \in B$ then x is related to y, and if $(x,y) \notin B$ then it is not. However, rather that writing $(x,y) \in B$, we usually write $x\,B\,y$, and we commonly use a symbol resembling $\geq$ or $>$, like $\succsim$, $\succ$, $\trianglerighteq$, or $\triangleright$, rather than a letter for a relation.

Definition 16.1: Binary relation

For any set X, a *binary relation on* X is a subset of $X \times X$, the set of ordered pairs (x, y) with $x \in X$ and $y \in X$. If $\succsim$ is a binary relation, we write $x \succsim y$ to mean $(x, y) \in \succsim$.

Here are two key properties that a binary relation may or may not possess.

Definition 16.2: Properties of binary relation

For any set X, a binary relation $\succsim$ on X is

- *complete* if for every $x \in X$ and $y \in X$ (with x and y not necessarily distinct) we have either $x \succsim y$ or $y \succsim x$ (or both)
- *transitive* if for every $x \in X$, $y \in X$, and $z \in X$ with $x \succsim y$ and $y \succsim z$ we have $x \succsim z$

The binary relation $\geq$ is complete, whereas $>$ and $=$ are not; all three are transitive.

We model an individual's preferences over a set of alternatives as a complete transitive binary relation, say $\succsim$, interpreting $x \succsim y$ to mean that the individual likes x at least as much as y.

Definition 16.3: Preference relation

For any set X, a *preference relation on* X is a complete transitive binary relation on X.

Given any preference relation, two associated binary relations are defined as follows.

Definition 16.4: Binary relations associated with preference relation

Let X be a set and and $\succsim$ a preference relation on X. The *strict preference relation* and *indifference relation* associated with $\succsim$ are the binary relations $\succ$ and $\sim$ defined by

$$\begin{aligned} x \succ y \quad &\Leftrightarrow \quad x \succsim y \text{ and not } y \succsim x \\ x \sim y \quad &\Leftrightarrow \quad x \succsim y \text{ and } y \succsim x. \end{aligned}$$

The strict preference relation $\succ$ and indifference relation $\sim$ associated with any preference relation $\succsim$ are transitive, and if $x \succsim y$ and $y \succ z$ then $x \succ z$. If $\succsim$ is the preference relation of an individual then $x \succ y$ may be interpreted to mean that the individual prefers x to y, and $x \sim y$ may be interpreted to mean that she regards x and y as equally appealing. Whenever I introduce a preference relation denoted $\succsim$, I implicitly introduce also the strict preference relation and indifference relation associated with $\succsim$, which I denote by $\succ$ and $\sim$, and whenever I introduce a preference relation denoted $\trianglerighteq$, I implicitly introduce the strict preference relation associated with $\trianglerighteq$, which I denote $\vartriangleright$.

If all the members of a set may be arranged in order, with no distinct members at the same position in the ordering, then the associated binary relation is called a linear order. ("Linear" because we can order the members along a line.)

Definition 16.5: Linear order

For any set X, a *linear order on* X is a complete transitive binary relation $\succsim$ on X with the property that whenever $x \in X$, $y \in X$, $x \succsim y$, and $y \succsim x$ we have $x = y$.

An example of a linear order on the set of real numbers is $\geq$ (greater than or equal to). An example of a linear order on the set of n-vectors is the lexicographic order, defined as follows. ("Lexicographic" because for languages written using an alphabet it is the way the words in a dictionary are ordered.)

Definition 16.6: Lexicographic order on $\mathbb{R}^n$

For any positive integer n, the *lexicographic order* on $\mathbb{R}^n$ is the binary relation $\geq_L$ defined by $x \geq_L y$ if and only if either $x_1 > y_1$ or for some $k \in \{1, \ldots, n-1\}$ we have $x_i = y_i$ for $i = 1, \ldots, k$ and $x_{k+1} > y_{k+1}$. This binary relation is complete and transitive, and is hence a preference relation.

In some models, an individual is assumed to not be indifferent between any two distinct alternatives. That is, her preference relation is assumed to be a linear order, so that it coincides with the strict preference relation associated with it. For convenience, in this case I say that her preference relation is strict.

Definition 16.7: Strict preference relation

For any set X, a preference relation $\succsim$ on X is *strict* if it is a linear order.

When analyzing an individual's behavior, it is often convenient to work with a function that represents her preference relation in the sense that it attaches a higher number to alternative x than to alternative y if and only if the individual prefers x to y.

Definition 16.8: Payoff function that represents preference relation

Let X be a set and let $\succsim$ be a preference relation on X. The function $u : X \to \mathbb{R}$ *represents* $\succsim$ if

$$x \succsim y \text{ if and only if } u(x) \geq u(y).$$

Not every preference relation can be represented by a payoff function. For example, a lexicographic preference ordering cannot be so represented. A sufficient condition for a preference relation to be represented by a payoff function is that it is continuous.

Definition 16.9: Continous preference relation

Let X be a subset of a Euclidean space. A preference relation $\succsim$ on X is *continuous* if for every sequence $(x^j, y^j)_{j=1}^{\infty}$ for which $x^j \in X$, $y^j \in X$, $x^j \succsim y^j$ for all j, and $\lim_{j\to\infty} x^j$ and $\lim_{j\to\infty} y^j$ exist, we have $\lim_{j\to\infty} x^j \succsim \lim_{j\to\infty} y^j$, or, equivalently, for every $x^* \in X$ the sets $\{x \in X : x \succsim x^*\}$ and $\{x \in X : x^* \succsim x\}$ are closed.

Proposition 16.1: Payoff function that represents preference relation

Let X be a subset of a Euclidean space and let $\succsim$ be a continuous preference relation on X. Then there is a continuous function $u : X \to \mathbb{R}$ that represents $\succsim$.

For a proof of a more general result see Debreu (1954), and for a proof of a special case see Mas-Colell et al. (1995).

In models in which a decision-maker has preferences over the numbers in an interval, a common assumption is that these preferences are represented by a single-peaked function, defined precisely as follows.

Definition 16.10: Single-peaked function

Let X be a convex subset of $\mathbb{R}$. A function $f : X \to \mathbb{R}$ is *single-peaked* if it is continuous and strictly quasiconcave. Equivalently, f is continuous and satisfies one of the following conditions.

- f is increasing
- f is decreasing
- there exists a number $x^* \in X$ such that f is increasing on $\{x \in X : x \leq x^*\}$ and decreasing on $\{x \in X : x \geq x^*\}$.

Uncertainty

A lottery over a set is a function that assigns a positive probability to each of a finite number of members of the set and 0 to the remaining members, with the sum of the probabilities equal to 1.

Definition 16.11: Lottery

For any set Z, a *lottery* over Z is a function $p: Z \to \mathbb{R}$ that assigns a positive number (probability) $p(z)$ to a finite number of members of Z and 0 to all other members, with $\sum_{z \in Z} p(z) = 1$. The lottery p with $p(z_k) = p_k$ for $k = 1, \ldots, K$ and $p(z) = 0$ otherwise is denoted $p_1 \cdot z_1 \oplus p_2 \cdot z_2 \oplus \cdots \oplus p_k \cdot z_k$ and the lottery that assigns probability 1 to a single alternative $z \in Z$ is denoted $[z]$.

Whenever a preference relation on a set of lotteries over a set is used in this book, I assume that it may be represented by the expected value of a real-valued function on the set. Such preference relations were first systematically studied by von Neumann and Morgenstern (1944); for this reason they are referred to as vNM preference relations. A preference relation on a set of lotteries over a set is a vNM preference relation if, and only if, it is continuous and has the independence property, defined as follows.

Definition 16.12: Continuous preference relation on set of lotteries

For any set Z, a preference relation $\succsim$ on the set of lotteries over Z is *continuous* if for any $a \in Z$, $b \in Z$, and $c \in Z$ such that $[a] \succ [b] \succ [c]$ there is a number α with $0 < \alpha < 1$ such that $[b] \sim \alpha \cdot a \oplus (1-\alpha) \cdot c$.

Definition 16.13: Independence property for preference relation on set of lotteries

For any set Z, a preference relation $\succsim$ on the set of lotteries over Z satisfies the *independence property* if for any lotteries $\alpha_1 \cdot z_1 \oplus \cdots \oplus \alpha_k \cdot z_k \oplus \cdots \oplus \alpha_K \cdot z_K$ and $\beta \cdot a \oplus (1-\beta) \cdot b$ over Z we have

$$[z_k] \succsim \beta \cdot a \oplus (1-\beta) \cdot b$$

$$\Leftrightarrow$$

$$\alpha_1 \cdot z_1 \oplus \cdots \oplus \alpha_k \cdot z_k \oplus \cdots \oplus \alpha_K \cdot z_K \succsim \alpha_1 \cdot z_1 \oplus \cdots \oplus \alpha_k \cdot (\beta \cdot a \oplus (1-\beta) \cdot b) \oplus \cdots \oplus \alpha_K \cdot z_K.$$

Proposition 16.2: Representation of preferences on set of lotteries by expected value of payoff function

Let Z be a finite set. A preference relation $\succsim$ on the set $L(Z)$ of lotteries over Z is continuous and satisfies the independence property if and only if it is a *vNM preference relation*: that is, there is a function $u: Z \to \mathbb{R}$ such that $\succsim$ is represented by the function $U: L(Z) \to \mathbb{R}$ defined by

$$U(p) = \sum_{z \in Z} p(z)u(z) \quad \text{for all } p \in L(Z).$$

Another function $v: Z \to \mathbb{R}$ has this property if and only if there is a number α and a positive number β such that $v(z) = \alpha + \beta u(z)$ for all $z \in Z$. Any such function is called a *Bernoulli function* for $\succsim$.

For a proof for the first part this result, see Propositions 3.1 and 3.2 of Osborne and Rubinstein (2023). For a proof of the second part, see Proposition 6.B.2 of Mas-Colell et al. (1995).

16.2 Sets of individuals and profiles

In some of the models in this book the set of individuals is identified with an interval of real numbers, to capture situations in which the number of individuals is large enough that each individual's behavior is insignificant relative to the totality of all individuals. In these cases, I sometimes need to refer to the size of a subset of individuals. For a subset of the interval that is a countable union of disjoint intervals, we can use the sum of the lengths of these intervals for this purpose. But extending this notion of size to all subsets while maintaining appealing properties for it is impossible. (See, for example, Royden 1968, Chapter 3.) I follow standard practice and restrict attention to Lebesgue-measurable subsets, taking the size of such a subset to be its Lebesgue measure (which, roughly, is the smallest total length of a collection of intervals whose union contains the subset). In particular, whenever I refer to a set of individuals in such a model, I mean a Lebesgue-measurable subset of the interval.

Given a set of individuals, I use the term *profile* to refer to a collection of objects, one for each individual. For example, if each individual $i \in N$ has a preference relation $\succsim_i$, then the *preference profile* for the set of individuals is the collection of these preference relations; if each individual $i \in N$ is associated with the action $a_i \in A_i$, then the *action profile* for the set of individuals is the collection of these actions. We can think of any such collection as a function that associates an object with each individual $i \in N$.

Definition 16.14: Profile

For a set N of individuals and any set S, a *profile* of members of S is a function that associates with each $i \in N$ a member of S.

One way to denote the profile that associates with each $i \in N$ the member s_i of S is $(s_i)_{i \in N}$. For any profile $(s_i)_{i \in N}$, (x_i, s_{-i}) denotes the profile that differs from $(s_i)_{i \in N}$ only in that the element for individual i is x_i rather than s_i (so that, in particular, $(s_i, s_{-i}) = (s_i)_{i \in N}$).

The notation $(s_i)_{i \in S}$ is most appealing if N is countable, but I use it also when N is uncountable (for example, an interval).

16.3 Brouwer's fixed-point theorem

The following result, due to Luitzen Egbertus Jan Brouwer (1881–1966), is used in some proofs that models have equilibria. For a proof of the result, see Smart (1974, Theorem 2.1.11).

Proposition 16.3: Brouwer's fixed point theorem

Let X be a compact convex subset of a Euclidean space and let $f : X \to X$ be a continuous function. Then f has a fixed point: there exists $x \in X$ with $f(x) = x$.

16.4 Strategic games

A strategic game is a model of the interaction among the members of a set of decision-makers. Each decision-maker, called a player, chooses an action and cares about the actions chosen by all decision-makers.

Definition 16.15: Strategic game

A *strategic game* $\langle N, (A_i)_{i \in N}, (\succsim_i)_{i \in N} \rangle$ consists of

players
a set N

and for each player $i \in N$

actions
a set A_i

preferences
a preference relation $\succsim_i$ over the set $\times_{j \in N} A_j$ of *action profiles*.

For any $i \in N$, a function $u_i : \times_{j \in N} A_j \to \mathbb{R}$ that represents $\succsim_i$ is a *payoff function* for player i.

A Nash equilibrium of a strategic game is an action profile with the property that no player is better off choosing a different action, given the actions of the remaining players. One interpretation of a Nash equilibrium is that it corresponds to a steady state in an environment in which each decision-maker plays the game many times against other decision-makers chosen randomly from populations of potential players. No decision-maker observes the identity of any particular player, so no decision-maker can condition her action in any play of the game on the actions chosen previously by any other particular decision-maker. But every decision-maker knows, from her long experience playing the game, the actions that the other players will take in any occurrence of the game.

Definition 16.16: Nash equilibrium of strategic game

For a strategic game $\langle N, (A_i)_{i \in N}, (\succsim_i)_{i \in N} \rangle$, a *Nash equilibrium* is an action profile $(a_i)_{i \in N} \in \times_{i \in N} A_i$ for which for every player $i \in N$

$$(a_i, a_{-i}) \succsim_i (x_i, a_{-i}) \text{ for all } x_i \in A_i.$$

The following result gives sufficient conditions for a strategic game to have a Nash equilibrium. For a proof of the result, see Osborne and Rubinstein (1994, Proposition 20.3).

Proposition 16.4: Existence of Nash equilibrium in strategic game

The strategic game $\langle N, (A_i)_{i \in N}, (\succsim_i)_{i \in N} \rangle$ in which N is finite has a Nash equilibrium if for every $i \in N$

- the set A_i of actions of player i is a nonempty compact convex subset of a Euclidean space

and the preference relation $\succsim_i$

- may be represented by a continuous payoff function
- is quasiconcave on A_i:

$$\{a_i \in A_i : (a_i, a^*_{-i}) \succsim_i a^*\} \text{ is convex for every } a^* \in \times_{j \in N} A_j.$$

A strategic game may have more than one Nash equilibrium, but if the game has two players and their interests are opposed, the players' payoffs in every Nash equilibrium are the same.

Definition 16.17: Strictly competitive strategic game

The two-player strategic game $\langle\{1,2\},(A_i)_{i\in N},(\succsim_i)_{i\in N}\rangle$ is *strictly competitive* if $a \succsim_1 b$ if and only if $b \succsim_2 a$ for all $a \in A_1 \times A_2$ and $b \in A_1 \times A_2$.

Proposition 16.5: Unique Nash equilibrium payoffs in strictly competitive strategic game

Every Nash equilibrium of a strictly competitive strategic game yields the same pair of payoffs.

For a proof of this result, see Osborne and Rubinstein (1994, Proposition 22.2).

The action a'_i of player i in a strategic game weakly dominates the action a_i if for *all* actions of the other players, i likes the outcome in which she chooses a'_i at least as much as the outcome when she chooses a_i, and for at least *one* collection of actions for the other players she prefers the outcome when she chooses a'_i to the outcome when she chooses a_i.

Definition 16.18: Weak domination in strategic game

Let $\langle N,(A_i)_{i\in N},(\succsim_i)_{i\in N}\rangle$ be a strategic game. The action $a'_i \in A_i$ of player $i \in N$ *weakly dominates* her action $a_i \in A_i$ if

$$(a'_i, x_{-i}) \succsim_i (a_i, x_{-i}) \text{ for every action profile } x$$
$$(a'_i, x_{-i}) \succ_i (a_i, x_{-i}) \text{ for some action profile } x.$$

A notion of equilibrium closely related to Nash equilibrium models the set of options of each individual as the set of probability distributions over a set of actions. The outcome of such an equilibrium is a probability distribution over action profiles, so that to define it we need to include in the description of the game the players' preferences over such probability distributions, not only their preferences over deterministic action profiles. We assume that these preference relations are vNM preference relations.

Definition 16.19: Strategic game with vNM preferences

A *strategic game with vNM preferences* $\langle N,(A_i)_{i\in N},(u_i)_{i\in N}\rangle$ consists of

players
a set N

and for each player $i \in N$

actions
a set A_i

preferences
a function $u_i : \times_{j\in N} A_j \to \mathbb{R}$ (a Bernoulli function whose expected value represents individual i's preferences over lotteries on the set $\times_{j\in N} A_j$ of action profiles).

To define the notion of mixed strategy equilibrium, it is convenient to first define the notion of a mixed strategy.

Definition 16.20: Mixed strategy of player in strategic game with vNM preferences

Let $\langle N, (A_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a strategic game with vNM preferences. For any player $i \in N$, a *mixed strategy* of player i in the game is a probability distribution over A_i.

Definition 16.21: Mixed strategy equilibrium of strategic game with vNM preferences

Let $\langle N, (A_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a strategic game with vNM preferences for which N and each set A_i are finite. For each player i define the function U_i by

$$U_i(x) = \sum_{a\in A} \left(\prod_{j\in N} x_j(a_j) \right) u_i(a) \quad \text{for every profile } x \text{ of mixed strategies,}$$

where $A = \times_{j\in N} A_j$, the set of action profiles, and $x_j(a_j)$ is the probability that j's mixed strategy x_j assigns to the action a_j. (Thus $U_i(x)$ is i's expected payoff for x.) A *mixed strategy equilibrium* of the game is a profile α of mixed strategies for which

$$U_i(\alpha_i, \alpha_{-i}) \geq U(x_i, \alpha_{-i}) \text{ for every mixed strategy } x_i \text{ of player } i.$$

The next result asserts that every finite strategic game has a mixed strategy equilibrium.

Proposition 16.6: Existence of mixed strategy equilibrium in finite strategic game with vNM preferences

Every strategic game with vNM preferences in which the set of players and the set of actions of each player are finite has a mixed strategy equilibrium.

For a proof of this result, see Proposition 33.1 of Osborne and Rubinstein (1994).

One interpretation of a mixed strategy equilibrium is an extension of the interpretation of a Nash equilibrium that I mention before Definition 16.16: it corresponds to a *stochastic* steady state in an environment in which each decision-maker plays the game many times against other decision-makers chosen randomly from populations of potential players. This interpretation and others are discussed in Section 3.2 of Osborne and Rubinstein (1994).

Consider a mixed strategy equilibrium in which some player's mixed strategy assigns positive probabilities to two actions, say a and b. Her expected payoff if she chooses a must equal her expected payoff if she chooses b, given the equilibrium mixed strategies of the other players, because if these payoffs differ she can increase her expected payoff by increasing the probability she assigns to the action that yields the higher payoff. That is, in the equilibrium she is indifferent between a and b, and has no positive incentive to choose them with the probabilities required by the equilibrium. The equilibrium probabilities are determined not by her optimization process, but by the equilibrium requirement that the *other* players must be indifferent between the actions to which their mixed strategies assign positive probability. This property of a mixed strategy equilibrium makes it more difficult to interpret than a strict Nash equilibrium, in which each player's payoff decreases if she deviates from her equilibrium action.

16.5 Bayesian games

We may model players' uncertainty about each other's characteristics by using the notion of a Bayesian game. The uncertainty is modeled by specifying a set of states. For each state, each player observes a signal; for any given signal, she cannot distinguish among the states that generate that signal for her. A player who observes the same signal for every state, for example, has no information about the state, and a player who observes a different signal for every state has perfect information. Each player has a prior belief about the probability of each state, which she updates, using the rule due to the eponymous Thomas Bayes, after observing her signal. (This formulation follows Osborne and Rubinstein (1994) in taking the players' prior beliefs as primitive, from which the posterior belief following the observation of a signal may be derived; Osborne (2004) takes the latter as primitive.)

Definition 16.22: Bayesian game

A *Bayesian game* $\langle N, \Omega, (A_i)_{i\in N}, (T_i)_{i\in N}, (\tau_i)_{i\in N}, (p_i)_{i\in N}, (u_i)_{i\in N}\rangle$ consists of

players
a set N

states
a set Ω

and for each player $i \in N$

actions
a set A_i

signals
a set T_i and a function $\tau_i : \Omega \to T_i$ that associates a signal with every state

prior beliefs
a probability measure p_i on Ω with $p_i(\tau_i^{-1}(t_i)) > 0$ for all $t_i \in T_i$

payoffs
a Bernoulli payoff function u_i over the set $(\times_{i\in N} A_i) \times \Omega$ of pairs (a, ω) consisting of an action profile a and a state ω (with i's preferences over pairs (a, ω) represented by the expected value of u_i).

The signals that a player may receive are referred to also as her possible *types*.

The condition $p_i(\tau_i^{-1}(t_i)) > 0$ for all $t_i \in T_i$ on the prior beliefs says that each signal of every player is possible: each player assigns positive probability to the set of states that generate each of her signals. (If a signal of a player is not possible, we can omit it from the description of the game.)

We define a Nash equilibrium of a Bayesian game as a Nash equilibrium of an associated strategic game G^* that has one player for each type of each player in the Bayesian game. That is, the set of players in G^* is the set of pairs (i, t_i) for $i \in N$ and $t_i \in T_i$. The set of actions of each player (i, t_i) is A_i. That is, for all the types of each player i the set of actions is A_i. To specify the payoffs of the players in G^*, consider player (i, t_i). The set of states that generate the signal t_i for player i in the Bayesian game is $\tau_i^{-1}(t_i)$, so the probability that (i, t_i) assigns to each state ω, derived from her prior belief using Bayes' rule, is

$$\Pr(\omega \mid t_i) = \begin{cases} p_i(\omega)/p_i(\tau_i^{-1}(t_i)) & \text{if } \omega \in \tau_i^{-1}(t_i) \\ 0 & \text{otherwise.} \end{cases} \tag{16.1}$$

Let a^* be an action profile in G^*. That is, a^* assigns a member of A_i to each pair (i, t_i) with $i \in N$ and $t_i \in T_i$. Denote the action assigned by a^* to (i, t_i) by $a^*(i, t_i)$ (rather than $a^*_{(i,t_i)}$, for readability). Then the action that player i takes in the Bayesian game when the state is ω is $a^*(i, \tau_i(\omega))$ and hence the expected payoff of player (i, t_i) in G^* for the action profile a^* is

$$u^*_{(i,t_i)}(a^*) = \sum_{\omega \in \Omega} \Pr(\omega \mid t_i) u_i((a^*(j, \tau_j(\omega)))_{j \in N}, \omega), \tag{16.2}$$

where $\Pr(\omega \mid t_i)$ is given by (16.1).

Definition 16.23: Nash equilibrium of Bayesian game

A *Nash equilibrium* of a Bayesian game $\langle N, \Omega, (A_i)_{i \in N}, (T_i)_{i \in N}, (\tau_i)_{i \in N}, (p_i)_{i \in N}, (u_i)_{i \in N} \rangle$ is a Nash equilibrium of the strategic game $\langle N^*, (A^*_j)_{j \in N^*}, (u^*_j)_{j \in N^*} \rangle$ in which

players
$N^* = \{(i, t_i) : i \in N \text{ and } t_i \in T_i\}$

and for all $i \in N$ and $t_i \in T_i$

actions
$A^*_{(i,t_i)} = A_i$

payoffs
$u^*_{(i,t_i)}$ is given by (16.2).

Like a Nash equilibrium of a strategic game, a Nash equilibrium of a Bayesian game may be interpreted as a steady state in an environment in which the decision-makers interact anonymously. In this interpretation, we assume that from her long experience playing the game, each decision-maker knows the action of each type of every other decision-maker.

16.6 Extensive games

An extensive game is a model of interaction among decision-makers that includes a specification of the sequential structure of the decision-making. At the start of the game, the members of a subset of the players (consisting possibly of a single player) simultaneously choose actions. This list of actions determines the subset of players who move next. Play continues in the same manner until the game ends. The resulting sequence of lists of actions is called a *terminal history*. The structure of the decision-making in the game is specified by the set of possible terminal histories. (This formulation follows Osborne 2004 in taking terminal histories as primitive, from which histories are derived; Osborne and Rubinstein 1994 takes histories as primitive and derives terminal histories from them.)

A sequence of lists of actions is a *history*. For any finite history $h = (a^1, \ldots, a^k)$, the *subhistories* of h are $\varnothing$ (the empty sequence) and all sequences of the form $(a^1, a^2, \ldots, a^m)$ with $1 \leq m \leq k$. (Note that h is a subhistory of itself.) A sequence $(a^1, a^2, \ldots, a^m)$ with $m \leq k-1$ is a *proper subhistory* of h. For any infinite history $h = (a^1, a^2, \ldots)$, the subhistories are $\varnothing$, all sequences of the form $(a^1, a^2, \ldots, a^m)$ with $m \geq 1$ (proper subhistories), and h itself.

16.6.1 Extensive game with perfect information

I start with the most general definition of an extensive game with perfect information, which allows for both simultaneous moves and chance moves.

Definition 16.24: Extensive game with perfect information, simultaneous moves, and chance moves

An *extensive game with perfect information, simultaneous moves, and chance moves* $\langle N, Z, P, (A_i(h))_{\{(i,h): i \in P(h)\}}, (q^h)_{\{h: c = P(h)\}}, (\succsim_i)_{i \in N} \rangle$ consists of

players
a set N

terminal histories
a set Z of sequences with the property that no member of Z is a proper subhistory of any other member of Z; the set of all subhistories of members of Z, proper or not, is the set H of *histories*, and $H \setminus Z$ is the set of *nonterminal histories*

player function
a function P that assigns either c (*chance*) or a subset of N to every nonterminal history

actions
for each nonterminal history h with $P(h) \subseteq N$ and each player $i \in P(h)$, a set $A_i(h)$ (the set of *actions* available to player i after the history h)

chance probabilities
for each nonterminal history h for which $P(h) = c$, a probability measure q^h on $\{a : (h, a) \in H\}$, with each such measure independent of every other such measure (q^h gives the probabilities with which chance selects actions after the history h)

preferences

for each player $i \in N$, a preference relation $\succsim_i$ on the set of lotteries over the set Z of terminal histories

such that the set H of histories, player function P, and sets $(A_i(h))_{\{(i,h):i\in P(h)\}}$ of actions are consistent in the sense that for every (nonterminal) history $h \in H \setminus Z$ for which $P(h) \subseteq N$ we have $\{a : (h,a) \in H\} = \times_{i\in P(h)} A_i(h)$.

Special cases in which chance moves are absent and/or no players ever move simultaneously are defined as follows.

Definition 16.25: Special cases of extensive game with perfect information, simultaneous moves, and chance moves

An extensive game with perfect information, simultaneous moves, and chance moves $\langle N, Z, P, (A_i(h))_{\{(i,h):i\in P(h)\}}, (q^h)_{\{h:c=P(h)\}}, (\succsim_i)_{i\in N}\rangle$ is

- an *extensive game with perfect information* if P assigns a single player (member of N) to every nonterminal history
- an *extensive game with perfect information and chance moves* if P assigns either c (chance) or a single player (member of N) to every nonterminal history
- an *extensive game with perfect information and simultaneous moves* if P assigns a subset of N to every nonterminal history.

In the first two cases, $P(h)$ is a singleton for all h, and I write $P(h) = i$ rather than $P(h) = \{i\}$. For such games, the set of actions available to each player when it is her turn to move does not need to be specified explicitly as part of the description of the game, but instead can be deduced from the set of terminal histories: if $P(h) = \{i\}$ (player i moves after the history h) then the set of actions available to i at h is $\{a : (h,a) \in H\}$.

An extensive game with perfect information that has finitely many terminal histories, each of finite length, may be represented in a diagram. An example is given in Figure 16.1. In any such diagram, the start of the game is indicated by a small circle. In this example, the start is located at the top; for some games, it is conveniently located at another position.

A key concept in the analysis of an extensive game is that of a strategy. The definition of a strategy is straightforward, but its interpretation is not. For discussions of the interpretation, see Osborne and Rubinstein (1994, Section 6.1.2),

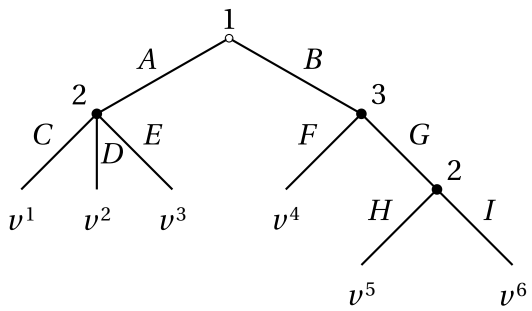


Figure 16.1 An example of an extensive game with perfect information. The start of the game (the empty history) is indicated by the small circle. Each line segment represents an action. The number near the empty history and the end of each nonterminal history is the player who moves after that history. The symbols $v^1, \ldots, v^6$ are profiles of payoffs that represent the players' preferences over terminal histories.

Osborne (2004, Section 5.2.1), or Osborne and Rubinstein (2023, Section 16.2).

Definition 16.26: Strategy in extensive game with perfect information, simultaneous moves, and chance moves

Let $\langle N, Z, P, (A_i(h))_{\{(i,h):i\in P(h)\}}, (q^h)_{\{h:c=P(h)\}}, (\succsim_i)_{i\in N}\rangle$ be an extensive game with perfect information, simultaneous moves, and chance moves. For any $i \in N$, a *strategy of player* i is a function that assigns to each history $h \in H$ for which $i \in P(h)$ a member of $A_i(h)$.

The definition of a Nash equilibrium of an extensive game is the analogue of the definition for a strategic game: a strategy profile with the property that no player prefers the terminal history that results from any change in her strategy, given the other players' strategies. This notion of equilibrium does not restrict the actions of players following histories that are inconsistent with the equilibrium. The main solution concept for an extensive game with perfect information, subgame perfect equilibrium, does restrict these actions: it requires that each player's strategy is optimal in the remainder of the game whenever the player moves, given the other players' strategies. To define this solution concept, I first define the subgame following any history to be the part of the game that remains after the history has occurred.

Definition 16.27: Subgame of extensive game with perfect information, simultaneous moves, and chance moves

Let $\langle N, Z, P, (A_i(h))_{\{(i,h):i\in P(h)\}}, (q^h)_{\{h:c=P(h)\}}, (\succsim_i)_{i\in N}\rangle$ be an extensive game with perfect information, simultaneous moves, and chance moves. For any nonterminal history $h \in H$, the *subgame following* h is the extensive game with perfect information, simultaneous moves, and chance moves with the following components.

Players
The set N.

Terminal histories
The set of all sequences h' such that $(h, h') \in Z$.

Player function
The player assigned to each proper subhistory h' of a terminal history is $P(h, h')$.

Actions
For all sequences h' such that $P(h, h') \subseteq N$, the set of actions of each player $i \in P(h, h')$ is $A_i(h, h')$.

Chance probabilities
For all sequences h' such that $P(h, h') = c$, the probability measure that determines the action selected by chance after h' is $q^{h,h'}$.

Preferences
Each player $i \in N$ prefers the lottery l over sequences h' such that $(h, h') \in Z$ to the lottery l' over such sequences if and only if according to $\succsim_i$ she prefers the lottery over Z generated by h followed by l to the lottery generated by h followed by l'.

Definition 16.28: Subgame perfect equilibrium of extensive game with perfect information, simultaneous moves, and chance moves

A *subgame perfect equilibrium* of an extensive game with perfect information, simultaneous moves, and chance moves $\langle N, Z, P, (A_i(h))_{\{(i,h): i \in P(h)\}}, (q^h)_{\{h: c = P(h)\}}, (\succsim_i)_{i \in N}\rangle$ is a strategy profile s^* such that for every player $i \in N$ and every history h with $i \in P(h)$,

$$L_h(s^*) \succsim_i L_h(r_i, s^*_{-i}) \text{ for every strategy } r_i \text{ of player } i,$$

where for any strategy profile s, $L_h(s)$ is the lottery over Z that assigns to each terminal history (h, h') the probability assigned to h' by the lottery over the terminal histories of the subgame following h that results when the players follow the prescriptions of s in the subgame.

A subgame perfect equilibrium of an extensive game with perfect information in which every terminal history is finite may be found by using backward induction. (For a precise description of the procedure for a game without simultaneous or chance moves, see Osborne and Rubinstein 2023, Section 16.3.)

In a subgame perfect equilibrium, for every subgame no player can generate an outcome in the subgame that she prefers by changing her strategy in the subgame. In particular, for every subgame the player who moves first cannot generate an outcome in the subgame that she prefers by changing her action at the start of the subgame. This second property is called the one-deviation property.

Definition 16.29: One-deviation property

A strategy profile in an extensive game with perfect information, simultaneous moves, and chance moves satisfies the *one-deviation property* if, for each player i and each history h after which i moves, i does not prefer any lottery over terminal histories generated by changing only her action at the start of the subgame following h, given the other players' strategies, to the lottery over terminal histories generated by the strategy profile in the subgame.

For a game in which the number of players is finite and every terminal history is finite, a strategy profile is a subgame perfect equilibrium if and only if it satisfies this property. This result reduces considerably the complexity of checking that a strategy profile is a subgame perfect equilibrium. For a proof of the result, see Osborne and Rubinstein (1994, Lemma 98.2, Exercise 102.1, and Exercise 103.3).

Proposition 16.7: Subgame perfect equilibrium of finite horizon extensive game and the one-deviation property

A strategy profile in an extensive game with perfect information, simultaneous moves, and chance moves in which the number of players is finite and every terminal history is finite is a subgame perfect equilibrium if and only if it satisfies the one-deviation property.

An implication of this result is that for such a game a subgame perfect equilibrium may be found (if one exists) by using the procedure of backward induction. We find a Nash equilibrium of the last subgame in each terminal history, replace the subgame with the outcome of the Nash equilibrium, and then repeat the process for the resulting game, working back to the start of the game. For a game without simultaneous or chance moves in which the player who moves at the start of each subgame has an optimal action, the procedure of backward induction generates at least one strategy profile, so that such a game has a subgame perfect equilibrium. A sufficient condition for each player to have an optimal action whenever she moves is that her preferences are represented by a payoff function that takes finitely many values, implying the next result. (A stronger condition is that the number of terminal histories is finite.)

Proposition 16.8: Existence of subgame perfect equilibrium for finite extensive game

Every extensive game with perfect information in which the number of players is finite, every terminal history is finite, and every player's preference relation is represented by a payoff function that takes finitely many values has a subgame perfect equilibrium.

A version of Proposition 16.7 holds for games in which the terminal histories are not finite, but for every player the difference between the payoffs of pairs of terminal histories whose first t components coincide converges to zero as t increases without bound. For a proof of this result see Theorem 4.2 of Fudenberg and Tirole (1991, 110).

Proposition 16.9: Subgame perfect equilibrium of extensive game with perfect information, simultaneous moves, and chance moves and the one-deviation property

Let G be an extensive game with perfect information, simultaneous moves, and chance moves in which the set of players is finite. Denote the set of terminal histories by Z and suppose that the preferences of each player i over lotteries over Z are represented by the expected value of a function u_i for which

$$\lim_{t\to\infty} \sup_{h,\tilde{h}\in Z} \{|u_i(h) - u_i(\tilde{h})| : h^t = \tilde{h}^t\} = 0,$$

where for any terminal history h, h^t consists of the first t components of h. A strategy profile in G is a subgame perfect equilibrium if and only if it satisfies the one-deviation property.

16.6.2 Bayesian extensive game with observable actions

An extensive game with perfect information, simultaneous moves, and chance moves models a situation in which each player knows the structure of the interaction (who moves when, and which actions they can choose) and all players' characteristics. A Bayesian extensive game with observable actions models a situation in which each player knows the structure of the interaction but does not know the other players' characteristics. For each player i there is a set Θ_i of possible types and a probability measure p_i over this set. The probability measures $(p_i)_{i\in N}$ are independent; the profile of the players' types is drawn according to these measures. Each player knows her own type, but not the type of any other individual. The type profile $\theta \in \times_{j\in N}\Theta_j$ determines each player's payoff function over terminal histories. We can think of the game as one in which chance first determines a type profile, then the players engage in an extensive game with perfect information and simultaneous moves in which the payoffs are determined by the type profile.

Definition 16.30: Bayesian extensive game with observable actions

A *Bayesian extensive game with observable actions* $\langle N, Z, P, (A_i(h))_{\{(i,h):i\in P(h)\}}, (\Theta_i)_{i\in N}, (p_i)_{i\in N}, (u_i)_{i\in N}\rangle$ consists of

players
a set N

terminal histories
a set Z of sequences with the property that no member of Z is a proper subhistory of any other member of Z; the set of all subhistories of members of Z, proper or not, is the set H of *histories*, and $H \setminus Z$ is the set of *nonterminal histories*

player function
a function P that assigns a subset of N to every nonterminal history

and for each player $i \in N$

actions
for each nonterminal history h a set $A_i(h)$ (the set of *actions* available to player i after the history h)

types
a set Θ_i

probabilities
a probability measure p_i on Θ_i with $p_i(\theta_i) > 0$ for all $\theta_i \in \Theta_i$ ($p_i(\theta_i)$ is the probability that i's type is θ_i)

payoff function
$u_i : \Theta \times Z \to \mathbb{R}$, a Bernoulli function over pairs consisting of a profile of types and a terminal history (the expected value of $u_i(\theta, h)$ represents i's preferences on the set of lotteries over $\Theta \times Z$)

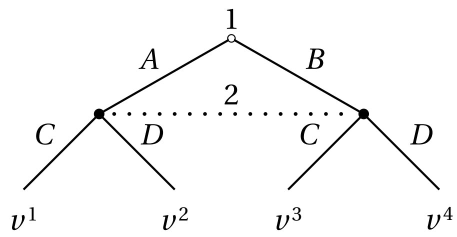


Figure 16.2 An example of an extensive game with imperfect information. The dotted line indicates that the histories A and B are in the same information set; when player 2 moves, she does not know whether player 1 chose A or B at the start of the game.

such that the measures $(p_i)_{i\in N}$ are independent and the set H of histories, player function P, and sets $(A_i(h))_{\{(i,h):i\in P(h)\}}$ of actions are consistent in the sense that for every (nonterminal) history $h \in H \setminus Z$ we have $\{a : (h,a) \in H\} = \times_{i\in P(h)} A_i(h)$.

A strategy of each player i in a Bayesian extensive game with observable actions specifies, for each type $\theta_i \in \Theta_i$, a strategy for i in the extensive game with perfect information and simultaneous moves. That is, a strategy of player i is a function that associates with each type $\theta_i \in \Theta_i$ a function that assigns to each history $h \in H$ for which $i \in P(h)$ a member of $A_i(h)$. A notion of equilibrium may be defined for a general Bayesian extensive game with observable actions (see Definition 232.1 in Osborne and Rubinstein 1994), but for the specific model I analyze (see Definition 8.7) a simpler notion, which I specify in Definition 8.8, suffices.

16.6.3 *Extensive game with imperfect information*

An extensive game with imperfect information allows for the possibility that each player, when choosing an action, does not know the actions chosen previously by the other players. This lack of information is modeled by assuming that each player i, when choosing an action, knows only that the history is a member of some set I_i, called an information set. We assume that following every history in a given information set, the set of actions available to the player who moves is the same, so that the set of actions she faces gives her no information about the history that has led to the information set. In a diagrammatic representation of a game, I connect the ends of all the histories in each information set with a dotted line, as in Figure 16.2.

Definition 16.31: Extensive game with imperfect information

An *extensive game with imperfect information* $\langle N, Z, P, (q^h)_{\{h:c=P(h)\}}, (\mathscr{I}_i)_{i \in N}, (\succsim_i)_{i \in N} \rangle$ consists of

players
a set N

terminal histories
a set Z of sequences with the property that no member of Z is a proper subhistory of any other member of Z; the set of all subhistories of members of Z, proper or not, is the set H of *histories*, $H \setminus Z$ is the set of *nonterminal histories*, and for any nonterminal history h, the set of actions available following h is $A(h) = \{a : (h, a) \in H\}$

player function
a function P that assigns either c (*chance*) or a member of N to every nonterminal history

chance probabilities
for each nonterminal history h for which $P(h) = c$, a probability measure q^h on $A_c(h) = \{a : (h, a) \in H\}$, with each such measure independent of every other such measure ($q^h(a)$ is the probability with which chance selects a after the history h)

information partitions
for each player $i \in N$ a partition $\mathscr{I}_i$ of $\{h \in H : P(h) = i\}$ with the property that for any $I_i \in \mathscr{I}_i$, $A(h) = A(h')$ whenever $h \in I_i$ and $h' \in I_i$; the common value of $A(h)$ for all $h \in I_i$ is denoted $A(I_i)$ ($\mathscr{I}_i$ is the *information partition* of player i, and a member of $\mathscr{I}_i$ is an *information set*)

preferences
for each player $i \in N$ a preference relation $\succsim_i$ on the set of lotteries over the set Z of terminal histories.

The solution concept I use for an extensive game with imperfect information allows for the possibility that players' actions are probabilistic. Specifically, a behavioral strategy assigns to each of a player's information sets a probability distribution over the set of actions available at that information set, with the probability distribution for each information set independent of the probability distribution for all of the player's other information sets.

Definition 16.32: Behavioral strategy in extensive game with imperfect information

A *behavioral strategy* of player $i \in N$ in an extensive game with imperfect information $\langle N, Z, P, (q^h)_{\{h:c=P(h)\}}, (\mathscr{I}_i)_{i\in N}, (\succsim_i)_{i\in N}\rangle$ is a function that assigns to each of i's information sets $I_i \in \mathscr{I}_i$ a probability distribution over the actions in $A(I_i)$, with the property that the probability distribution for any given information set I_i of player i is independent of the distributions for all her other information sets.

We assume that each player's choice at each of her information sets is based on her belief about the history in the information set that has occurred. For an information set reached with positive probability given the strategy profile, this belief may be derived from the strategy profile via Bayes' law, but the same is not true for an information set reached with probability zero given the strategy profile. We finesse this issue by making each player's belief part of the equilibrium, with the requirement that for information sets reached with positive probability given the strategy profile, the probability assigned by the belief to each history in the information set is the one derived from the strategy profile using Bayes' law.

Definition 16.33: Belief system

A *belief system* in an extensive game with imperfect information is a function that assigns to each information set a probability distribution over the histories in that information set.

Definition 16.34: Assessment

An *assessment* in an extensive game with imperfect information is a pair consisting of a profile of behavioral strategies and a belief system.

The solution concept that I use imposes two conditions on an assessment. First, for each information set of each player, the player's strategy is required to be optimal in the part of the game that follows the information set, given the strategy profile and the player's belief regarding the history that has occurred. Second, for each information set reached with positive probability given the strategy profile, the belief system is required to assign to each history in the information set the probability that the history occurs conditional on the information set's being reached, given the strategy profile. This requirement is called weak consistency of beliefs with strategies; the word "weak" honors the fact that the requirement puts no restriction on the probabilities assigned by the belief system to histories in information sets that are not reached if the players adhere to the strategy profile. The solution concept is called weak sequential equilibrium; the concept of sequential equilibrium, which I do not use in this book, imposes an additional condition on an assessment. (The notion of weak sequential equilibrium is sometimes called weak perfect Bayesian equilibrium, although no related notion of perfect Bayesian equilibrium is defined for the class of all extensive games with imperfect information.)

Definition 16.35: Weak sequential equilibrium

An assessment (β, μ) in an extensive game with imperfect information $\langle N, Z, P, (q^h)_{\{h:c=P(h)\}}, (\mathscr{I}_i)_{i\in N}, (\succsim_i)_{i\in N}\rangle$, where β is a behavioral strategy profile and μ is a belief system, is a *weak sequential equilibrium* if it satisfies the following two conditions.

Sequential rationality

For each player $i \in N$ and each information set $I_i \in \mathscr{I}_i$,

$$O_{I_i}(\beta, \mu) \succsim_i O_{I_i}((\gamma_i, \beta_{-i}), \mu) \text{ for each behavioral strategy } \gamma_i \text{ of player } i,$$

where for any profile σ of behavioral strategies $O_{I_i}(\sigma, \mu)$ is the probability distribution over terminal histories conditional on play reaching I_i, given σ and μ.

Weak consistency of beliefs with strategies

For each player $i \in N$ and every information set $I_i \in \mathscr{I}_i$ reached with positive probability given the strategy profile β, the probability assigned by the belief system μ to each history $h^* \in I_i$ is

$$\frac{\Pr(h^* \text{ according to } \beta)}{\sum_{h\in I_i} \Pr(h \text{ according to } \beta)}.$$

16.7 Coalitional games

A coalitional game with transferable payoff models a situation in which each group of players can obtain a certain total payoff, independently of the behavior of the remaining players, and payoff may be distributed in any way among the players. A subset of the set of players is called a coalition, and the total payoff available to it is called its worth. I restrict attention to games in which the set of players is finite.

Definition 16.36: Coalitional game with transferable payoff

A *coalitional game with transferable payoff* $\langle N, v \rangle$ consists of a finite set N (of players) and a function v that assigns a real number $v(S)$ (the *worth* of S) to every nonempty subset S of N (*coalition*). A *payoff profile* for the game is a profile $(x_i)_{i \in N}$ of real numbers; it is *feasible* if $\sum_{i \in N} x_i = v(N)$.

16.7.1 Core

One solution concept for coalitional games with transferable payoff is the core, the set of feasible payoff profiles with the property that no coalition can by itself make all its members better off.

Definition 16.37: Core of coalitional game with transferable payoff

Let $\langle N, v \rangle$ be a coalitional game with transferable payoff. A coalition S can *improve upon* the payoff profile $(x_i)_{i \in N}$ if $\sum_{i \in S} x_i < v(S)$. The *core* of $\langle N, v \rangle$ is the set of feasible payoff profiles upon which no coalition can improve.

The core of a coalitional game with transferable payoff may be empty. For example, the game $\langle N, v \rangle$ with $N = \{1, 2, 3\}$, $v(\{1, 2, 3\}) = v(\{1, 2\}) = v(\{1, 3\}) = v(\{2, 3\}) = 1$, and $v(\{1\}) = v(\{2\}) = v(\{3\}) = 0$ (in which a majority rules) has an empty core: for each player $i \in N$, the coalition $N \setminus \{i\}$ can improve upon any feasible payoff profile $(x_i)_{i \in N}$ with $x_i > 0$.

16.7.2 Shapley value

Another solution concept for coalitional games with transferable payoff is the Shapley value (due to Lloyd S. Shapley, 1923–2016). Unlike the core, the Shapley value assigns a single payoff profile to every game. I first define a value to be a solution of this type.

Definition 16.38: Value of coalitional game with transferable payoff

A *value* for coalitional games with transferable payoff is a function that associates with every such game a unique feasible payoff profile.

The Shapley value assigns to each player the average of the amount by which the player's presence increases the worth of the coalition that precedes her in a random ordering of the players.

Definition 16.39: Shapley value of coalitional game with transferable payoff

Let $\langle N, v\rangle$ be a coalitional game with transferable payoff and let n be the number of members of N. The *Shapley value* assigns to $\langle N, v\rangle$ the payoff profile $(x_i)_{i\in N}$ for which

$$x_i = \frac{1}{n!}\sum_{R\in\mathscr{R}}(v(S_i^R \cup \{i\}) - v(S_i^R)) \quad \text{for each } i \in N, \tag{16.3}$$

where $\mathscr{R}$ is the set of all $n!$ orderings of N, S_i^R is the set of players who precede i in the ordering R, and $v(\varnothing) = 0$.

The Shapley value satisfies the following properties, and is the only value that does so.

Definition 16.40: Properties of value of coalitional game with transferable payoff

A value ψ for coalitional games with transferable payoff is

symmetric
if for every game $\langle N, v\rangle$ and players $i \in N$ and $j \in N$ for which $v(S\cup\{i\}) = v(S\cup\{j\})$ for every coalition S that includes neither i nor j we have $\psi_i(N, v) = \psi_j(N, v)$

null-consistent
if for every game $\langle N, v\rangle$ and player $i \in N$ for which $v(S\cup\{i\}) = v(S)$ for every coalition S of which i is not a member we have $\psi_i(N, v) = 0$

additive
if for all games $\langle N, v\rangle$ and $\langle N, v'\rangle$, $\psi_i(N, w) = \psi_i(N, v) + \psi_i(N, v')$ for all $i \in N$, where $\langle N, w\rangle$ is the game defined by $w(S) = v(S) + v'(S)$ for every coalition S.

Proposition 16.10: Axiomatic characterization of Shapley value

Let N be a finite set. The Shapley value is the only value for coalitional games with transferable payoff with player set N that is symmetric, null-consistent, and additive.

For a proof of this result, see Osborne and Rubinstein (1994, Proposition 293.1).

Suppose that the players' payoffs in a coalitional game with transferable payoff $\langle N, v\rangle$ are determined by the value ψ. If player i leaves the game, the payoff of any player $j \neq i$ changes from $\psi_j(N, v)$ to $\psi_j(N \setminus \{i\}, v^{N \setminus \{i\}})$, where $v^{N \setminus \{i\}}$ is the restriction of v to coalitions in $N \setminus \{i\}$: $v^{N \setminus \{i\}}(S) = v(S)$ for all nonempty $S \subseteq N \setminus \{i\}$. That is, $\psi_j(N, v) - \psi_j(N \setminus \{i\}, v^{N \setminus \{i\}})$ is the amount j loses when i departs. The Shapley value has the property that this amount is the same for all players i and j, and is the only value with this property. One motivation for the property is that for every objection of a certain type by one player to the payoff profile there is a valid counterobjection by another player (Osborne and Rubinstein 1994, 290–291).

Proposition 16.11: Characterization of Shapley value in terms of balanced contributions

The Shapley value is the only value for coalitional games with transferable payoff that satisfies the condition

$$\psi_i(N, v) - \psi_i(N \setminus \{j\}, v^{N \setminus \{j\}}) = \psi_j(N, v) - \psi_j(N \setminus \{i\}, v^{N \setminus \{i\}})$$

for every coalitional game $\langle N, v \rangle$ and all $i \in N$ and $j \in N$.

For a proof of this result, see Osborne and Rubinstein (1994, Proposition 291.3).

The following result is used in the proof of Proposition 11.3.

Lemma 16.1: Shapley value of dual of coalitional game with transferable payoff

Let $\langle N, v \rangle$ be a coalitional game with transferable payoff and let $v^\#(S) = v(N) - v(N \setminus S)$ for each subset S of N. The Shapley value assigns the same payoff profile to $\langle N, v^\# \rangle$ as it does to $\langle N, v \rangle$.

Proof

Let R be an ordering of N, let R' be the reverse ordering, and let $i \in N$. Then the set of individuals who come before i in R' is the set of individuals who come after i in R: $S_i^{R'} = N \setminus (S_i^R \cup \{i\})$. Thus

$$\begin{aligned} v^\#(S_i^{R'} \cup \{i\}) - v^\#(S_i^{R'}) &= (v(N) - v(N \setminus (S_i^R \cup \{i\}))) - (v(N) - v(N \setminus S_i^R)) \\ &= v(N \setminus S_i^R) - v(N \setminus (S_i^R \cup \{i\})) \\ &= v(S_i^R \cup \{i\}) - v(S_i^R). \end{aligned}$$

Hence by (16.3) i's payoff in the Shapley value of $\langle N, v^\# \rangle$ is equal to her payoff in the Shapley value of $\langle N, v \rangle$.

16.8 Optimization

The results in this section are used in Propositions 8.8 and 11.1.

Proposition 16.12: Necessary conditions for solution of unconstrained optimization problem

Let $S \subseteq \mathbb{R}^n$ and let $f : S \to \mathbb{R}$. If the point x^* in the interior of S is a local maximizer or minimizer of f and the partial derivative of f with respect to its jth argument exists at x then $f'_j(x^*) = 0$. In particular, if all the partial derivatives of f exist at x^* then

$$f'_i(x^*) = 0 \text{ for } i = 1, \ldots, n.$$

For a proof, see Sydsæter (1981, Theorem 5.7).

Proposition 16.13: Conditions under which first-order conditions are necessary and sufficient for solution of unconstrained optimization problem

Let $S \subseteq \mathbb{R}^n$ be convex and let $f : S \to \mathbb{R}$ be differentiable.

- If f is concave then a point x^* in the interior of S is a (global) maximizer of f in S if and only if it is a stationary point of f (i.e. $f'_i(x^*) = 0$ for $i = 1, \ldots, n$).
- If f is convex then a point x^* in the interior of S is a (global) minimizer of f in S if and only if it is a stationary point of f (i.e. $f'_i(x^*) = 0$ for $i = 1, \ldots, n$).

For a proof, see Sydsæter (1981, Theorem 5.18). (Sydsæter's result assumes that f is continuously differentiable, but that assumption is unnecessary because every differentiable concave or convex function is continuously differentiable by Rockafellar 1970, Corollary 25.5.1.)

Proposition 16.14: Necessary conditions for solution of optimization problem with equality constraint

Let $S \subseteq \mathbb{R}^n$, let $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$ be continuously differentiable, let $c \in \mathbb{R}$, and let x^* be an interior point of S that solves the problem

$$\max_{x \in S} f(x) \text{ subject to } g(x) = c$$

or the problem

$$\min_{x \in S} f(x) \text{ subject to } g(x) = c$$

or is a local maximizer or minimizer of $f(x)$ subject to $g(x) = c$. Suppose also that $g_i'(x^*) \neq 0$ for some $i \in \{1, \ldots, n\}$.

Then there is a unique number λ such that

$$f_i'(x^*) - \lambda g_i'(x^*) = 0 \text{ for } i = 1, \ldots, n.$$

In addition, $g(x^*) = c$.

For a proof of a more general result (for problems with many constraints), see Sydsæter (1981, Theorem 5.20).

Proposition 16.15: Conditions under which first-order conditions are sufficient for solution of optimization problem with equality constraint

Let $S \subseteq \mathbb{R}^n$ be open and convex, let $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$ be differentiable, and let $c \in \mathbb{R}$. Suppose that there exists a number λ and an interior point x^* of S such that

$$f_i'(x^*) - \lambda g_i'(x^*) = 0 \text{ for } i = 1, \ldots, n.$$

Suppose further that $g(x^*) = c$.

Define the function $\mathscr{L} : S \to \mathbb{R}$ by

$$\mathscr{L}(x) = f(x) - \lambda(g(x) - c) \text{ for all } x \in S.$$

- If $\mathscr{L}$ is concave—in particular if f is concave and λg is convex—then x^* solves the problem $\max_{x \in S} f(x)$ subject to $g(x) = c$.
- If $\mathscr{L}$ is convex—in particular if f is convex and λg is concave—then x^* solves the problem $\min_{x \in S} f(x)$ subject to $g(x) = c$.

For a proof of a more general result (for problems with many constraints), see Sydsæter (1981, Theorem 5.21).

References

Acemoglu, Daron, and James A. Robinson (2000), "Why did the West extend the franchise? Democracy, inequality, and growth in historical perspective." *Quarterly Journal of Economics*, 115, 1167–1199, 10.1162/003355300555042. [517]

Ali, S. Nageeb, B. Douglas Bernheim, and Xiaochen Fan (2019), "Predictability and power in legislative bargaining." *Review of Economic Studies*, 86, 500–525, 10.1093/restud/rdy013. [490]

Anesi, Vincent, and Daniel J. Seidmann (2015), "Bargaining in standing committees with an endogenous default." *Review of Economic Studies*, 82, 825–867, 10.1093/restud/rdv009. [490]

Ansolabehere, Stephen, and James M. Snyder (2000), "Valence politics and equilibrium in spatial election models." *Public Choice*, 103, 327–336. [280]

Aragonès, Enriqueta, and Thomas R. Palfrey (2002), "Mixed equilibrium in a Downsian model with a favored candidate." *Journal of Economic Theory*, 103, 131–161, 10.1006/jeth.2001.2821. [280]

Arrow, Kenneth J. (1951), *Social choice and individual values*, first edition. Wiley, New York, 10.12987/9780300186987. [68]

Ashworth, Scott (2006), "Campaign finance and voter welfare with entrenched incumbents." *American Political Science Review*, 100, 55–68, 10.1017/s0003055406062009. [430]

Aumann, Robert J., and Mordecai Kurz (1977), "Power and taxes." *Econometrica*, 45, 1137–1161, 10.2307/1914063. [404]

Austen-Smith, David (1984), "Two-party competition with many constituencies." *Mathematical Social Sciences*, 7, 177–198, 10.1016/0165-4896(84)90004-0. [280]

Austen-Smith, David, and Jeffrey S. Banks (1996), "Information aggregation, rationality, and the Condorcet Jury Theorem." *American Political Science Review*, 90, 34–45, 10.2307/2082796. [227]

Austen-Smith, David, and Jeffrey S. Banks (1999), *Positive political theory I*. University of Michigan Press, Ann Arbor, 10.3998/mpub.14228. [ix, 68]

Austen-Smith, David, and Jeffrey S. Banks (2005), *Positive political theory II*. University of Michigan Press, Ann Arbor, 10.3998/mpub.105376. [ix, 93, 279, 318, 478, 490]

Banks, Jeffrey S. (1985), "Sophisticated voting outcomes and agenda control." *Social Choice and Welfare*, 1, 295–306, 10.1007/bf00649265. [183]

Banks, Jeffrey S., and John Duggan (2000), "A bargaining model of collective choice." *American Political Science Review*, 94, 73–88, 10.2307/2586381. [483, 490]

Baron, David P. (1994), "Electoral competition with informed and uninformed voters." *American Political Science Review*, 88, 33–47, 10.2307/2944880. [430]

Baron, David P., and John A. Ferejohn (1989), "Bargaining in legislatures." *American Political Science Review*, 83, 1181–1206, 10.2307/1961664. [490]

Bell, David E. (1982), "Regret in decision making under uncertainty." *Operations Research*, 30, 961–981, 10.1287/opre.30.5.961. [124]

Bernhardt, Dan, Peter Buisseret, and Sinem Hidir (2020), "The race to the base." *American Economic Review*, 110, 922–942, 10.1257/aer.20181606. [277, 280]

Bernhardt, Dan, John Duggan, and Francesco Squintani (2007), "Electoral competition with privately-informed candidates." *Games and Economic Behavior*, 58, 1–29, 10.1016/j.geb.2006.03.004. [279]

Bernhardt, Dan, John Duggan, and Francesco Squintani (2009), "Private polling in elections and voter welfare." *Journal of Economic Theory*, 144, 2021–2056, 10.1016/j.jet.2008.05.013. [255, 279]

Besley, Timothy, and Stephen Coate (1997), "An economic model of representative democracy." *Quarterly Journal of Economics*, 112, 85–114, 10.1162/003355397555136. [360]

Black, Duncan (1948a), "The decisions of a committee using a special majority." *Econometrica*, 16, 245–261, 10.2307/1907278. [182]

Black, Duncan (1948b), "On the rationale of group decision-making." *Journal of Political Economy*, 56, 23–34, 10.1086/256633. [93, 182]

Black, Duncan (1958), *The theory of committees and elections*. Cambridge University Press, Cambridge. [68, 182]

Blackorby, Charles, David Donaldson, and John A. Weymark (1984), "Social choice with interpersonal utility comparisons: a diagrammatic introduction." *International Economic Review*, 25, 327–356, 10.2307/2526200. [68]

Bleichrodt, Han, and Peter P. Wakker (2015), "Regret theory: a bold alternative to the alternatives." *Economic Journal*, 125, 493–532, 10.1111/ecoj.12200. [124]

Börgers, Tilman (2004), "Costly voting." *American Economic Review*, 94, 57–66, 10.1257/000282804322970706. [124]

Bossert, Walter, and John A. Weymark (2004), "Utility in social choice." In *Handbook of utility theory* (Salvador Barberà, Peter J. Hammond, and Christian Seidl, eds.), volume 2, 1099–1177, Kluwer, Boston, 10.1007/978-1-4020-7964-1_7. [68]

Bouton, Laurent, Micael Castanheira, and Aniol Llorente-Saguer (2017), "Multicandidate elections: aggregate uncertainty in the laboratory." *Games and Economic Behavior*, 101, 132–150, 10.1016/j.geb.2015.10.003. [148]

Brennan, Geoffrey, and Alan Hamlin (1998), "Expressive voting and electoral equilibrium." *Public Choice*, 95, 149–175, 10.1023/A:1004936203144. [280]

Brett, Craig, and John A. Weymark (2017), "Voting over selfishly optimal nonlinear income tax schedules." *Games and Economic Behavior*, 101, 172–188, 10.1016/j.geb.2016.02.008. [404]

Brett, Craig, and John A. Weymark (2020), "Majority rule and selfishly optimal nonlinear income tax schedules with discrete skill levels." *Social Choice and Welfare*, 54, 337–362, 10.1007/s00355-019-01188-4. [404]

Brusco, Sandro, Marcin Dziubiński, and Jaideep Roy (2012), "The Hotelling–Downs model with runoff voting." *Games and Economic Behavior*, 74, 447–469, 10.1016/j.geb.2011.08.004. [360]

Calvert, Randall L. (1985), "Robustness of the multidimensional voting model: candidate motivations, uncertainty, and convergence." *American Journal of Political Science*, 29, 69–95, 10.2307/2111212. [279, 318]

Camion, Paul (1959), "Chemins et circuits hamiltoniens des graphes complets." *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 249, 2151–2152. [182]

Campbell, Colin M. (1999), "Large electorates and decisive minorities." *Journal of Political Economy*, 107, 1199–1217, 10.1086/250094. [124]

Chamberlain, Gary, and Michael Rothschild (1981), "A note on the probability of casting a decisive vote." *Journal of Economic Theory*, 25, 152–162, 10.1016/0022-0531(81)90022-3. [124]

Cho, In-Koo, and David M. Kreps (1987), "Signaling games and stable equilibria." *Quarterly Journal of Economics*, 102, 179–221, 10.2307/1885060. [438]

Cho, Seok-Ju (2014), "Voting equilibria under proportional representation." *American Political Science Review*, 108, 281–296, 10.1017/s0003055414000136. [149]

Cho, Seok-ju, and John Duggan (2009), "Bargaining foundations of the median voter theorem." *Journal of Economic Theory*, 144, 851–868, 10.1016/j.jet.2008.06.011. [490]

Cho, Seok-ju, and John Duggan (2015), "A folk theorem for the one-dimensional spatial bargaining model." *International Journal of Game Theory*, 44, 933–948, 10.1007/s00182-014-0460-8. [490]

Coate, Stephen (2004), "Pareto-improving campaign finance policy." *American Economic Review*, 94, 628–655, 10.1257/0002828041464443. [430]

Coate, Stephen, and Michael Conlin (2004), "A group rule-utilitarian approach to voter turnout: theory and evidence." *American Economic Review*, 94, 1476–1504, 10.1257/0002828043052231. [195]

Condorcet, Marquis de (Marie Jean Antoine Nicolas de Caritat) (1785), *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, Paris, 10.1017/cbo9781139923972. [67]

Coughlin, Peter J. (1986), "Elections and income distribution." *Public Choice*, 50, 27–91, 10.1007/bf00124925. [404]

Dasgupta, Partha, and Eric S. Maskin (2008), "On the robustness of majority rule." *Journal of the European Economic Association*, 6, 949–973, 10.1162/JEEA.2008.6.5.949. [67]

d'Aspremont, Claude, and Louis Gevers (1977), "Equity and the informational basis of collective choice." *Review of Economic Studies*, 44, 199–209, 10.2307/2297061. [68]

Davis, Otto A., Morris H. DeGroot, and Melvin J. Hinich (1972), "Social preference orderings and majority rule." *Econometrica*, 40, 147–157, 10.2307/1909727. [68]

De Sinopoli, Francesco, and Giovanna Iannantuoni (2007), "A spatial model where proportional rule leads to two-party equilibria." *International Journal of Game Theory*, 35, 267–286, 10.1007/s00182-006-0056-z. [149]

Debreu, Gerard (1954), "Representation of a preference ordering by a numerical function." In *Decision processes* (Robert M. Thrall, Clyde H. Coombs, and Robert L. Davis, eds.), chapter 11, 159–165, Wiley, New York. [526]

Demange, Gabrielle (1982), "Single-peaked orders on a tree." *Mathematical Social Sciences*, 3, 389–396, 10.1016/0165-4896(82)90020-8. [68]

Dhillon, Amrita, and Ben Lockwood (2004), "When are plurality rule voting games dominance-solvable?" *Games and Economic Behavior*, 46, 55–75, 10.1016/s0899-8256(03)00050-2. [148, 149]

Downs, Anthony (1957), *An economic theory of democracy*. Harper and Row, New York. [279]

Droop, Henry R. (1881), "On methods of electing representatives." *Journal of the Statistical Society of London*, 44, Part II, 141–196, 10.2307/2339223. [360]

Duggan, John (2013), "Uncovered sets." *Social Choice and Welfare*, 41, 489–535, 10.1007/s00355-012-0696-9. [183]

Duggan, John (2014), "A survey of equilibrium analysis in spatial models of elections." [Unpublished paper, University of Rochester] [280, 318]

Duggan, John, and César Martinelli (2015), "The political economy of dynamic elections: a survey and some new results." [Unpublished paper] [456]

Duggan, John, and César Martinelli (2017), "The political economy of dynamic elections: accountability, commitment, and responsiveness." *Journal of Economic Literature*, 55, 916–984, 10.1257/jel.20150927. [456]

Duggan, John, and César Martinelli (2020), "Electoral accountability and responsive democracy." *Economic Journal*, 130, 675–715, 10.1093/ej/uez070. [456]

Duggan, John, and Yoji Sekiya (2009), "Voting equilibria in multi-candidate elections." *Journal of Public Economic Theory*, 11, 875–889, 10.1111/j.1467-9779.2009.01433.x. [135, 136, 148, 149]

Duverger, Maurice (1951), *Les partis politiques*. Armand Colin, Paris, 10.3917/lsrel.duver.1981.01. [360]

Duverger, Maurice (1964), *Political parties*, third edition. Methuen, London. [360]

Eguia, Jon X. (2007), "Citizen candidates under uncertainty." *Social Choice and Welfare*, 29, 317–331, 10.1007/s00355-006-0207-y. [353]

Elkind, Edith, Martin Lackner, and Dominik Peters (2022), "Preference restrictions in computational social choice: a survey.", 10.48550/arXiv.2205.09092. [Unpublished paper, arXiv:2205.09092v1] [69]

Enelow, James M., and Melvin J. Hinich (1984), *The spatial theory of voting*. Cambridge University Press, Cambridge. [68]

Eraslan, Hülya (2002), "Uniqueness of stationary equilibrium payoffs in the Baron–Ferejohn model." *Journal of Economic Theory*, 103, 11–30, 10.1006/jeth.2001.2820. [484, 490]

Evren, Özgür (2012), "Altruism and voting: a large-turnout result that does not rely on civic duty or cooperative behavior." *Journal of Economic Theory*, 147, 2124–2157, 10.1016/j.jet.2012.09.005. [121]

Farquharson, Robin (1969), *Theory of voting*. Yale University Press, New Haven. [For some corrections, see Richard G. Niemi, "An exegesis of Farquharson's *Theory of voting*", *Public Choice* , 40 (1983), 323-328.] [124, 182]

Fearon, James D. (1999), "Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance." In *Democracy, Accountability, and Representation* (Adam Przeworski, Susan C. Stokes, and Bernard Manin, eds.), chapter 2, 55–97, Cambridge University Press, 10.1017/cbo9781139175104.003. [456]

Fearon, James D. (2011), "Self-enforcing democracy." *Quarterly Journal of Economics*, 126, 1661–1708, 10.1093/qje/qjr038. [517]

Feddersen, Timothy J., and Wolfgang Pesendorfer (1996), "The swing voter's curse." *American Economic Review*, 86, 408–424. [226]

Feddersen, Timothy J., and Wolfgang Pesendorfer (1998), "Convicting the innocent: the inferiority of unanimous jury verdicts under strategic voting." *American Political Science Review*, 92, 23–35, 10.2307/2585926. [226, 227]

Feddersen, Timothy J., and Alvaro Sandroni (2005), "The calculus of ethical voting." [Unpublished paper.] [193, 195]

Feddersen, Timothy J., and Alvaro Sandroni (2006a), "The calculus of ethical voting." *International Journal of Game Theory*, 35, 1–25, 10.1007/s00182-006-0047-0. [190, 192, 195]

Feddersen, Timothy J., and Alvaro Sandroni (2006b), "A theory of participation in elections." *American Economic Review*, 96, 1271–1282, 10.1257/aer.96.4.1271. [195]

Feddersen, Timothy J., Itai Sened, and Stephen G. Wright (1990), "Rational voting and candidate entry under plurality rule." *American Journal of Political Science*, 34, 1005–1016, 10.2307/2111469. [148, 359]

Ferejohn, John A. (1986), "Incumbent performance and electoral control." *Public Choice*, 50, 5–25, 10.1007/bf00124924. [456]

Ferejohn, John A., and Morris P. Fiorina (1974), "The paradox of not voting: a decision theoretic analysis." *American Political Science Review*, 68, 525–536, 10.1017/s0003055400117368. [124]

Fishburn, Peter C. (1973), *The theory of social choice*. Princeton University Press, 10.1515/9781400868339. [33]

Fishburn, Peter C. (1974), "Social choice functions." *SIAM Review*, 16, 63–90, 10.1137/1016005. [68]

Fishburn, Peter C. (1977), "Condorcet social choice functions." *SIAM Journal on Applied Mathematics*, 33, 469–489, 10.1137/0133030. [183]

Forand, Jean Guillaume (2014), "Two-party competition with persistent policies." *Journal of Economic Theory*, 152, 64–91, 10.1016/j.jet.2014.03.014. [318]

Fudenberg, Drew, and Jean Tirole (1991), *Game theory*. MIT Press, Cambridge, Massachusetts. [541]

Gans, Joshua S., and Michael Smart (1996), "Majority voting with single-crossing preferences." *Journal of Public Economics*, 59, 219–237, 10.1016/0047-2727(95)01503-5. [68, 404]

Geanakoplos, John (2005), "Three brief proofs of Arrow's impossibility theorem." *Economic Theory*, 26, 211–215, 10.1007/s00199-004-0556-7. [68]

Gehrlein, William V., and Peter C. Fishburn (1976), "The probability of the paradox of voting: a computable solution." *Journal of Economic Theory*, 13, 14–25, 10.1016/0022-0531(76)90063-6. [21]

Gibbard, Allan (1973), "Manipulation of voting schemes: a general result." *Econometrica*, 41, 587–601, 10.2307/1914083. [93]

Goodman, Leo A. (1954), "On methods of aggregation." In *Decision processes* (Robert M. Thrall, Clyde H. Coombs, and Robert L. Davis, eds.), 39–48, Wiley, New York. [68]

Grossman, Gene M., and Elhanan Helpman (1996), "Electoral competition and special interest politics." *Review of Economic Studies*, 63, 265–286, 10.2307/2297852. [430]

Haan, Marco, and Bjørn Volkerink (2001), "A runoff system restores the principle of minimum differentiation." *European Journal of Political Economy*, 17, 157–162, 10.1016/s0176-2680(00)00033-1. [360]

Hägele, Günter, and Friedrich Pukelsheim (2001), "Llull's writings on electoral systems." *Studia Lulliana*, 41, 3–38. [68]

Hamlin, Alan, and Colin Jennings (2011), "Expressive political behaviour: foundations, scope and implications." *British Journal of Political Science*, 41, 645–670, 10.1017/s0007123411000020. [195]

Hamlin, Alan, and Colin Jennings (2019), "Expressive voting." In *Oxford Handbook of Public Choice* (Roger D. Congleton, ed.), volume 1, 333–350, Oxford University Press, 10.1093/oxfordhb/9780190469733.013.16. [196]

Hammond, Peter J. (1976), "Equity, Arrow's conditions, and Rawls' difference principle." *Econometrica*, 44, 793–804, 10.2307/1913445. [68]

Harsanyi, John C. (1977a), "Morality and the theory of rational behavior." *Social Research*, 44, 623–656. [195]

Harsanyi, John C. (1977b), "Rule utilitarianism and decision theory." *Erkenntnis*, 11, 25–53, 10.1007/BF00169843. [195]

Harsanyi, John C. (1980), "Rule utilitarianism, rights, obligations and the theory of rational behavior." *Theory and Decision*, 12, 115–133, 10.1007/BF00154357. [195]

Herrera, Helios, David K. Levine, and César Martinelli (2008), "Policy platforms, campaign spending and voter participation." *Journal of Public Economics*, 92, 501–513, 10.1016/j.jpubeco.2007.11.001. [430]

Hinich, Melvin J., John O. Ledyard, and Peter C. Ordeshook (1972), "Nonvoting and the existence of equilibrium under majority rule." *Journal of Economic Theory*, 4, 144–153, 10.1016/0022-0531(72)90145-7. [280]

Horan, Sean (2021), "Agendas in legislative decision-making." *Theoretical Economics*, 16, 235–274, 10.3982/TE2618. [182]

Horan, Sean, Martin J. Osborne, and M. Remzi Sanver (2019), "Positively responsive collective choice rules and majority rule: a generalization of May's theorem to many alternatives." *International Economic Review*, 60, 1489–1504, 10.1111/iere.12394. [67]

Hotelling, Harold (1929), "Stability in competition." *Economic Journal*, 39, 41–57, 10.2307/2224214. [246, 279]

Humphreys, Macartan, and Michael Laver (2010), "Spatial models, cognitive metrics, and majority rule equilibria." *British Journal of Political Science*, 40, 11–30, 10.1017/S0007123409990263. [68]

Imai, Haruo, and Hannu Salonen (2012), "Bargaining and rent seeking." Discussion Paper 80, Aboa Centre for Economics, Turku, Finland. [485]

Indriðason, Indriði H. (2011), "Proportional representation, majoritarian legislatures, and coalitional voting." *American Journal of Political Science*, 55, 954–970, 10.1111/j.1540-5907.2011.00530.x. [149]

Itsumi, Yoshitaka (1974), "Distributional effects of linear income tax schedules." *Review of Economic Studies*, 41, 371–381, 10.2307/2296756. [404]

Kartik, Navin, Richard Van Weelden, and Stephane Wolton (2017), "Electoral ambiguity and political representation." *American Journal of Political Science*, 61, 958–970, 10.1111/ajps.12310. [307, 318]

Kleiner, Andreas, and Benny Moldovanu (2017), "Content-based agendas and qualified majorities in sequential voting." *American Economic Review*, 107, 1477–1506, 10.1257/aer.20160277. [182, 183]

Krasa, Stefan, and Mattias K. Polborn (2009), "Is mandatory voting better than voluntary voting?" *Games and Economic Behavior*, 66, 275–291, 10.1016/j.geb.2008.05.004. [124]

Laffond, Gilbert, and Jean-François Laslier (1991), "Slater's winners of a tournament may not be in the Banks set." *Social Choice and Welfare*, 8, 365–369, 10.1007/bf00183047. [178]

Ledyard, John O. (1981), "The paradox of voting and candidate competition: a general equilibrium analysis." In *Essays in contemporary fields of economics* (George Horwich and James P. Quirk, eds.), 54–80, Purdue University Press, West Lafayette, Indiana. [124]

Ledyard, John O. (1984), "The pure theory of large two-candidate elections." *Public Choice*, 44, 7–41, 10.1007/bf00124816. [264, 279]

Lindbeck, Assar, and Jörgen W. Weibull (1987), "Balanced-budget redistribution as the outcome of political competition." *Public Choice*, 52, 273–297, 10.1007/bf00116710. [404]

Loomes, Graham, and Robert Sugden (1982), "Regret theory: an alternative theory of rational choice under uncertainty." *Economic Journal*, 92, 805–824, 10.2307/2232669. [124]

Martinelli, César, and Akihiko Matsui (2002), "Policy reversals and electoral competition with privately informed parties." *Journal of Public Economic Theory*, 4, 39–61, 10.1111/1467-9779.00087. [318]

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green (1995), *Microeconomic theory*. Oxford University Press, New York. [526, 528]

Maskin, Eric S., and Jean Tirole (2004), "The politician and the judge: accountability in government." *American Economic Review*, 94, 1034–1054, 10.1257/0002828042002606. [456]

May, Kenneth O. (1952), "A set of independent necessary and sufficient conditions for simple majority decision." *Econometrica*, 20, 680–684, 10.2307/1907651. [67]

McKelvey, Richard D., and Richard G. Niemi (1978), "A multistage game representation of sophisticated voting for binary procedures." *Journal of Economic Theory*, 18, 1–22, 10.1016/0022-0531(78)90039-x. [182]

McKelvey, Richard D., and Richard E. Wendell (1976), "Voting equilibria in multidimensional choice spaces." *Mathematics of Operations Research*, 1, 144–158, 10.1287/moor.1.2.144. [68]

McMurray, Joseph C. (2013), “Aggregating information by voting: the wisdom of the experts versus the wisdom of the masses.” *Review of Economic Studies*, 80, 277–312, 10.1093/restud/rds026. [227]

Meirowitz, Adam, and Kenneth W. Shotts (2009), “Pivots versus signals in elections.” *Journal of Economic Theory*, 144, 744–771, 10.1016/j.jet.2008.08.008. [280]

Miller, Nicholas R. (1977), “Graph-theoretical approaches to the theory of voting.” *American Journal of Political Science*, 21, 769–803, 10.2307/2110736. [183]

Mirrlees, James A. (1971), “An exploration in the theory of optimum income taxation.” *Review of Economic Studies*, 38, 175–208, 10.2307/2296779. [384]

Moulin, Hervé (1980), “On strategy-proofness and single peakedness.” *Public Choice*, 35, 437–455, 10.1007/bf00128122. [93]

Moulin, Hervé (1986), “Choosing from a tournament.” *Social Choice and Welfare*, 3, 271–291, 10.1007/bf00292732. [183]

Moulin, Hervé (1988), *Axioms of cooperative decision making*. Cambridge University Press, Cambridge, 10.1017/ccol0521360552. [ix, 68, 93, 403]

Mueller, Dennis C. (2003), *Public choice III*. Cambridge University Press, Cambridge, 10.1017/cbo9780511813771. [ix]

Myatt, David P. (2007), “On the theory of strategic voting.” *Review of Economic Studies*, 74, 255–281, 10.1111/j.1467-937x.2007.00421.x. [148]

Myatt, David P. (2015), “A theory of voter turnout.” [Unpublished paper] [124]

Nash, John F. (1950), “The bargaining problem.” *Econometrica*, 18, 155–162, 10.2307/1907266. [24, 68]

Osborne, Martin J. (1993), “Candidate positioning and entry in a political competition.” *Games and Economic Behavior*, 5, 133–151, 10.1006/game.1993.1007. [359]

Osborne, Martin J. (2004), *An introduction to game theory*. Oxford University Press, New York. [254, 261, 523, 533, 535, 538]

Osborne, Martin J., Jeffrey S. Rosenthal, and Matthew A. Turner (2000), “Meetings with costly participation.” *American Economic Review*, 90, 927–943, 10.1257/aer.90.4.927. [93]

Osborne, Martin J., and Ariel Rubinstein (1994), *A course in game theory*. MIT Press, Cambridge, Massachusetts. [93, 226, 403, 464, 523, 530, 531, 533, 536, 537, 540, 543, 548, 549]

Osborne, Martin J., and Ariel Rubinstein (2003), "Sampling equilibrium, with an application to strategic voting." *Games and Economic Behavior*, 45, 434–441, 10.1016/s0899-8256(03)00147-7. [149]

Osborne, Martin J., and Ariel Rubinstein (2023), *Models in microeconomic theory*, second edition. Open Book Publishers, Cambridge, 10.11647/OBP.0361. [68, 93, 523, 528, 538, 540]

Osborne, Martin J., and Al Slivinski (1996), "A model of political competition with citizen-candidates." *Quarterly Journal of Economics*, 111, 65–96, 10.2307/2946658. [360]

Osborne, Martin J., and Rabee Tourky (2008), "Party formation in single-issue politics." *Journal of the European Economic Association*, 6, 974–1005, 10.1162/jeea.2008.6.5.974. [93]

Osborne, Martin J., and Matthew A. Turner (2010), "Cost benefit analyses versus referenda." *Journal of Political Economy*, 118, 156–187, 10.1086/650305. [226, 227]

Palfrey, Thomas R. (1989), "A mathematical proof of Duverger's law." In *Models of strategic choice in politics* (Peter C. Ordeshook, ed.), 69–91, University of Michigan Press, Ann Arbor. [145, 148]

Palfrey, Thomas R., and Howard Rosenthal (1983), "A strategic calculus of voting." *Public Choice*, 41, 7–53, 10.1007/bf00124048. [124]

Palfrey, Thomas R., and Howard Rosenthal (1985), "Voter participation and strategic uncertainty." *American Political Science Review*, 79, 62–78, 10.2307/1956119. [124]

Prat, Andrea (2006), "Rational votes and political advertising." In *The Oxford Handbook of Political Economy* (Barry R. Weingast and Donald A. Wittman, eds.), chapter 3, 50–63, Oxford University Press, Oxford, 10.1093/oxfordhb/9780199548477.003.0003. [430]

Puppe, Clemens (2018), "The single-peaked domain revisited: a simple global characterization." *Journal of Economic Theory*, 176, 55–80, 10.1016/j.jet.2018.03.003. [69]

Rae, Douglas, and Michael Taylor (1971), "Decision rules and policy outcomes." *British Journal of Political Science*, 1, 71–90, 10.1017/s0007123400008954. [68]

Rasch, Bjørn Erik (2000), "Parliamentary floor voting procedures and agenda setting in Europe." *Legislative Studies Quarterly*, 25, 3–23. [182]

Riker, William H. (1982), "The two-party system and Duverger's Law: an essay on the history of political science." *American Political Science Review*, 76, 753–766, 10.2307/1962968. [360]

Roberts, Kevin W. S. (1977), "Voting over income tax schedules." *Journal of Public Economics*, 8, 329–340, 10.1016/0047-2727(77)90005-6. [68, 404]

Rockafellar, R. Tyrrell (1970), *Convex analysis*. Princeton University Press, 10.1515/9781400873173. [550]

Röell, Ailsa A. (2012), "Voting over nonlinear income tax schedules." [Unpublished paper (earlier version 1996)] [404]

Roemer, John E. (1994), "A theory of policy differentiation in single issue electoral politics." *Social Choice and Welfare*, 11, 355–380, 10.1007/bf00183303. [318]

Roemer, John E. (1997), "Political-economic equilibrium when parties represent constituents: the unidimensional case." *Social Choice and Welfare*, 14, 479–502, 10.1007/s003550050082. [318]

Romer, Thomas (1975), "Individual welfare, majority voting, and the properties of a linear income tax." *Journal of Public Economics*, 4, 163–185, 10.1016/0047-2727(75)90016-x. [404]

Rosenthal, Jeffrey S. (2020), "Maximum binomial probabilities and game theory voter models." *Advances and Applications in Statistics*, 64, 75–85, 10.17654/as064010075. [117]

Rothstein, Paul (1990), "Order restricted preferences and majority rule." *Social Choice and Welfare*, 7, 331–342, 10.1007/bf01376281. [68, 69, 404]

Rothstein, Paul (1991), "Representative voter theorems." *Public Choice*, 72, 193–212, 10.1007/bf00123744. [404]

Royden, Halsey L. (1968), *Real analysis*, second edition. Macmillan, New York. [528]

Rubinstein, Ariel (1991), "Comments on the interpretation of game theory." *Econometrica*, 59, 909–924, 10.2307/2938166. [478]

Saporiti, Alejandro (2009), "Strategy-proofness and single-crossing." *Theoretical Economics*, 4, 127–163, https://econtheory.org/ojs/index.php/te/article/view/20090127. [93]

Satterthwaite, Mark A. (1975), "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions." *Journal of Economic Theory*, 10, 187–217, 10.1016/0022-0531(75)90050-2. [93]

Schuessler, Alexander A. (2000a), "Expressive voting." *Rationality and Society*, 12, 87–119, 10.1177/104346300012001005. [194, 195]

Schuessler, Alexander A. (2000b), *A logic of expressive choice*. Princeton University Press, 10.1515/9780691222417. [195]

Schultz, Christian (1996), "Polarization and inefficient policies." *Review of Economic Studies*, 63, 331–344, 10.2307/2297855. [318]

Schwartz, Thomas (2008), "Parliamentary procedure: principal forms and political effects." *Public Choice*, 136, 353–377, 10.1007/s11127-008-9300-3. [182]

Sen, Amartya K. (1970), *Collective choice and social welfare*. Holden-Day, San Francisco. [68]

Shepsle, Kenneth A., and Barry R. Weingast (1984), "Uncovered sets and sophisticated voting outcomes with implications for agenda institutions." *American Journal of Political Science*, 28, 49–74, 10.2307/2110787. [183]

Smart, D. R. (1974), *Fixed point theorems*. Cambridge University Press. [529]

Sydsæter, Knut (1981), *Topics in mathematical analysis for economists*. Academic Press, London. [550, 551]

Taylor, Curtis R., and Huseyin Yildirim (2010), "A unified analysis of rational voting with private values and group-specific costs." *Games and Economic Behavior*, 70, 457–471, 10.1016/j.geb.2010.02.008. [124]

von Neumann, John, and Oskar Morgenstern (1944), *Theory of games and economic behavior*, first edition. Princeton University Press, Princeton. [527]

Ward, Benjamin (1961), "Majority rule and allocation." *Journal of Conflict Resolution*, 5, 379–389, 10.1177/002200276100500405. [182]

Wendell, Richard E., and Stuart J. Thorson (1974), "Generalizations of social decisions under majority rule." *Econometrica*, 42, 893–912. [44, 68]

Wittman, Donald (1977), "Candidates with policy preferences: a dynamic model." *Journal of Economic Theory*, 14, 180–189, 10.1016/0022-0531(77)90091-6. [318]

Wittman, Donald (1983), "Candidate motivation: a synthesis of alternative theories." *American Political Science Review*, 77, 142–157, 10.2307/1956016. [318]