

STRATEGIC EQUILIBRIUM IN SOCIAL NETWORKS AND GAMES: THEORY  
AND APPLICATIONS

By

Valeska Groenert

Dissertation

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

August, 2009

Nashville, Tennessee

Approved:

Myrna Wooders (chair)

Andrew Daughety

Paul Edelman

Quan Wen

## ACKNOWLEDGEMENTS

I am truly indebted to my advisor Myrna Wooders who has guided me in all aspects of academic life and always been exceptionally supportive of my research. For their valuable input to this dissertation, I am also very grateful to Professors Rabah Amir, Andrew Daughety, Paul Edelman, Jennifer Reinganum, Quan, Wen, and Ben Zissimos.

The financial support of the Kirk Dornbush summer research grant and of the summer research grant of the College of Arts and Science at Vanderbilt University is gratefully acknowledged.

During my graduate studies, and especially while working on this dissertation, I have benefited from interactions with graduate students, faculty members, and staff at the Departments of Economics at Vanderbilt University and at other places. I would like to express my gratitude to each of them. In particular, I am very grateful to Rosa and Ali Sina for their precious input and for being true friends. I would also like to thank Alper, Bob, Demet, Jean-Francois, Jisong, Julio, Katie, Linda, Mostafa, Naoto, Rohit, Sangsoo, Scott, Tamara, Tommaso, Willemien, and Yariv. For their help at various stages of my graduate studies I thank Bill Collins, Kathleen Finn, Greg Huffman, and John Weymark.

Finally, I am indebted to my family and friends in Germany and elsewhere for their continuous encouragement and love.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	ii
LIST OF FIGURES .....	v
LIST OF TABLES .....	vi
Chapter	
I. INTRODUCTION.....	1
II. TRIMMED STRATEGIES: ACHIEVING SEQUENTIALLY RATIONAL EQUILIBRIA WITH ONLY PARTIALLY SPECIFIED STRATEGIES .....	7
Introduction.....	7
Preliminaries.....	12
Extensive form games .....	12
Behavioral Strategies.....	13
Notation concerning the structure of a game tree .....	14
Trimmed SPE.....	14
An Example.....	22
Trimmed WPBE .....	24
Conclusion .....	31
III. A CHARACTERIZATION OF WEAKLY PAIRWISE STABLE NETWORKS.....	33
Introduction.....	33
The Linking Game with Player Pairs.....	35
The linking game with player pairs.....	41
An Example.....	44
The Two-Stage Linking Game .....	50
Conclusion .....	54
IV. COMPETITION OVER STANDARDS AND TAXES.....	56
Introduction.....	56
The Model .....	60
Firms .....	60
Governments .....	63
Efficiency .....	65
Competition over Standards and Taxes .....	67
The developing country's best response function.....	72
The developed country's best response function.....	80
Conclusion .....	84
Appendix .....	87
Indifference Set.....	87
Proofs .....	88
V. LOBBYING IN A NETWORK.....	105
Introduction.....	105
The Model .....	110
Lobbying in a Threshold Network.....	113

The main result .....	115
Bringing about unanimous decisions .....	121
Opinion Formation in General Networks .....	125
Conclusion .....	130
Appendix .....	131
A behavioral model of voters .....	131
Proofs .....	132
VI. REFERENCES.....	139

## LIST OF FIGURES

Figure		Page
1.	Nonexistence of SPE due to nonexistence of NE in some subgames.....	9
2.	Information sets off the unilateral deviation path .....	17
3.	A trimmed SPE that is not a SPE .....	20
4.	A trimmed SPE can exist when a SPE does not .....	24
5.	Information sets off the believed unilateral deviation path .....	27
6.	A trimmed WPBE that is not a WPBE.....	29
7.	A supernetwork .....	38
8.	A cycle in a supernetwork .....	40
9.	Four networks.....	48
10.	Network $g^3$ does not belong to a singleton basin of attraction .....	49
11.	Firms' location decision .....	62
12.	Strategy Sets for Government F .....	64
13.	Efficient allocation for $k > 1$ .....	66
14.	Equilibrium standard levels depending on $k$ .....	70
15.	Sharing tax limit for $k < 1$ .....	75
16.	Undercutting dominates for small $k$ .....	78
17.	Sharing tax limit for $k > 1$ .....	80
18.	Government L's payoff function .....	81
19.	A threshold network .....	114
20.	Relabeling of nodes in a threshold network.....	119
21.	Threshold network of Example 2.....	120
22.	A stable state might not exist .....	127
23.	A network with asymmetric influences .....	129

## LIST OF TABLES

Table		Page
1	Combinations of $s_i, s_j,$ and $s_{ij}$ .....	42
2	Nine cases .....	138

## CHAPTER I

### INTRODUCTION

This dissertation comprises four chapters on networks and games. The first chapter, "Trimmed Strategies: Achieving Sequentially Rational Equilibria With Only Partially Specified Strategies," examines notions of sequential rationality underlying two popular equilibrium refinements: subgame perfect equilibrium and weak perfect Bayesian equilibrium. I start my analysis by showing that each strategy profile, together with a notion of sequential rationality, induces a partition of a game into two parts: a "relevant" part and an "irrelevant part." It is shown that whether play on the relevant part is sequentially rational (according to whichever notion of sequential rationality is considered) is independent from the play on the irrelevant part. The trimmed version of an equilibrium concept is then defined as a profile that satisfies the equilibrium concept's conditions on the relevant part that the profile induces. The independence of sequentially rational behavior in the relevant part from play in the irrelevant part ensures that a trimmed equilibrium is never sustained by sequentially irrational play.

There are games for which a subgame perfect equilibrium does not exist because Nash equilibrium does not exist in some of their subgames. The concept of trimmed equilibrium is of interest for the analysis of such games. It demonstrates that even if a certain equilibrium does not exist in a game, one might nonetheless be able to identify an outcome that is consistent with the notion of rationality underlying that equilibrium concept. Another motivation for this chapter is to establish robustness properties of equilibria. Such robustness properties have recently received much interest in the literature (see for example Kalai 2004 and the references therein). The irrelevant part of an equilibrium profile is by definition such that the equilibrium is robust to any changes made in that part.

The second chapter, "A Characterization of Weakly Pairwise Nash Stable Networks," studies various aspects of pairwise Nash stability, a widely used concept in the network literature. First, I characterize weakly pairwise Nash stable networks as the Nash equilibrium networks of a well-defined game in strategic form. Second, I define a two-stage game of network formation that models how cooperation can arise as an outcome of noncooperative play.

The starting point for the characterization of weakly pairwise Nash stable networks is Myerson's linking game of network formation (Myerson 1991), which has received much attention in the literature (see Jackson 2001 for a survey). In Myerson's game, each player announces a list of players, and links are formed if and only if two players announce each other. A network is weakly pairwise Nash stable if it is supported by a Nash equilibrium of Myerson's linking game and there are no two players that are not linked for whom it would be mutual beneficial to form a link. I extend Myerson's linking game by adding pairs of players to it and I call the game the *linking game with player pairs*. The rules of the game and the payoffs for player pairs are designed so that the set of its Nash equilibrium networks is equivalent to the set of weakly pairwise Nash stable networks.

A benefit of the linking game with player pairs is that it allows to define mixed strategies on the coordinated moves of pairs of players. The analysis of the mixed extension of the linking game with player pairs is of interest because weakly pairwise Nash stable networks need not exist.<sup>1</sup> Examining these mixed strategy equilibria, I find that networks that receive positive probability in equilibrium (a) might not be part of any absorbing state of a process of dynamic network formation, and (b) might not be supported by a Nash equilibrium of the linking game. Thus, the set of networks supported by a mixed strategy equilibrium of the linking game with player pairs is larger than the set of Nash equilibrium networks.

---

<sup>1</sup>This approach might be applicable to other games with cooperative elements. A solution often does not exist for such games, but mixed strategies cannot be defined in any obvious way.



To provide intuition for the presence of pairs of players, the next part of the chapter defines and analyzes a game in which network formation occurs in two stages. In the first stage, noncooperatively acting players can form links with others that allow them to coordinate (at least to some degree) their moves in the second stage. In the second stage, a constrained version of the linking game with player pairs is played. In the constrained version, only pairs who formed links in the first stage can coordinate on their links. I argue that the two-stage game displays natural features of real world network formation. The game is used to show that, even if two players benefit from forming a link, a network might be stable, simply because the two players have not formed a "first stage link," which can be interpreted as not knowing much about each other. In this vein, it is shown that the set of subgame perfect equilibrium networks of the two-stage linking game is larger than the set of weakly pairwise Nash stable networks. However, refining the set of equilibria by requiring undominated strategies ensures that the corresponding outcomes of the two-stage game are weakly pairwise Nash stable networks.

The third and fourth chapters turn to more applied questions of strategic equilibrium in networks and games.

The third chapter, "Competition Over Standards and Taxes," a joint work with Myrna Wooders and Ben Zissimos, develops a model of interjurisdictional competition, in which governments choose standards and taxes to attract mobile firms. In a setting in which firms have varying requirements for standards, we show that a country that sets its standards and tax levels first provides higher standard levels and sets higher taxes.

In our model, countries use revenue from taxation to enforce standards, such as property rights, environmental standards, and child labor regulations. The standard level in a firm's chosen jurisdiction inflicts a nonnegative cost on the firm. That cost could be directly or indirectly incurred by the firm. For example, the firm might have

to install a filter to comply with an environmental standard, but there could also be an indirect cost incurred by the firm through negative externalities from other firms. The cost is heterogeneous across firms and non-monotonic in the level of standard: Each firm has a unique level of standard at which its cost is minimized. This contrasts with public good models where, all else equal, firms unanimously prefer higher levels of public goods. Jurisdictions move sequentially rather than simultaneously. We have in mind the case of a well developed country and a less developed country or an emerging economy. In equilibrium the first-moving government sets a higher standard level and charges a higher tax than the second-moving government. The majority of firms locates in the jurisdiction of the second-moving government.

On the technical side, our model poses difficult problems for existence of subgame perfect equilibrium in pure strategies since some subgames of the game may not have an equilibrium in pure strategies. This difficulty is resolved by showing that these subgames are never reached because they are in some sense dominated by other subgames. The argument uses concepts that are in spirit similar to the ones introduced in Chapter 1.

The fourth chapter, "Lobbying in Networks," attempts to shed light on diffusion of behavior and on optimal strategies to promote or prevent diffusion in a population. In particular, the chapter examines how an interested agent would go about influencing a group of individuals, exploiting their group structure. Examples can be found in politics (lobbying, election campaigns), marketing (word-to-mouth marketing, viral marketing), and in numerous other situations. The chapter focuses on group decisions.<sup>2</sup> Voting decisions are at least partially influenced by interactions with others, and we frequently observe that decided individuals (individuals with clear-cut interests) try to influence the vote of others. So far, this observation has received little attention in the literature.<sup>3</sup> Here, I model a lobbyist who tries to influence the

---

<sup>2</sup>But the model is general enough to fit other applications.

<sup>3</sup>Even though group decisions are pervasive throughout most societies (group decisions are made

decision of a group.

In the model, a group votes on a proposal and the lobbyist tries to persuade the group to vote for the proposal. Interactions between individuals are represented by a network (but in contrast to Chapter 2, here the network is taken as given). The main idea, assuming individuals are influenced by their interactions, is that the lobbyist can use information about group structure to engineer "persuading cascades." The general goal of the chapter is to gain insight into diffusion. The more specific goal is to identify key sets of influential group members and to show how the results of lobbying depend on the network structure.

Finding the key sets of influential players in arbitrary networks is a difficult problem (see Ballester et al. 2009 for a discussion). So far the literature has approached the problem by obtaining algorithms that approximate the solution to the problem or by addressing related questions instead. Here, I take a different approach. By restricting the class of networks in the first part of the chapter, I am able to provide an exact solution to the problem: In networks with a core-periphery structure, the lobbyist's optimal strategy is to target those group members that have high degrees and oppose the proposal. I also compute bounds on the number of voters that have to be convinced to eventually reach unanimous support and show that more tightly connected groups are harder to convince.

The second part of the chapter examines the process of opinion formation in arbitrary networks. Using a result from Goles and Olivos (1980), I show that the process becomes periodic after a finite number of periods, independent of the initial opinions held in the network. Once the periodic state is reached, voters either do not change their opinions anymore or switch back and forth between two different opinions. The chapter concludes with a discussion of how diffusion in a network can

---

by congress, committees, corporations, families, juries, etc.), most of the literature has focused on the case where one individual tries to persuade another individual. As Caillaud and Tirole (2007) write "surprisingly little has been written on group persuasion."

be interpreted in terms of hierarchies of groups within the network with some groups being opinion leaders while others are followers.

## CHAPTER II

# TRIMMED STRATEGIES: ACHIEVING SEQUENTIALLY RATIONAL EQUILIBRIA WITH ONLY PARTIALLY SPECIFIED STRATEGIES

### Introduction

Many refinements of Nash equilibrium (NE) incorporate some notion of sequential rationality. Such refinements serve to rule out outcomes sustained by irrational play off the equilibrium path, in particular by noncredible threats. For example, a subgame-perfect equilibrium (SPE) incorporates sequential rationality by requiring NE in each subgame, and a weak perfect Bayesian equilibrium (WPBE) incorporates sequential rationality by requiring best responses at each information set given the belief at this information set, and given future play. This paper establishes that both SPE and WPBE can be more restrictive than necessary to ensure outcomes that are consistent with their particular notions of sequential rationality. More precisely, given a strategy profile (and a system of beliefs), this paper characterizes the maximal collection of information sets off the equilibrium path such that choices within this set never affect the rationality as imposed by a SPE (a WPBE) outside of the set.<sup>4</sup> This collection will be called the maximal collection of SPE-irrelevant (WPBE-irrelevant) information sets.

Let  $V$  be a collection of information sets and let  $b$  be a profile of behavioral strategies. If  $V$  is the maximal collection of SPE-irrelevant (WPBE-irrelevant) information sets, then the requirement of sequentially rational play in  $V$  can be dropped while making sure that the outcome of  $b$  is never sustained by sequentially irrational play on  $V$ . A strategy profile  $b$  that satisfies sequential rationality as imposed by SPE (WPBE) outside the maximal collection of irrelevant sets will be called a trimmed

---

<sup>4</sup>In principle the characterization of inessential game parts could be done for other equilibrium concepts as well, but is not examined here further.

SPE (trimmed WPBE). This terminology is chosen to emphasize that strategies in a trimmed equilibrium can be viewed as smaller strategies, because the play at irrelevant sets can essentially be left unspecified.<sup>5</sup> It will be shown that a trimmed SPE (WPBE) can not be relaxed further. Its conditions are necessary and sufficient to ensure that outcomes are not sustained by their respective notions of irrational play.

There are several motivations for the study of maximal collections of irrelevant information sets and for the concept of trimmed equilibrium. First, there are games in which a SPE does not exist, because on a subset of their subgames NE does not exist. In such cases, knowing which parts of a game are irrelevant for a SPE can be helpful. As Section 4 demonstrates via an example, the results of this paper can help find an outcome which is, nonetheless, consistent with sequential rationality.<sup>6</sup> For a first illustration of this example (more details are provided in Section 4), consider two profit-maximizing firms,  $i$  and  $j$ . Both firms produce the same homogenous good and compete against each other in a Bertrand-Edgeworth world. In stage one, they simultaneously build capacities  $x_i$  and  $x_j$  at a cost of  $K$  dollars per unit. In stage two, they simultaneously announce prices  $p_i$  and  $p_j$ , and demand is realized.<sup>7</sup> Suppose only pure strategies are available to the firms and that we are interested in the SPE of this game. Each capacity pair chosen in the first stage induces a pricing subgame in the second stage. As Kreps and Scheinkman (1983) show, pure strategy equilibria exist only for a subset of these subgames, as illustrated in Figure 1. There, the functions  $r_i(x_j)$  and  $r_j(x_i)$  are the best response functions derived from a Cournot game in which two firms simultaneously choose quantities  $x_i$  and  $x_j$  and each firm has a unit

---

<sup>5</sup>Another interpretation goes as follows. A maximal collection of irrelevant information sets for some profile  $b$  and some notion of sequential rationality identifies an equivalence class of strategy profiles. The class consists of all profiles that coincide with  $b$  on the relevant part of the game and induce arbitrary play on the part of the game that is irrelevant for  $b$ .

<sup>6</sup>Theorems 7 and 14 and their corollaries show that the trimmed version of an equilibrium, even though it is less restrictive, does not give up on the degree of rationality imposed: Provided the original equilibrium exists, the sets of outcomes for the original concept and for its trimmed version are the same.

<sup>7</sup>Demand is determined by a rationing rule.

cost of 0 dollars per unit. Pure strategy pricing equilibria do not exist in subgames where either firm  $i$  has chosen a capacity above  $r_i(x_j)$  or firm  $j$  has chosen a capacity above  $r_j(x_i)$ . Since a SPE requires NE play in all subgames, a pure strategy SPE does not exist. In contrast, the definition of a trimmed SPE does not require Nash equilibrium play in every subgame and as a consequence a pure strategy trimmed SPE might exist. Section 4 demonstrates the application of a trimmed SPE in this particular example.

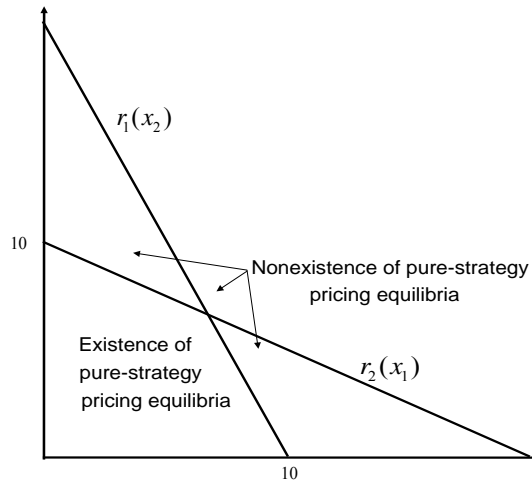


Figure 1: Nonexistence of SPE due to nonexistence of NE in some subgames

Second, determining the maximal collection of irrelevant information sets of an equilibrium strategy profile, also determines the parts of the game whose specifications, be they players, payoffs, order of moves, or any other specification, are not relevant for the equilibrium. In other words, the equilibrium is robust to any kind of change in this part of the game. Such a robustness result is of interest in situations where some specifications are uncertain in a way that does not allow for a probability assessment (see Kalai 2004). For example, players at an initial stage of the game might not know how many other players will enter the game at later stages.

Third, the results in this paper call into question the standard game-theoretic assertion that equilibrium play is the outcome of complete contingent plans made by rational agents. As is shown, players can reach an equilibrium that is consistent with sequentially rational play, via trimmed strategies. Indeed, one would expect individuals to form a complete contingent plan only in simple strategic situations. In more complex situations, it seems unlikely that individuals actually form such plans.<sup>8</sup> Moreover, in a trimmed equilibrium players might form incorrect expectations about some of the other players' future moves. Neither are these incorrect expectations ever uncovered when players follow their trimmed equilibrium strategies, nor do they sustain a player's strategy choice.<sup>9</sup> By showing that the play at some information sets is irrelevant, this paper provides a theoretical rationale of why players might not form complete contingent plans or might not revise incorrect expectations.

While there is no directly related prior literature, a somewhat related strand of the existing literature examines robustness of equilibrium towards a game's specifications.<sup>10</sup> Since strategic situations are often not completely specified, it is of interest to know which properties of an equilibrium guarantee its robustness towards the game's specifications. One part of the literature deals with robust equilibria in large games, where the number of players is uncertain (see for example Kalai 2004, 2005). Games

---

<sup>8</sup>It lies in the nature of game-theoretic experiments that the experimenter only observes the participants' actual play, but not their complete strategies. Selten et al. (1997) describe a method, developed by Selten in previous work, which allows experimenters to observe complete strategies. However, this method "forces" participants to form complete contingent plans. It is not clear whether participants in experiments actually make complete contingent plans. To my knowledge, an experimental study of this issue has not been conducted.

<sup>9</sup>A similar idea lies behind the concept of a self-confirming equilibrium (Fudenberg and Levine 1993). However, a self-confirming equilibrium need not even be a Nash equilibrium.

<sup>10</sup>A paper that seems related at first sight is Kalai and Neme (1992). They introduce the concept of a  $p$ -subgame perfect equilibrium, which requires subgame perfection after histories with no more than  $p$  deviations from the equilibrium path. However,  $p$ -subgame perfect equilibrium rationalizes behavior beyond SPE play. After a certain number of deviations players do not expect rational behavior by other players. Accordingly, a  $p$ -subgame perfect equilibrium can result into outcomes not sustainable by a SPE. This difference to a trimmed SPE also shows that the maximal collection of irrelevant information sets cannot be characterized by simply looking at numbers of deviations. In particular, the maximal collection of *SPE*-irrelevant information sets for a profile  $b$  is *not* the set of information sets that cannot be reached by a unilateral deviation.



with uncertain features are usually called "partially specified." Instead of identifying equilibrium properties that guarantee robustness of equilibrium to changes in certain specifics of a game, this paper characterizes entire parts of a game, the specifics of which do not affect the equilibrium.

Recently, and independently from my work, Briata et al. (2007) have addressed similar questions. They identify what they call the "essential collection" of information sets. However, our concepts do not coincide. In particular the complement of an essential collection is not equivalent to a the maximal collection of irrelevant information sets.<sup>11</sup> Also, while they focus on providing a general and unified framework for what they call essentializing equilibria, I focus on two widely used equilibrium concepts, for which irrelevant sets can be particularly large.

The reader familiar with the repeated game literature might be aware that this literature already uses the idea of trimmed strategies in the context of SPE. However, to my knowledge, the concept has not been formalized, nor has it been generalized to a wider class of games (the task of characterizing maximal collections of irrelevant information sets for general games is not trivial). In addition, this paper introduces a way of trimming strategies for belief-based concepts.

In summary, this paper (1) characterizes maximal collections of irrelevant information sets for both SPE and WPBE, (2) defines trimmed versions of SPE and WPBE and shows that their outcomes are never sustained by their respective notions of sequentially irrational play, and (3) demonstrates that a trimmed equilibrium can exist, even though the original equilibrium does not exist.

The rest of the paper is organized as follows. Section 2 introduces basic notations. Section 3 deals with the maximal collections of irrelevant information sets for SPE

---

<sup>11</sup>Basically, this is because they require essential collections to be closed under  $\leq$  (roughly meaning that if an information set is in the essential collection, so are all its predecessors). Another difference originates in their definition of essential collections for belief-based concepts. An information set belongs to the essential collection if it is relevant under *some* belief, while here irrelevant sets depend on a specific belief.

and the trimmed SPE. Section 4 provides a detailed example. Section 5 parallels Section 3, but applies to WPBE. Section 6 concludes.

## Preliminaries

This section introduces basic notations. Since the information structure of a game plays an important role, I analyze extensive form games.<sup>12</sup> The class of extensive form games examined satisfies perfect recall and complete information. In addition, while the game might be infinite, each player has a finite number of choices whenever the player moves. At least the main results concerning SPE can be extended to games of incomplete information and with an infinite number of choices. I restrict the class of games mainly to avoid lading the exposition with more notation. All terminology not introduced explicitly (for example, a path, a rooted tree etc.) is used in the standard game-theoretic or graph-theoretic sense.<sup>13</sup>

### Extensive form games

An *extensive form* is a tuple  $\Gamma = (T, \mathcal{P}, \mathcal{W}, \mathcal{C})$ , where

1.  $T = (X, E)$  is a rooted tree with  $X$  being a countable set of vertices and  $E$  being a set of (unordered) pairs from  $X$ . The origin (root) of the tree is denoted by  $x_0$ . For every vertex  $x$ , the sets of its immediate predecessors and its immediate successors are denoted by  $s(x)$  and  $p(x)$ .<sup>14</sup> The (possibly empty) set of terminal nodes is the set  $Z = \{x \in X : s(x) = \emptyset\}$ .
2.  $\mathcal{P} = (P_1, \dots, P_n)$  is a partition of the set  $X \setminus Z$  into  $n$  sets, one for each player  $i \in N = \{1, \dots, n\}$ .

---

<sup>12</sup>An interesting topic for future reseach would be to examine what are the analogs of collections of irrelevant information sets and trimmed equilibrium for normal form games. The remark in Footnote 5 points in that direction.

<sup>13</sup>The description of an extensive form game follows Selten (1975) and van Damme (1981).

<sup>14</sup>The order is naturally given by the distance to the origin.

3.  $\mathcal{W} = (W_1, \dots, W_n)$  is an information partition, where  $W_i$  is a partition of  $P_i$  into information sets of player  $i$  so that
- (a) every path from the origin intersects the information set at most once, and
  - (b) nodes in the same information sets have the same number of immediate successors.

Let  $W = \cup W_i$ .

4.  $\mathcal{C} = \{\mathcal{C}_w\}_{w \in W}$  is a collection of partitions. Each partition corresponds to choices at  $w$ . The partition  $\mathcal{C}_w$  divides nodes in  $\cup_{x \in w} s(x)$  into the finite number of choices available at information set  $w$ , so that every choice contains exactly one element of  $s(x)$  for every  $x \in w$ . A generic choice at  $w$  (a member of the partition  $\mathcal{C}_w$ ) is denoted by  $c_w$  and the set of choices at  $w$  is denoted by  $C_w$ .

An *extensive form game* is a pair  $\Sigma = (\Gamma, u)$ , where  $u = (u_1, \dots, u_n)$  are  $n$  real-valued von Neumann-Morgenstern expected utility functions, one for each player  $i \in N$ . The domain of each  $u_i$  is the set of probability distributions over terminal histories (see below for the definition of terminal histories).

### Behavioral strategies

A behavioral strategy for player  $i$  is the collection  $b_i = (b_i(w))_{w \in W_i}$  such that  $b_i(w)(c_w)$  denotes the probability  $b_i$  attaches to choice  $c_w$  at  $w$ . If mixed strategies are not available at  $w$ , the mapping  $b_i(w)$  assigns either probabilities zero or one. A profile of strategies is denoted by  $b$  and a system of beliefs is denoted by  $\mu$ . Let  $\mu(w)$  denote the probability distribution over nodes in  $w$  induced by the system of beliefs  $\mu$  and let  $\mu(x)$  denote the belief induced by  $\mu$  that node  $x$  is reached given that the information set to which  $x$  belongs is reached. A subform of  $\Gamma$  is denoted by  $\gamma$ , and  $G$  is the set of subforms of  $\Gamma$ .<sup>15</sup> A subgame of  $\Sigma = (\Gamma, u)$  is denoted by  $(\gamma, u|_\gamma)$  where

---

<sup>15</sup>A subform of  $\Gamma$  is the analog to a subgame of  $\Sigma$ , that is it is an extensive form that can be obtained by restriction of  $\Gamma$  on a subset  $X'$  of  $X$ .

$u|_\gamma$  is the restriction of  $u$  to terminal histories in  $\gamma$ . The profile  $b^\gamma$  is the strategy profile for subform  $\gamma$  induced by strategy profile  $b$ .

### Notation concerning the structure of a game tree

The player who moves at information set  $w$  is denoted by  $\iota(w)$ . Let  $h$  be a finite or infinite path in  $T$  starting at  $x_0$ . That is,  $h = (x_0, x_1, \dots, x_K)$  for some  $K < \infty$  such that  $x_k = s(x_{k-1})$  for  $k = 1, \dots, K$ , or  $h = (x_0, x_1, \dots)$  such that  $x_k = s(x_{k-1})$  for  $k = 1, 2, \dots$ . Such a path  $h$  is called a *history* of  $\Gamma$ . With a slight abuse of notation, I will write  $h \subseteq h'$  if  $h' = (h, x_k, x_{k+1}, \dots, x_K)$  or  $h' = (h, x_k, x_{k+1}, \dots)$  and  $x_k \in h$  if  $x_k$  is part of the sequence  $h$ . The set of all histories is denoted  $\mathcal{H}$ . A history  $h$  is *terminal* if there is no other history  $h' \in \mathcal{H}$  such that  $h \subset h'$ . The set of terminal histories is denoted by  $\bar{\mathcal{H}}$ . For  $w \in W$ , let  $H(w)$  denote the set of histories ending at some  $x \in w$ . The subpath of  $h$  with initial node  $x$  is denoted  $h_x$ . Let  $\leq$  be a partial order, defined over both  $X$  and  $W$ , where (1)  $x \leq x'$  if and only if  $x'$  is accessible from  $x$ , that is there exists a history  $h$  such that  $x' \in h_x$ , and (2)  $w \leq w'$  if and only if there exists a pair  $(x, x')$  with  $x \in w$ , and  $x' \in w'$  such that  $x \leq x'$ .

The information set that contains  $x$  is denoted by  $w(x)$ . The set of information sets that have, respectively, positive and zero probability to occur under  $b$  are denoted by  $A(b)$  and  $B(b)$  (so  $\{A(b), B(b)\}$  is a partition of  $W$ ). The probability distribution that  $b$  induces on  $\bar{\mathcal{H}}$  is denoted by  $o(b)$ .

### Trimmed SPE

The first part of this section gives a precise definition of maximal collections of irrelevant information sets and shows their existence and uniqueness. Let  $\rho_{SPE}$  stand for "sequential rationality as imposed by SPE." For extensive form game  $\Sigma$ , strategy profile  $b$ , and a set of information sets  $V \subseteq W$ , say that  $b$  *satisfies*  $\rho_{SPE}$  on  $V$  in  $\Sigma$  if  $b$  induces NE in all subgames  $(\gamma, u|_\gamma)$  with origin in  $V$ . For extensive form  $\Gamma$  and strategy profile  $b$ , let  $V \subseteq B(b)$  be such that whether  $b$  satisfies  $\rho_{SPE}$  on  $W \setminus V$  is

independent of the play on  $V$  in *any* game  $\Sigma = (\Gamma, u)$ . Call  $V$  a *collection of SPE-irrelevant information sets for  $b$* . A collection of *SPE-irrelevant information sets* divides a game into two parts. Whether the play on the "relevant" part is consistent with subgame-perfect rationality is independent from the play on the irrelevant part. Let  $\mathcal{W}_{irr}(b, \rho_{SPE})$  be the set of all collections of *SPE-irrelevant information sets for  $b$* . The set  $\mathcal{W}_{irr}(b, \rho_{SPE})$  is partially ordered (under inclusion). Notice that it follows directly from the definition that  $\mathcal{W}_{irr}(b, \rho_{SPE})$  is closed under union. Hence, every chain in  $\mathcal{W}_{irr}(b, \rho_{SPE})$  has an upper bound in  $\mathcal{W}_{irr}(b, \rho_{SPE})$ , namely the union of all members of the chain. It follows from Zorn's Lemma that there exists a unique, though possibly empty, maximal collection of irrelevant information sets, which will be denoted by  $W_{irr}(b, \rho_{SPE})$ . I show that whenever a strategy profile satisfies  $\rho_{SPE}$  on the set of all relevant information sets,  $W \setminus W_{irr}(b, \rho_{SPE})$ , one can find a strategy profile that induces the same outcome and satisfies the original equilibrium conditions, provided that equilibrium exists for the game.

To economize on notation, for the remainder of this section " $b$  satisfies  $\rho$ " shall always mean " $b$  satisfies  $\rho_{SPE}$ ." The first result in this section shows that the maximal collection of *SPE-irrelevant information sets for  $b$*  consists of those sets that can be reached from  $b$ 's outcome path only if at least two players deviate in the "first stage" of some subform (the emphasis on "only" is made because a non-singleton information set can be reached by different deviations). To make precise what is meant by the "first stage" of a subform, let  $\gamma^w$  denote the minimal subform to which  $w$  belongs, that is there is no proper subform of  $\gamma^w$  to which  $w$  belongs as well. Let  $\{W_\gamma\}_{\gamma \in G}$  be a partition of the set of information sets  $W$  where  $W_\gamma = \{w \in W : \gamma^w = \gamma\}$ . The set  $W_\gamma$  is referred to as the *first stage* of  $\gamma$ . Let  $b^V$  denote the restriction of  $b$  to the set  $V \subseteq W$ .

**Definition 1** (*Information sets on and off the unilateral deviation path*).

*Fix an extensive form game  $\Gamma$  and a strategy profile  $b$ . The set of information sets*

on the unilateral deviation path of  $b$  is denoted by  $B_1(b)$ . Information set  $w \in B_1(b)$  if and only if

1.  $w \in B(b)$ , and

2. there exists  $b'$  such that

(a)  $w \in A(b')$ , and

(b) for each  $W_\gamma$ , for at most one player  $i$ ,  $b_i^{W_\gamma} \neq b_i^{W'_\gamma}$ .

The set  $B_2(b) \equiv B(b) \setminus B_1(b)$  is the set of information sets off the unilateral deviation path of  $b$ .

Furthermore, let  $W_{SPE}(b) = A(b) \cup B_1(b)$  (so  $W \setminus W_{SPE}(b) = B_2(b)$ ). The sets  $A(b)$  and  $B(b)$  partition  $W$  into information sets on the outcome path of  $b$ , and information sets off the outcome path of  $b$ . The criterion of a unilateral deviation path leads to a further partition of  $B(b)$  into two sets, the set of information sets on the unilateral deviation path,  $B_1(b)$ , and the set of information sets off the unilateral deviation path,  $B_2(b)$ . Figure 2 illustrates a set of information sets off the unilateral deviation path for the profile that is indicated in the figure by the bold edges. The arrows indicate the corresponding outcome path. The set of information sets off the unilateral deviation path is  $B_2(b) = \{w'\}$ . Note that the figure shows a game form instead of a fully specified game with payoffs. Indeed, whether an information set is off the unilateral deviation path depends only on the game form and a strategy profile. As in this example,  $B_2(b)$  does not necessarily coincide with the set of information sets that can not be reached by a unilateral deviation from  $b$ . Here, both  $w$  and  $w'$  can not be reached by a unilateral deviation. The reason why  $w$  is not in  $B_2(b)$  is that it can be reached by a sequence of unilateral deviations (that is via a unilateral deviation path), one by player 2, and one by player 3 at her left information set. These deviations occur in the first stages of different subforms.

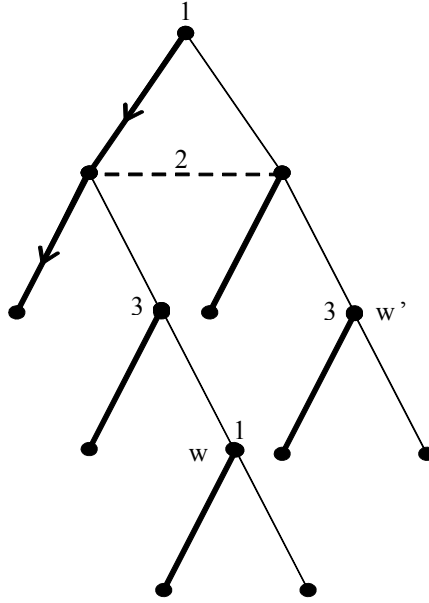


Figure 2: Information sets off the unilateral deviation path

The set  $W_{SPE}(b)$  will be shown to contain all information sets that are possibly relevant for  $b$  being consistent with subgame-perfect rationality. Notice the dependency of this statement on the strategy profile  $b$ . While a particular equilibrium might not be sensitive to specifics in some part of a game, the *set* of equilibria consistent with a certain concept might be sensitive to these specifics.

To prove Theorem 3, the following observation concerning the collection of information sets which are relevant for a NE is useful. Suppose one wants to check whether the profile  $b$  is a NE for  $\Sigma$ . Because a NE only requires that there are no beneficial unilateral deviations, it is sufficient to compare the outcome of  $b$  with outcomes obtained from a unilateral deviation from  $b$ . In terms of information sets, one only needs to consider play at information sets in  $A(b)$  and the information sets in  $B(b)$  that are reachable from  $A(b)$  by a unilateral deviation from  $b$ . The following lemma is also useful for the proof of Theorem 3.

**Lemma 2** *Let  $b$  be a strategy profile for  $\Gamma$ . For any pair  $w, \gamma$  such that  $w \in B_2(b)$ ,*

$w$  belongs to  $\gamma$ , and  $\{x_{0\gamma}\} \in W_{SPE}(b)$ , it holds that  $w \in B_2(b^\gamma)$ , that is,  $w$  is off the unilateral deviation path of  $b^\gamma$  in  $\gamma$ .

**Proof.** Because  $\{x_{0\gamma}\} \in W_{SPE}(b)$ , it can be reached through a sequence of unilateral deviations from  $b$ , at most one on each set  $W_\gamma$ ,  $\gamma \in G$ . Because  $w \in B_2(b)$ , it can *only* be reached from  $b$  if there are at least two players deviating at information sets in the first stage of some subform. Because  $\{x_{0\gamma}\}$  is a singleton information set and  $w$  belongs to  $\gamma$ , (a) only one history leads to  $x_{0\gamma}$  and (b) any history leading to  $w$  must pass through  $x_{0\gamma}$ . Hence,  $w \in B_2(b^\gamma)$  for otherwise, the concatenation of the unilateral deviation path leading to  $x_{0\gamma}$  and the path leading from  $x_{0\gamma}$  to  $w$ , is a unilateral deviation path as well, contradicting that  $w \in B_2(b)$ . ■

**Theorem 3** For any  $b$ ,  $W_{irr}(b, \rho) = B_2(b)$ , that is the maximal collection of SPE-irrelevant information sets for  $b$  is the set of information sets off the unilateral deviation path of  $b$ .

**Proof.**

1.  $W_{irr}(b, \rho) \subseteq B_2(b)$ .

To the contrary, suppose that  $W_{irr}(b, \rho) \not\subseteq B_2(b)$ . By definition  $W_{irr}(b, \rho) \subseteq B(b)$ , so we must have  $W_{irr}(b, \rho) \cap B_1(b) \neq \emptyset$ . Case (1): There exists  $w \in W_{irr}(b, \rho) \cap B_1(b)$  such that the origin of  $\gamma^w$ , denoted by  $x_{0\gamma^w}$ , is not an element of  $W_{irr}(b, \rho)$  and  $\{x_{0\gamma^w}\} \neq w$ . Since  $w \in B_1(b)$ , it can be reached from  $x_{0\gamma^w}$  by a unilateral deviation from  $b$ . Thus, there exists  $u_{|\gamma}$  such that the play at  $w$  matters for whether the play at  $x_{0\gamma^w}$  is a best response in the game  $(\gamma^w, u_{|\gamma})$ , contradicting that  $w \in W_{irr}(b, \rho)$  and  $\{x_{0\gamma^w}\} \notin W_{irr}(b, \rho)$ . Case (2): There must exist a  $w \in W_{irr}(b, \rho) \cap B_1(b)$  such that  $w = \gamma^w$ . In this case, consider the origin of  $\gamma'$ , the smallest subgame of which  $\gamma$  is a subgame. Since  $w \in B_1(b) \subseteq W_{SPE}(b)$ , so is  $x_{0\gamma'}$ , the origin of  $\gamma'$ . Moreover,  $w$  can be reached from  $x_{0\gamma'}$  by a unilateral deviation from  $b^{\gamma'}$ . Thus, there exists  $u_{|\gamma'}$  such that the play at  $w$  matters for whether the play at  $x_{0\gamma'}$  is a best response in



subgame  $(\gamma', u_{|\gamma'})$ , contradicting that  $w \in W_{irr}(b, \rho)$  and  $\{x_{0\gamma'}\} \notin W_{irr}(b, \rho)$ . Since  $W_{irr}(b, \rho) \cap B_1(b) \subseteq B(b)$  either Case 1 or Case 2 holds.

2.  $B_2(b) \subseteq W_{irr}(b, \rho)$

I will show that the play on the set  $B_2(b)$  is irrelevant for whether  $b$  satisfies  $\rho$  in any subform  $\gamma$  with origin in  $W_{SPE}(b)$ . Let  $\gamma$  be a subform with origin in  $W_{SPE}(b)$ .

Case 1)  $B_2(b) \cap \{w \in W : w \text{ belongs to } \gamma\} = \emptyset$ , so whether some profile  $b^\gamma$  is a NE for a subgame  $(\gamma, u_{|\gamma})$  is independent of the play on  $B_2(b)$ .

Case 2)  $B_2(b) \cap \{w \in W : w \text{ belongs to } \gamma\} \neq \emptyset$ . By Lemma 3 no  $w \in B_2(b) \cap \{w \in W : w \text{ belongs to } \gamma\}$  can be reached from  $b^\gamma$  by unilateral deviation, so whether  $b^\gamma$  is a NE of a game  $(\gamma, u_{|\gamma})$  does not depend on the play on  $B_2(b)$ . ■

We are ready to define a trimmed SPE.

**Definition 4** (*Trimmed SPE*). Strategy profile  $b^*$  is a trimmed SPE for  $\Sigma$  if it induces a NE in all subgames  $(\gamma, u_{|\gamma})$  with origin  $\{x_{0\gamma}\} \in W_{SPE}(b^*)$ .

The following result relates trimmed SPE to SPE.

**Theorem 5** (1) Every SPE is a trimmed SPE, but the converse does not hold. (2) In a game of perfect information,  $B_2(b) = \emptyset$  for all  $b$ .

**Proof.**

(1) This follows trivially from the definitions of a SPE and a trimmed SPE.

Figure 3 shows an example of a trimmed SPE that is not a SPE. Bold edges indicate the profile. Bold edges with arrows indicate the outcome path the profile induces.

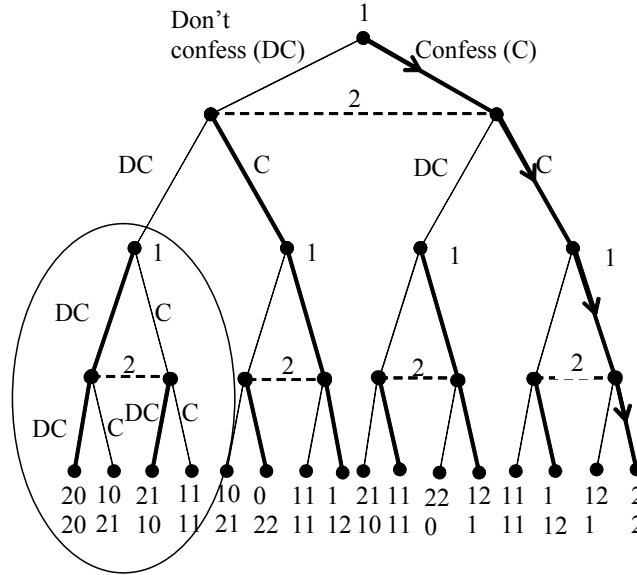


Figure 3: A trimmed SPE that is not a SPE

(2) In a game of perfect information every  $x \in X$  is the origin of some subform. Therefore, we have  $\{W_\gamma\}_{\gamma \in G} = W$ , implying that every  $w \in W$  can be reached from any  $b$  by a sequence of unilateral deviations. Thus  $B_2(b) = \emptyset$  for any  $b$ . ■

**Corollary 6** *In a game of perfect information, every trimmed SPE is a SPE.*

Since in perfect information games every move is the first move of some subgame, every information set can be reached from any strategy profile through a sequence of unilateral deviations, at most one per first stage of the game's subgames. Thus,  $W_{irr}(b, \rho) = \emptyset$  and so a trimmed SPE requires NE in all subgames of a game of perfect information.

A trimmed SPE is, by definition, less or at most as restrictive than a SPE. The following theorem shows that it is just as "strong" as a SPE, in the sense that it does not allow for equilibrium outcomes that cannot be supported by a SPE - provided NE exists for all subgames. Together with Theorem 3, this is the main result concerning

SPE. Let  $TSPE$  and  $SPE$  denote the sets of, respectively, trimmed subgame perfect equilibria and subgame perfect equilibria for some game  $\Sigma$ . Let  $O(TSPE)$  and  $O(SPE)$  denote the sets of outcomes (probability distributions of the set of terminal histories  $\tilde{\mathcal{H}}$ ) induced by, respectively, some  $b \in TSPE$  and some  $b \in SPE$ . Given a set of information sets  $V \subseteq W$ , and two strategy profiles  $b$  and  $b'$ , let  $(b^V, b^{W \setminus V})$  denote the profile obtained by playing  $b$  on  $V$ , and  $b'$  on  $W \setminus V$ .

**Theorem 7** *If  $SPE = \emptyset$  for  $\Sigma$ , then  $O(TSPE) = O(SPE)$ .*

**Proof.**

1.  $O(SPE) \subseteq O(TSPE)$

This follows from the fact that  $SPE \subseteq TSPE$ .

2.  $O(TSPE) \subseteq O(SPE)$

Let  $o \in O(TSPE)$  be induced by  $\tilde{b} \in TSPE$ . If  $\tilde{b} \in SPE$ , then  $o \in O(SPE)$ . If not, pick some  $b \in SPE$  and consider the strategy profile  $b^{SPE} = (\tilde{b}^{W_{SPE}(\tilde{b})}, b^{B_2(\tilde{b})})$ . Notice that  $b$  and  $\tilde{b}$  induce the same outcomes because they only differ at information sets in  $B(\tilde{b})$ . Next, I will show that  $b^{SPE}$  is a subgame perfect equilibrium for  $\Sigma$ . First, consider subgame  $(\gamma, u_{|\gamma})$  with  $\{x_{0\gamma}\} \in W_{SPE}(\tilde{b})$ . By Lemma 3, all changes made in subgame  $\gamma$  when moving from  $\tilde{b}$  to  $b^{SPE}$  were made at information sets  $w \in B_2(\tilde{b}^\gamma)$ . Hence these changes do not affect play on the unilateral deviation path of  $\tilde{b}^\gamma$  in  $\gamma$ , and so, since  $\tilde{b}$  is a trimmed SPE,  $b^{SPE}$  induces a NE on  $(\gamma, u_{|\gamma})$ . Second, consider any subgame  $(\gamma, u_{|\gamma})$  with  $\{x_{0\gamma}\} \in B_2(\tilde{b})$ . All information sets belonging to  $\gamma$  are elements of  $B_2(\tilde{b})$  because all histories leading to them pass through  $x_{0\gamma}$ . Hence  $b^{SPE}$  and  $b$  induce the same play on  $\gamma$ , implying that  $b^{SPE}$  induces a NE on  $(\gamma, u_{|\gamma})$ . Thus  $b^{SPE}$  induces NE on all subgames and therefore  $b^{SPE} \in SPE$ , showing that the outcome induced by  $\tilde{b}$  is also an outcome of some subgame perfect equilibrium. ■

**Corollary 8** *For  $b \in TSPE$  and any  $b'' = (b^{W_{SPE}(b)}, b^{B_2(b)})$  where  $b^{B_2(b)}$  induces NE on all subgames  $(\gamma, u_{|\gamma})$  with origin in  $B_2(b)$ , the profile  $b''$  is a SPE.*

To return to the motivations for the concept of trimming strategies, consider again the example in Figure 3. It follows from the results in this section that: (1) Even if the subgame at the left was such that a NE for it did not exist, the outcome path of the trimmed SPE, indicated by the arrows, would be consistent with sequential rationality as imposed by SPE; (2) Whatever changes are made to the subgame at the left, the outcome remains consistent with SPE and is thus robust to such changes; and (3) If a player considers to deviate from the indicated play, expecting, as is consistent with subgame perfection, Nash equilibrium play at any subgame reached through the deviation, the subgame at the left does not enter this player's considerations.

### An Example

The following example demonstrates the potential usefulness of a trimmed SPE. Suppose two firms,  $i$  and  $j$ , produce the same homogenous good and compete against each other in Bertrand-Edgeworth fashion. First, they simultaneously build capacities,  $x_i$  and  $x_j$ , at a cost of  $K$  dollars per unit. Second, they simultaneously announce prices,  $p_i$  and  $p_j$ . After that, demand is realized. Market demand is given by  $D = 20 - P$ . Due to its limited capacity, the low price firm might not be able to serve everyone who demands to buy at the price it charges. Hence, a rationing rule is needed. With a surplus maximizing rationing rule (Levitan and Shubik 1972 and Shubik 1955) the lower price firm serves the high demand consumers.<sup>16</sup> That is, if  $p_i < p_j$ , demand for firm  $i$  is  $20 - p_i$ . For simplicity, assume variable production costs are zero. Firm  $i$  then produces  $\text{Min}\{20 - p_i, x_i\}$ . If the constraint binds for firm  $i$ , firm  $j$  might serve some consumers as well, but only up to its capacity, that is it produces  $\text{Min}\{\text{Max}\{20 - p_j - x_i, 0\}, x_j\}$ . If  $p_i = p_j = p$ , firm  $i$ 's demand is given by  $\text{Min}\{x_i, \frac{D}{2} + \text{Max}\{0, \frac{D}{2} - x_j\}\}$ , and similarly for firm  $j$ , which means that if  $p$  is such that  $D(p) = x_i + x_j$ , firms simply produce up to their capacities.

---

<sup>16</sup>Suppose there is a mass of consumers of measure one, who all demand one unit of the good and whose willingness to pay is uniformly distributed on the interval  $[0, 20]$ .

Suppose only pure strategies are available to the firms, and that we are interested in the SPE of this game. Each capacity pair chosen in the first stage induces a pricing subgame in the second stage. As Kreps and Scheinkman (1983) show, pure strategy equilibria exist only for a subset of these subgames. More specifically, let  $r_i(x_j)$  and  $r_j(x_i)$  denote the best reply correspondences of the Cournot game in which firms  $i$  and  $j$  simultaneously choose how much to produce and both have no capacity constraints and zero unit costs. Pure strategy price equilibria exist only for the set  $M = \{x \in \mathbb{R}_+^2 : x_i \leq r_i(x_j) \text{ and } x_j \leq r_j(x_i), \text{ or } x_i = x_j = 20\}$ .<sup>17</sup> Hence, a SPE does not exist. Kreps and Scheinkman compute the equilibrium for the mixed extension of the game.<sup>18</sup> In that equilibrium firms choose capacities equal to the equilibrium quantities of the Cournot game when firms have unit costs of  $K$ . Thus, the equilibrium capacities are given by  $x_i = x_j = \frac{20-K}{3}$ . Except if  $x_i = x_j = 20$ , all pricing equilibria in  $M$  are such that both firms charge  $D^{-1}(x_i + x_j)$ . If  $x_i = x_j = 20$ , both firms charge zero.

Now suppose that, in addition, the cost of capacity is discontinuous and capacity higher than some  $\bar{x}$  is prohibitively expensive. Suppose that  $\frac{20-K}{3} < \bar{x} \leq \frac{20}{3} + \frac{K}{6}$ . For this value, there remains a subset of subgames for which pure strategy equilibria do not exist, as Figure 4 illustrates. However, given that firm  $j$  builds a capacity of  $\frac{20-K}{3}$ , all feasible capacity levels for firm  $i$  are in the set  $M$ . Hence, for all subgames reachable for firm  $i$ , conditional on firm  $j$  producing  $\frac{20-K}{3}$ , the expected pricing equilibrium will have both firms charging  $P(x_i + x_j)$  (where  $P(\cdot)$  is the inverse demand function), and serving half of  $20 - P(x_i + x_j)$ . It is easy to verify that then a deviation for firm  $i$  is not worthwhile. Simply solve

$$\max_{x_i \leq \bar{x}} x_i \left( 20 - \left( \frac{20-K}{3} \right) - x_i - K \right)$$

---

<sup>17</sup>For simplicity, assume that firms never build capacities beyond 20.

<sup>18</sup>They conduct a rather complicated analysis, owed to the continuous strategy set and because they demonstrate uniqueness of the equilibrium.

which yields  $x_i^* = \frac{20-K}{3}$ . By symmetry the same holds for firm  $j$ .

Here,  $x_i^*$  and  $x_j^*$  together with the pricing equilibria on the subset of relevant subgames (the cross in the figure) and arbitrary price choices on all other subgames constitute a trimmed SPE. Note how the trimmed SPE rules out noncredible threats here.

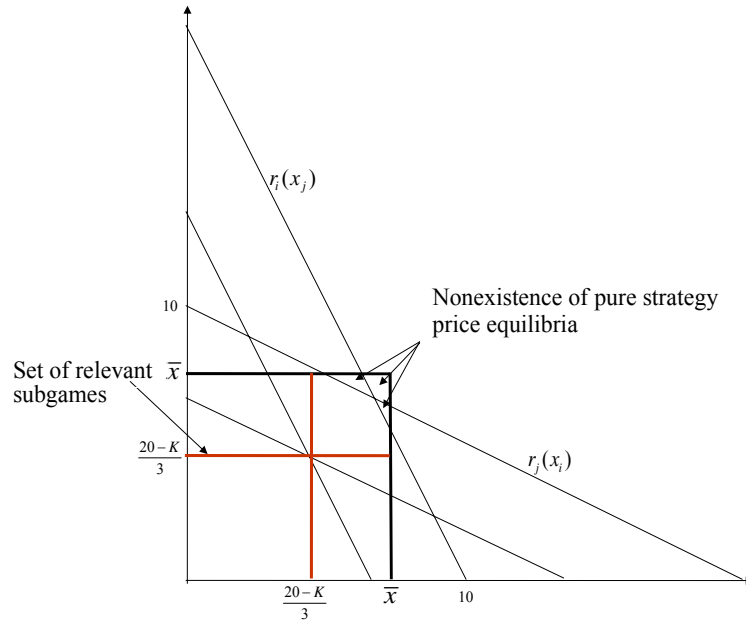


Figure 4: A trimmed SPE can exist when a SPE does not.

### Trimmed WPBE

The task remains in principle the same as before. Given a strategy profile  $b$  and now also a system of beliefs  $\mu$ , determine the maximal subset of  $B(b)$  such that the play inside this set does not affect the sequential rationality outside of the set. First, we need a precise definition of maximal collections of  $WPBE$ -irrelevant information sets. Let  $\rho_{WPBE}$  stand for "sequential rationality as imposed by WPBE." For extensive

form game  $\Sigma$ , strategy profile and system of beliefs  $(b, \mu)$ , and a set of information sets  $V \subseteq W$ , say that  $(b, \mu)$  *satisfies*  $\rho_{WPBE}$  on  $V$  in  $\Sigma$  if at each  $w \in V$ , the play  $b$  specifies at  $w$  is optimal given  $\mu(w)$  and given future play induced by  $b$ . In addition,  $\mu(w)$  has to be derived using Bayes' rule if possible. For  $(b, \mu)$  and extensive form  $\Gamma$ , let  $V \subseteq B(b)$  be such that whether  $b$  satisfies  $\rho_{WPBE}$  on  $W \setminus V$  is independent of the play on  $V$  in *any* game  $\Sigma = (\Gamma, u)$ . Call  $V$  a *collection of WPBE-irrelevant information sets for  $(b, \mu)$* . As before, a collection of WPBE-irrelevant information sets divides a game into two parts. Whether the play and beliefs on the "relevant" part are consistent with weak-perfect Bayesian rationality is independent from the play on the irrelevant part. Let  $\mathcal{W}_{irr}(b, \mu, \rho_{WPBE})$  be the set of all collections of WPBE-irrelevant information sets for  $(b, \mu)$  and let  $W_{irr}(b, \mu, \rho_{WPBE})$  be a maximal collection of WPBE-irrelevant information sets for  $(b, \mu)$ . The argument why  $W_{irr}(b, \mu, \rho_{WPBE})$  exists and is unique is the same as for  $W_{irr}(b, \rho_{SPE})$ .

It will be shown that whenever a strategy profile and a system of beliefs satisfy  $\rho_{WPBE}$  on the set of all relevant information sets,  $W \setminus W_{irr}(b, \mu, \rho_{WPBE})$ , one can find a strategy profile and a system of beliefs with the same outcome path satisfying the original equilibrium conditions, provided that equilibrium exists for the game.

In this section, " $(b, \mu)$  satisfies  $\rho$ " shall always mean " $(b, \mu)$  satisfies  $\rho_{WPBE}$ ." How can the set  $W_{irr}(b, \mu, \rho)$  be characterized? It turns out that the characterization is similar to the one for SPE, but incorporates beliefs. The idea is that for any  $w \notin W_{irr}(b, \mu, \rho)$  and any history leading from  $w$  to  $W_{irr}(b, \mu, \rho)$ , a player whose move it is some time before  $W_{irr}(b, \mu, \rho)$  is reached does not believe that  $W_{irr}(b, \mu, \rho)$  can be reached. This belief is either due to his, possibly incorrect, belief about past play or due to the fact that  $b$  attaches zero probability to another player's future move in the history. In other words, players do not believe that  $W_{irr}(b, \mu, \rho)$  can be reached by a unilateral deviation, but these beliefs need not be correct.

For  $x' = s(x)$ , let  $b(x, x')$  denote the probability that  $b$  attaches to the choice

leading from  $x$  to  $x'$ .

**Definition 9** (*Information sets on and off the believed unilateral deviation path*).

Fix an extensive form  $\Gamma$ , a strategy profile  $b$ , and a system of beliefs  $\mu$ . The set of information sets off the believed unilateral deviation path of  $(b, \mu)$  is denoted  $B_2(b, \mu)$ .

Information set  $w \in B_2(b, \mu)$  if and only if

(i)  $w \in B(b)$ , and

(ii)  $h = (x_0, x_1, \dots, x_K) \in H(w)$  implies that for each  $x_k \in h$

(a)  $w(x_k) \in B_2(b, \mu)$ , or

(b)  $\mu(x_k) = 0$ , or

(c)  $\exists x_{k'} \in h$  with  $k < k' < K$ ,  $w(x_{k'}) \notin B_2(b, \mu)$ , and  $\iota(x_{k'}) \neq \iota(x_k)$  such that  $b(x_{k'}, x_{k'+1}) = 0$ .

Also  $B_1(b, \mu) = B(b) \setminus B_2(b, \mu)$  is the set of information sets off the believed unilateral deviation path of  $(b, \mu)$  and  $W_{WPBE}(b, \mu) = A(b) \cup B_1(b, \mu)$ .

Figure 5 illustrates the set of information sets off the believed unilateral deviation path of the profile and beliefs indicated in the figure. Again, bold edges represent the strategy profile and arrows represent the corresponding outcome path. If player 3 believes that the left node at his information set in the middle is reached with probability  $\mu = 0$ , then  $B_2(b, \mu) = \{w\}$ .<sup>19</sup> As will be shown,  $\{w\}$  is the maximal collection of *WPBE*-irrelevant information sets for the indicated profile and beliefs. The information set  $w''$  for example is not irrelevant because the play there matters for the rationality of player 2's choice at her right information set, which in turn matters for the rationality of the choice by player 1 at the origin.

---

<sup>19</sup>Here,  $\mu$  is used with a slight abuse of notation as it was previously defined as a system of beliefs.



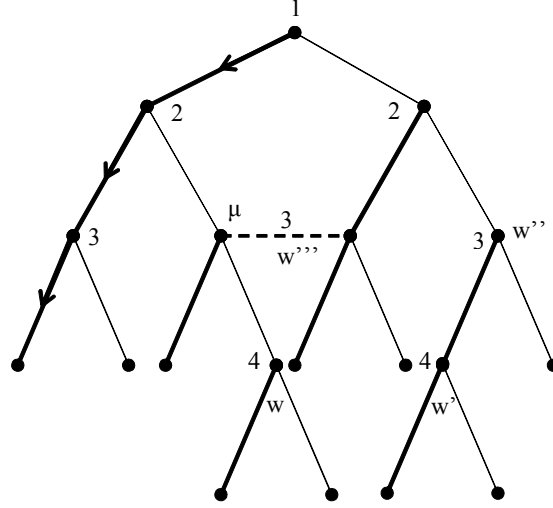


Figure 5: Information sets off the believed unilateral deviation path

The first result in this section parallels Theorem 3. It states that the set of information sets off the believed unilateral deviation path of  $(b, \mu)$  is indeed the maximal collection of  $WPBE$ -irrelevant information sets for  $(b, \mu)$ .

**Theorem 10** *For any  $(b, \mu)$ ,  $W_{irr}(b, \mu, \rho) = B_2(b, \mu)$ , that is the maximal collection of  $WPBE$ -irrelevant information sets for  $(b, \mu)$  is the set of information sets off the believed unilateral deviation path of  $(b, \mu)$ .*

**Proof.**

1.  $W_{irr}(b, \mu, \rho) \subseteq B_2(b, \mu)$ .

To the contrary, suppose that  $M \equiv W_{irr}(b, \mu, \rho) \cap W_{WPBE}(b, \mu) \neq \emptyset$ . Because  $M \subseteq W_{irr}(b, \mu, \rho) \subseteq B(b)$ , we can pick  $\hat{w} \in M$  such that for all  $w \leq \hat{w}$ , we have that  $w \notin M$  (because  $\{x_0\} \notin M$ ). Since  $M \subseteq W_{WPBE}(b, \mu) = W \setminus B_2(b, \mu)$ , the information set  $\hat{w} \notin B_2(b, \mu)$ . However,  $\hat{w} \in B(b)$  so there exists  $\hat{h} \in H(\hat{w})$  for which none of the conditions in part (ii) of Definition 9 holds. Write  $\hat{h} = (x_0, x_1, \dots, x_K)$  with  $x_K \in \hat{w}$ . By failure of (a), (b), and (c) in Definition 9, there exists a node  $x_{k'} \in \hat{h}$  such that  $w(x_{k'}) \notin B_2(b, \mu)$ ,  $\mu(x_{k'}) > 0$  and  $b(x_{k''}, x_{k''+1}) > 0$  for all  $(x_{k''}, x_{k''+1}) \subseteq \hat{h}$  for

which  $k'' > k'$  and  $\iota(k'') \neq \iota(k')$ . Therefore, for some  $u$  the play at  $\hat{w}$  matters at  $x_{k'}$  (given the other players' strategies,  $\iota(x_{k'})$  believes that he can choose a strategy so that  $\hat{w}$  is reached). However,  $w(x_{k'}) \notin B_2(b, \mu)$  and  $w(x_{k'}) \notin M$  implies that  $w(x_{k'}) \notin W_{irr}(b, \mu, \rho)$ , contradicting that  $\hat{w} \in W_{irr}(b, \mu, \rho)$ .

2.  $B_2(b, \mu) \subseteq W_{irr}(b, \mu, \rho)$

It suffices to show that  $B_2(b, \mu)$  is a collection of *WPBE*-irrelevant information sets for  $(b, \mu)$ . Pick  $w' \in B_2(b, \mu)$  and  $w \notin B_2(b, \mu)$ .

Case 1) not  $(w \leq w')$ .

In this case, no node in  $w$  has a choice that leads to  $w'$ . Then, given  $\mu$ , the choice at  $w'$  does not matter for the sequential rationality at  $w$  in any game  $\Sigma = (\Gamma, u)$ .

Case 2)  $w \leq w'$ . Let  $L \subseteq H(w')$  such that  $h \in L$  implies  $h \cap w \neq \emptyset$ , that is  $L$  contains the histories that can access  $w'$  from  $w$ . Pick any  $x_{k'} \in w$  such that  $x_{k'} \in h$  for some  $h \in L$ . Because  $w \notin B_2(b, \mu)$ , either (b) or (c) in part (ii) of the above definition must hold. Therefore, either  $\iota(w)$  does not believe to be at  $x_{k'}$  or believes that  $w'$  cannot be reached from  $x_{k'}$  because of the future play of other players. Therefore the play at  $w'$  does not matter for the sequential rationality of play at  $w$  in any game  $\Sigma = (\Gamma, u)$ .

This shows that  $B_2(b, \mu)$  is a collection of *WPBE*-irrelevant sets for  $(b, \mu)$ . ■

In view of  $B_2(b, \mu)$  being the maximal collection of *WPBE*-irrelevant information sets for  $(b, \mu)$ , a trimmed WPBE is defined as follows

**Definition 11** *The pair  $(b, \mu)$  is a trimmed WPBE of  $\Sigma$  if  $(b, \mu)$  satisfies  $\rho$  on  $W_{WPBE}(b, \mu)$ .*

The following result relates trimmed WPBE to WPBE. It parallels the result in Theorem 5.

**Theorem 12** (1) *Every WPBE is a trimmed WPBE, but the converse does not hold.*  
(2) *In a game of perfect information,  $B_2(b, \mu) = \emptyset$  for all  $(b, \mu)$ .*

**Proof.**

(1) This follows trivially from the definitions of a WPBE and a trimmed WPBE.

Figure 6 shows an example of a trimmed WPBE that is not a WPBE.

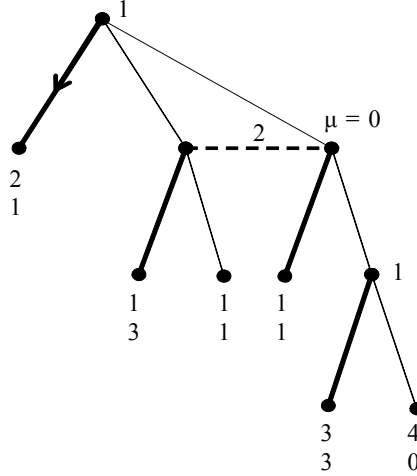


Figure 6: A trimmed WPBE that is not a WPBE

(2) Fix some  $(b, \mu)$ . Suppose  $B_2(b, \mu) \neq \emptyset$ . Pick  $w' \in B_2(b, \mu)$  with  $w \leq w'$  implying that  $w \notin B_2(b, \mu)$ . Let  $x \in p(w')$ , so  $w(x) \notin B_2(b, \mu)$ , so (a) does not hold. By perfect information,  $\mu(x) = 1$ , so (b) does not hold. Because  $x$  is an immediate predecessor of  $w'$ , (c) cannot hold either, a contradiction. ■

**Corollary 13** *In a game of perfect information, every trimmed WPBE is a WPBE.*

The next result parallels Theorem 7. It establishes that every trimmed WPBE can be matched with a WPBE that induces the same outcome, *provided that a WPBE exists*. Let  $WPBE$  and  $TWPBE$  be the sets of pairs  $(b, \mu)$  that are, respectively, a WPBE and a trimmed WPBE. Let  $O(WPBE)$  and  $O(TWPBE)$  be the sets of outcomes induced by some  $(b, \mu)$  in, respectively,  $WPBE$  and  $TWPBE$ .

**Theorem 14** *If  $WPBE \neq \emptyset$  (the features of the game are such that a WPBE exists), then  $O(TWPBE) = O(WPBE)$ .*

**Proof.**

1.  $O(WPBE) \subseteq O(TWPBE)$ .

This follows from the fact that  $WPBE \subseteq TWPBE$  as shown in Theorem 12.

2.  $O(TWPBE) \subseteq O(WPBE)$

Pick any  $(b, \mu) \in TWPBE$ . Construct  $(b', \mu')$  as follows.

(a) For each  $w \in W_{WPBE}(b, \mu)$ , let  $b'(w) = b(w)$  and let  $\mu'(w) = \mu(w)$ .

(b) For information sets in  $B_2(b, \mu)$ , construct a reduced game from  $\Gamma$  and  $(b, \mu)$  as follows. Let  $X_{WPBE}$  and  $X_{B_2}$  denote the sets of nodes for which  $w(x)$  is an element of  $W_{WPBE}(b, \mu)$  and  $B_2(b, \mu)$ , respectively.

*Step 1.* Remove all  $x \in X_{WPBE}$  such that  $x' \leq x$  implies  $x' \in X_{WPBE}$ .

*Step 2.* For all  $x \in X_{WPBE}$  such that  $x \leq x'$  implies  $x' \in X_{WPBE}$ , if  $p(x) \in X_{B_2}$ , replace  $x$  by the expected payoff induced by  $b$ , and if  $p(x) \in X_{WPBE}$  delete  $x$ .

*Step 3.* For the remaining  $x \in X_{WPBE}$ , if  $p(x) \in X_{B_2}$ , replace  $x$  with a move by nature as follows. Each path emanating from  $x$  so that the path passes only through  $X_{WPBE}$  is replaced by a terminal node. Assign the payoff to the terminal node that equals the expected payoff that  $b$  induces at  $x$  when this path is taken. Assign nature a probability to choose this final node that equals the probability that  $b$  attaches to the path emanating from  $x$ . For each  $x' \geq x$  such that  $x' \in X_{B_2}$  and for which  $x' \geq x'' \geq x$  implies  $x'' \in X_{WPBE}$ , replace the path leading to  $x'$  by a move by nature leading directly to  $x'$  with the probability as induced by  $b$  (which is well defined because all nodes on this path belong to  $X_{WPBE}$ ).

After performing these three steps, all nodes in  $X_{WPBE}$  were either deleted, replaced by a final node, or replaced by a move by nature. Note that the resulting graph does not need to be connected and that there might be several initial nodes. Add an initial move by nature that attaches some positive probability to each initial node of the

components obtained. Call the new game  $\Gamma^*$ . Apart from nature's moves, the set of information sets for  $\Gamma^*$  is  $B_2(b, \mu)$ . Now, find a weak perfect equilibrium for  $\Gamma^*$ , denoted  $(b^*, \mu^*)$  and complete the specification of  $(b', \mu')$  by letting  $b'(w) = b^*(w)$  and  $\mu'(w) = \mu^*(w)$  for all  $w \in B_2(b, \mu)$ .

It remains to verify that  $(b', \mu')$  is a weak perfect Bayesian equilibrium. For any  $w \in W_{WPBE}(b, \mu)$  we know that  $b(w)$  satisfies  $\rho$  at  $w$  given  $b$ . Because we only changed strategies and beliefs at information sets in  $B_2(b, \mu)$ , those changes could not affect the sequential rationality at  $w$ .

For any  $w \in B_2(b, \mu)$ , sequential rationality follows from the fact that  $(b^*, \mu^*)$  is a weak perfect Bayesian equilibrium and that for each  $x \in w$  and each choice at  $x$ , the expected payoff from that choice given  $b^*$  is the same as the expected payoff given  $b'$ . Notice that, because  $B_2(b, \mu) \subseteq B(b)$ , Bayes' rule does not need to be satisfied on  $B_2(b, \mu)$ . ■

**Corollary 15** *For  $(b, \mu) \in TWPBE$  and any  $b'' = (b^{W_{WPBE}(b, \mu)}, b^{B_2(b, \mu)})$  and  $\mu'' = (\mu^{W_{WPBE}(b, \mu)}, \mu^{B_2(b, \mu)})$ , where  $(b^{B_2(b, \mu)}, \mu^{B_2(b, \mu)})$  induces sequentially rational play at all  $w \in B_2(b, \mu)$ , the pair  $(b'', \mu'')$  is a WPBE.*

## Conclusion

This paper introduces the concept of equilibrium in trimmed strategies. The concept is based on the notion of maximal collections of irrelevant information sets. Trimmed equilibrium potentially can provide resolution when some parts of the game are difficult to predict, be it due to nonexistence of equilibrium, uncertainties about the game's specifications, or computational difficulties. It is shown that the trimmed version of an equilibrium is sufficiently restrictive to capture the notion of sequential rationality of the original concepts. At the same time, it disposes of restrictions

on entire parts of the game. The characterization of maximal collections of irrelevant information sets for both SPE and WPBE provides insight as to the kinds of rationalities these concepts place on players.

It is pointed out that, while a trimmed equilibrium is invariant in the specifications on the maximal collection of irrelevant information sets, the set of trimmed equilibria is not. Also, since a trimmed equilibrium leaves strategies on the maximal collection of irrelevant information sets essentially unspecified, it is not clear how to solve for such an equilibrium. A process of backward induction does not seem suitable. Of course, it is possible to determine the entire set of Nash equilibria and then check whether they are trimmed equilibria. This, however, might not be practical as the set of Nash equilibria in extensive form games can be very large. Constructing an algorithm that solves for trimmed equilibria - or which at least narrows down the set of candidates - remains an open task.

## CHAPTER III

### A CHARACTERIZATION OF WEAKLY PAIRWISE STABLE NETWORKS

#### Introduction

Consider a dynamic process of network formation in which the network evolves according to successive modifications by myopic players. In each modification, either a single player severs a subset of his current links or a pair of players forms a link. A stable state of this process is a weakly pairwise Nash stable network: No player benefits from severing links and for no pair both players benefit from forming a link.<sup>20</sup> However, a stable state, and therefore a weakly pairwise Nash stable network, might not exist. This paper characterizes the set of weakly pairwise Nash stable networks as the set of pure strategy Nash equilibrium networks of a network formation game in strategic form. Since a weakly pairwise Nash stable network might not exist, it is then natural to consider the equilibrium of the mixed extension of this strategic form game. Because the game is finite, this equilibrium exists. Thus, my characterization result makes possible a prediction of the network formation process in cases when a weakly pairwise Nash stable network does not exist.

The network formation game I define is a variant of Myerson's linking game (Myerson 1991). In Myerson's linking game all players simultaneously announce sets of players with whom they wish to form links. Links are formed if and only if two players announce each other. While in equilibrium no player wishes to cut links, equilibrium networks do permit situations in which two players would both benefit from forming a link.

---

<sup>20</sup>This paper uses the concept of weak pairwise Nash stability instead of the more familiar concept of pairwise Nash stability for technical reasons, which will be explained in due course.

To eliminate this coordination failure in Myerson's linking game, I introduce pairs of players as additional players to the game. The so modified game is called the linking game with player pairs. Each player pair announces whether it wishes to form a link. However, the strategy of a pair  $ij$  has an impact on the outcome of the game only if neither of the two players announces the other. In this case, if  $ij$  announces it wishes to form a link, the rules of the game are such that the link  $ij$  is added. The payoffs for player pairs are defined so that a pair benefits from a link if and only if both players benefit from it. Together with the rules of the game, this payoff specification ensures that the set of equilibrium outcomes is precisely the set of weakly pairwise Nash stable networks.

To provide intuition for the presence of pairs of players coordinating on the formation of their link, I define and analyze a game in which network formation occurs in two stages. In the first stage, players play Myerson's linking game. In the second stage, they play a constrained version of the linking game with player pairs in which only pairs that have formed a link in the first stage are added as players to the game. The outcome of the game is the network formed in the second stage. Thus, the only purpose of a first stage link is that it allows players to coordinate in the second stage. A pair that has formed a link in the first stage ensures its (actual) link is formed in the second stage whenever both players benefit from doing so.

The two-stage linking game provides a natural model of link formation. Consider friendship formation as an example. There, the first stage might correspond to something like joining a club or getting to know the friends of friends and the second stage corresponds to actual friendship formation. The interpretation of stage one links eliminating coordination failure in stage two is that once two people get to know each other they will become friends provided they both like each other. Similarly, two people might never become friends, simply because they do not know (much about) each other. As a consequence, real-world networks might be stable even though they do not



satisfy the definition of pairwise stability. For example, there are likely many people whose friendship we would enjoy if we only knew them. In this vein, Proposition 17 shows that the set of subgame perfect equilibrium networks of the two-stage linking game is larger than the set of pairwise Nash stable networks. However, as Proposition 18 shows, refining the equilibria of the two-stage linking game (by requiring undominated strategies) leads to weakly pairwise Nash stable equilibrium networks.

This paper relates to the literature of network formation that goes back to Myerson (1991) and the seminal work by Jackson and Wolinsky (1996). Here, I address the issue of existence of weakly pairwise Nash stable networks. The prior literature on network formation that uses pairwise stability or closely related concepts and addresses existence issues has either focused on establishing conditions that guarantee existence or shown existence in special settings (see e.g., Belleflamme and Bloch, 2004; Calvo-Armengol 2004; Goyal and Joshi, 2006; Jackson and Watts, 2001). A novelty in this paper is that it presents a way to incorporate mixed strategies into a network formation game with coordinated moves. The paper thereby addresses the existence of pairwise Nash stable equilibrium.

The next section defines the linking game with player pairs and shows that its equilibrium outcomes are equivalent to the set of weakly Nash stable networks. At the end of the section, I show that the mixed strategy equilibrium might put weight on networks that are not part of an absorbing state of a dynamic process of network formation. Section 3 defines the two-stage linking game and shows that networks supported by its undominated subgame perfect equilibria are weakly pairwise Nash stable. Section 4 concludes.

### **The Linking Game with Player Pairs**

Let  $N = \{1, 2, \dots, n\}$  be a finite set of players, let  $g \subseteq \{ij : i, j \in N, i \neq j\}$  be a social network, and let  $G$  be the set of all networks. The payoff for individual  $i \in N$

is given by the function  $u_i : G \rightarrow R$ . A network that is derived from network  $g$  by adding (deleting) link  $ij$  is denoted by  $g + ij$  ( $g - ij$ ). The set of player  $i$ 's neighbors in  $g$  is  $N_i(g) = \{j \in N : ij \in g\}$ .

In the *linking game*, each player  $i \in N$  has the strategy set  $S_i = P(N \setminus \{i\})$  (the power set of  $N \setminus \{i\}$ ) with typical element  $s_i$ . Given a strategy profile  $s = (s_1, s_2, \dots, s_n)$ , the outcome of the game is the network  $g(s)$  defined by  $ij \in g(s)$  if and only if  $i \in s_j$  and  $j \in s_i$ . If  $g = g(s)$ , the profile  $s$  *supports*  $g$ . A network is *Nash stable* if it is supported by a Nash equilibrium of the linking game. In other words, a network is Nash stable if there exists a strategy profile  $s$  for the linking game such that

$$u_i(g(s)) \geq u_i(g(s'_i, s_{-i})) \text{ for all } s'_i \in S_i, \text{ for all } i \in N.$$

As has been pointed out in the literature, Nash stability is an unsatisfactory concept in the context of network formation. Its predictive power is limited as the set of Nash stable networks tends to be large. In particular, the empty network, a network where no link is formed, is always Nash stable. This is due to the fact that a network in which two players are not linked but both would benefit from the addition of the link can be Nash stable if neither of the two players indicates the other player. To overcome this coordination failure, Jackson and Wolinsky (1996) introduced the concept of pairwise stability. A network is *pairwise stable* if

$$\begin{aligned} u_i(g) &\geq u_i(g - ij) \text{ for all } i, \text{ for all } ij \in g, \text{ and} \\ u_i(g + ij) &> u_i(g) \Rightarrow u_j(g + ij) < u_j(g), \text{ for all } ij \notin g. \end{aligned}$$

In a pairwise stable network no single player wishes to sever a single link and no pair of players wishes to add a link.

A network is *pairwise Nash stable* if it is pairwise stable and Nash stable. Pairwise Nash stability captures the idea that adding a link requires mutual consent but

cutting a link is at the discretion of either of the linked players. Note that one could decompose the concept into a noncooperative component, Nash stability, and a cooperative component, the second condition of pairwise stability. In this paper I consider a slightly weaker version of pairwise stability: A network is *weakly pairwise stable* if

$$\begin{aligned}
 u_i(g) &\geq u_i(g - ij) \text{ for all } i, \text{ for all } ij \in g, \text{ and} \\
 u_i(g + ij) &> u_i(g) \Rightarrow u_j(g + ij) \leq u_j(g), \text{ for all } ij \notin g.
 \end{aligned}$$

Here, the idea is that a coordinated deviation by two players requires that both players benefit from that deviation. A network is *weakly pairwise Nash stable* if it is Nash stable and weakly pairwise stable. Jackson and Wolinsky (1996) informally discuss this notion in the last section of their paper. They point out that most of their results are not sensitive to which notion of pairwise stability is used.

To see that a weak pairwise Nash stable network might not exist it is useful to consider the following dynamic network formation process.<sup>21</sup> The process starts out with an arbitrary network. If the network is pairwise Nash stable, the process ends. If not, there exists at least one pair of players who wishes to be linked or a at least one single player who wishes to sever at least one of his links. One of the pairs who wish to add links (if any) or one of the players who wish to sever links (if any) is selected and the network is modified accordingly. If the modified network is pairwise Nash stable, the process ends; if not, the procedure is repeated. This process itself can be depicted as a (directed) network, called a *supernetwork*.<sup>22</sup> The set of nodes of the supernetwork is the set of networks  $G$ . A *supernetwork* is a directed network

---

<sup>21</sup>Such processes are examined by, e.g., Bala and Goyal (2000), Goyal and Vega-Redondo (2005), Jackson and Watts (2002), and Watts (2001).

<sup>22</sup>Supernetworks were introduced by Page, Wooders, and Kamat, 2005.

$\mathbf{G} \subseteq G \times G$  such that  $(g, g') \in \mathbf{G}$  if and only if

- (i)  $g' = g - \{ik\}_{k \in B}$  for some  $B \subseteq N_i(g)$  and  $u_i(g') > u_i(g)$ , or
- (ii)  $g' = g + ij$  and  $u_i(g') > u_i(g)$  and  $u_j(g') > u_j(g)$ .

An arc  $(g, g')$  in the supernetwork precisely means that network  $g$  is not weakly pairwise Nash stable because either (i) or (ii) holds. The usual terminology is to say that network  $g'$  *defeats* network  $g$ . Given supernetwork  $\mathbf{G}$ , it is straightforward to identify weakly pairwise Nash stable networks. One simply searches for networks without outgoing arcs. Such a network is not defeated by any other network and must be weakly pairwise Nash stable. If a weakly pairwise Nash stable network does not exist, by the finiteness of  $G$ , the process must eventually end up in a cycle which has no outgoing arc (see Lemma 1 in Jackson and Watts, 2002). Figure 7 shows how a supernetwork could look like. Here,  $g^8$  is weakly pairwise Nash stable. In addition, the networks  $g^0, g^2, g^3$ , and  $g^1$  form a cycle.<sup>23</sup>

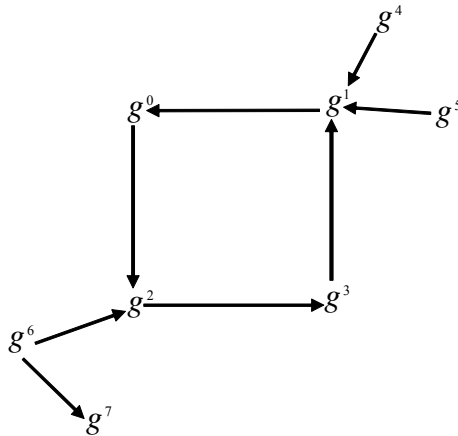


Figure 7: A supernetwork

<sup>23</sup>One could just as well define a supernetwork where a network is defeated by another network if and only if it is not pairwise Nash stable. In that case a network without outgoing arcs would be pairwise Nash stable.

The possibility of a cycle in the supernetwork is demonstrated in more detail in Panel (a) of Figure 8.<sup>24</sup> The example is due to Jackson and Watts (2002), who also provide more details. Deviations by the four individual players and the pairs of players lead to a cycle. At  $g^0$ , players 2 and 3 benefit from forming a link. So  $g^0$  is defeated by  $g^1$ . There, player 3 has an incentive to sever his link to player 4, so  $g^2$  defeats  $g^1$ . Network  $g^3$  defeats  $g^2$  because player 2 has an incentive to sever her link to player 3. Finally, at  $g^3$  players 3 and 4 benefit from forming a link. If they do so, the process is back to  $g^0$ , closing the cycle.

Such cycles are symptomatic not only for the nonexistence of pairwise Nash stable networks. For example, when a pure strategy equilibrium of a finite noncooperative game does not exist, there will be a cycle of outcomes such that each outcome in the cycle can be reached from the previous outcome through a unilateral deviation that is beneficial for the deviating player. The standard example to illustrate this point is the game of Matching Pennies. Similarly, if the core of a finite cooperative game is empty, there will be a cycle of outcomes that are linked through coalitional deviations that are beneficial for the deviating coalition.

---

<sup>24</sup>The significance of Panel (b) will be explained later.

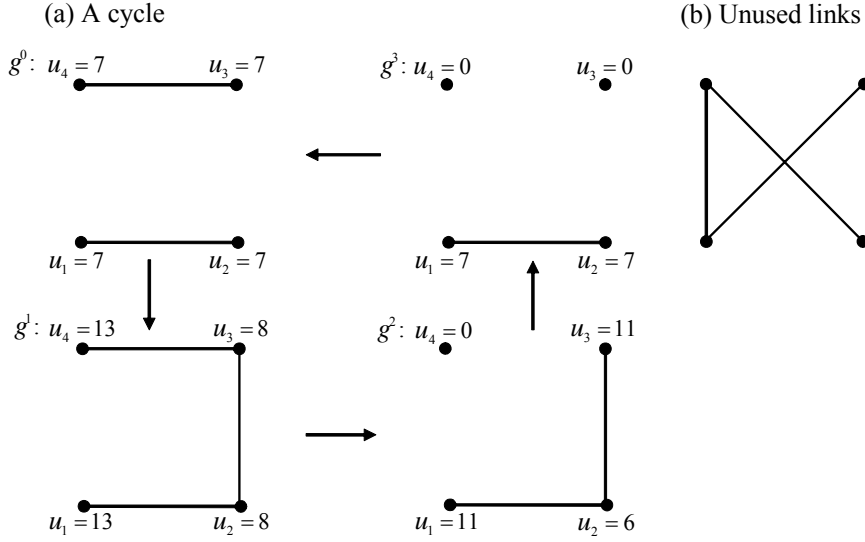


Figure 8: A cycle in a supernetwork

In noncooperative games the problem of existence is resolved by reverting to mixed strategies. Mixed strategies are probability distributions over sets of pure strategies. However, since no well-defined strategic form network formation game underlies the concept of weak pairwise Nash stability, it is not clear what a mixed strategy means in this context.<sup>25</sup> In the following, I define such a strategic form network formation game. With that game at hand, a natural way to introduce mixed strategies is via the mixed extension of that network formation game.

Formally, the linking game with player pairs is a noncooperative game with players acting in their own interest. In fact though, the game incorporates cooperation. This is done by treating pairs of players as players of the game, with strategies that are independent from the individual players' strategies. Each pair indicates whether it wishes to form a link. The game is designed so that a pair's indication of wanting to form a link has an effect on the outcome only when the corresponding individual players do not indicate each other. In addition, a pair's payoff is specified so that a

<sup>25</sup>Bloch and Jackson (2006) define a pairwise Nash equilibrium as the Nash equilibrium profile  $s$  of the linking game that satisfies  $u_i(g(s) + ij) > u_i(g(s)) \Rightarrow u_j(g(s) + ij) < u_j(g(s))$ . This, however, does not permit the introduction of mixed strategies either.

pair wishes to add their link if and only if both players benefit from it. This design eliminates the possibility of two players failing to form a link even though both would benefit from it.

### **The linking game with player-pairs**

The set of players is  $N \cup \{ij : i, j \in N, i \neq j\}$ .<sup>26</sup> The set of pure strategies available to each  $i \in N$  is the same as in the linking game,  $S_i = P(N \setminus \{i\})$ . A pair  $ij$  has only two pure strategies,  $S_{ij} = \{Y, N\}$ , with the interpretation that  $s_{ij} = Y$  indicates the pair  $ij$  wants to form the link  $ij$  and  $s_{ij} = N$  that it does not want to form the link. A strategy profile is denoted by  $(s^s, s^p)$  where  $s^s$  is the subprofile for individual players and  $s^p$  is the subprofile for pairs. Whether or not the link  $ij$  is formed for a given strategy profile depends on  $s_i$ ,  $s_j$ , and  $s_{ij}$ . Table 1 lists the possible combinations of  $s_i$ ,  $s_j$ , and  $s_{ij}$ , showing that a pair's strategy only affects the outcome if neither of the individual players indicates the other player. Note also that, at profiles for which the link is formed, individual players always have the power to cut the link. In Cases 1 and 2, the individual player can switch to not indicating the other player. In Case 7, he could switch to indicating the other player. This last feature might seem somewhat paradoxical, but is necessary to avoid an equilibrium in which a player would like to sever several of his connections.

---

<sup>26</sup>Compared to the linking game, there are  $n(n-1)/2$  additional players, one for each pair of players.

Table 1: Combinations of  $s_i$ ,  $s_j$ , and  $s_{ij}$

Case	$s_i$	$s_j$	$s_{ij}$	Outcome
1)	$j \in s_i$	$i \in s_j$	$Y$	$ij \in g$
2)	$j \in s_i$	$i \in s_j$	$N$	$ij \in g$
3)	$j \notin s_i$	$i \in s_j$	$Y$	$ij \notin g$
4)	$j \notin s_i$	$i \in s_j$	$N$	$ij \notin g$
5)	$j \in s_i$	$i \notin s_j$	$Y$	$ij \notin g$
6)	$j \in s_i$	$i \notin s_j$	$N$	$ij \notin g$
7)	$j \notin s_i$	$i \notin s_j$	$Y$	$ij \in g$
8)	$j \notin s_i$	$i \notin s_j$	$N$	$ij \notin g$

The individual players' payoffs are the same as in the original game. The payoff for a pair  $ij$  is

$$u_{ij}(g) = \begin{cases} 0 & \text{if } ij \notin g \\ \min \{u_i(g) - u_i(g - ij), u_j(g) - u_j(g - ij)\} & \text{if } ij \in g. \end{cases}$$

Thus, if the current network is  $g - ij$  and both players benefit from the addition of  $ij$ , then  $u_{ij}(g) > u_{ij}(g - ij) = 0$ . Whenever possible, the pair would then switch to a strategy that induces the link in the network.<sup>27</sup>

If  $g$  is supported by a Nash equilibrium of the game with player pairs, I will say that  $g$  is *Nash stable for the linking game with player pairs*.

**Proposition 16** *A network is Nash stable for the linking game with player pairs if and only if it is weakly pairwise Nash stable.*

---

<sup>27</sup>Note that, to induce the deviation, the payoff from  $u_{ij}(g)$  has to be strictly positive, which requires that each of the individual players benefit from adding the link. This feature is what makes it necessary to use the weak notion of pairwise stability to obtain the result in Proposition 16.



(1) Let  $g$  be a Nash stable network for the game with player pairs with  $(s^s, s^p)$  being the supporting Nash equilibrium strategy profile. The proof of this part proceeds by first showing that  $g$  must be Nash stable, and then that  $g$  must be weakly pairwise stable.

(i) Nash stability

Let  $s'$  be a strategy profile for the linking game in which, for all  $i, j \in N$ ,  $j \in s'_i$  if and only if  $ij \in g$ , that is  $s'$  supports  $g$ . No player can gain by indicating additional players because this would not change the outcome. Moreover, since in the linking game with player pairs, every individual player can sever any number of his links in  $g(s^s, s^p)$  by changing his strategy, and since  $(s^s, s^p)$  is a Nash equilibrium, severing links cannot be beneficial.<sup>28</sup> Thus  $g$  is Nash stable.

(ii) Weak pairwise stability

To see that  $g$  must be weakly pairwise stable, note that the first requirement of weak pairwise stability,  $u_i(g) \geq u_i(g - ij)$  for all  $ij \in g$ , is implied by the already established fact that  $g$  is Nash stable. Next, consider some  $ij \notin g$  (if any), and suppose that  $u_i(g + ij) > u_i(g)$  and that  $u_j(g + ij) > u_j(g)$ , contrary to the requirement of weak pairwise stability. In this case,  $u_{ij}(g + ij) = \min \{u_i(g + ij) - u_i(g), u_j(g + ij) - u_j(g)\} > 0$ . There are two kinds of cases, in which  $ij \notin g(s^s, s^p)$ . Each case leads to a contradiction:

(1)  $j \notin s_i$ ,  $i \notin s_j$ , and  $s_{ij} = N$ , a contradiction because  $ij$  can deviate to  $s_{ij} = Y$  and increase the pair's payoff from 0 to  $\min \{u_i(g + ij) - u_i(g), u_j(g + ij) - u_j(g)\} > 0$ .

(2)  $j \notin s_i$ ,  $i \in s_j$ , and  $s_{ij} = N$  or  $s_{ij} = Y$ , a contradiction because  $i$  can deviate to  $j \in s_i$  and induce the network  $g + ij$  (the same holds true for the cases  $j \in s_i$ ,  $i \notin s_j$ , and  $s_{ij} = N$  or  $s_{ij} = Y$ ).<sup>29</sup>

---

<sup>28</sup>If player  $i$  wishes to sever a link to player  $j$ , he can, if  $j \in s_i$  and  $i \in s_j$ , simply change to a strategy  $s'_i$  such that  $j \notin s_i$ , and if  $j \notin s_i$ ,  $i \notin s_j$ , and  $s_{ij} = Y$ , to  $s'_i$  such that  $j \in s'_i$ .

<sup>29</sup>The case  $j \in s_i$ ,  $i \notin s_j$ , and  $s_{ij} = N$  is the one that requires a weakening of pairwise stability. Here,  $j$  has to deviate in order to have the link added to the network. However, if one allows  $j$  to be indifferent he would not necessarily want to deviate, not yielding the desired contradiction.

Thus,  $g$  is Nash stable and weakly pairwise stable, implying that  $g$  is weakly pairwise Nash stable.

(2) Conversely, suppose that  $g$  is weakly pairwise Nash stable. Let  $(s^s, s^p)$  be a strategy profile for the linking game with player pairs, in which  $j \in s_i, i \in s_j$ , and  $s_{ij} = Y$  for all  $i, j$  such that  $ij \in g$ , and  $j \notin s_i, i \notin s_j$ , and  $s_{ij} = N$  for all  $i, j$  such that  $ij \notin g$ . The profile  $(s^s, s^p)$  supports  $g$ . By deviating, an individual player  $i$  can only sever links, which is not beneficial for him because  $g$  is Nash stable. The deviation of a pair of players  $ij$  can only lead to the addition of a link, which is not beneficial both players in the pair because  $g$  is weakly pairwise stable. Thus, the profile  $(s^s, s^p)$  is a Nash equilibrium, showing that  $g$  is Nash stable for the linking game with player pairs.

Proposition 16 shows that a weakly pairwise Nash stable network can be viewed as the outcome of a "noncooperative" game. Because this game is finite, a Nash equilibrium of its mixed extension exists. The following example illustrates such a mixed strategy equilibrium. The example is based on the one depicted in Figure 8.

### An Example

To simplify the analysis, let the payoffs for networks containing any of the links, 14, 13, or 24 (the "unused" links in Panel (b) of Figure 8) be as follows. For all  $i \in N$ , if  $g$  contains all three of these links,  $u_i(g) = -3$ , if  $g$  contains exactly two of these links,  $u_i(g) = -2$ , and if  $g$  contains exactly one such link,  $u_i(g) = -1$ . All other networks, except for the ones depicted in Panel (a), yield a payoff of zero to the individual players. For this specification of payoffs, no network is pairwise stable.

Let us further simplify the analysis by eliminating weakly dominated strategies. It is clear that for any player  $i$ , if  $s_i$  indicates the willingness to form any of the links 14, 13, or 24, then  $s_i$  is weakly dominated by  $s'_i = s_i \setminus \{14, 13, 24\}$ . Similarly, for every player pair  $ij = 14, 13, 24$ , the strategy  $s_{ij} = N$  weakly dominates  $s_{ij} = Y$ . Furthermore, note that for player 3, a strategy  $s_3$  involving  $2 \notin s_3$  is weakly dominated

by a strategy  $s'_3 = s_3 \cup \{2\}$ . Similarly, a strategy for player 4 with  $3 \notin s_4$  is weakly dominated by  $s'_4 = s_4 \cup \{3\}$ , and a strategy for player 1 with  $2 \notin s_1$  is weakly dominated by  $s'_1 = s_1 \cup \{2\}$ . Lastly, for player 2, the strategy  $s_2$  such that  $3 \in s_2$  but  $1 \notin s_2$  is weakly dominated by a strategy  $s'_2$  with  $3 \in s'_2$  and  $1 \in s'_2$ , and for player 3 the strategy  $s_3$  such that  $4 \in s_3$  but  $2 \notin s_3$  is weakly dominated by a strategy  $s'_3$  with  $4 \in s'_3$  and  $2 \in s'_3$ . Hence, in a Nash equilibrium in undominated strategies, the following strategies are played with probability one:  $s_1 = \{2\}$ ,  $s_4 = \{3\}$ , and  $s_{13} = s_{14} = s_{24} = N$ . In addition, for players 2 and 3, the only strategies that are not weakly dominated are  $s_2 = \{1\}$ ,  $s'_2 = \{1, 3\}$ ,  $s_3 = \{2\}$ , and  $s'_3 = \{2, 4\}$ .

Note that every combination of the individual players' undominated pure strategies is such that the strategies of the pairs 12, 23, or 34 will not affect the outcome (recall that a pair  $ij$  only affects the outcome if  $i \notin s_j$  and  $j \notin s_i$ ). Therefore, if  $\rho_{ij}$  denotes the probability for  $s_{ij} = Y$ , then any  $\rho_{ij} \in [0, 1]$ , for  $ij = 12, 23$ , or  $34$ , supports the equilibrium. Let  $\alpha$  denote the probability that player 2 chooses  $s_2$  and let  $\beta$  denote the probability that player 3 attaches to  $s_3$ . In a mixed strategy equilibrium,  $\alpha$  and  $\beta$  must solve the following optimization problems for players 2 and 3.

Player 2:

$$\max_{\alpha} \alpha (\beta 7 + (1 - \beta) 7) + (1 - \alpha) (\beta 6 + (1 - \beta) 8)$$

*Foc* (interior solution)

$$\beta 7 + (1 - \beta) 7 = \beta 6 + (1 - \beta) 8 \Leftrightarrow \beta^* = \frac{1}{2}.$$

Player 3:

$$\max_{\beta} \beta (\alpha 0 + (1 - \alpha) 11) + (1 - \beta) (\alpha 7 + (1 - \alpha) 8)$$

*Foc* (interior solution)

$$(1 - \alpha) 11 = \alpha 7 + (1 - \alpha) 8 \Leftrightarrow \alpha^* = \frac{3}{10}$$

In summary, the set of equilibrium strategy profiles in undominated strategies for the linking game with player pairs is as follows:

$$\begin{aligned}
s_1 &= \{2\}; \\
s_2 &= \{1\} \text{ and } s'_2 = \{1, 3\} \text{ with probability } \frac{1}{2} \text{ each}; \\
s_3 &= \{2\} \text{ with probability } \frac{3}{10}, s'_3 = \{2, 4\} \text{ with probability } \frac{7}{10}; \\
s_4 &= \{3\}; \\
s_{ij} &= N, \text{ for } ij = 13, 14, \text{ and } 24; \\
s_{ij} &= Y \text{ with probability } \rho_{ij} \in [0, 1], \text{ for } ij = 12, 23, \text{ and } 34.
\end{aligned}$$

Only the four networks in Panel (a) occur with positive probability. Under the induced distribution,  $g^0$ ,  $g^1$ ,  $g^2$ , and  $g^3$  occur with probabilities  $\frac{7}{20}$ ,  $\frac{7}{20}$ ,  $\frac{3}{20}$ , and  $\frac{3}{20}$ , respectively.

It is worth noting that there are other specifications for the player pairs' payoffs that yield the result in Proposition 16. In particular, any specification such that  $u_{ij}(g + ij) > u_{ij}(g)$  if and only if  $u_i(g + ij) > u_i(g)$  and  $u_j(g + ij) > u_j(g)$  yields the same set of pure strategy equilibria of the linking game with player pairs. This fact might lead to the concern that other payoff specifications result yield different mixed strategy equilibrium outcomes. For affine transformations of the payoff function used here, a standard result from expected utility theory guarantees that the set of mixed strategy equilibria remains the same. All other kinds of transformations (that preserve the above condition), might lead to different mixed strategy equilibria. However, I believe that the specification chosen here is the most natural one. Moreover, it can be applied to any set of preferences, while other specifications will likely work only for a subset of preferences.

As already discussed, in the supernetwork  $\mathbf{G}$ , the nonexistence of a weakly pairwise Nash stable network implies at least one cycle with no outgoing arc to a network outside the cycle. Such cycles of networks have been labeled basins of attraction (Page

and Wooders, 2008).<sup>30</sup> Formally, a set of networks  $A \subseteq G$  is a *basin of attraction* if

- (i)  $g \in A, g' \notin A$  implies that  $(g, g') \notin \mathbf{G}$ , and
- (ii)  $g, g' \in A, g \neq g'$ , implies that  $\exists g_0, g_1, \dots, g_L \in G$  such that  $(g_l g_{l+1}) \in \mathbf{G}$  for  $l = 0, 1, \dots, L - 1$ , where  $g_0 = g$  and  $g_L = g'$ .

Basins of attractions are "absorbing states." Once a basin is reached, the process remains within the basin. Weakly pairwise Nash stable networks are precisely the networks that belong to a singleton basin of attraction. Let  $\mathcal{A}$  be the collection of basins of attraction for the supernetwork  $\mathbf{G}$ . A natural conjecture is that networks that are supported by a mixed strategy equilibrium of the linking game with player pairs must belong to a basin of attraction, or, more formally: If  $g$  occurs with positive probability under some mixed strategy Nash equilibrium of the linking game with player pairs, then  $g \in A$  for some  $A \in \mathcal{A}$ .

However, the following example refutes this conjecture. Let  $N = \{1, 2, 3, 4\}$  and consider the following four networks, depicted in Figure 9:

$$\begin{aligned} g^0 &= \{14, 23\} \\ g^1 &= \{14, 23, 12\} \\ g^2 &= \{14, 23, 34\} \\ g^3 &= \{14, 23, 12, 34\}. \end{aligned}$$

---

<sup>30</sup>The term basin of attraction is used for similar concepts in mathematics and the sciences. Page and Wooders (2008) have introduced it to the context of strategic network formation.

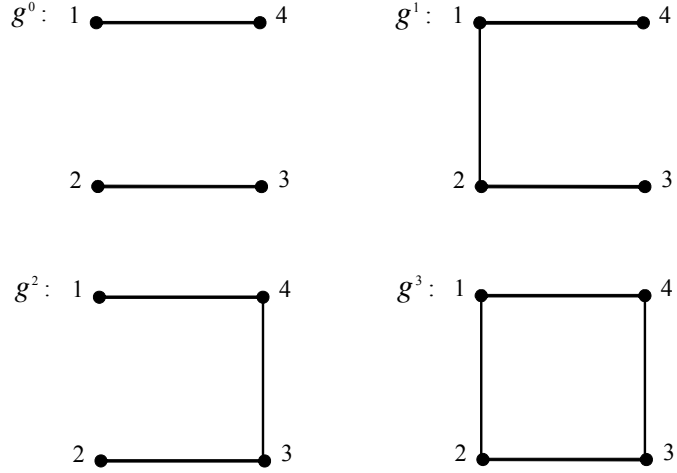


Figure 9: Four networks

Suppose that the individual players' preferences over networks are represented by the following utility functions

$$u_1(g^0) = u_1(g^1) = 1 \text{ and } u_1(g) = -1 \text{ for } g \in G \setminus \{g^0, g^1\};$$

$$u_2(g^0) = u_2(g^1) = 1 \text{ and } u_2(g) = -1 \text{ for } g \in G \setminus \{g^0, g^1\};$$

$$u_3(g^0) = 0, u_3(g^2) = u_3(g^3) = 1, u_3(g^1) = 2, \text{ and}$$

$$u_3(g) = -1 \text{ for } g \in G \setminus \{g^0, g^1, g^2, g^3\};$$

$$u_4(g^0) = 0, u_4(g^2) = u_4(g^3) = 1, u_4(g^1) = 2, \text{ and}$$

$$u_4(g) = -1 \text{ for } g \in G \setminus \{g^0, g^1, g^2, g^3\},$$

implying utilities for player pairs:

$$\begin{aligned}
 u_{14}(g^0) &= 1, u_{14}(g^1) = 2, \text{ and } u_{14}(g) = 0 \text{ for } g \in G \setminus \{g^0, g^1\}; \\
 u_{23}(g^0) &= 1, u_{23}(g^1) = 2, \text{ and } u_{23}(g) = 0 \text{ for } g \in G \setminus \{g^0, g^1\}; \\
 u_{12}(g) &= 0 \text{ for } g \in G; \\
 u_{34}(g^2) &= 1, u_{34}(g^3) = -1, \text{ and } u_{34}(g) = 0 \text{ for } g \in G \setminus \{g^2, g^3\}; \\
 u_{13}(g) &\leq 0 \text{ for } g \in G; \\
 u_{24}(g) &\leq 0 \text{ for } g \in G.
 \end{aligned}$$

The supernetwork  $\mathbf{G}$  restricted to the networks in  $\{g^0, g^1, g^2, g^3\}$  contains only two arcs:  $(g^0, g^2)$  and  $(g^3, g^1)$ . Moreover, for any  $g \in \{g^0, g^1, g^2, g^3\}$  and  $g' \in G \setminus \{g^0, g^1, g^2, g^3\}$ , we have  $(g, g') \notin \mathbf{G}$ , as illustrated in Figure 10. Therefore, networks  $g^1$  and  $g^2$  each constitute a singleton basin of attraction while  $g^3$  does not belong to any basin of attraction because it is defeated by  $g^1$ .

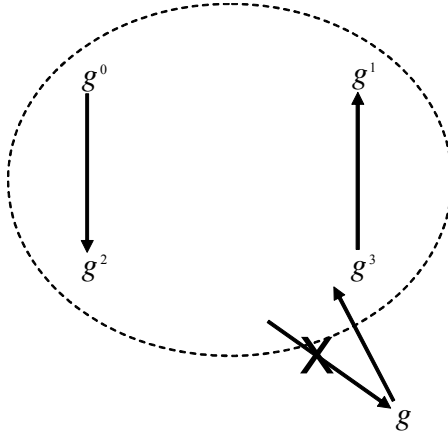


Figure 10: Network  $g^3$  does not belong to a singleton basin of attraction

It is straightforward (though somewhat tedious) to verify that a strategy profile satisfying the following conditions constitutes a mixed strategy Nash equilibrium of

the linking game with player pairs:

$$\begin{aligned}
 s_1 &= \{4\}; \\
 s_2 &= \{3\}; \\
 s_3 &= \{2, 4\}; \\
 s_4 &= \{1, 3\}; \\
 s_{12} &= Y \text{ and } s_{12} = N \text{ with probability } \frac{1}{2} \text{ each}; \\
 s_{13} &= N; \\
 s_{24} &= N.^{31}
 \end{aligned}$$

No player (individual or pair) has an incentive to deviate. Under such a profile networks  $g^2$  and  $g^3$  occur with equal probability  $\frac{1}{2}$  each. Since  $g^3$  does not belong to any basin of attraction, the example shows that the above conjecture is false.

### The Two-Stage Linking Game

This section models explicitly how coordination between two players might arise in a noncooperative two-stage game. In the first stage, players play the linking game. A stage-one link constitutes a coordination possibility in stage two. Given a network of coordination possibilities formed in the first stage, players play a constrained version of the linking game with player pairs in the second stage. In the constrained version, only pairs that have formed a link in the first stage can coordinate on whether they form a link. Thus, if the network of coordination possibilities formed in the first stage is given by  $h \subseteq \{ij : i, j \in N, i \neq j\}$ , then the set of players in the constrained linking game with player pairs is  $N \cup h$ . The corresponding constrained game is denoted by  $\hat{\Gamma}(h)$ .

Intuitively, the two stages capture the idea that coordination as assumed by pairwise stability can arise between two individuals only if they have some sort of con-



nection to each other; therefore the terminology of a coordination possibility. For example, to form friendships, first people become acquainted with each other (through work, by joining a club, because they have common friends, etc.). Then they might form a closer friendship, which, arguably, will happen if and only if both sides wish to establish that friendship.

I show that any network that can be supported by a Nash equilibrium of the linking game can also be supported by a subgame perfect Nash equilibrium of the two-stage linking game. However, if this network is not weakly pairwise stable, then the subgame perfect Nash equilibrium that supports it will have at least one player playing a weakly dominated strategy. Thus, a network supported by an undominated subgame perfect Nash equilibrium of the two-stage linking game, must be weakly pairwise stable.

As in the linking game with player pairs, the set of players in the two-stage linking game is  $N \cup \{ij : i, j \in N, i \neq j\}$ . A pure strategy for player  $i$  is denoted by  $t_i$  and a pure strategy for a player pair  $ij$  is denoted by  $q_{ij}$ . A strategy  $t_i$  for player  $i$  consists of a set  $r_i \in R_i = P(N \setminus \{i\})$  and a function  $f_i : G \rightarrow R_i$ . The set  $r_i$  indicates with whom player  $i$  would like to form a coordination possibility. The function  $f_i$  indicates, for each first stage "network" formed in the first stage, the set of players player  $i$  wishes to form a link with. Let  $G_{ij} = \{g : ij \in g\}$ , that is  $G_{ij}$  is the set of networks in which players  $i$  and  $j$  are directly linked to each other. A strategy  $q_{ij}$  for player pair  $ij$  is a function  $f_{ij} : G_{ij} \rightarrow \{Y, N\}$ . The corresponding mixed strategies for a player (individual or pair) are probability distributions over the player's set of pure strategies. For ease of notation, mixed strategies are not formally introduced. A profile of strategies and sets of strategies are denoted in the usual way.

**Proposition 17** *A network  $\hat{g}$  is Nash stable if and only if it is supported by a subgame perfect Nash equilibrium of the two-stage linking game.*

**Proof.**

(1) Let  $s$  be a Nash equilibrium of the linking game that supports  $\hat{g}$ . The proof proceeds by constructing a profile  $(t, q) = ((r_i, f_i)_{i \in N}, q)$  that is subgame perfect and supports  $\hat{g}$ . Let  $(t, q)$  be as follows.

Stage 1:

For all  $i, j \in N$ , let  $j \in r_i$  if and only if  $ij \in \hat{g}$ .

Stage 2:

(a) Subgame  $\hat{g}$ : For all  $i \in N$ , let  $f_i(\hat{g}) = r_i$ . For all  $ij \in \hat{g}$ , let  $f_{ij}(\hat{g}) = N$ .

Let  $B \subseteq \{ik : ik \in \hat{g}\}$ .

(b) Subgames  $g = \hat{g} \setminus B$ : For all  $i \in N$ , let  $f_i(g) = r_i$ . For all  $ij \in g$ , let  $f_{ij}(g) = N$ .

(c) All other subgames  $g$ : Pick a Nash equilibrium of the constrained linking game with player pairs  $\hat{\Gamma}(g)$ . (Note that this might be a Nash equilibrium in mixed strategies).

The outcome of this profile is  $\hat{g}$ . It remains to verify that the profile is a subgame perfect equilibrium. Because  $\hat{g}$  is Nash stable, the play induced on  $\hat{\Gamma}(\hat{g})$  is a Nash equilibrium: No individual player wants to deviate, and since player pairs only have the power to add links but not to delete them and there is no player pair in  $\hat{\Gamma}(\hat{g})$  whose link is not in  $\hat{g}$ , no pair can deviate either. The same logic applies to the play induced on subgames in (b). Lastly, the profile induces a Nash equilibrium on subgames in (c). Next, we need to verify that the specified profile also induces a Nash equilibrium on the entire game. It was already shown that no player can benefit from a deviation in any of the subgames. A unilateral deviation by a single player in stage 1 can only lead to a subgame of type (b). However, the Nash equilibrium outcome  $(t, q)$  induces on these subgames is the same as the outcome of subgame  $\hat{g}$ . Thus, the profile is a Nash equilibrium for the entire game.

(2) Conversely, let  $\hat{g}$  be supported by a subgame perfect Nash equilibrium of the two-stage linking game. Let  $g'$  be the set of potential links formed in stage 1. Because  $\hat{g}$

is the outcome of a subgame perfect equilibrium, it is supported by a Nash equilibrium of the constrained linking game with player pairs  $\Gamma(g')$ . Therefore no individual player wishes to delete any subset of his links in  $\hat{g}$ , showing that  $\hat{g}$  is Nash stable. ■

Proposition 17 shows that, even when individuals can form links to foster cooperation, the result can be a complete coordination failure. For example, as in the linking game, the empty network is trivially supported by a subgame perfect Nash equilibrium of the two-stage linking game. However, Proposition 18 shows that if a subgame perfect equilibrium of the two-stage linking game supports a network that is not weakly pairwise Nash stable, at least one player plays a weakly dominated strategy. In other words, any undominated subgame perfect equilibrium of the two-stage linking game will only support weakly pairwise Nash stable outcomes. This is not true for the linking game. Jackson (2008, p. 374) provides an example of a network which is supported by a Nash equilibrium in undominated strategies of the linking game but is not weakly pairwise Nash stable.

**Proposition 18** *If a network is supported by a subgame perfect equilibrium in undominated strategies of the two-stage linking game then it is weakly pairwise Nash stable.*

**Proof.** I prove the contrapositive. Let  $\hat{g}$  be a network that is not weakly pairwise Nash stable. If  $\hat{g}$  is not Nash stable, Proposition 17 implies that it cannot be supported by a subgame perfect equilibrium of the two-stage linking game. If  $\hat{g}$  is not weakly pairwise stable, pick a pair  $\{i, j\}$  such that  $ij \notin \hat{g}$  but  $u_i(\hat{g}+ij) > u_i(\hat{g})$  and  $u_j(\hat{g}+ij) > u_j(\hat{g})$ . Let  $(t, q)$  be a subgame perfect equilibrium profile of the two-stage linking game that supports  $\hat{g}$ , and let  $h$  be the first stage links formed under the profile  $(t, q)$ . Since  $(t, q)$  supports  $\hat{g}$ , it must hold that  $ij \notin h$  for otherwise  $\hat{g}$  could not be supported by a Nash equilibrium of the constrained game  $\hat{\Gamma}(h)$  (a deviation by  $ij$  would make sure that the link  $ij$  was formed). So either  $j \notin r_i$  or  $i \notin r_j$ . Without loss of generality, suppose that  $j \notin r_i$ . Now, consider the following strategy  $t'_i = (r'_i, f'_i)$ .

Let  $t'_i$  be a strategy for player  $i$  that coincides with  $t_i$  except that  $j \in r'_i$  and that  $f'_i(\hat{g} + ij) = f_i(\hat{g})$ . To show that  $t'_i$  weakly dominates  $t_i$ , we need to show that there exists at least one play of all other players, say  $(t'_{-i}, q')$  to which  $t'_i$  is a better response than  $t_i$ , and so that  $t'_i$  yields at least as much as  $t_i$  to every other play by the remaining players.

Let  $(t'_{-i}, q')$  coincide with  $(t_{-i}, q)$  except that  $i \in r'_j$ ,  $f'_j(\hat{g} + ij) = f_j(\hat{g})$ , and  $f'_{ij}(\hat{g} + ij) = Y$ . The outcome of the profile  $(t_i, t'_{-i}, q')$  is  $\hat{g}$ , while the outcome of the profile  $(t'_i, t'_{-i}, q')$  is  $\hat{g} + ij$ . Thus  $t'_i$  is a better response to  $(t'_{-i}, q')$  than  $t_i$ .

Now, let  $(t'_{-i}, q')$  be an arbitrary play. If  $t'_j$  is such that  $i \notin r'_j$ , then the same subgame is reached under  $(t_i, t'_{-i}, q')$  and  $(t'_i, t'_{-i}, q')$ , and it is not the subgame  $\hat{g} + ij$ . On all these subgames the play induced by the two strategy profiles coincides, and therefore  $i$  obtains the same utility from both profiles. If  $t'_j$  is such that  $i \in r'_j$ , then  $\hat{g} + ij$  is reached under  $(t'_i, t'_{-i}, q')$  and  $\hat{g}$  is reached under  $(t_i, t'_{-i}, q')$ . In this case, again, the two profiles lead to the same outcome because  $f'_i(\hat{g} + ij) = f_i(\hat{g})$ .

Thus  $t'_i$  weakly dominates  $t_i$ , which is what we wanted to show. ■

## Conclusion

This paper demonstrates that weakly pairwise Nash stable networks are the equilibrium networks a modified version of Myerson's linking game, in which pairs of players are treated as additional players. Two features of the game bring about the type of coordination required by weak pairwise stability. First, pairs have the discretion to add links whenever individual players fail to coordinate on the formation of a link. Second, a pair's payoff is designed so that the pair benefits from the addition of the link if and only if both players benefit from it. My characterization allows to introduce mixed strategies to coordinated moves.

The paper also defines a two-stage game of network formation which provides a natural model of how coordination can arise in a noncooperative setting. The

equilibrium of the two-stage game provides a rationale for networks in which players are not linked even though both would benefit from being linked. In the game, such networks can arise in equilibrium if players fail to form coordination links in the first stage. In a real-world network, such networks might be stable simply because two players do not know enough about each other (or do not know each other at all).

The nonexistence of solutions to games with cooperative elements has been studied extensively in the literature with a focus on finding conditions when such a solution exists. In all these games, reverting to mixed strategies is infeasible because they have no well-defined strategic form. To resolve this, it might be feasible to define equivalent strategic form noncooperative games, using methods similar to the ones used in this paper.

## CHAPTER IV

### COMPETITION OVER STANDARDS AND TAXES

#### Introduction

The recent integration of countries in Eastern Europe to the European Union (EU) has provoked renewed concern about the aggressive competition by new members for firms and other mobile factors.<sup>32</sup> To investigate this issue, our paper develops a model of international competition over standards and taxes. By a ‘standard’ we have in mind such things as labor regulations, pollution control and property rights enforcement. Firms who locate in a country are required to pay taxes which are used, at least in part, to enforce the standard in that country. The main purpose of this paper is to show that, through competition in standards and taxes, a developing/transition country may indeed have a ‘second-mover advantage’ over a developed country in attracting firms and extracting rents. While this concern has circulated in policy discussions for some time now, to our knowledge it has not been studied formally before in the literature on fiscal competition.<sup>33</sup>

Although often modeled as a type of local public good, standards have an important distinguishing feature. A reasonable assumption in the context of most public goods is that (for a given tax outlay) all firms at least weakly prefer a higher level of public good provision. On the other hand all firms do not unanimously prefer higher

---

<sup>32</sup>For example, although EU accession requirements demand moves towards harmonization of environmental standards and some measures have been put onto statute books, there appears to be widespread skepticism about the actual implementation of such measures. Citing the incentive not to raise standards in order to attract firms, Post (2002) states that ‘there is a “deception gap” between what is said on paper and what is done in practice’ with regard to environmental policy. Andonova (2003) provides extensive details of these environmental standards. Although environmental standards provide a good motivating example, our concern will be with standards more broadly defined as we shall explain.

<sup>33</sup>This issue has been raised both with respect to developing countries and to countries from the former Soviet Union often referred to as ‘transition countries/economies.’ For brevity, throughout paper we will use ‘developing country’ as a catch-all term.

standard levels. For example, a high level of property rights enforcement may benefit a firm engaged in research and production of new pharmaceutical products while it may hurt a firm engaged in the mass-production of generic drugs. To an otherwise familiar model of fiscal competition, we introduce the assumption that firms have differing ideal standard levels. It is this assumption that gives rise to the second-mover advantage in standard setting that we identify.

As the discussion so far suggests, we model competition for mobile firms as a sequential game between governments who choose standards and taxes. Due to monitoring costs, the higher the standard set by a country the more costly it is to implement. Following a common hypothesis in the literature (with Niskanen 1977 as its source) national governments are run by bureaucrats who seek to maximize their budgets (tax revenue minus the cost of implementing the standard). There is a continuum of firms (while consumers are not explicitly considered). We refer to the difference between a firm's ideal standard level and the level actually set in a country as the 'standard mismatch' for that firm. A key parameter in the model is the 'marginal cost of standard mismatch' which parameterizes how a given standard mismatch affects a firm's costs of production. Each firm (being small and behaving non-strategically) chooses its location to maximize profits, taking as given the tax levels and its standard mismatches in the two countries.

Our simple framework yields a surprisingly rich set of equilibrium predictions which depend on the cost of standard mismatch. There are three possible outcomes. (1) If the cost of standard mismatch is low then tax competition leads to an efficient equilibrium outcome (as in Brennan and Buchanan's 1980 model of tax competition). (2) If the cost of standard mismatch is in an intermediate range then the developed country sets standards inefficiently high and the developing country becomes a standard haven; a place where firms that prefer a low standard locate in order to escape the high standard set in the developed country. It is especially interesting that inef-

ficiently high standards in the developed country arise in equilibrium purely through strategic interaction between governments in their competition for firms and not as a result of attempts by governments to win the favor of a voting public. (3) If the cost of standard mismatch is high then there is a race to the top; both governments set standards inefficiently high and, because countries are differentiated by their standard levels, the intensity of tax competition is reduced as well. The precise set of interactions that gives rise to these equilibrium outcomes will be described in due course.

Much of the literature that examines fiscal competition where the public good in question is a standard assumes that (for a given tax take) citizens at least weakly prefer higher standards and that the standard in question is environmental. As a result, insights from the literature on tax competition with local public goods extend naturally; see Wilson (1996) for a survey. Broadly, the literature can be categorized into three areas. The first category, following Tiebout (1956), focuses on situations where competition among independent governments is like competition among firms and enhances efficiency. The second category concerns the presence of a policy-failure that allows or induces governments to set taxes on capital too high. This in turn induces governments to try to offset the depressive effects of capital taxes on investment by setting environmental standards too lax; this outcome is popularly known as a ‘race to the bottom.’ See Oates and Schwab (1988) for further details, as well as a discussion by Wilson (1996) of Oates and Schwab plus the related literature. The third category considers situations in which there is strategic interaction, over standards and taxes, between governments and a small number of firms. See for examples Markusen et al. (1995) and Davies and Ellis (2007). In such settings, strategic interactions over the market power held by firms and the policy failures of governments are the source of inefficient policy choices.

Our model combines features of models from papers in the first two categories: on



the one hand competition between governments introduces efficiency enhancing incentives; on the other hand the broader environment in which these incentives operate is one of market - or policy - failures that preclude the attainment of a fully efficient equilibrium. As in the literature that follows Tiebout, governments in our model are rent (or profit) maximizing but are constrained by competition. For example, Fischel (1975) and White (1975) share with the present paper the assumption that there is variation over firms' preferences for standards. In contrast to our model, Fischel (1975) and White (1975) assume that individual firms can be targeted for transfers and there is 'free entry' of jurisdictions, none of which has sufficient market power to extract rents from firms. As a result, within such a setting, an efficient outcome can be demonstrated in which firms 'vote with their feet.' In our model firms cannot be targeted for transfers. Moreover, there is policy failure in the sense that once the levels of public goods – in our model, standards – are fixed they cannot be altered. Another difference is that we fix the number of countries (at two) which enables their rent-maximizing governments to make positive rents and thus allows inefficiencies to arise.<sup>34</sup>

Rent-maximizing governments are a source of policy inefficiency for Oates and Schwab (1988) as well. Again, if governments are able to earn rents from taxation of mobile resources (in their case, capital) then there is an incentive to simultaneously set standards inefficiently low. Other papers in the literature build on these basic features in various ways. Interestingly, although the source of excessive taxation put forward by Oates and Schwab is the same as ours, their outcome in terms of environmental standards is starkly different. In their setting the outcome is a race to the bottom; in our setting, if the marginal cost of standard mismatch is sufficiently

---

<sup>34</sup>Our focus is on national governments while in much of the literature on standards and tax competition governments preside over jurisdictions more broadly defined. The reason we interpret the context of our model as international is that the range of policy options under consideration is more limited than in a domestic or federal context. In particular, the feature of our model that transfers between jurisdictions are not allowed appears to mirror more closely an international setting.

large, the outcome is a race to the top.<sup>35</sup>

The remainder of this paper is organized as follows. Section 2 introduces the model, then defines strategies and the subgame perfect equilibrium. Section 3 solves for the efficient allocation. Section 4 presents the equilibrium outcome, which is defined in terms of the location decisions of firms and policies set by the developing country and the developed country respectively. Conclusions are drawn in Section 5.

## The Model

The governments of two countries, a developed country,  $L$  (for ‘leader’), and a developing country,  $F$  (for ‘follower’), compete over standard levels and taxes in their attempts to induce firms to locate in their respective countries. The governments are assumed to be rent maximizers. There is a set of firms, each of which is able to sell a single unit of a good. The production costs of a firm depend on the level of taxation and the level of the standard in the country where it locates. We will first specify the behavior of firms, and then we will turn to the governments. This is the natural sequence of exposition given that we solve for equilibrium using backwards induction.

### Firms

The world price of the unit that each firm sells is  $p$ , and each firm pays a private per-unit production cost,  $c$ .<sup>36</sup> The tax levied on the firm is  $\tau_L$  if it locates in  $L$  and  $\tau_F$  if it locates in  $F$ . The value  $s \in [0, 1]$  uniquely identifies a firm and its ideal standard level.<sup>37</sup> The *standard mismatch* for a firm  $s$  is given by the difference between  $s$  and the standard level actually set in the country where the firm locates. The impact of standard mismatch on production costs is parameterized by  $k$ ; we refer to  $k$  as the

---

<sup>35</sup>Wilson (1996) insightfully conjectures that, under certain parameterizations, it may be possible to show that Oates and Schwab’s framework could motivate a race to the top as well.

<sup>36</sup>To increase realism, the price that each firm receives for the good that it sells could be made to vary across firms without affecting the results.

<sup>37</sup>We choose the interval  $[0, 1]$  to simplify the exposition. The same qualitative results may be obtained using an arbitrary interval  $[a, b]$ .

*marginal cost of standard mismatch.* If we let the variables  $l_L, l_F \in [0, 1]$  denote the standard levels set by  $L$  and  $F$  respectively then we can express the profit function for firm  $s$  as follows:

$$\pi(s) \equiv \begin{cases} p - c - \tau_L - k|l_L - s| & \text{if the firm locates in } L; \\ p - c - \tau_F - k|l_F - s| & \text{if the firm locates in } F. \end{cases}$$

To focus the analysis on location decisions, it will be assumed throughout that  $p$  is sufficiently high to ensure that all firms make positive profits. Also,  $p$  will serve as an upper bound for the tax that a government can set.

A firm  $s$  makes equal profits in both countries if and only if

$$\tau_L + k|l_L - s| = \tau_F + k|l_F - s|$$

in which case the firm is indifferent between the two countries. If there is a single indifferent firm,  $\hat{s}$ , then it holds that  $\hat{s}$  lies between  $l_L$  and  $l_F$ . Solving for  $\hat{s}$  in this case we obtain:

$$\hat{s} \equiv \hat{s}(l_L, \tau_L, l_F, \tau_F) = \begin{cases} \frac{\tau_L - \tau_F}{2k} + \frac{l_L + l_F}{2} & \text{if } l_F < l_L \\ \frac{\tau_F - \tau_L}{2k} + \frac{l_L + l_F}{2} & \text{if } l_F > l_L. \end{cases}$$

Firm  $s$  may prefer one country, say  $F$ , in terms of the tax that it sets;  $\tau_F < \tau_L$ . But if  $L$ 's standard is sufficiently close to  $s$  (i.e.  $|s - l_L| < |s - l_F|$ ) then  $L$  can attract  $s$  to its country.<sup>38</sup> If there is more than one firm that is indifferent between the two countries, then it must hold that for any such firm  $s$ , either  $s \leq \min\{l_L, l_F\}$  or  $s \geq \max\{l_L, l_F\}$ . If all firms are indifferent, then  $\tau_L + kl_L = \tau_F + kl_F$ . If no firms are indifferent then clearly all firms locate in one country or the other. These cases

---

<sup>38</sup>Firms' location decisions and hence the sizes of the countries, in terms of the measure of firms in each country, are determined strictly by the interaction of policy choices with firms' preferences. Additional features could be introduced to make the model more realistic including, for example, infrastructure and an 'attachment to home' but this would obscure the effects we want to focus on.

are treated in the rent functions of the countries defined in Section 2.2.<sup>39</sup>

Three more assumptions are needed to obtain clear-cut solutions for firm locations:

- (1) Given taxes and standards, firms that are indifferent between the two countries locate in the country with the lower standard mismatch, i.e. firms care more about the persistence of an established standard level than the constancy of a given tax level;
- (2) If all firms are indifferent between the two countries, then half locate in one country and half locate in the other;
- (3) If a government has multiple best responses, it chooses the best response that maximizes its share of firms.

These assumptions will be discussed further when we derive equilibrium in Section

4. The location decisions of firms described above are illustrated in Figure 11.

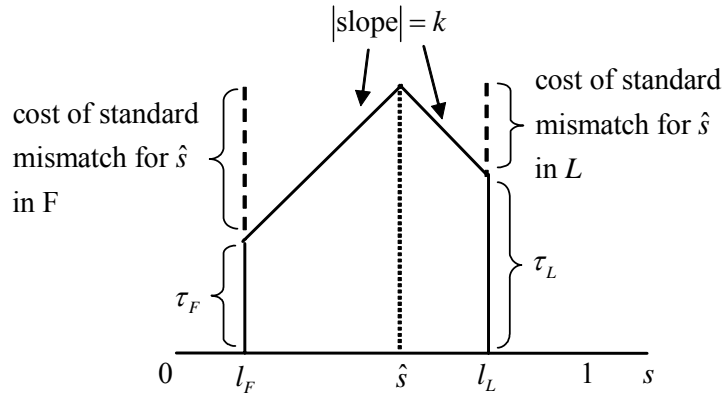


Figure 11: Firms' location decisions

Figure 11 is reminiscent of 'Hotelling's umbrella,' and reflects the underlying structure of our model which has some Hotelling features (see Hotelling 1929). The figure shows illustrative levels of standards and taxes set by governments  $F$  and  $L$ . For

<sup>39</sup>See the Appendix for additional details.

standards and taxes as shown, the point  $\hat{s}$  represents the ideal standard level of the indifferent firm  $\hat{s}$ . For  $\hat{s}$ , the absolute cost of standard mismatch is lower in  $L$ , but the tax in  $F$  is lower than the tax in  $L$ .

### **Governments**

Rents are given by tax revenues minus the cost of standard provision. A government's cost of enforcing a standard level  $l \in [0, 1]$  is  $l$  per firm that is located in its country. Thus the cost of enforcing a standard is assumed to be proportional to the level of the standard and the number of firms over which it must be enforced. Government  $F$  takes  $l_L$  and  $\tau_L$  as parameters and chooses  $l_F$  and  $\tau_F$  to maximize its rents. Discontinuities arise in the rent function at points where, given  $L$ 's strategy,  $F$ 's strategy is such that  $\hat{s} = l_F$  or  $\hat{s} = l_L$ , and additionally when  $l_F = l_L$  and  $\tau_F = \tau_L$ . Below is the rent function for  $F$ . The rent function for  $L$  is symmetric:

$$r_F(l_F, \tau_F; l_L, \tau_L) = \left\{ \begin{array}{ll} (\tau_F - l_F) \frac{1}{2} & \text{if } \tau_F = \tau_L \text{ and } l_F = l_L \quad \text{Case 1.} \\ (\tau_F - l_F) & \text{if } \tau_F < \tau_L - k |l_L - l_F| \quad \text{Case 2.} \\ (\tau_F - l_F) \hat{s} & \text{if } |\tau_F - \tau_L| \leq k(l_L - l_F) \\ & \text{and } l_F < l_L \quad \text{Case 3.} \\ (\tau_F - l_F)(1 - \hat{s}) & \text{if } |\tau_F - \tau_L| \leq k(l_F - l_L) \\ & \text{and } l_F > l_L \quad \text{Case 4.} \\ 0 & \text{if } \tau_F > \tau_L + k |l_L - l_F|. \quad \text{Case 5.} \end{array} \right.$$

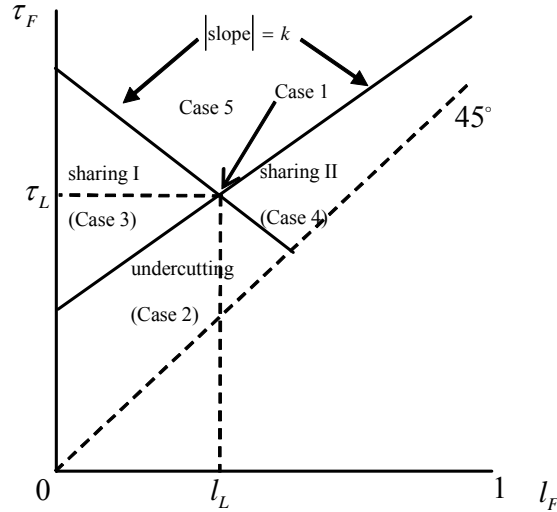


Figure 12: Strategy sets for Government  $F$

Figure 12 depicts the sets in the strategy space of  $F$  corresponding to the different cases of  $r_F(\cdot)$ . Case 1 arises when both governments choose the same standard and tax levels. By assumption, half of the firms then locate in  $F$ . In Case 2, which we will refer to as *undercutting*, the combination of standard levels and taxes induces all firms to locate in  $F$ . Cases 3 and 4 arise when strategies result in a positive fraction

of firms locating in each of the countries, with  $F$  setting a lower standard than  $L$  in Case 3 and a higher standard than  $L$  in Case 4. We will refer to these third and fourth cases, where firms are shared between the two countries, as *sharing I* and *sharing II*. Finally, Case 5 arises when  $F$  chooses its strategy so that it attracts no firms.

### Efficiency

Within the context of our model, an allocation is efficient if it maximizes the aggregate surplus realized by firms plus the governments' rents. An allocation consists of two ES levels and an assignment of firms to countries, denoted by  $(l_F, l_L, \hat{s})$ . Formally, the allocation  $(l_F, l_L, \hat{s})$  is efficient if it solves<sup>40</sup>

$$\begin{aligned} & \max_{\{l_F, l_L, \hat{s}\}} \int_0^{\hat{s}} (p - c - \tau_F - k |l_F - s|) ds + (\tau_F - l_F) \hat{s} \\ & + \int_{\hat{s}}^1 (p - c - \tau_L - k |l_L - s|) ds + (\tau_L - l_L) (1 - \hat{s}) \\ & \text{s.t.} \quad l_F \in [0, 1], l_L \in [l_F, 1], \text{ and } \hat{s} \in [0, 1]. \end{aligned}$$

The integrals are the profits of firms that are allocated to the two countries. The other two terms are the rents of the two governments. The problem can be simplified to

$$\begin{aligned} & \min_{\{l_F, l_L, \hat{s}\}} \int_0^{\hat{s}} k |l_F - s| ds + l_F \hat{s} + \int_{\hat{s}}^1 k |l_L - s| ds + l_L (1 - \hat{s}) \\ & \text{s.t.} \quad l_F \in [0, 1], l_L \in [l_F, 1], \text{ and } \hat{s} \in [0, 1]. \end{aligned}$$

Thus the efficient allocation minimizes the sum of the aggregate costs of standard mismatch and the costs of ES setting. We use superscript  $e$  to denote an efficient allocation. To express dependencies on  $k$ , we write  $l_F^e(k)$ ,  $l_L^e(k)$ , and  $\hat{s}^e(k)$ .

---

<sup>40</sup>If it is efficient that the two countries set different standard levels, it does not matter for the efficiency of the allocation whether  $F$  or  $L$  sets the higher standard. Here, we pose the problem so that  $L$  sets a standard not lower than  $F$ . Since the roles of  $F$  and  $L$  can be exchanged, the results in this section are unique only up to a relabelling of countries.

It is immediate that, if  $k < 1$ , the set of efficient outcomes is given by  $l_F^e(k) = 0$ ,  $l_L^e(k) = 0$ , and  $\hat{s}^e(k) \in [0, 1]$ . That is, for  $k < 1$  it is efficient to set a zero ES with the share of firms that locates in each country being indeterminate. Even for the firm  $s = 1$ , it is more efficient to incur the costs of standard mismatch,  $k$ , than to pay for a positive ES level  $l$  that would lower mismatch costs:  $k < l + k(1 - l) = k + l(1 - k)$ . If  $k = 1$ , any allocation for which  $l_F^e = 0$  and  $l_L^e = \hat{s}^e \in [0, 1]$  is efficient. In addition, for  $l_F^e = l_L^e = 0$  any  $\hat{s}^e \in (0, 1]$  is efficient as well. Since the *mcsm* and the marginal cost of enforcing the standard for an additional firm are equal if  $k = 1$ , there exists a continuum of efficient allocations.

For  $k > 1$ , solving the minimization problem above yields the efficient allocation:

$$\begin{aligned} l_F^e(k) &= \frac{k-1}{4k}; \\ l_L^e(k) &= \frac{3k-1}{4k}; \\ \hat{s}^e(k) &= \frac{1}{2}. \end{aligned}$$

The efficient standard levels are increasing in  $k$ . Figure 13 illustrates the efficient ES levels and the allocation of firms to countries depending on  $k$  for the case  $k > 1$ .

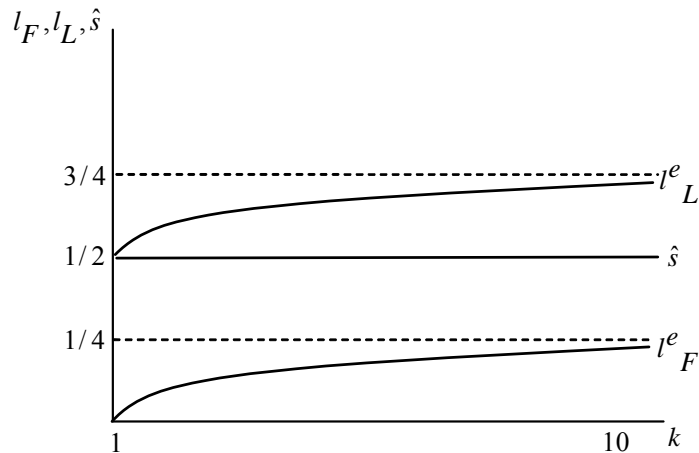


Figure 13: Efficient allocation for  $k > 1$



Given the Hotelling features of our underlying model, one might have expected the efficient solution to have the form  $l_F^e(k) = \frac{1}{4}$  and  $l_L^e(k) = \frac{3}{4}$  familiar from Hotelling (1929). In our model the efficient levels of enforcement are lower, starting at just above  $l_F^e(k) = 0$  and just above  $l_L^e(k) = \frac{1}{2}$  respectively for  $k \rightarrow 1$  (from above) and converging towards  $l_F^e(k) = \frac{1}{4}$  and  $l_L^e(k) = \frac{3}{4}$  respectively as  $k$  becomes large. To understand why our efficient ES levels are lower than they would have been in a direct application of Hotelling, recall that in our model one has to take into account the costs of enforcing the ES for each firm assigned to a country as well as the costs of standard mismatch. If our model were a direct application of Hotelling then the level of the ES would not have affected its cost of enforcement. Efficient ES levels in our model approach the efficient levels that would have arisen in a direct application of Hotelling's model as  $k$  becomes large because the cost of standard mismatch becomes large relative to the cost of enforcement. Finally, as in Hotelling's model, in our model the share of firms between countries is equal. This efficient solution will serve as a benchmark against which to compare the equilibrium outcome.

### **Competition over Standards and Taxes**

In this section we will derive and discuss the equilibrium outcome. Our approach will be to first define equilibrium and then state our main theorem in which equilibrium is characterized. The derivation of equilibrium will be undertaken subsequently.

As mentioned above, standard provision and tax setting are modeled as a two-stage game. The sequence of events is as follows. Government  $L$  sets its standard level and tax and then, observing  $L$ 's choices, Government  $F$  sets its standard level and tax. Taking government policies as given, firms then make location decisions to maximize profits. As usual, a *subgame perfect Nash equilibrium* is a strategy profile

with the property that the governments' strategies constitute a Nash equilibrium in every subgame of the game.

A strategy for Government  $L$  is a pair consisting of a standard level and a tax. A strategy is *feasible* if the tax is high enough to cover the cost of standard provision.<sup>41</sup> Formally, the set of feasible strategies is

$$S_L = \{(l_L, \tau_L) \in [0, 1] \times [0, p] \mid \tau_L \geq l_L\}.$$

A strategy for Government  $F$  is a mapping that assigns a pair, consisting of a standard level and a tax, to each possible strategy choice made by Government  $L$  in the first stage of the game. Formally, this mapping is described by  $f : S_L \rightarrow [0, 1] \times [0, p]$  where  $f(l_L, \tau_L) = (l_F, \tau_F)$ . Let  $F$  be the set that contains all such mappings. The set of feasible strategies for Government  $F$  consists of those members of  $F$  with the property that tax revenue covers the cost of the associated standard level; that is,

$$S_F = \{f \in F \mid \text{for all } (l_L, \tau_L) \in S_L, f(l_L, \tau_L) \text{ satisfies } \tau_F \geq l_F\}.$$

We are interested in the pure strategy subgame-perfect Nash equilibrium of the game, which can be viewed as a Stackelberg game.<sup>42</sup>

**Definition 19** *A pure strategy subgame-perfect Nash equilibrium in taxes and standard levels is a pair of strategies  $((l_L^*, \tau_L^*), f^*)$  such that*

1.  $(l_L^*, \tau_L^*) \in S_L$  is a best response to  $f^*$ .
2.  $f^* \in S_F$  and  $f^*(l_L, \tau_L)$  is a best response to  $(l_L, \tau_L)$  for all  $(l_L, \tau_L) \in S_L$ .

---

<sup>41</sup>Thus we make the simplifying assumption that there are no other sources of government revenue and no international capital market which governments can tap. We do not think that allowing such a possibility would change our results, wherein governments make positive rents in equilibrium.

<sup>42</sup>It will be assumed throughout that mixed strategies in tax rates are not available to governments. This is generally deemed to be an acceptable assumption in the applied literature on policy setting in a perfect information environment.

With the structure of the model in place and equilibrium defined, we are now ready to state our main theorem which characterizes equilibrium.

**Theorem 20** *The outcome of the subgame-perfect equilibrium.*<sup>43</sup>

*The subgame-perfect equilibrium is as follows.*

(a) *(Efficient outcome) If  $k \leq \frac{1}{3}$ , both countries set the minimum standard level and set zero taxes. Firms split equally between the two countries; that is,  $(l_L^*, \tau_L^*) = (0, 0)$  and  $(l_F^*, \tau_F^*) = (0, 0)$ , and  $\hat{s}^* = \frac{1}{2}$ .*

(b) *(Standard haven) If  $\frac{1}{3} < k \leq 1$ , the differentiation in standard levels between the two countries is high; the developed country sets a standard close to the maximum level and the developing country sets a zero standard level. Both countries set taxes that lead to positive rents, and rents are always higher for the developing country than for the developed country. The majority of firms locates in the developing country. Specifically, it holds that  $l_L^* \geq \frac{8}{9}$ ,  $\tau_L^* \in (l_L^*, 2l_L^*)$ , and  $l_F^* = 0$ ,  $\tau_F^* \in (\frac{3}{4}, \frac{4}{3})$ , and  $\hat{s}^* > \frac{2}{3}$ .*

(c) *(Race to the top) If  $k > 1$ , the standard level is above  $\frac{1}{2}$  in both countries, with the developed country setting a higher standard than the developing country. The standard levels do not vary with  $k$ . Both governments make positive rents, requiring firms to pay more than twice the cost of standard provision. The developing country sets a higher tax than the developed country and earns higher rents. Two-thirds of the firms locate in the developing country, and every firm with strictly higher ideal standard level than set in the developing country locates in the developed country. Specifically, it holds that  $l_L^* = \frac{8}{9}$ ,  $\tau_L^* = \frac{4}{3} + \frac{4}{9}k > 2l_L^*$ , and  $l_F^* = \frac{2}{3}$ ,  $\tau_F^* = \frac{4}{3} + \frac{2}{3}k > 3l_F^*$ , and  $\hat{s}^* = \frac{2}{3}$ .*

Figure 14 shows the equilibrium standard levels set in the two countries depending on  $k$ . The subgame-perfect standard and tax levels differ considerably across the three regions of  $k$ : A small  $k$  leads to an efficient outcome; for  $k$  in an intermediate range there is almost maximum differentiation in standards; for large  $k$  there is some

---

<sup>43</sup>The theorem is restated in the Appendix with formulae for all the equilibrium values shown explicitly.

differentiation but it is substantially smaller than for  $k$  in the intermediate range. The reason is that  $F$  sets two-thirds of the maximum standard instead of zero as in the ranges where  $k$  is low and high. For low  $k$ , taxes are the same in both countries. For  $\frac{1}{3} < k \leq 1$ , the developed country sets a higher tax than the developing country, whereas for  $k > 1$ , the developing country sets a higher tax than the developed country.

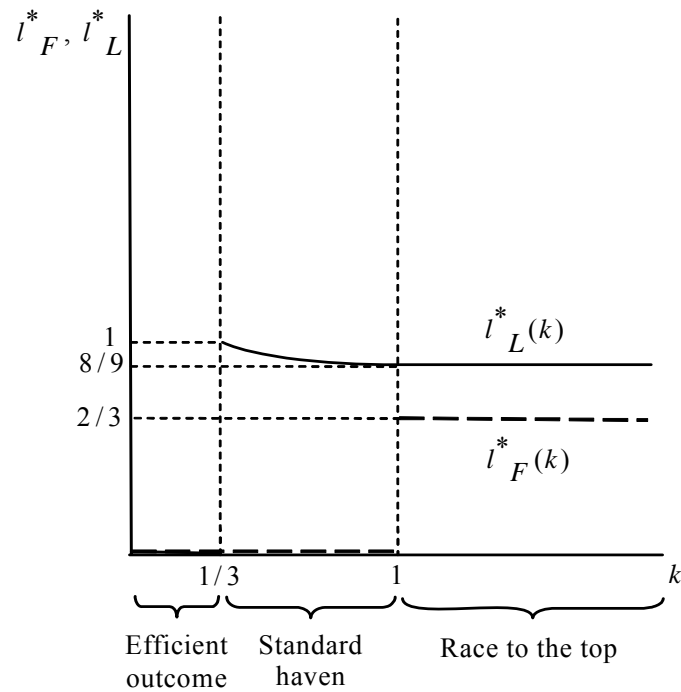


Figure 14: Equilibrium standard levels depending on  $k$

The common characteristic of equilibrium across all levels of  $k$  is that the developing country attracts at least as many firms as the developed country. Also, for  $k > \frac{1}{3}$ , both the developed and developing country are able to extract rents.<sup>44</sup> This arises as a result of the monopolistic power that each government has over location

<sup>44</sup>The result is particularly striking for the country that supplies zero standard even though it levies a positive tax.

within its country. Each firm must locate in one country or the other in order to produce, and the government of the country where it does locate is able to exploit its resultant power when setting taxes. An additional interesting aspect is that  $F$ , who sets a lower standard, makes more rents because it both attracts more firms and makes more rents per firm. Except for  $k \leq \frac{1}{3}$ , countries set inefficiently high standard levels.

The intuition behind the result for low marginal cost of standard mismatch, case (a), is straightforward. For  $k \leq \frac{1}{3}$ , the costliness of standard mismatch is so low that countries do not succeed in differentiating themselves via standard levels. This is due to the fact that firms do not perceive countries with different standard levels as sufficiently distinct from each other. Therefore a country cannot extract a monopolistic rent by setting a standard level different from the one set in another country. All competition occurs in taxes, which brings about an efficient outcome.

Turning to case (b), the intuition behind the maximum differentiation in standards that occurs when the marginal cost of standard mismatch is in an intermediate range is as follows. The developing country has a second-mover advantage and so creates a standard haven for firms whose costs are affected more by taxes than by standard mismatch. The developed country can extract some rents (because  $k$  is not too small), but only by differentiating itself substantially (because  $k$  is not too large) from the developing country. Because it is a dominant strategy for the developing country to become a standard haven, the developed country can only differentiate itself by setting its standard at a high level. As a result there is close to maximum differentiation between the two countries.

Regarding case (c), when the cost of standard mismatch is high relative to taxes, both countries offer inefficiently high standard levels. (Recall from Section 3 that the efficient outcome calls for the countries offering up to, respectively, 12% and 35% of the maximum standard level.) Because firms value a lower standard mismatch more

than lower taxes, the developing country has an incentive to choose a standard level close to the standard level that the developed country sets. Whether the developing country chooses a standard level that is lower or higher than the one in the developed country depends on whether the developed country sets a relatively high standard level (in which case the developing country would set a lower standard level) or whether it sets a relatively low standard level (in which case the developing country would set a higher standard level). In equilibrium, the developed country chooses a high standard level even though a lower standard level would be less costly. This is because the developed country has to allow the developing country to extract high rents to prevent the developing country from undercutting.

Notice that case (c) is the case which one would expect to be least stable among the three cases. Because  $F$  sets the highest tax that still attracts a positive fraction of firms to its country (all firms ‘to the left of  $F$ ’ with an ideal standard level not higher than the one  $F$  sets), Government  $L$  - if able to do so - could marginally lower its tax, and by doing so attract all firms to its country. An additional fraction of two-thirds of all firms would be attracted, from which  $L$  could extract rents.

Now that we have stated our main result and given the basic intuition behind it, we will next provide a detailed analysis of its derivation. To do so, the next subsection provides a characterization of  $F$ ’s best response, and this is followed by a characterization of  $L$ ’s best response in the subsection that follows. All proofs are given in the Appendix.

### **The developing country’s best response function**

In this section, we analyze Government  $F$ ’s best response to a given strategy  $(l_L, \tau_L)$  of Government  $L$ . We can ignore Case 1 since setting the same standard level and tax as  $L$  is never a best response for  $F$  except if  $(l_L, \tau_L) = (0, 0)$  and  $k \leq 1$ , which is treated below. We can also ignore Case 5 since choosing a response that does not attract any firm is never a best response for  $F$ .

To find Government  $F$ 's best response to a given strategy of  $L$ , we proceed in two steps. First, we maximize  $F$ 's rents separately over the three response subsets, *sharing I*, *sharing II*, and *undercutting*.

Government  $F$ 's optimization problem:

(a) Maximize rents over sharing I

$$\begin{aligned} & \max_{\tau_F, l_F} (\tau_F - l_F) \hat{s} \\ & \text{s.t.} \\ & l_F \in [0, l_L) \\ & \tau_F \in [l_F, p] \\ & \tau_F \in [\tau_L - k(l_L - l_F), \tau_L + k(l_L - l_F)] \end{aligned}$$

(b) Maximize rents over sharing II

$$\begin{aligned} & \max_{\tau_F, l_F} (\tau_F - l_F)(1 - \hat{s}) \\ & \text{s.t.} \\ & l_F \in (l_L, 1] \\ & \tau_F \in [l_F, p] \\ & \tau_F \in [\tau_L - k(l_F - l_L), \tau_L + k(l_F - l_L)] \end{aligned}$$

(c) Maximize rents over undercutting

$$\begin{aligned} & \max_{\tau_F, l_F} (\tau_F - l_F) \\ & \text{s.t.} \\ & l_F \in [0, 1] \\ & \tau_F < \tau_L - k|l_L - l_F|. \end{aligned}$$

Second, given the solutions to (a), (b), and (c), the best response is found by comparing the maximized rents across the three sets of possible solutions.

There are two issues that can arise when solving for the developing country's best response to  $(l_L, \tau_L)$ : First, a best response might not exist; second, a best response might not be unique. The existence of a best response to  $(l_L, \tau_L)$  is not guaranteed because an optimal undercutting strategy does not exist. The reason is that the rent function does not have a well-defined maximum on the set of undercutting strategies. That is, for each undercutting strategy with  $\tau_F = \tau_L - k|l_L - l_F| - \varepsilon$  where  $\varepsilon > 0$ , we can find a slightly higher tax (i.e., a smaller  $\varepsilon$ ) that still undercuts  $L$ 's strategy. Because such a tax would yield higher rents, the optimal undercutting strategy is not well defined.<sup>45</sup>

In our model this difficulty can be resolved in a straightforward way. Even though one cannot determine an optimal undercutting strategy, one can determine when Government  $F$  will undercut  $L$  and when it will share firms with  $L$ . Because Government  $L$  will avoid strategies that induce  $F$  to undercut (i.e., undercutting happens only off the equilibrium path), we can solve our model without determining the specific undercutting strategy. We determine which of  $L$ 's strategies lead  $F$  to undercut by assuming that  $F$  undercuts whenever there exists some undercutting strategy that yields more rents than the best sharing strategy.

To be more specific, let  $r_F^s(l_L, \tau_L)$  be  $F$ 's rent from an optimal sharing strategy after  $L$  has chosen  $(l_L, \tau_L)$ , and, given  $\varepsilon > 0$ , let  $r_F^u(l_L, \tau_L; \varepsilon)$  be  $F$ 's rent from undercutting where  $\tau_F = \tau_L - k|l_L - l_F| - \varepsilon$ . Let  $r_F^u(l_L, \tau_L) = \lim_{\varepsilon \rightarrow 0} r_F^u(l_L, \tau_L; \varepsilon)$ . Note that, by choosing  $\varepsilon$  sufficiently small,  $F$  can obtain a rent arbitrarily close to  $r_F^u(l_L, \tau_L)$ , but still  $r_F^u(l_L, \tau_L; \varepsilon) < r_F^u(l_L, \tau_L)$  no matter how small is  $\varepsilon$ . By solving  $r_F^s(l_L, \tau_L) = r_F^u(l_L, \tau_L)$  we obtain a critical tax  $\tau_L$  that depends on  $l_L$ . We denote

---

<sup>45</sup>The literature on entry deterrence through pricing strategy has also had to broach the issue of what constitutes a best response when payoff functions defined by the game are discontinuous and might not have a well defined maximum. This issue carries over to the present setting.



this tax by  $\hat{\tau}_L(l_L)$  and will refer to it as the *sharing tax limit*. The sharing tax limit can be used to classify payoffs to  $F$ 's standard and tax as follows:

- If  $\tau_L \leq \hat{\tau}_L(l_L)$  then for all  $\varepsilon > 0$ , it holds that  $r_F^s(l_L, \tau_L) > r_F^u(l_L, \tau_L; \varepsilon)$ ;
- if  $\tau_L > \hat{\tau}_L(l_L)$  then there exists an  $\varepsilon > 0$  such that  $r_F^s(l_L, \tau_L) < r_F^u(l_L, \tau_L; \varepsilon)$ .

In other words, if  $L$ 's tax is higher than the sharing tax limit, then  $F$  can find an  $\varepsilon$  small enough to make the rents earned from undercutting higher than the rents earned by sharing. However, if  $L$  sets its tax no higher than the sharing tax limit,  $F$  finds that sharing yields strictly higher rents than undercutting, no matter how small is  $\varepsilon$ . Figure 15 depicts the situation (the significance of  $\hat{l}_L$  in the figure will be explained later).

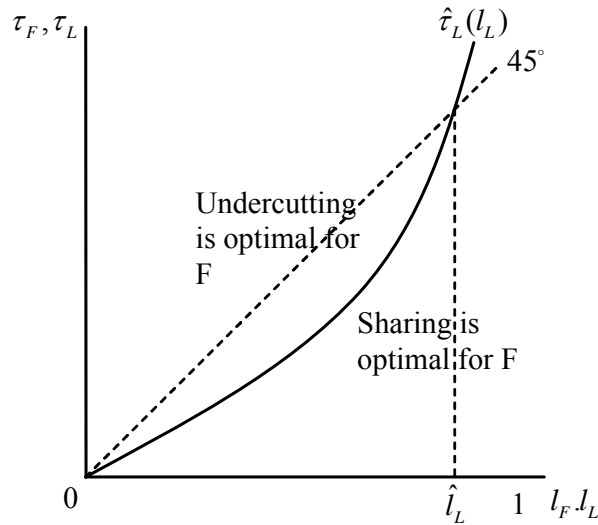


Figure 15: Sharing tax limit for  $k < 1$

To deal with the fact that Government  $F$  might have multiple best responses recall our assumption that, if a government has multiple best responses, it chooses

the best response that maximizes its share of firms. This implies that of the best responses available,  $F$  chooses the one that requires the lowest level of standard. Moreover, if in addition  $(l_L, \tau_L) = (0, 0)$ , we assume that  $F$  sets  $\tau_F = 0$ . We only require these properties in two situations. First, if  $k < 1$  and  $(l_L, \tau_L) = (0, 0)$ , there is no response that yields positive rents for  $F$ . Our assumption then implies that  $F$  chooses  $(l_F, \tau_F) = (0, 0)$ . Notice that any other feasible strategy for  $F$  would induce all firms to locate in  $L$ . Second, if  $k = 1$  then for any  $(l_L, \tau_L)$  Government  $F$  has a whole range of best responses. More specifically, there is a best response at each standard level  $l_F$ . The reason is that standard mismatch and taxes are equally costly for firms. Therefore if a country decides, for example, to set a lower standard and use the resources that it saves to reduce its tax, it will attract the same share of firms as before and it will make the same rents per firm. For this case, our assumption implies that  $l_F = 0$ .

Before stating our first result, to keep track of the different kinds of sets characterizing our results, we introduce the following notation. The responses that maximize  $r_F(l_F, \tau_F; l_L, \tau_L)$  over undercutting, sharing I, and sharing II are denoted by  $(l_F^u, \tau_F^u)$ ,  $(l_F^{s1}, \tau_F^{s1})$ , and  $(l_F^{s2}, \tau_F^{s2})$ , respectively. The corresponding rents are denoted by  $r_F^u$ ,  $r_F^{s1}$ , and  $r_F^{s2}$ , respectively. The responses and revenues all depend on  $l_L$  and  $\tau_L$ . For notational ease, we will use  $(l_F^*, \tau_F^*)$  to denote the response that maximizes  $r_F(l_F, \tau_F; l_L, \tau_L)$  over  $\{(l_F^u, \tau_F^u), (l_F^{s1}, \tau_F^{s1}), (l_F^{s2}, \tau_F^{s2})\}$ .

The nature of the results we obtain differs across three intervals,  $k \in (0, \frac{1}{3}]$ ,  $k \in (\frac{1}{3}, 1]$ , and  $k \in (1, \infty)$ . For each of the three regions of  $k$ , Proposition 21 summarizes the best response of Government  $F$  to any standard level and tax that Government  $L$  has chosen in the first stage.

**Proposition 21** *(The developing country's best response)*

(a) *If the marginal cost of standard mismatch for firms is low ( $k \leq \frac{1}{3}$ ), Government  $F$ 's best response to any of Government  $L$ 's feasible strategies is to set zero standard,*

and to set an undercutting tax if  $\tau_L > 0$  and to set  $\tau_F = 0$  if  $\tau_L = 0$ . Specifically, if  $\tau_L > 0$  then  $(l_F^*, \tau_F^*) = (0, \tau_F^u(l_L, \tau_L))$ , and if  $\tau_L = 0$  then  $(l_F^*, \tau_F^*) = (0, 0)$ .

(b) If the marginal cost of standard mismatch is at an intermediate level ( $\frac{1}{3} < k \leq 1$ ) there exists, for each standard level set by  $L$ , a corresponding sharing tax limit. If  $L$  sets its tax above (equal to or below) the sharing tax limit, then Government  $F$ 's best response is to set no standard and to set the corresponding optimal undercutting tax (optimal sharing tax). Specifically, for each  $l_L$  there exists a sharing tax limit,  $\hat{\tau}_L(l_L)$ , such that if  $\tau_L > \hat{\tau}_L(l_L)$  then  $(l_F^*, \tau_F^*) = (0, \tau_F^u(l_L, \tau_L))$  and if  $\tau_L \leq \hat{\tau}_L(l_L)$  then  $(l_F^*, \tau_F^*) = (0, \tau_F^{s1}(l_L, \tau_L))$ .  $F$ 's optimal sharing tax is given by  $\tau_F^{s1}(l_L, \tau_L) = \frac{1}{2}\tau_L + \frac{k}{2}l_L$ .

(c) If the marginal cost of standard mismatch is high ( $k > 1$ ) there exists, for each standard level set by  $L$ , a corresponding sharing tax limit. If  $L$  sets its tax above (equal to or below) the sharing tax limit, then Government  $F$ 's best response is to set the optimal undercutting tax while setting the same standard level as  $L$  (set the optimal sharing tax and set either a lower or higher standard than  $L$ ). Specifically, for each  $l_L$  there exists a sharing tax limit,  $\hat{\tau}_L(l_L)$ , such that if  $\tau_L > \hat{\tau}_L(l_L)$  then  $(l_F^*, \tau_F^*) = (l_L, \tau_F^u(l_L, \tau_L))$  and if  $\tau_L \leq \hat{\tau}_L(l_L)$  then  $(l_F^*, \tau_F^*) \in \{(l_F^{s1}, \tau_F^{s1}(l_L, \tau_L)), (l_F^{s2}, \tau_F^{s2}(l_L, \tau_L))\}$ .

If the marginal cost of standard mismatch is low or at an intermediate level ( $k \leq 1$ ), it does not pay for Government  $F$  to compete in the standard at all. Thus  $l_F^* = 0$  in parts (a) and (b). However, if the marginal cost of standard mismatch is high ( $k > 1$ ),  $F$  has an incentive to set a positive standard level. Moreover, the cheapest way to attract *all* firms is to set exactly the same level of standard as  $L$ . In this way  $F$  does not need to compensate any of the firms for a higher standard mismatch. The optimal sharing I and sharing II taxes for case (c) are both boundary solutions. Government  $F$  sets the highest tax that still attracts some firms to its country (the firms in the intervals  $[0, l_F]$  and  $[l_F, 1]$ , respectively).

Part (a) of Proposition 21 shows that for small  $k$  undercutting dominates sharing. All firms can be attracted without having to set the tax much below  $L$ 's tax. At the

same time, the area in policy space over which firms are shared is reduced - as  $k$  is reduced, a given tax set by  $F$  will induce all firms to locate in country  $L$ . Figure 16 illustrates the situation. A reduction of  $k$  increases undercutting possibilities while at the same time it reduces sharing possibilities. In particular, as  $k \rightarrow 0$ , the set of sharing possibilities shrinks to the empty set. For  $\tau_L = 0$  there is no strategy for  $F$  that yields a positive rent, meaning that  $F$  is indifferent among all feasible strategies. Thus, (by assumption)  $F$  sets a zero standard and sets  $\tau_F = 0$ .

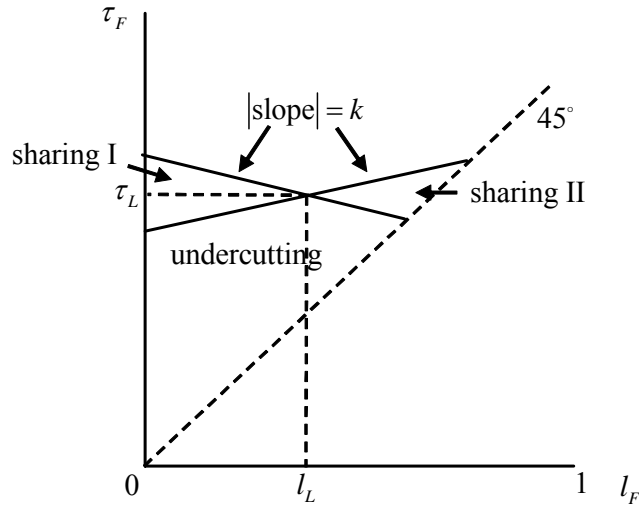


Figure 16: Undercutting dominates for small  $k$

Part (b) of Proposition 21 illustrates the role of the sharing limit tax,  $\hat{\tau}_L(l_L)$ , in the model. Government  $F$  shares if  $L$ 's tax does not exceed  $\hat{\tau}_L(l_L)$  but undercuts if  $L$ 's tax is above it. To see why suppose that, for some standard and tax levels,  $L$  and  $F$  are sharing firms. If  $L$  increases its tax,  $F$  will raise its own tax by only half the amount ( $\tau_F^{s1} = \frac{1}{2}\tau_L + \frac{k}{2}l_L$ ). When raising its own tax,  $F$  has to consider a 'tax level effect' -  $F$  will earn more rent per firm - and a 'tax base effect' - fewer firms will locate in  $F$ . Since, when  $k \leq 1$ , firms' location decisions are relatively elastic

with respect to taxes (recall that  $\hat{s} = \frac{\tau_L - \tau_F}{2k} + \frac{l_L + l_F}{2}$ ) the tax base effect dominates the tax level effect and  $F$  increases its tax less than  $L$  does. Thus the share of firms locating in  $F$  increases. The more  $L$  raises its tax, the more firms it will lose to  $F$ . Eventually all firms with  $s \leq l_L$  will locate in  $F$ . At this point,  $F$  will want to switch to an undercutting strategy because  $F$  has to lower  $\tau_F$  only marginally to induce an additional share of  $(1 - l_L)$  firms to locate in its country.

Similarly, there exists a sharing tax limit in part (c). In Figure 17 the sharing tax limit is given by  $\hat{\tau}_L(l_L) = \max \{ \hat{\tau}_L^1(l_L), \hat{\tau}_L^2(l_L) \}$ , where  $\hat{\tau}_L^1(l_L)$  is the tax up to which an optimal sharing I strategy is better for  $F$  than any undercutting strategy and  $\hat{\tau}_L^2(l_L)$  is the tax up to which an optimal sharing II strategy is better for  $F$  than any undercutting strategy. Notice that the proposition only states that  $F$  shares firms up to that tax level, but not whether it does so by setting a lower or higher standard than  $L$ . It is possible (as shown in the proof) to identify two subsets of  $S_L$  so that  $F$  chooses  $(l_F^{s1}, \tau_F^{s1})$  or  $(l_F^{s2}, \tau_F^{s2})$  if  $(l_L, \tau_L)$  is in the first or second of the subsets respectively. Intuitively, if  $L$  sets a relatively low standard level then sharing with a higher standard level tends to yield higher rents for  $F$ ; if  $L$  sets a relatively high standard level then sharing with a lower standard level will yield higher rents for  $F$ . More specifically, we show in the appendix that if  $l_L \geq \frac{1}{2}$  then setting an even higher standard and sharing is never a best response for Government  $F$ . In this case, instead of setting a standard that exceeds  $L$ 's standard by  $x$ , i.e.,  $l_F = l_L + x$ , and set some tax,  $F$  can set  $l_F = l_L - x$ , without changing the tax. Doing so will increase  $F$ 's rent per firm, and the share of firms attracted will be at least as large.

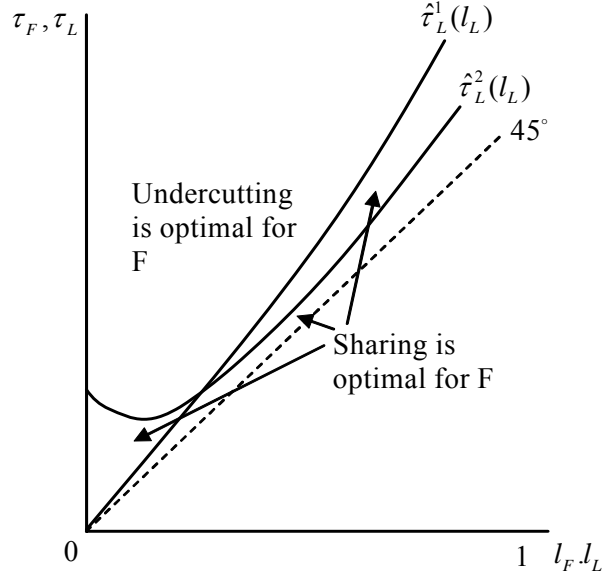


Figure 17: Sharing tax limit for  $k > 1$

### The developed country's best response

Government  $L$  takes Government  $F$ 's subgame-perfect strategy  $f^*(l_L, \tau_L)$  as given and maximizes its rent function over  $S_L$ . Formally  $L$ 's problem is

$$\begin{aligned} & \max_{\{l_L, \tau_L\}} r_L(l_L, \tau_L, f^*(l_L, \tau_L)) \\ & \text{s.t.} \\ & l_L \in [0, 1] \\ & \tau_L \in [l_L, p] \end{aligned}$$

Just like Government  $F$ 's rent function,  $L$ 's rent function evaluated at  $f^*$  is not continuous. For example, discontinuities arise at the sharing tax limit  $\hat{\tau}_L(l_L)$ . But given  $f^*$ , we can safely exclude from the set of candidates for best response all strategies with  $\tau_L > \hat{\tau}_L(l_L)$  (except  $(l_L, \tau_L) = (0, 0)$ ), because such strategies would induce  $F$  to undercut and hence would leave  $L$  with zero rents. Accounting for  $F$ 's response in the second stage, those taxes will yield zero rents for  $L$ , while a tax that induces  $F$

to share firms yields positive rents. Thus, we can formulate a reduced problem for  $L$ .

Government  $L$ 's reduced optimization problem:

$$\begin{aligned} & \max_{\{l_L, \tau_L\}} r_L(l_L, \tau_L, f^*(l_L, \tau_L)) \\ & \text{s.t.} \\ & l_L \in [0, 1] \\ & \tau_L \in [l_L, \hat{\tau}_L(l_L)] \end{aligned}$$

Figure 18 depicts  $L$ 's rent function,  $r_L(l_L, \tau_L, f^*(l_L, \tau_L))$ , depending on  $\tau_L$  and fixing some standard level  $l_L$ . Rents are zero when  $L$  sets  $\tau_L = l_L$ , but then increase at low levels of  $\tau_L$  when  $F$  is willing to share firms. Rents jump to zero at  $\tau_L = \hat{\tau}_L(l_L)$ .

As with Government  $F$ , Government  $L$ 's best response to  $f$  might not be unique. Government  $L$ 's best response to  $F$ 's equilibrium strategy  $f^*$  is not unique if and only if  $k \leq \frac{1}{3}$ . In this case  $L$  cannot make any positive rents because  $F$  always undercuts  $L$ . Paralleling our assumption for Government  $F$ , we assume that  $L$  then chooses  $(l_L, \tau_L) = (0, 0)$ . As before, this assumption reflects a preference for strategies that attract larger shares of firms.

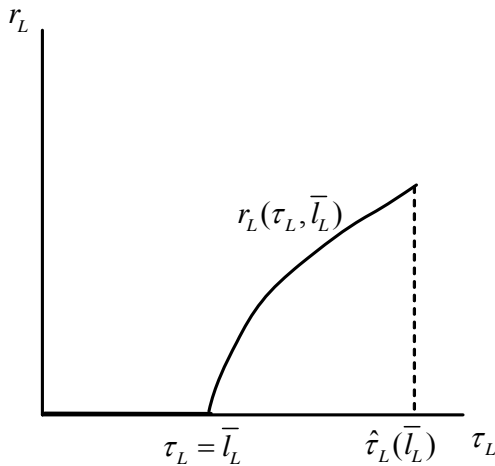


Figure 18: Government  $L$ 's payoff function

Since the nature of  $f^*$  depends on  $k$ , so will the nature of Government  $L$ 's optimal strategy; Proposition 22 summarizes. We use  $\hat{l}_L$  to denote the critical standard level so that the sharing tax limit is at least as large as the cost to cover the standard if and only if  $l_L \geq \hat{l}_L$ . See Figure 15 for an illustration.<sup>46</sup>

**Proposition 22** (*The developed country's best response to  $f^*$* ).

(a) *If the marginal cost of standard mismatch is low ( $k \leq \frac{1}{3}$ ), Government  $L$ 's best response to  $f^*$  is to set no standard and set zero tax. Specifically,  $(l_L^*, \tau_L^*) = (0, 0)$ .*

(b) *If the marginal cost of standard mismatch is at an intermediate level ( $\frac{1}{3} < k \leq 1$ ), Government  $L$ 's best response to  $f^*$  is to set a standard strictly larger than  $\hat{l}_L$  and set its tax at the sharing tax limit,  $\hat{\tau}_L(l_L)$ , the highest tax that induces  $F$  to share firms. This tax is higher than the tax set by  $F$ . As  $k$  is increased, standard provision by  $L$  decreases from  $l_L^* \lesssim 1$  to  $l_L^* \simeq \frac{22}{25}$ , and rents per firm increase.*

(c) *If the marginal cost of standard mismatch is high ( $k > 1$ ), Government  $L$ 's best response to  $f^*$  is to set a standard of  $l_L^* = \frac{8}{9}$  and to set  $\tau_L = \hat{\tau}_L^1(l_L)$ , the highest tax that induces  $F$  to share firms and which exceeds costs by at least a factor of 2. Specifically,  $(l_L^*, \tau_L^*) = (\frac{8}{9}, \frac{4}{3} + \frac{4}{9}k)$ .*

If  $k \leq \frac{1}{3}$ , Government  $F$  chooses an undercutting strategy for each tax that exceeds the cost of the standard (Proposition 21). Thus each of  $L$ 's strategies yields zero rents and, by assumption, Government  $L$  picks  $(l_L^*, \tau_L^*) = (0, 0)$ . Rents for both governments are zero.

To derive the results for  $\frac{1}{3} < k \leq 1$ , we show that for  $k \leq 1$ , there exists a ( $k$ -dependent) critical standard level  $\hat{l}_L$  such that  $\hat{\tau}_L(l_L) = l_L$  for all  $l_L \leq \hat{l}_L$  and  $\hat{\tau}_L(l_L) > l_L$  for all  $l_L > \hat{l}_L$ , as a result of which we can focus on standard levels

---

<sup>46</sup>Proposition 22 is restated in the Appendix with the exact expressions for the optimal strategies.



$l_L \in [\hat{l}_L, 1]$ . From this we can see immediately that  $L$  chooses the tax  $\hat{\tau}_L(l_L)$  since, as shown in Figure 15,  $L$ 's rents are increasing in  $\tau_L$  up to  $\hat{\tau}_L(l_L)$ .

At first sight it seems surprising that for  $\frac{1}{3} < k \leq 1$  only a standard of more than  $\hat{l}_L = \frac{8k(1-k)}{(1+k)^2}$  allows  $L$  to earn positive rents (see Lemma 27 in the Appendix). On the face of it, there is of course an incentive for  $L$  to set a low standard level since this saves monitoring costs and would increase the share of firms locating in  $L$ . However, the lower the standard level that  $L$  sets, the greater the incentive for  $F$  to switch to an undercutting strategy because switching from a sharing strategy to an undercutting strategy induces all firms located in  $L$  to move to  $F$ . Therefore, so that it does not induce  $F$  to undercut,  $L$  puts itself into a situation in which it attracts only a relatively small share of firms by setting a high standard level. This happens despite the fact that standard mismatch is not very important to firms.

When  $k > 1$ ,  $L$  sets a high standard level. Intuitively, for  $L$ , setting a low standard level in order to induce  $F$  to set a higher standard level might seem a better strategy. However, it is in fact better for  $L$  to let  $F$  be the country that sets a low standard. This guarantees  $F$  higher rents from sharing, which means that  $F$  accommodates higher taxes by  $L$  without undercutting. For example, suppose that  $L$  chooses  $l_L = \frac{1}{9}$  and sets a tax  $\tau_L = \hat{\tau}_L(\frac{1}{9})$  instead of its actual equilibrium choice  $l_L^* = \frac{8}{9}$  and  $\tau_L^* = \hat{\tau}_L(\frac{8}{9})$ . Government  $F$ 's best response would be to set  $l_F = \frac{1}{3}$  instead of  $l_F^* = \frac{2}{3}$ . As with the actual equilibrium strategies,  $L$  attracts one third of the firms. But the tax  $L$  is able to set,  $\hat{\tau}_L(\frac{1}{9})$ , is so much lower than  $\hat{\tau}_L(\frac{8}{9})$  that rents per firm are only  $\frac{4}{9}k - \frac{4}{9}$  compared to  $\frac{4}{9}k + \frac{4}{9}$  with the equilibrium strategy. In order to obtain the ability to set a higher tax without losing firms,  $L$  accepts that it has to set a costlier standard level.

Notice also that, for  $k > 1$ , in contrast to the situation where  $k \leq 1$ , the standard level set by  $L$  does not vary with  $k$ . As standard mismatch becomes more costly for firms,  $L$  extracts more rents through an increase in taxes.

It should now be clear that Propositions 21 and 22 can be used to solve for the mutual best responses of the strategies of  $F$  and  $L$ , thus yielding the subgame-perfect Nash equilibrium in pure strategies presented in Theorem 20 above.

## Conclusion

We began this paper by noting concerns in policy circles that developing countries resembling those of recent entrants to the EU may, under certain circumstances, have a second mover advantage in setting standards and taxes. This paper sets out a formal framework which makes precise a set of circumstances under which such a second mover advantage may arise. Three possible predictions are made about the outcome of fiscal competition when the public good in question is a standard. The particular prediction that emerges in equilibrium depends on the marginal cost of standard mismatch. The model focuses on the interplay between governments' incentives to manipulate policy - standards and taxes - in order to maximize rents and firms' incentives to locate where these policies have the most favorable impact on profits. The key point is that the government of the developed country wants to avoid inducing the developing country to undercut because being undercut implies losing all firms and hence all rents. If the marginal cost of standard mismatch is low, then standards are not important enough to firms for governments to be able to use them strategically. In this case, the forces of tax competition envisaged by Brennan and Buchanan are strong enough to dominate, and the outcome is efficient. If the marginal cost of standard mismatch is high enough, the developed country government successfully induces sharing by setting a sufficiently high standard relative to the tax. A proportion of firms will then find it beneficial to locate in each country. Governments are able to use policy to make rents, and the resulting outcome is inefficient in that either the developed country government or both governments set standards too high.

It is worth drawing parallels between our work and the large literature, primarily in the field of international trade, that has focused on pollution havens. While our work addresses the issue of ‘standard havens’ more broadly defined, it is in the area of the environment that the idea of a haven has attracted the most attention and so it seems worth evaluating the contribution of our work in that context. The pollution haven hypothesis is that, as economies open up to each other, dirty industry will tend to become concentrated in the country with the weakest environmental standards. Standard international trade theory provides a natural explanation for this, which explains why it forms the cornerstone of the main explanation that is put forward for the possible existence of pollution havens. The idea is that, all else equal, thinking of pollution as an ‘input’ to the production process, lax environmental standards are a source of comparative advantage since they make the opportunity cost of pollution low. Antweiler, Copeland and Taylor (2001) construct a model around this idea and present cross-country empirical evidence that provides some support for the existence of pollution havens (also see Taylor 2004). More recent empirical work calls into question the existence of pollution havens on the basis that the pollution content of trade flows do not appear to support the predictions of the trade model; see Ederington et al. (2004). Part (b) of our Theorem 20 is helpful in this regard since it presents an alternative strategic motivation for the existence of pollution havens in developing countries based solely on the feature that developed countries have tended to introduce environmental standards earlier than developing countries.

Inevitably, the theoretical framework developed here simplifies the situation in a number of key respects. For example, to keep the analysis manageable we have not explicitly treated consumers in our analysis and we have restricted the number of countries to just two. A promising direction for future research would be to extend our model to give consumers a more prominent role. One potential limitation to our conclusions is that the government in the developing country does not set standards

‘too low.’ While it seems reasonable to argue that developed countries may set standards too high, a concern is that developing countries actually set their standards too low from the perspective of consumers. The introduction of consumers to the model could make it possible for standards to be set too low in the developing country.

Another promising direction for future research would be to ask how robust our results would be to the introduction of a larger number of countries to the model. From our analysis of the present framework it is not obvious how the outcome would be changed by the introduction of more countries. One conjecture would be that the  $n$ th country to move would always have the greatest advantage, with prior countries being constrained by those that would set policy subsequently. A different conjecture about the outcome would be that only two countries could make positive rents and that the presence of more countries would be irrelevant. If the analysis of a larger number of countries turned out to be analytically intractable then it might be possible to obtain characterizations through numerical simulation.<sup>47</sup>

Finally, a question that could be addressed in the future is whether incentives exist for governments to coordinate/harmonize policy within our framework. Under perfect collusion in our model, governments would simply agree that neither of them would set a positive standard level and they would set taxes at the level of prices, thereby extracting all surplus. Such an outcome would be efficient in our framework in the case where  $k \leq 1$  because in that case the efficient outcome has zero standards; for  $k > 1$  the efficient outcome does have a positive level of standard provision. However, such perfect collusion would require a strong enforcement mechanism and, in the absence of an international enforcement body, the incentives to break such an agreement may be overwhelming. This may explain why in practice proposals for collusion have tended to be weaker, entailing for example the introduction of

---

<sup>47</sup>It is tempting to think that one could analyze a model in which a ‘core’ country sets policy first and a larger number of periphery countries set policy subsequently (but at the same time as each other). However, the difficulty here is that in the present framework in general there may not exist an equilibrium in pure strategies when countries set policies simultaneously.

minimum standards. A surprising implication of our framework is that it is not in the interest of the developed country to introduce a binding minimum standard. The reason is that the developed country benefits from being able to differentiate itself from the developing country and putting in place a minimum standard would limit the scope for doing so. Thus our model presents a possible way of understanding situations in which standards have been called for but none have actually emerged.

## Appendix

### Indifference set

The following is an application of the approach taken by d'Aspremont et al. (1979) to the present setting. Given  $(l_L, \tau_L, l_F, \tau_F)$ , there may be more than one firm that is just indifferent between the two countries. To deal with this possibility, we define the *indifferent set* of firms and denote it by  $I(l_L, \tau_L, l_F, \tau_F)$ . If the Indifferent Set is not a singleton, a tie breaking rule is needed to determine where indifferent firms locate. With two exceptions, the indifferent set  $I(l_L, \tau_L, l_F, \tau_F)$  will be a singleton set, i.e.,  $\hat{s}(l_L, \tau_L, l_F, \tau_F)$  is the only member of  $I(l_L, \tau_L, l_F, \tau_F)$ . The two exceptions are as follows.

(1) Suppose that  $l_F < l_L$  so that  $F$  sets a lower standard than  $L$ . For  $s$  satisfying  $s = l_F$ , if  $s \in I(l_L, \tau_L, l_F, \tau_F)$  then for all  $s' < s$ , it holds that  $s' \in I(l_L, \tau_L, l_F, \tau_F)$ . To see this, first note that for firm  $s \in I(l_L, \tau_L, l_F, \tau_F)$ ,  $s = l_F$ , the extent to which the tax in  $F$  exceeds the tax in  $L$  exactly matches the cost of standard mismatch in  $L$ , i.e.  $\tau_F - \tau_L = k(l_L - l_F)$ . Compared to the costs the firm  $s = l_F$  has in  $F$  and  $L$ , respectively, a firm  $s < l_F$  has an additional cost of standard mismatch of  $k(l_F - s)$  in either  $F$  or  $L$ , implying that those firms must be indifferent as well and that  $I(l_L, \tau_L, l_F, \tau_F) = [0, l_F]$ . By analogous reasoning, if firm  $s = l_L$  is indifferent, then  $I(l_L, \tau_L, l_F, \tau_F) = [l_L, 1]$ . The case  $l_L < l_F$  is symmetric.

(2) Suppose that  $l_F = l_L$ ; in this case a firm's choice of location is determined by

taxes. If  $\tau_F = \tau_L$  then all firms are indifferent and again  $I(l_L, \tau_L, l_F, \tau_F)$  is not a singleton set but equals  $[0, 1]$ .

Note that it might also be the case that no firm is indifferent. For example if  $l_L = l_F$  and  $\tau_F \neq \tau_L$ , all firms prefer whichever country sets the lower tax; consequently  $I(l_L, \tau_L, l_F, \tau_F)$  is the empty set. More generally, whenever one country undercuts the tax of the other country by more than the cost of the standard difference between the two countries, the indifferent set will be empty.

### Proofs

The proof of Proposition 21 uses a sequence of auxiliary results, which are stated and proven separately in the following Lemmas.

**Lemma 23** 1. *If  $k < 1$ , undercutting is feasible if and only if  $(l_L, \tau_L) \in S_L \setminus \{(0, 0)\}$ .*

*Undercutting with  $l_F > 0$  is never a best response.*

2. *If  $k = 1$ , undercutting is feasible if and only if  $\tau_L > l_L$ . For every undercutting strategy with  $l_F > 0$ , there exists an undercutting strategy with  $l_F = 0$  that yields the same rent for  $F$ .*

3. *If  $k > 1$ , undercutting is feasible if and only if  $(l_L, \tau_L) \in S_L$  such that  $\tau_L > l_L$ . Undercutting with  $l_F \neq l_L$  is never a best response.*

### Proof.

1. We will first show that undercutting I is non-empty if  $k < 1$  and  $(l_L, \tau_L) \neq (0, 0)$ . Let  $(l_L, \tau_L)$  be any strategy in  $S_L \setminus \{(0, 0)\}$ . Set  $l_F = 0$ . Then for small enough  $\varepsilon$ ,  $l_F$  together with the tax  $\tau_F = \tau_L - kl_L - \varepsilon \geq 0$  is a feasible undercutting strategy. If  $(l_L, \tau_L) = (0, 0)$  undercutting is not feasible, because it requires to set a tax strictly below zero, which is not feasible. Next, we show that  $(l_F^u, \tau_F^u) = (0, \tau_L - kl_L - \varepsilon)$  for some  $\varepsilon > 0$ . Take any undercutting strategy with  $l_F > 0$  and a corresponding undercutting tax  $\tau_F = \tau_L - k|l_L - l_F| - \varepsilon$ . Using the same  $\varepsilon$  to undercut, the strategy  $l_F' = 0$  with undercutting tax  $\tau_F' = \tau_L - kl_L - \varepsilon$  is feasible (i.e.,  $\tau_L - kl_L - \varepsilon > 0$ ) and

yields more rents per firm because it saves costs of  $l_F$  per firm and reduces revenue per firm by at most  $kl'_F$ . Thus, undercutting with  $l_F > 0$  is never a best response.

2. If  $k = 1$ , it is obvious that undercutting is feasible if  $\tau_L > l_L$ : simply let  $l_F = l_L$  and choose  $\varepsilon$  such that  $\tau_F = \tau_L - \varepsilon \geq l_L$ . To see that the reverse implication holds, suppose that  $\tau_L = l_L$ . In this case  $F$  cannot find a strategy so that the firm  $s = l_L$  prefers the tax and the standard level offered in  $F$  to the ones offered in  $L$  because if  $l_F \neq l_L$ ,  $F$  will have to compensate  $s = l_L$  for more than its standard mismatch meaning that  $F$ 's tax would have to undercut  $L$ 's tax by more than  $|l_L - l_F|$  which is not feasible. Next, fix  $(l_L, \tau_L)$  and let  $(l_F, \tau_F)$  be a feasible undercutting strategy with  $l_F > 0$ . The strategy  $(l'_F, \tau'_F)$  with  $l'_F = 0$  and  $\tau'_F = \tau_F - l_F$  yields the same rent per firm as  $(l_F, \tau_F)$ . Moreover, the fact that every firm preferred  $(l_F, \tau_F)$  to  $(l_L, \tau_L)$  implies that every firm also prefers  $(l'_F, \tau'_F)$  to  $(l_L, \tau_L)$  (all firms  $s < l_F$  strictly prefer  $(l'_F, \tau'_F)$  to  $(l_F, \tau_F)$  and all other firms are indifferent).

3. Suppose that  $k > 1$ . If  $\tau_L > l_L$ , undercutting is obviously feasible. If  $\tau_L = l_L$ , undercutting is not feasible. Take any  $l_F \in [0, 1]$ . We have

$$\tau_F = \tau_L - k|l_L - l_F| - \varepsilon < l_F.$$

Next, we show that if undercutting is feasible then  $(l'_F, \tau'_F) = (l_L, \tau_L - \varepsilon)$  for some  $\varepsilon > 0$ . Assume that  $\tau_L > l_L$ . Take any undercutting strategy with  $l_F \neq l_L$  and a corresponding undercutting tax  $\tau_F = \tau_L - k|l_L - l_F| - \varepsilon$ . Using the same  $\varepsilon$  to undercut, the strategy  $l'_F = l_L$  with undercutting tax  $\tau'_F = \tau_L - \varepsilon$  is feasible. Comparing rents per firm if  $l_F < l_L$ , we get that

$$\begin{aligned} \tau_L - \varepsilon - l_L &> \tau_L - k(l_L - l_F) - \varepsilon - l_F \iff \\ l_L(k - 1) &> l_F(k - 1), \end{aligned}$$

which is true for  $k > 1$ . If  $l_F > l_L$ , we get that

$$\begin{aligned}\tau_L - \varepsilon - l_L &> \tau_L - k(l_F - l_L) - \varepsilon - l_F \iff \\ l_L(1 - k) &> l_F(1 - k),\end{aligned}$$

which is true as well, showing that for any undercutting strategy  $l_F \neq l_L$  there exists another undercutting strategy yielding more rents. ■

In the following, we will deal with the case  $(l_L, \tau_L) \neq (0, 0)$ . If  $(l_L, \tau_L) = (0, 0)$ , by Lemma 23, undercutting is not feasible, and any feasible strategy for  $F$  yields zero rents. By assumption,  $F$  chooses  $(l_F, \tau_F) = (0, 0)$ .

**Lemma 24** 1. *If  $k < 1$ , for any  $(l_L, \tau_L)$ , a sharing strategy is optimal among strategies in sharing I and sharing II only if  $l_F = 0$ .*

2. *If  $k = 1$ , for any  $(l_L, \tau_L)$ , there exists a best response  $(l_F, \tau_F)$  to  $(l_L, \tau_L)$  such that  $l_F = 0$ .*

**Proof.**

1. Take any sharing strategy  $(l_F, \tau_F)$  such that  $0 < l_F \leq l_L$ . Let  $(l'_F, \tau'_F) = (0, \tau_F - kl_F)$ . This strategy is feasible, attracts the same fraction of firms, and  $F$  makes strictly higher rents per firm. Next, take any sharing strategy  $(l_F, \tau_F)$  such that  $l_F > l_L$ . The strategy  $(l'_F, \tau'_F) = (l_F - \varepsilon, \tau_F - \varepsilon)$  such that  $l_F - \varepsilon > l_L$  is feasible for small enough  $\varepsilon$  and yields strictly higher rents for jurisdiction  $F$ .

2. We will proof the statement by showing that for any sharing strategy with  $l_F > 0$  there exists a sharing strategy with  $l_F = 0$  that yields the same rent. Fix  $(l_L, \tau_L)$  and let  $(l_F, \tau_F) \in \text{sharing I}$ . Consider the strategy  $(l'_F, \tau'_F)$  with  $l'_F = 0$  and  $\tau'_F = \tau_F - l_F$ . This strategy yields the same rent per firm, so it suffices to show that the same firms locate in  $F$  under  $((l_L, \tau_L), (l_F, \tau_F))$  as under  $((l_L, \tau_L), (l'_F, \tau'_F))$ . Suppose  $s$  (weakly)



preferred  $F$  to  $L$  under  $((l_L, \tau_L), (l_F, \tau_F))$ . If  $s \leq l_F$ , then

$$|l'_F - s| + \tau'_F = s + \tau_F - l_F \leq s - l_F + \tau_F,$$

so  $s$  (at least weakly) prefers  $(l'_F, \tau'_F)$  to  $(l_F, \tau_F)$ , implying that  $s$  also prefers  $(l'_F, \tau'_F)$  to  $(l_L, \tau_L)$ . If  $s > l_F$ , then

$$|l'_F - s| + \tau'_F = s - l_F + \tau_F,$$

so  $s$  is indifferent between  $(l'_F, \tau'_F)$  to  $(l_F, \tau_F)$ . The proof for  $(l_F, \tau_F) \in \text{sharing II}$  is analogous. ■

**Lemma 25** 1. If  $k < 1$ , the unique rent maximizing sharing strategy for  $F$  is

$$(l_F^s, \tau_F^s) = \begin{cases} (0, \frac{1}{2}\tau_L + \frac{k}{2}l_L) & \text{if } \tau_L \leq 3kl_L \\ (0, \tau_L - kl_L) & \tau_L > 3kl_L \end{cases}.$$

2. If  $k = 1$ , the sharing strategy

$$(l_F^s, \tau_F^s) = \begin{cases} (0, \frac{1}{2}\tau_L + \frac{1}{2}l_L) & \text{if } \tau_L \leq 3l_L \\ (0, \tau_L - l_L) & \tau_L > 3l_L \end{cases}.$$

maximizes rents.

**Proof.**

1. From Lemmas 23 and 24 we know that  $l_F = 0$  at any best response of  $F$ . We will derive the optimal sharing tax and show that there always exists an  $\varepsilon$  such that undercutting yields more rents. Given  $(\tau_L, l_L)$ , government  $F$  faces the following

optimization problem for sharing,

$$\begin{aligned} & \max_{\tau_F} \left\{ \tau_F \left( \frac{\tau_L - \tau_F}{2k} + \frac{l_L}{2} \right) \right\} & ((*) \\ \text{s.t. } & \hat{s}(l_L, \tau_L, 0, \tau_F) \in [0, l_L] \\ & \tau_F \geq 0. \end{aligned}$$

We will ignore the constraints for the moment. The revenue function is strictly concave in  $\tau_F$ , so our solution will be unique and we only need to consider first-order conditions

$$\begin{aligned} \frac{\partial}{\partial \tau_F} \left( \tau_F \left( \frac{\tau_L - \tau_F}{2k} + \frac{l_L}{2} \right) \right) &= \frac{1}{2}l_L - \frac{1}{2k}\tau_F^* + \frac{1}{2k}(\tau_L - \tau_F^*) = 0 \\ \iff \tau_F^* &= \frac{1}{2}\tau_L + \frac{k}{2}l_L. \end{aligned}$$

Obviously,  $\tau_F^* \geq 0$ , so we only need to verify whether  $\hat{s}(l_L, \tau_L, 0, \tau_F^*) \in [0, l_L]$ . We have

$$\hat{s}(l_L, \tau_L, 0, \tau_F^*) = \frac{1}{4k}\tau_L + \frac{l_L}{4},$$

which is strictly larger than zero. But

$$\hat{s}(l_L, \tau_L, 0, \tau_F^*) \leq l_L \iff \tau_L \leq 3kl_L.$$

If  $\tau_L \leq 3kl_L$  one of the constraints binds. Strategies with  $\tau_F = 0$  or  $\hat{s}(\tau_L, l_L, \tau_F, 0) = 0$  yield zero rents. A strategy with  $\hat{s}(l_L, \tau_L, 0, \tau_F) = l_L$ , i.e.  $\tau_F = \tau_L - kl_L$ , yields  $r_F(l_L, \tau_L) = (\tau_L - kl_L)l_L > 0$  if  $l_L > 0$  (if  $l_L = 0$ , then  $\tau_L \leq 3kl_L$  implies  $\tau_L = 0$ , and we do not consider such strategies here).

2. The proof is analogous to the proof of Part 1, except that we do not get uniqueness.

■

**Lemma 26** *If  $k \leq \frac{1}{3}$ , the rent maximizing undercutting strategy yields higher rents than the rent maximizing sharing strategy for all  $(l_L, \tau_L)$  such that  $\tau_L > l_L$ .*

**Proof.** For  $k \leq \frac{1}{3}$ , we have  $\tau_L > 3kl_L$  for all strategies with  $\tau_L > l_L$ . so by Lemma 25 the optimal sharing strategy is  $(l_F^s, \tau_F^s) = (0, \tau_L - kl_L)$ . The corresponding rents are  $r_F^s(l_L, \tau_L) = (\tau_L - kl_L)l_L$  (note that our assumptions imply that, for  $\tau_F$  such that  $\hat{s}(l_L, \tau_L, 0, \tau_F) = l_L$ , all firms  $s \geq l_L$  locate in  $L$ ) Comparing this to the rents from undercutting shows that, for  $\varepsilon$  small enough (notice  $\varepsilon$  depends on  $l_L$ ), undercutting rents are better. If  $l_L < 1$ ,

$$r_F^v(\tau_L, l_L) = (\tau_L - kl_L - \varepsilon) > (\tau_L - kl_L)l_L = r_F^{s1}(\tau_L, l_L)$$

for  $\varepsilon$  sufficiently small. If  $l_L = 1$ , the optimal sharing strategy is in fact an undercutting strategy (it attracts all firms but a set of firms of measure zero). ■

**Lemma 27** *Let  $\frac{1}{3} < k \leq 1$ . For each  $l_L \in [0, 1]$ , there exists a  $\hat{\tau}_L(l_L; k)$  such that  $r_F^s > r_F^u$  for all  $\tau_L \leq \hat{\tau}_L(l_L; k)$  and  $r_F^u > r_F^s$  for all  $\tau_L > \hat{\tau}_L(l_L; k)$ .*

**Proof.** By Lemma 25, if  $\tau_L > 3kl_L$ , then the optimal sharing strategy is not interior, and the proof of Lemma 26 shows that undercutting is better than sharing.<sup>48</sup> It only remains to consider the case  $\tau_L \leq 3kl_L$ . Optimal sharing revenues are given by

$$r_F^s(\tau_L, l_L) = \left( \frac{1}{2}\tau_L + \frac{k}{2}l_L \right) \left( \frac{1}{4k}\tau_L + \frac{l_L}{4} \right).$$

For  $l_L < 1$ , sharing yields more rents than undercutting if and only if

$$\left( \frac{1}{2}\tau_L + \frac{k}{2}l_L \right) \left( \frac{1}{4k}\tau_L + \frac{l_L}{4} \right) > (\tau_L - kl_L - \varepsilon).$$

We now set  $\varepsilon = 0$  and solve for the tax at which both sides are equal. This tax will be the highest tax that  $L$  can set so that  $F$  does not undercut. No matter how small

---

<sup>48</sup>Notice that at  $l_L = 0$ , this always holds so that undercutting is always better.

$F$  sets  $\varepsilon$ , the right hand side will be smaller than the left hand side at this tax. On the other hand, for a tax that is larger than the tax at which both sides are equal,  $F$  can find an  $\varepsilon$  sufficiently small that undercutting yields higher rents than sharing. We solve

$$\begin{aligned} \left(\frac{1}{2}\tau_L + \frac{k}{2}l_L\right) \left(\frac{1}{4k}\tau_L + \frac{l_L}{4}\right) &= (\tau_L - kl_L) \iff \\ kl_L - \tau_L + \frac{1}{8k} (2kl_L\tau_L + \tau_L^2 + k^2l_L^2) &= 0. \end{aligned}$$

The left hand side expression is a quadratic function of  $\tau_L$ . Solving the equation yields two solutions, which we denote by  $\tilde{\tau}_L^1(l_L, k)$  and  $\tilde{\tau}_L^2(l_L, k)$ . They are given by

$$\begin{aligned} \tilde{\tau}_L^1(l_L, k) &= k \left(4 - 4\sqrt{1 - l_L} - l_L\right) \\ \tilde{\tau}_L^2(l_L, k) &= k \left(4 + 4\sqrt{1 - l_L} - l_L\right). \end{aligned}$$

Notice that the factor in front of  $\tau_L^2$  is positive. Sharing revenues are therefore larger than undercutting revenues for  $\tau_L \leq \tilde{\tau}_L^1(l_L, k)$ . Because  $\tilde{\tau}_L^2(l_L, k) > 3kl_L > \tilde{\tau}_L^1(l_L, k)$  undercutting revenues are higher for all  $\tau_L > \tilde{\tau}_L^1(l_L, k)$ . It can be verified that  $\tilde{\tau}_L^1(l_L, k) \leq l_L$  for  $l_L \leq 8k\frac{1-k}{(1+k)^2}$ , and  $\tilde{\tau}_L^1(l_L, k) > l_L$  for  $l_L > 8k\frac{1-k}{(1+k)^2}$  (we omit the derivation). Therefore the critical tax beyond which  $F$  will undercut is given by

$$\hat{\tau}_L(l_L, k) = \begin{cases} l_L & \text{if } l_L \leq 8k\frac{1-k}{(1+k)^2} \\ k(4 - 4\sqrt{1 - l_L} - l_L) & \text{otherwise} \end{cases}.$$

See also Figure 15 in Section 4. ■

**Lemma 28** *Let  $k > 1$ . The strategy that maximizes  $r_F(l_F, \tau_F; l_L, \tau_L)$  over sharing I is given by  $l_F^{s1} = \frac{kl_L + \tau_L}{2(k+1)}$  and  $\tau_F^{s1} = \frac{(\tau_L + kl_L)(k+2)}{2(k+1)}$ . The strategy that maximizes  $r_F(l_F, \tau_F; l_L, \tau_L)$  over sharing II is given by  $l_F^{s2} = \frac{k(1+l_L) - (1+\tau_L)}{2(k-1)}$  and  $\tau_F^{s2} = \frac{(\tau_L - kl_L)(k-2) + k(k-1)}{2(k-1)}$ .*

**Proof.** We start with deriving the optimal sharing strategy over sharing I. Government  $F$ 's problem is

$$\begin{aligned} & \max_{(l_F, \tau_F)} \left\{ (\tau_F - l_F) \left( \frac{\tau_L - \tau_F}{2k} + \frac{l_L + l_F}{2} \right) \right\} \\ & \text{s.t.} \\ & \tau_F \geq l_F \\ & l_F \in [0, l_L) \\ & \tau_F \in [\tau_L - k(l_L - l_F), \tau_L + k(l_L - l_F)]. \end{aligned}$$

Without doing the calculus, we will reduce the optimization problem by first showing that a necessary condition for  $(l_F, \tau_F)$  being a solution to the problem is that  $\tau_F = \tau_L + k(l_L - l_F)$ , i.e., given some  $l_F$ , Government  $F$  will set the highest tax that possibly attracts some firms to its jurisdiction. Take any strategy  $(l_F, \tau_F)$  with  $\tau_F < \tau_L + k(l_L - l_F)$  (notice that these are the strategies that are not at the upper bound of the sharing I set, see also Figure 12). Compare this strategy to another strategy  $(l'_F, \tau'_F)$  with  $l'_F = l_F + \delta$  and  $\tau'_F = \tau_F + \delta$ , where  $\delta > 0$ . For  $\delta$  small enough,  $(l'_F, \tau'_F)$  is in sharing I. This strategy yields the same rents per firm but attracts more firms to  $F$  because

$$\begin{aligned} \hat{s}(l_F, \tau_F, l_L, \tau_L) &< \hat{s}(l'_F, \tau'_F, l_L, \tau_L) \iff \\ \frac{\tau_L - \tau_F}{2k} + \frac{l_L + l_F}{2} &< \frac{\tau_L - \tau'_F}{2k} + \frac{l_L + l'_F}{2} \iff \\ -\tau_F + kl_F &< -\tau_F - \delta + k(l_F + \delta) \iff \\ 1 &< k, \end{aligned}$$

for  $\delta > 0$ .

Therefore, we can reduce  $F$ 's problem to

$$\begin{aligned} & \max_{l_F} \{(\tau_L + k(l_L - l_F) - l_F) l_F\} \\ & \text{s.t.} \\ & l_F \in [0, l_L]. \end{aligned}$$

The objective function is strictly concave in  $l_F$ , so second order conditions will be satisfied, and the maximizer is unique. Ignoring the constraint for the moment and solving for an interior solution yields

$$\begin{aligned} \frac{\partial}{\partial l_F} ((\tau_L + k(l_L - l_F) - l_F) l_F) &= -2l_F k + l_L k - 2l_F + \tau_L = 0 \iff \\ l_F^{s1} &= \frac{k l_L + \tau_L}{2(k+1)}. \end{aligned}$$

Obviously,  $l_F^{s1} \geq 0$ . But

$$\begin{aligned} l_F^{s1} &\leq l_L \iff \\ k l_L + \tau_L &\leq l_L 2(k+1) \iff \\ \tau_L &\leq l_L (k+2). \end{aligned}$$

For higher  $\tau_L$ , sharing with less standard is not the optimal strategy. At the boundary solution  $l_F^{s1} = l_L$  undercutting yields more than sharing. The corresponding tax  $F$  would set would be  $\tau_F = \tau_L = l_L (k+2)$ . By assumption, it would attract half of the firms and therefore

$$\begin{aligned} r_F^{s1}(l_L, \tau_L) &= (\tau_L - l_L) \frac{1}{2} \\ &< (\tau_L - \varepsilon - l_L) \\ &= r_F^u(l_L, \tau_L), \end{aligned}$$

for small enough  $\varepsilon$ . Therefore undercutting is better than the optimal sharing I strategy if  $\tau_L > l_L(k+2)$ .

The strategy that maximizes  $r_F$  over sharing II can be derived analogously. We omit this derivation here, but notice that

$$\begin{aligned} l_F^{s2} &\geq l_L \iff \\ 1 - l_L k - k + \tau_L &\leq l_L(2 - 2k) \iff \\ \tau_L &\leq -1 + 2l_L + k - kl_L. \end{aligned}$$

So again, we get a bound for  $\tau_L$  so that the optimal undercutting strategy yields higher rents than the strategy that maximizes  $r_F$  over sharing II if  $\tau_L$  is larger than this bound. ■

**Lemma 29** *Let  $k > 1$ . For each  $l_L \in [0, 1]$ , there exists a  $\hat{\tau}_L^1(l_L; k)$  such that  $r_F^{s1} > r_F^u$  for all  $\tau_L \leq \hat{\tau}_L^1(l_L; k)$  and  $r_F^u > r_F^{s1}$  for all  $\tau_L > \hat{\tau}_L^1(l_L; k)$ , and a  $\hat{\tau}_L^2(l_L; k)$  such that  $r_F^{s2} > r_F^u$  for all  $\tau_L \leq \hat{\tau}_L^2(l_L; k)$  and  $r_F^u > r_F^{s2}$  for all  $\tau_L > \hat{\tau}_L^2(l_L; k)$ .*

**Proof.** We first derive  $\hat{\tau}_L^1(l_L; k)$ . Suppose  $\tau_L \leq (1 - l_L)(k + 2)$  (recall from the proof of Lemma 28 that this was the upper bound for  $\tau_L$ , so that the constraint  $l_F \leq 1 - l_L$  was not binding). We will derive  $\hat{\tau}_L^1(l_L; k)$  and then verify that it is indeed not larger than this bound, so that undercutting is better than the optimal sharing I strategy for all  $\tau_L > \hat{\tau}_L^1(l_L; k)$ . For given  $(l_L, \tau_L)$  rents from the optimal sharing I strategy are given by

$$r_F^{s1}(l_L, \tau_L) = \frac{1}{4}(k+1)^{-1}(l_L k + \tau_L)^2.$$

The derivation of  $\hat{\tau}_L^1(l_L; k)$  is analogous to the derivation of  $\hat{\tau}_L(l_L; k)$  in the proof of Lemma 27, so we provide less detail. Let  $\varepsilon = 0$ , and set the difference of this rent and undercutting rents equal to zero. We can solve for the highest tax of government

$L$  depending on  $l_L$  such that  $F$  prefers sharing to undercutting<sup>49</sup>:

$$\begin{aligned} r_F^{s1}(l_L, \tau_L) - r_F^u(l_L, \tau_L) &= \frac{1}{4}(k+1)^{-1}(l_L k + \tau_L)^2 - (\tau_L - l_L) = 0 \\ \iff \hat{\tau}_L^1(l_L; k) &= 2 - l_L k + 2k - 2(k+1)\sqrt{1-l_L}. \end{aligned}$$

It can be verified that  $l_L \leq \hat{\tau}_L^1(l_L) \leq l_L(k+2)$ , and therefore, for all  $\tau_L \leq \hat{\tau}_L^1(l_L)$  Government  $F$  prefers the strategy with  $l_F^{s1}$  and the highest sharing tax to the optimal undercutting strategy, and prefers undercutting otherwise.

The derivation for  $\hat{\tau}_L^2(l_L)$  exactly parallels the one for  $\hat{\tau}_L^1(l_L)$ . The rent difference for  $\varepsilon = 0$  can be solved to obtain  $\hat{\tau}_L^2(l_L)$ , which can also be verified to be no smaller than  $l_L$

$$\begin{aligned} r_F^{s2}(l_L, \tau_L) - r_F^u(\tau_L, l_L) &= \frac{1}{4}(k-1)^{-1}(k - l_L k - 1 + \tau_L)^2 - (\tau_L - l_L) = 0 \\ \iff \hat{\tau}_L^2(l_L) &= -z + l_L k + k - 2(k-1)\sqrt{l_L}. \end{aligned}$$

Again we can verify that  $l_L \leq \hat{\tau}_L^2(l_L) \leq -1 + 2l_L + k - kl_L$ , and therefore, for all  $\tau_L \leq \hat{\tau}_L^2(l_L)$  Government  $F$  prefers the strategy with  $l_F^{s2}$  and the highest sharing tax to the optimal undercutting strategy, and prefers undercutting otherwise. ■

With Lemmas 23-29 at hand Proposition 21 can be proved as follows.

**Proof of Proposition 1.**

Part a follows from Lemmas 23, 24, 25, and 26.

Part b follows from Lemmas 23, 24, 25, and 27.

Part c follows from Lemmas 28 and 29.

Proposition 30 restates Proposition 22 with exact expressions for all variables. ■

**Proposition 30** (*The developed country's best response to  $f^*$* )

(a) *If  $k \leq \frac{1}{3}$ , then  $(l_L^*, \tau_L^*) = (0, 0)$ .*

---

<sup>49</sup>As in the proof of Lemma 27, we obtain two solutions but the second one will be larger than  $l_L(2+k)$



$$\begin{aligned}
& (b) \text{ If } \frac{1}{3} < k \leq 1, \text{ then } (l_L^*, \tau_L^*) \\
& = \left( \begin{array}{c} 1 - \left( \frac{1}{3k+3} (4k - \sqrt{3 - 6k + 7k^2}) \right)^2, \\ \frac{2k}{9} (k+1)^{-2} (k^2 + 6\sqrt{7k^2 - 6k + 3} + 2k\sqrt{7k^2 - 6k + 3} + 15) \end{array} \right). \\
& (c) \text{ If } k > 1, \text{ then } (l_L^*, \tau_L^*) = \left( \frac{8}{9}, \frac{4}{3} + \frac{4}{9}k \right).
\end{aligned}$$

**Proof.**

(a) By Proposition 21,  $f^*(l_L, \tau_L) = (0, \tau_F^u(l_L, \tau_L))$  for all  $(l_L, \tau_L) \in S_L \setminus \{(0, 0)\}$ . By assumption,  $f^*(0, 0) = (0, 0)$ . Therefore  $r_L(l_L, \tau_L, f(l_L, \tau_L)) = 0$  for all  $(l_L, \tau_L) \in S_L$ . Using our assumptions again, we obtain  $(l_L^*, \tau_L^*) = (0, 0)$ .

(b) From Proposition 21, we know that, for each level  $l_L$ , Government  $F$  is going to locate at  $l_F = 0$  and undercut if  $\tau_L > \hat{\tau}_L(l_L)$ . Such strategies can therefore not be optimal for government  $L$ , because it can assure itself of positive rents by setting  $l_L = 1$  and  $\tau_L \in (1, 3k)$  (by Lemma 27,  $F$  would choose a sharing strategy in this case). We can also exclude strategies with  $l_L = 0$  as  $F$  is going to undercut then for every positive tax. The reduced optimization problem for  $L$  is therefore

$$\begin{aligned}
& \max_{(\tau_L, l_L)} \{(\tau_L - l_L) (1 - \hat{s}(\tau_L, l_L, \tau_F^s, 0))\} \\
& \text{s.t. } l_L \in [0, 1] \\
& \tau_L \in [l_L, \hat{\tau}_L(l_L)].
\end{aligned}$$

The objective function is continuous and the feasible set is compact. Hence, there exists a solution to the problem. As previously, we will first ignore the constraints, which yields

$$l_L = 4k(k+1)^{-1} > 1.$$

So, an interior solution does not exist. At least one of the four constraints is binding. We can exclude  $\tau_L = l_L$  and  $l_L = 0$  as both strategies yield zero rents.

Case 1) Suppose  $\tau_L = \hat{\tau}_L(l_L)$ . We will derive the optimal  $l_L$  by considering the two cases,  $l_L = 1$  and  $l_L \in (0, 1)$ , separately and then compare the corresponding rents.

(i)  $l_L = 1$

This yields rents of  $r_L(\hat{\tau}_L(1), 0) = 0$  (because  $\hat{s} = 1$ ).

(ii)  $l_L \in (0, 1)$

The maximization problem is

$$\begin{aligned} & \max_{l_L} \\ & \left\{ \left( k \left( 4 - 4\sqrt{1-l_L} - l_L \right) - l_L \right) \left( 1 - \frac{\frac{1}{2} \left( k \left( 4 - 4\sqrt{1-l_L} - l_L \right) - \frac{k}{2}l_L - \frac{l_L}{2} \right)}{2k} \right) \right\} \\ & \text{s.t. } l_L \in (0, 1). \end{aligned}$$

The solution to which is  $l_L^* = 1 - \left( \frac{1}{3k+3} (4k - \sqrt{3 - 6k + 7k^2}) \right)^2 \in (0, 1)$ . It can be verified that at  $l_L^*$  indeed  $\hat{\tau}(l_L^*, k) > l_L^*$  (i.e.,  $l_L^* > \hat{l}_L$ ). We denote the corresponding rents by  $r_L^1(\tau_F^*, 0)$ . They are given by

$$\begin{aligned} & r_L^1(\tau_F^*, 0) \\ & = \left( \frac{2}{27} \right) (k+1)^{-2} \left( 6k + k^2 + 2k\sqrt{7k^2 - 6k + 3} - 3 \right) \left( 4k - \sqrt{7k^2 - 6k + 3} \right) > 0. \end{aligned}$$

Case 2) Consider a strategy with  $l_L = 1$ . Maximizing rents with respect tax yields  $\tau_L^* = \frac{3}{2}k + \frac{1}{2}$ , which is indeed less than  $\hat{\tau}_L(1) = 3k$ . We denote the corresponding rents by  $r_L^2(\tau_F^*, 0)$ . They are given by  $r_L^2(\tau_F^*, 0) = \frac{1}{16}k^{-1} (3k - 1)^2$ .

It can be verified that the inequality

$$r_L^1(\tau_F^*, 0) > r_L^2(\tau_F^*, 0) \iff$$

$$\begin{aligned} & \left( \frac{2}{27} \right) (k+1)^{-2} \left( 6k + k^2 + 2k\sqrt{7k^2 - 6k + 3} - 3 \right) \left( 4k - \sqrt{7k^2 - 6k + 3} \right) \\ & > \frac{1}{16}k^{-1} (3k - 1)^2 \end{aligned}$$

holds. We omit the details.

The corresponding tax for government  $L$  is

$$\tau_L^* = \frac{2}{9} (k+1)^{-2} \left( k^2 + 6\sqrt{7k^2 - 6k + 3} + 2k\sqrt{7k^2 - 6k + 3} + 15 \right) k.$$

(c) By Lemma 29, we know that  $r_F^u(l_L, \tau_L) > r_F^{s1}(l_L, \tau_L)$  if  $\tau_L > \hat{\tau}_L^1(l_L)$  and  $r_F^u(l_L, \tau_L) > r_F^{s2}(l_L, \tau_L)$  if  $\tau_L > \hat{\tau}_L^2(l_L)$ . It follows that  $f^*(l_L, \tau_L) = (l_F^u, \tau_F^u)$  if  $\tau_L > \max[\hat{\tau}_L^1(l_L), \hat{\tau}_L^2(l_L)]$ . If not we, we need to compare  $r_F^{s1}$  with  $r_F^{s2}$ . If  $r_F^{s1} \geq r_F^{s2}$ , the optimal strategy must be  $f^*(l_L, \tau_L) = (l_F^{s1}, \tau_F^{s1})$ , as stated in Lemma 28. If  $r_F^{s2} \geq r_F^{s1}$ , the optimal strategy must be  $f^*(l_L, \tau_L) = (l_F^{s2}, \tau_F^{s2})$ , as stated in Lemma 28.

Turning to  $L$ , we take  $f^*$  as given and first exclude strategies such that  $\tau_L > \max[\hat{\tau}_L^1(l_L), \hat{\tau}_L^2(l_L)]$  as those yield zero rents, while the strategy  $(l_L, \tau_L) = (0, \hat{\tau}_L^2(0))$  yields strictly positive rents (for this choice,  $F$ 's best response is  $(l_F^{s2}, \tau_F^{s2})$ ,  $L$  attracts a positive fraction of firms, and  $\hat{\tau}_L^2(0) > 0$ ). From the reduced set of possibly optimal strategies for  $L$ , we proceed as follows to determine the rent maximizing strategy. First, we show that for  $(l_L, \tau_L)$  with  $l_L \geq \frac{1}{2}$ ,  $(l_F^{s2}, \tau_F^{s2})$  is not a best response for  $F$ . We then separately derive the optimal strategies for Government  $L$  under two different assumptions:

1.  $r_F^{s1} \geq r_F^{s2}$  so that, in the second stage,  $F$  chooses  $(l_F^{s1}, \tau_F^{s1})$  if  $\tau_L \leq \hat{\tau}_L^1(l_L)$  and undercuts otherwise.
2.  $l_L \in [0, \frac{1}{2}]$ ,  $r_F^{s2} \geq r_F^{s1}$ , so that, in the second stage,  $F$  chooses  $(l_F^{s2}, \tau_F^{s2})$  if  $\tau_L \leq \hat{\tau}_L^2(l_L)$  and  $l_L \in [0, \frac{1}{2}]$ , and undercuts otherwise.

We will then show that the optimal strategy under supposition 1 yields more rents than the one under supposition 2, and verify that, under this optimal strategy, government  $F$  indeed sets less standard and sets the highest sharing tax.

To see that, if  $l_L \geq \frac{1}{2}$ , setting more standard and setting the highest sharing tax

can never be the best response for government  $F$ , observe that any strategy  $(l_F, \tau_F)$  with  $l_F > l_L$  and  $\tau_F = \tau_L + k(l_F - l_L)$  is dominated by the strategy  $(l'_F, \tau'_F)$  with  $l'_F = l_L - (l_F - l_L) = 2l_L - l_F$  and  $\tau'_F = \tau_F$ .

1. Now suppose that  $r_F^{s1} \geq r_F^{s2}$ . Under this supposition, government  $L$ 's problem is

$$\begin{aligned} & \max_{(\tau_L, l_L)} \{(\tau_L - l_L) (1 - l_F^{s1})\} \\ & \max_{(\tau_L, l_L)} \left\{ (\tau_L - l_L) \left( 1 - \frac{kl_L + \tau_L}{2(k+1)} \right) \right\} \\ & \text{s.t. } l_L \in [0, 1] \\ & \tau_L \in [l_L, \hat{\tau}_L^1(l_L)]. \end{aligned}$$

Solving for an interior solution yields  $\tau_L = 1 + \frac{1}{2}l_L - \frac{1}{2}l_Lk + k$ . But  $1 + \frac{1}{2}l_L - \frac{1}{2}l_Lk + k \leq \hat{\tau}_L^1(l_L)$  if and only if  $l_L \geq 4\sqrt{3} - 6 \approx 0.9282$ . Hence, the tax constraint binds for all  $l_L \leq 4\sqrt{3} - 6$ . Substituting  $\hat{\tau}_L^1(l_L)$  into the objective function, we solve the following problem

$$\begin{aligned} & \max_{l_L} \left\{ (\hat{\tau}_L^1(l_L) - l_L) \left( 1 - \frac{kl_L + \hat{\tau}_L^1(l_L)}{2(k+1)} \right) \right\} \\ & \text{s.t. } l_L \in [0, 4\sqrt{3} - 6]. \end{aligned}$$

Solving this for  $l_L^*$  yields two solutions,  $l_L^{*1} = \frac{8}{9}$  and  $l_L^{*2} = 0$ . Checking the second-order condition clarifies that only  $l_L^{*1} = \frac{8}{9}$  is a maximizer. For simpler notation we write  $l_L^{*1} = l_L^*$ . Notice that, indeed,  $\frac{8}{9} \leq 4\sqrt{3} - 6$ . The corresponding revenues are given by  $r_L(\hat{\tau}_L^1, l_L^*, \tau_F^{s1}, l_F^{s1}) = \frac{4}{27}(k+1)$ . We also need to verify whether a strategy with  $l_L > 4\sqrt{3} - 6$  and no binding tax constraint yields more revenue. The partial derivative with respect to  $l_L$  is always positive, and therefore government  $L$  wants to set  $l_L$  as high as possible. We only need to check  $l_L = 4\sqrt{3} - 6$ . It can be verified that this strategy does not yield higher rents. The derivation is omitted.

2. Next, suppose that  $r_F^{s2} \geq r_F^{s1}$  for  $l_L \in [0, \frac{1}{2}]$ . Under this supposition, government

$L$ 's problem is

$$\begin{aligned} & \max_{(\tau_L, l_L)} \{(\tau_L - l_L) (1 - l_F^{s2})\} \\ & \max_{(\tau_L, l_L)} \left\{ (\tau_L - l_L) \left( \frac{1 - l_L k - k + \tau_L}{2 - 2k} \right) \right\} \\ & \text{s.t. } l_L \in \left[ 0, \frac{1}{2} \right] \\ & \tau_L \in [l_L, \hat{\tau}_L^2(l_L)] \end{aligned}$$

Solving for an interior solution yields  $\tau_L = \frac{1}{2}k + \frac{1}{2}l_L k - \frac{1}{2} + \frac{1}{2}l_L$ . But  $\frac{1}{2}k + \frac{1}{2}l_L k - \frac{1}{2} + \frac{1}{2}l_L \leq \hat{\tau}_L^2(l_L)$  if and only if  $l_L \leq 7 - 4\sqrt{3} \approx 0.072$ . Hence, the tax constraint binds for all  $l_L \geq 7 - 4\sqrt{3}$ . Substituting  $\hat{\tau}_L^2(l_L)$  into the objective function, we solve the following problem

$$\begin{aligned} & \max_{l_L} \left\{ (\hat{\tau}_L^2(l_L) - l_L) \left( \frac{1 - l_L k - k + \hat{\tau}_L^2(l_L)}{2 - 2k} \right) \right\} \\ & \text{s.t. } l_L \in \left[ 7 - 4\sqrt{3}, \frac{1}{2} \right]. \end{aligned}$$

Solving this for an interior solution yields two solutions,  $l_L^{**1} = 1$  and  $l_L^{**2} = \frac{1}{9}$ . Only the second is a maximizer. Indeed, we have that  $l_L^{**2} = \frac{1}{9} \geq 7 - 4\sqrt{3}$ . For simpler notation, we write  $l_L^{**2} = l_L^{**}$ . Corresponding profits are given by  $r_L(\hat{\tau}_L^2, l_L^{**}, \tau_F^{s2}, l_F^{s2}) = \frac{4}{27}(k - 1)$ . One can also verify that a strategy with  $l_L < 7 - 4\sqrt{3}$  and no binding tax constraint does not yield more revenue. Again, the derivation is omitted.

It is immediate to see that  $L$  prefers the strategy with high standard-provision to the one with low standard provision. It only remains to verify that at this strategy choice of  $L$ , Government  $F$  indeed wants to set less standard level and set the highest firm sharing tax. Since the tax  $L$  sets is, by derivation, the highest one at which  $F$  prefers sharing and less provision to undercutting, we only need to verify that  $F$  does not want to set more standard and share. But we showed already that this cannot be the case since  $l_L^* \geq \frac{1}{2}$ . The optimal strategy for  $L$  is therefore  $(l_L^*, \tau_L^*) = \left(\frac{8}{9}, \frac{4}{3} + \frac{4}{9}k\right)$ .

The derivations showing that the strategy choices, if  $k$  varies, behave in the way as stated in the proposition are omitted (all omitted parts of the proofs are available upon request). ■

Theorem 31 restates Theorem 20 with the exact expressions for all variables.

**Theorem 31** *The outcome of a subgame-perfect equilibrium.*

*The subgame-perfect equilibrium is as follows*

(a) *(Efficient outcome)* If  $k \leq \frac{1}{3}$ , then  $(l_L^*, \tau_L^*) = (0, 0)$ , and  $(l_F^*, \tau_F^*) = (0, 0)$ , and  $\hat{s}^* = \frac{1}{2}$ .

(b) *(Standard haven)* If  $\frac{1}{3} < k \leq 1$ , then  $l_L^* \geq \frac{8}{9}$ ,  $\tau_L^* \in (l_L^*, 2l_L^*)$ , and  $l_F^* = 0$ ,  $\tau_F^* = \frac{2}{3} (\sqrt{7k^2 - 6k + 3} - k + 3) (k + 1)^{-1}$   $k \in (\frac{3}{4}, \frac{4}{3})$ , and  $\hat{s}^* = \frac{1}{3} (\sqrt{7k^2 - 6k + 3} - k + 3) (k + 1)^{-1} > \frac{2}{3}$ .

(c) *(Race to the top)* If  $k > 1$ , then  $l_L^* = \frac{8}{9}$ ,  $\tau_L^* = \frac{4}{3} + \frac{4}{9}k > 2l_L^*$ , and  $l_F^* = \frac{2}{3}$ ,  $\tau_F^* = \frac{4}{3} + \frac{2}{3}k > 3l_F^*$ , and  $\hat{s}^* = \frac{2}{3}$ .

**Proof.** The subgame perfect equilibrium strategy for Government  $L$  is the one derived in Proposition 22. For Government  $F$  the outcome is obtained by plugging  $(l_L^*, \tau_L^*)$  into  $f^*$  as specified in Proposition 21. It is straightforward to verify that, for part b, the taxes lie indeed in the specified range. The equilibrium marginal type of firm is obtained by plugging the equilibrium strategies into  $\hat{s}(l_L, \tau_L, l_F, \tau_F)$ . Plugging all values into the rent functions yields the corresponding rents. For parts b and c, simple comparison shows that the follower makes higher rents. It is straightforward to verify that  $\hat{s}^* > \frac{2}{3}$  in part b. ■

## CHAPTER V

### LOBBYING IN A NETWORK

#### Introduction

Our actions and decisions are influenced by the actions and decisions of those around us. This effect is empirically well documented. A doctor's adoption of a new drug depends on her connections to other doctors who have already adopted the drug (Coleman et al., 1957). People adopt new technologies that their friends and colleagues have adopted (Rogers 2003). They are more likely to commit crimes if many around them commit crimes (Glaeser et al., 1996). Teenagers are more likely to use drugs, drink alcohol, smoke cigarettes, drop out of school, and go to church if their peers do so (Gaviria and Raphael, 2001). Even a woman's contraceptive decision can be influenced by contraceptive decisions of other women she knows (Kohler et al., 2001). Finally, our views and opinions are shaped by interactions with family, friends, colleagues, and others with whom we interact (Zuckerman et al., 1994). At the same time, there are agents (individuals, groups, institutions, etc.) that seek to reduce the number of crimes committed, influence which technology is adopted or which party we vote for. This paper examines optimal strategies of such an interested agent, assuming that the agent anticipates how individuals affect each other (or, how behavior is diffused). I assume that the agent is able to convince individuals through direct interaction and that resource constraints limit the number of individuals that he can convince. Thus, the agent's problem is to determine a set of individuals with maximum impact on others.

For focus, I assume that there is a lobbyist who wants to convince a group of voters to vote for a proposal. Each voter has an initial opinion. Over time, opinions evolve through interactions with others as well as with the lobbyist. Interactions

between voters are represented by a social network. If two voters are linked they have an impact on each other's views on the proposal. More specifically, if a sufficient number of a voter's neighbors are in favor of the proposal at period  $t - 1$ , the voter will be in favor of the proposal in period  $t$ . Which number is sufficient depends on an idiosyncratic cutoff. For motivation, imagine that in each period voters poll their neighbors and determine their next-period opinion based on the poll and their own current opinions.

Due to the impact voters have on each other's opinions, the lobbyist's activities can have a direct as well as an indirect effect. A voter's opinion is directly affected if the lobbyist approaches her. A voter's opinion is indirectly affected if she interacts with voters whose opinions were affected by the lobbyist or through interactions with neighbors of voters whose opinions were affected by the lobbyist, and so on.

Taking the empirical observation that individuals are influenced by their interactions as a basic premise, to gain insight into the diffusion of behavior, the theoretical literature has primarily focused on network models. Typically, the literature studies models in which each player in the network chooses whether to adopt an action. The basic primitives of the models are the network, the diffusion rule at the local level (such as the voters' cutoff rule in this paper), and in some cases, the objectives of external agents (such as the lobbyist in this paper). The network models can be divided roughly according to two criteria: whether agents in the networks are strategic, and whether the model is dynamic. I will now discuss two closely related papers that address questions similar to those in this paper. More discussion as to how the results in this paper relate to the literature more broadly will be provided throughout the paper.

A closely related work that follows a nonstrategic-dynamic approach is Kempe et al. (2003, 2005). This work studies models with two types of diffusion rules: a



threshold model and a cascade model.<sup>50</sup> In the threshold model, the effect individual  $i$  has on individual  $j$  is given by a parameter  $b_{ij}$ . Each individual has a threshold, a random draw from a distribution, and adopts the action if the total effect of his neighbors who adopted the action exceeds his threshold. The cascade model is a model with a probabilistic diffusion rule in which the probability that  $j$  adopts the action given that  $i$  adopted it is given by a parameter  $p_{ij}$ . The authors are interested in finding a set of individuals, called a *seed*, of size  $k$  that maximizes the expected spread. However, they show that this is an *NP*-hard problem.<sup>51</sup> In view of this result, they develop an algorithm that finds a seed with an expected spread of at least 63% of the optimal spread. In each step of the algorithm the seed is enlarged by adding the individual with the largest marginal impact (a hill-climbing strategy). This paper takes a different approach. It shows that by considering a restricted class of networks, it is possible to obtain an exact characterization of the optimal seed.

Another closely related paper is Caillaud and Tirole (2008). As in my paper, a sponsor of a proposal tries to persuade a group to vote for the proposal by targeting voters that have maximum impact on others. However, in their paper voters affect each other through information flows. Voters are uncertain about their payoffs if the proposal is passed. The sponsor can resolve a voter's uncertainty by sending the voter a report. Voters are assumed to share reports with each other. Since the voters' payoffs are correlated, learning about other voters' payoffs leads a voter to update his belief about his own payoff. These features make indirect persuasion possible.<sup>52</sup> However, Caillaud and Tirole do not consider a network structure, implicitly assuming that all voters can and in fact do communicate with each other. Therefore, the sponsor's optimal strategy depends only on the voters' initial opinions. In contrast,

---

<sup>50</sup>See Kleinberg (2007) for a survey of these and related papers, for example by Morris (2000).

<sup>51</sup>NP-hardness measures the computational complexity of a problem. Loosely speaking, a problem is NP-hard if it is not possible to program an algorithm that can solve the problem within a reasonable time (at least no such algorithm has been found yet for any NP-hard problem).

<sup>52</sup>The appendix of this paper contains a behavioral background model that tells a similar story.

in this paper network structure plays an important role: whether a lobbyist targets a voter depends both on the voter's initial opinion as well as on his position in the network.

I also follow a nonstrategic-dynamic approach. Even though insights can be gained from a static setup, the dynamic aspect is important as diffusion is inherently a dynamic phenomenon. I use the label "nonstrategic" to refer to an approach where individuals in the network use a decision rule that is exogenous to the model. Here, this approach is chosen for reasons of tractability. The appendix contains a behavioral model that generates the "updating rule" voters are assumed to use. Thus the decision rule can be viewed as the reduced form of that behavioral model. A model of strategic interaction that would generate the same type of behavior appears in Morris (2000), where individuals in a network play coordination games with each of their neighbors. The corresponding best-response dynamics lead to a similar cutoff rule as the one used by voters in this paper (see the discussion at the end of Section 3 for a more details).

The focus of this paper is on so-called threshold networks. These are networks in which some individuals have many connections and others have relatively few connections. I show that the lobbyist's optimal strategy follows a simple rule. In any period in which the lobbyist's resources permit her to approach at most  $c$  voters, it is optimal to approach the  $c$  voters with the highest number of links that currently oppose the proposal (Proposition 33). This result confirms the general intuition that better connected people are more influential.

While intuitive, the result is not obvious. In arbitrary networks, to determine who is influential, it is not sufficient to take into account only the impact individuals have on their neighbors. Second-, third-, and higher-order effects of an individual's influence trickling through the network have to be considered as well. The idea is that it does not only matter how many connections a person has, but also whether

this person is connected to other important people. An additional complication is that the lobbyist's problem is to find a most influential group of individuals. Since a group's impact does not necessarily equal the sum of its members' impacts, the lobbyist's optimal strategy need not be monotone in the constraint. That is if  $c < c'$ , then the set of size  $c$  of the jointly most influential people need not be a subset of the set of size  $c'$  of the jointly most influential people. My results show that these kinds of considerations can be ignored in the class of threshold networks.

While the main result (Proposition 33) holds for a restricted class of networks, it is obtained for a class of the lobbyist's objective function where all that is assumed is that an additional supporter at any point in time does not hurt the lobbyist's payoff. Moreover, it also holds that the optimal strategy remains essentially the same, even if the lobbyist's resources change over time.

Section 3.2 focuses on the special case where the lobbyist's goal is to bring about a unanimous decision in favor of the proposal. Using the result on optimal lobbying strategies from the previous section, I obtain upper bounds on the number of voters that need to be convinced directly to obtain unanimous support (Proposition 34). If voters support the proposal whenever at least a certain number of their neighbors support it, the upper bound does not change with the size of the network. Thus, as the network grows the fraction of voters that need to be convinced shrinks. If voters support the proposal whenever at least a certain fraction of their neighbors support it, the bound grows with the size of the network. So in this case, as the network grows the fraction of voters that need to be convinced remains constant. I also show that it is more difficult for the lobbyist to convince more tightly connected groups (Proposition 35).

Section 4 examines the process of opinion formation in arbitrary networks (without any external agent trying to influence opinion). After a finite number of periods, opinions become periodic. Each individual's opinion either remains unchanged or

oscillates between two opinions. Thus there are voters who continuously support or oppose the proposal and there are "swing voters" who keep changing their minds.

The remainder of the paper is organized as follows: Section 2 describes the model; Section 3 examines lobbying in threshold networks; Section 4 examines the opinion formation process in arbitrary networks; Section 5 concludes.

## The Model

A lobbyist tries to convince a group of voters  $N = \{1, 2, \dots, n\}$  to vote for a proposal, on which the voters vote at time  $t = d < \infty$ . Voter  $i$ 's opinion at time  $t$  is denoted by  $p_i^t$ , an element of the set  $P = \{-1, +1\}$ , where if  $p_i^t = -1$ , voter  $i$  opposes the proposal, and if  $p_i^t = +1$ , voter  $i$  supports the proposal. The corresponding vector of opinions at time  $t$  is denoted by  $p^t$ , an element of  $P^n$ . Let  $\pi = \{p^0, p^1, p^2, \dots\} \in (P^n)^{\mathbb{N}_0}$  denote a sequence of opinions. The lobbyist's payoff is  $u(\pi)$ . Since the lobbyist pursues the acceptance of the proposal,  $u$  is assumed to be a non-decreasing function of  $\pi$ , that is  $\pi' \geq \pi$  implies  $u(\pi') \geq u(\pi)$ . This payoff function accommodates a number of different situations. Three examples are

$$\begin{aligned}
 (1) \quad u(\pi) &= f\left(\sum p_i^d\right), \text{ where } f \text{ is an increasing function;} \\
 (2) \quad u(\pi) &= \begin{cases} q & \text{if } p_1^d = p_2^d = +1, \text{ and } |\{i : p_i^d = +1\}| \geq k \\ 0 & \text{otherwise,} \end{cases} \\
 &\text{where } q > 0; \\
 (3) \quad u(\pi) &= \sum_{t=0}^{\infty} \phi_t \sum_{i=1}^n w_i p_i^t, \\
 &\text{where } 0 \leq \phi_t < 1, \text{ and } 0 \leq w_i < \infty.
 \end{aligned}$$

In the first two specifications the lobbyist cares only about the outcome at the time of the vote. In (1) the lobbyist's payoff depends solely on the difference between positive and negative votes. As an example, suppose the group decides by majority

vote and the lobbyist cares only about whether the proposal is accepted.<sup>53</sup> Because the number of voters is fixed and abstaining from voting is not permitted, (1) also fits cases where the lobbyist cares only about attaining a certain number of positive votes. In (2) the lobbyist's payoff depends on whether she succeeds to convince voters 1 and 2 in addition to convincing at least  $k$  voters (the UN security council, for example, adopts a proposal if it gets affirmative votes by each of the permanent members and at least nine affirmative votes in total). In (3) the lobbyist's payoff depends on the entire sequence of opinions. Each voter's opinion is weighted by some nonnegative factor. For example, some members of the group might not be allowed to vote, so their opinion has no direct influence on the lobbyist's payoff. In addition, the lobbyist weighs the outcome according to time.<sup>54</sup> For example, the lobbyist might only care about the opinions leading up to period  $d$  (so  $\phi_t = 0$  for  $t > d$ ).<sup>55</sup>

Voters' opinions evolve over time through interactions with the lobbyist and with each other. Each period until the vote takes place in  $t = d$ , the lobbyist can approach and directly influence the opinions of at most  $c < n$  voters.<sup>56</sup> Here,  $c$  can be thought of as the result of some kind of constraint, for example, a time constraint. A strategy for the lobbyist is a list  $(S^1, \dots, S^d) \in \mathcal{S}^d$  where  $\mathcal{S} = \{S \subseteq N : |S| \leq c\}^d$ . If opinions at the beginning of period  $t$  are given by  $p^{t-1}$  and the lobbyist approaches voters in

---

<sup>53</sup>So the function  $f$  would be of the form  $f(\sum v_i) = 0$  if  $\sum v_i \leq 0$  and  $f(\sum v_i) = q > 0$  if  $\sum v_i > 0$ .

<sup>54</sup>To make sure (3) is well-defined, assume that  $\phi_t$  is bounded away from zero.

<sup>55</sup>Having the payoff function depend not only on the outcome in period  $d$  is of particular interest for other applications of the model. For example, an organization that wishes to reduce the number of drug users in a community cares about drug usage at each point in time, not only at a specific date. The influencing agent (a government, a lobbyist, a teacher, a non-profit organization, etc.) might have a finite or infinite horizon and value short-term outcomes more than long-term outcomes or vice versa. All these cases can be accommodated.

<sup>56</sup>The constraint  $c$  could change over time, but that does not change the qualitative nature of any of the results. See Footnote 61.

$S^t$ , opinions adjust to

$$p^{t-1}[S^t] = \begin{cases} -1 & \text{if } p_i^{t-1} = -1 \text{ and } i \notin S^t \\ +1 & \text{else.} \end{cases}$$

Therefore, the lobbyist has the ability to sway a voter's opinion, at least temporarily, in favor of the proposal.

Interactions among voters are captured by a social network, a collection of links  $g \subseteq \{ij : i, j \in N, i \neq j\}$ . If  $ij \in g$ , voters  $i$  and  $j$  interact and therefore influence each other's opinions. Voter  $i$ 's *neighborhood* in the network is the set  $N_i = \{j : ij \in g\}$  and his *degree* is the number of his neighbors  $n_i = |N_i|$ ; his closed neighborhood is the set  $\bar{N}_i = N_i \cup \{i\}$ . In each period  $t$ , first the lobbyist approaches voters in  $S^t$ . After that voters exchange their opinions, which are at that point given by  $p^{t-1}[S^t]$ . The outcome of that opinion exchange is determined by the function  $\Delta_i : P^n \rightarrow P$ , where

$$p_i^t = \Delta_i(p^{t-1}[S^t]) = \begin{cases} -1 & \text{if } \sum_{j \in \bar{N}_i} p_j^{t-1}[S^t] < \alpha_i \\ +1 & \text{if } \sum_{j \in \bar{N}_i} p_j^{t-1}[S^t] \geq \alpha_i, \end{cases}$$

where  $\alpha_i \in \mathbb{R}$ . Also, let  $\Delta = (\Delta_i)_{i \in N}$ .<sup>57</sup> In the model presented here, voters are not optimizers. A model in which voters optimize a payoff function and which generates the decision rule  $\Delta_i$  is detailed in the appendix. There, voters are uncertain about their payoffs from implementing the proposal. Their payoffs could be higher or lower than their status quo payoffs (the payoffs they obtain if the proposal is rejected), depending on the future state of the world. The belief about the future state of the world is influenced by conversations with network neighbors. The more of a voter's neighbors root for the proposal, the stronger the belief that the state of the world

---

<sup>57</sup>In the setting of opinion formation it is appropriate to assume that individuals can costlessly change their opinions in each time period. In other settings, such as adoption of a technology or formation of habits, this assumption is only appropriate in the long-run (that is, when each time period is sufficiently long).

will be favorable for implementing the proposal.

Notice that the cutoffs  $\alpha_i$ , which determine when a voter switches from one opinion to another, can vary across individuals. One might interpret small cutoffs as leaning towards the proposal and large cutoffs as leaning against the proposal. For example, if  $\alpha_i$  is large, in order to be persuaded,  $i$  needs to have many more neighbors who agree with the proposal than neighbors who disagree with it. In the extreme cases,  $i$  never supports the proposal (if  $\alpha_i$  is larger than  $n_i + 1$ ) or never opposes it (if  $\alpha_i$  is smaller than  $-n_i$ ). Note also that the model is consistent with absolute cutoffs, which are independent of the size of a voter's neighborhood, as well as relative cutoffs, which depend on the size of a voter's neighborhood.

The lobbyist maximizes her payoff by choosing a feasible lobbying strategy. Thus, her optimization problem is

$$\max_{S \in \mathcal{S}^d} u(\pi).$$

The main parameters of this problem are the initial opinions  $p^0$ , the resource constraint  $c$ , the network  $g$ , and the cutoffs  $\alpha$ .

### **Lobbying in a Threshold Network**

This section examines lobbying in a class of networks called threshold networks. A *threshold network* is a network that can be constructed starting with a single vertex and sequentially adding vertices in one of the following two ways:

1. The added vertex becomes an isolated vertex of the existing network.
2. The added vertex forms a link with every existing vertex of the network.

Let  $\{I, K\}$  be a partition of the set of vertices  $N$  (or, here, the set of voters) so that  $I$  contains all vertices added using Step 1 and  $K$  contains all vertices added using Step 2. The "starting vertex" is contained in  $K$  if the next vertex is added by using

Step 2, and is contained in  $I$  otherwise. Note that no two vertices in  $I$  are linked, and every two vertices in  $K$  are linked ( $K$  is a *clique*). Also note that every individual in  $K$  has at least as many neighbors as any other individual in  $I$ . Loosely speaking, a group with the structure of a threshold network has two types of individuals, well connected ones and poorly connected ones. An example of a threshold network is given in Figure 19. Vertices are added in the order indicated by their labels. Starting from vertex 1, vertices 2, 3, 4, 6 and 7 are added as isolated vertices, and vertices 5 and 8 form links with every previously added vertex. Here,  $I = \{1, 2, 3, 4, 6, 7\}$  and  $K = \{5, 8\}$ .

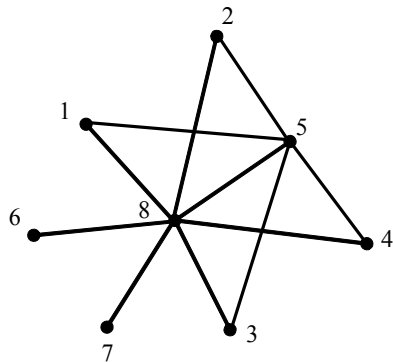


Figure 19: A threshold network

Special cases of threshold networks include the empty network, the complete network, the star network and the interlinked star network.<sup>58</sup> Except for the empty network, one would expect to see such networks in relatively small groups, where every pair of individuals either knows each other or has a common acquaintance. This seems also a reasonable assumption for groups with hierarchies that have only

---

<sup>58</sup>An interlinked star network has two sets of vertices. The first is completely connected and the second forms an independent set, in which everybody is connected to everyone in the first set.



two layers. Indeed, for a variety of settings, the theoretical literature on network formation has found that equilibrium outcomes of network formation games display core-periphery patterns with short average distances (see for example, Galeotti et al., 2006; Goyal and Joshi, 2003; Hojman and Szeidl, 2008).

Let  $I = \{1, 2, \dots, x\}$  and  $K = \{x + 1, x + 2, \dots, n\}$  so that in  $I$ , vertices with smaller labels were added later and in  $K$ , vertices with higher labels were added later. Henceforth, voters in  $N = I \cup K$  are assumed to be labeled in that way. The following facts about threshold networks will be useful later on. The proof of Fact 4 is given in Footnote 10. All other proofs are straightforward and therefore omitted.

**Fact 1.** If  $i < j$ , then  $n_i \leq n_j$ ;

**Fact 2.** If  $i \in I$ , then  $N_i \subseteq K$ ;

**Fact 3.** If  $i \in K$ , then  $K \subseteq \bar{N}_i$ ;

**Fact 4.**  $n_x < n_{x+1}$ .<sup>59</sup>

### The main result

As was mentioned, it will be shown that it is optimal for the lobbyist to approach each period the  $c$  best connected voters that currently oppose the proposal (Proposition 33). To formally describe the lobbyist's strategy, requires some additional notation. Let  $\delta_1 < \delta_2 < \dots < \delta_L$  be the distinct degrees occurring in  $g$ . Let  $D^l = \{i \in N : n_i = \delta_l\}$ . The collection  $\mathcal{D} = \{D^1, D^2, \dots, D^L\}$  is the *degree partition* of  $g$ .<sup>60</sup> Let  $N_-^t = \{i \in N : p_i^t = -1\}$  and let  $N_+^t = \{i \in N : p_i^t = +1\}$ . Let  $\mathcal{D}_-^t = \{D_-^{1,t}, D_-^{2,t}, \dots, D_-^{L,t}\}$  where  $D_-^{l,t} = D^l \cap N_-^t$ . Define  $\hat{l}_t$  to be the smallest integer such

---

<sup>59</sup>**Proof of Fact 4.**

Either  $x$  or  $x + 1$  is the starting vertex. Case 1:  $x$  is the starting vertex. Then  $n_{x+1} \geq n - (x + 1) + 2 = n - x + 1$ , because there were at least two existing vertices before  $x + 1$  was added. Because  $x$  is the starting vertex, the upper bound of neighbors of  $x$  is  $n - x$ , the number of vertices in  $K$ .

Case 2:  $x + 1$  is the starting vertex. In this case  $n_{x+1} = n - x - 1$ , the number of vertices in  $K$  minus 1. Because the second vertex added is added in Step 2 (vertex  $x + 2$ ), the first vertex added in Step 1, which is  $x$ , does not connect to  $x + 1$  or  $x + 2$ . Therefore,  $n_x \leq n - x - 2$ .

<sup>60</sup>Note that in a threshold network, if  $i, j \in D^l \subseteq I$ , then  $N_i = N_j$ , and if  $i, j \in D^l \subseteq K$ , then  $\bar{N}_i = \bar{N}_j$ .

that  $|D_-^{\hat{l}_t, t}| + \dots + |D_-^{L, t}| \leq c$ . For  $t = 0, 1, \dots, d - 1$ , let

$$\mathcal{S}^{t+1*} = \begin{cases} \left\{ S \in \mathcal{S} : S = D_-^{\hat{l}_t, t} \cup \dots \cup D_-^{L, t} \cup D_-^t \text{ where} \right. \\ \left. D_-^t \subseteq D_-^{\hat{l}_t, t} \text{ with } |D_-^t| = c - \left( |D_-^{\hat{l}_t, t}| + \dots + |D_-^{L, t}| \right) \right\} & \text{if } \hat{l}_t \geq 2 \\ N_-^t & \text{else} \end{cases}$$

$\mathcal{S}^{t+1*}$  is the collection of sets of  $c$  voters that have the largest number of connections among those who oppose the proposal at the beginning of period  $t + 1$ . Since individuals can have the same number of connections, there might be more than one such strategy. If less than  $c$  voters oppose the proposal the strategy simply prescribes to approach every voter who opposes the proposal.

As is shown in Proposition 33, a strategy  $(S^1, S^2, \dots, S^d) \in \mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  maximizes  $u(\pi)$  over  $\mathcal{S}^d$ . The key condition that leads to this result is that threshold networks have nested neighborhoods. Since this property is crucial, it is stated and proven in the next lemma.

**Lemma 32** *Let  $g$  be a threshold network with  $I = \{1, \dots, x\}$  and  $K = \{x + 1, \dots, n\}$ , where  $x \in N \cup \{0\}$  (with the understanding that if  $x = 0$ , then  $I = \emptyset$ , and if  $x = n$ , then  $K = \emptyset$ ). The neighborhoods of individuals in  $N$  are nested as follows*

$$N_1 \subseteq N_2 \subseteq \dots \subseteq N_x \subseteq \bar{N}_{x+1} \subseteq \dots \subseteq \bar{N}_n.$$

**Proof.**

Case (1):  $i, j \in I$

Suppose that  $i < j$ , and that  $k \in N_i$ . By Fact 2,  $N_i \subseteq K$ . This means that  $k \in K$  and that  $k$  is added to the network after  $i$ . However, since  $j$  is added before  $i$  (remember that in  $I$  individuals with smaller labels are added *later*),  $k$  is also added after  $j$ . Therefore,  $k \in N_j$ , showing that  $N_i \subseteq N_j$ .

Case (2):  $x \in I$  and  $x + 1 \in K$

The result follows from the fact that  $x \in I$ , so  $N_x \subseteq K$  (by Fact 2), and  $x + 1 \in K$ , so  $K \subseteq \bar{N}_{x+1}$  (by Fact 3).

Case (3):  $i, j \in K$

Suppose that  $i < j$ , and that  $k \in \bar{N}_i$ . If  $k \in K$ , then  $k \in \bar{N}_j$  because  $K \subseteq \bar{N}_j$  by Fact 3. If  $k \in I$ , then  $k$  must be added to the network before  $i$ . However, since  $j$  is added after  $i$ , this means that  $k$  is also added before  $j$ , so  $j$  connects to  $k$ . Thus  $k \in \bar{N}_j$ , showing that  $\bar{N}_i \subseteq \bar{N}_j$ . ■

The converse of Lemma 32 holds as well. If a network's neighborhoods can be nested in the way described above, then it is a threshold network. The fact that it is optimal for the lobbyist to convince a set of  $c$  voters that have the highest number of connections among the opposition relies crucially on the nestedness of neighborhoods. Intuitively, it ensures that convincing a well connected individual is always at least as effective as convincing a less well connected individual.

The following assumption on the cutoff  $\alpha_i$  is needed to show that strategies of the type described above are optimal.

**Assumption 1.** *For each  $i \in N$ ,  $\alpha_i \leq n_i - 1$*

The assumption says that if, after lobbying, all of a voter's neighbors are in favor of the proposal in period  $t - 1$ , the voter will be in favor of the proposal in period  $t$ .

**Proposition 33** *Under the above assumption, any strategy in  $\mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  is optimal for the lobbyist.*

The proof of Proposition 33 is long and appears in the appendix. Note that Proposition 33 does not rule out that other strategies than the ones in  $\mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  are optimal. However, this can only happen in two cases. First, the strategies in  $\mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  might be "stronger" than what is needed to achieve a certain outcome (the trivial case being that everyone already is in favor of the proposal to start with, so that every strategy yields the same result). Second, strategies in

$\mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  might achieve more than what the lobbyist wants to achieve (for example, suppose the lobbyist needs to convince 10 voters to get the proposal passed and does not care about additional votes).

The following two examples illustrate optimal lobbying strategies and indirect persuasion effects.

**Example 1.**

Figure 20 shows the same threshold graph as Figure 19 with the vertices labeled in ascending order of their degrees. Let  $p^0 = (-1, -1, +1, +1, -1, -1, +1, -1)$  be the voters' initial opinions. For simplicity assume that  $\alpha_i = 0$  for all  $i$ ,  $c = 2$ , and  $d = 2$ . It follows from Proposition 33 that  $(S^{1*}, S^{2*}) = (\{6, 8\}, \emptyset)$  is an optimal strategy for the lobbyist. Approaching voters 6 and 8 in  $t = 1$ , yields  $p^0[\{6, 8\}] = (-1, -1, +1, +1, -1, +1, +1, +1)$  and then  $p^1 = \Delta(p^0[\{6, 8\}]) = (+1, +1, +1, +1, +1, +1, +1, +1)$ . Thus there is no use for lobbying in  $t = 2$ , and the final outcome is  $p^2 = \Delta(p^1[\emptyset]) = (+1, +1, +1, +1, +1, +1, +1, +1)$ . Without lobbying, opinions in  $t = 2$  would be  $p^2 = (-1, -1, +1, +1, -1, -1, +1, -1)$ . Even though the lobbyist approaches only two voters in period 1, in period 2 five more voters are in favor of the proposal than would have been without any lobbying. Here, voters 1, 2, and 5 have been persuaded indirectly.

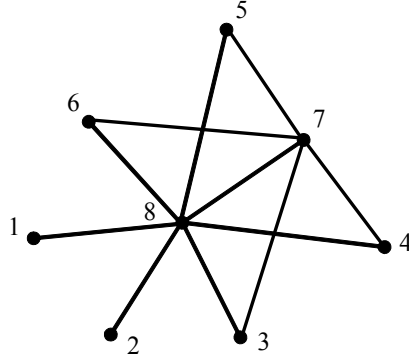


Figure 20: Relabeling of nodes in a threshold network

**Example 2.**

This example illustrates that the lobbyist might visit a voter multiple times. This can only happen if a voter keeps switching back to opposing the proposal. However, lobbying need not necessarily be ineffective in this case. At first, a lobbyist might be able to convince a particular voter only temporarily, but this can translate into long-term indirect persuasion effects on this and other voters. For example, suppose that in the threshold network in Figure 21 (for which  $I = \{1, 2, 3, 4, 5, 6\}$  and  $K = \{7, 8, 9\}$ ) all voters initially oppose the proposal and that  $c = 2$  and  $d = 3$ . Let the cutoffs be  $\alpha_9 = -5$ ,  $\alpha_1 = \alpha_2 = 0$ ,  $\alpha_3 = \alpha_4 = 1$ ,  $\alpha_5 = \alpha_6 = \alpha_7 = 2$ , and  $\alpha_8 = 4$ . According to Proposition 33, the lobbyist's optimal strategy for the first three periods

is  $(S^{1*}, S^{2*}, S^{3*}) = (\{8, 9\}, \{7, 8\}, \{7, 8\})$ , yielding the following sequence of opinions:

$$\begin{aligned}
 p^0 &= (-1, -1, -1, -1, -1, -1, -1, -1, -1), \\
 p^0 [S^{1*}] &= (-1, -1, -1, -1, -1, -1, -1, +1, +1), \\
 p^1 &= (+1, +1, +1, +1, -1, -1, -1, -1, +1), \\
 p^1 [S^{2*}] &= (+1, +1, +1, +1, -1, -1, +1, +1, +1), \\
 p^2 &= (+1, +1, +1, +1, +1, +1, -1, -1, +1), \\
 p^2 [S^{3*}] &= (+1, +1, +1, +1, +1, +1, +1, +1, +1), \\
 p^3 &= (+1, +1, +1, +1, +1, +1, +1, +1, +1).
 \end{aligned}$$

Voters 7 and 8 are visited multiple times, voter 8 in each of the three periods. Even though lobbying has no lasting effect on them at first, it has an indirect and lasting effect on voters 3, 4, and 9 and then also 5 and 6. In turn, the changes in their opinions help persuading 7 and 8 permanently.

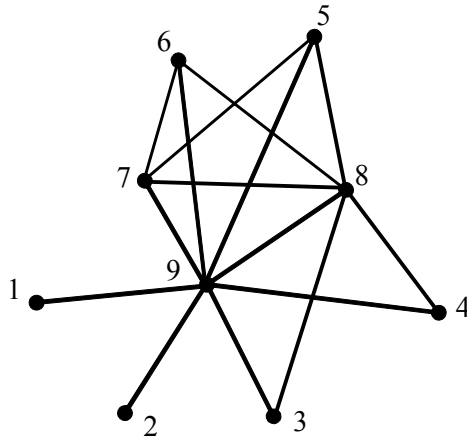


Figure 21: Threshold network of Example 2

Note again that Proposition 33 holds for a very broad class of the lobbyist's

objective. Moreover, the result can be generalized to a case where the lobbyist’s resource changes over time.<sup>61</sup> The main contribution of Proposition 33, however, is that it explicitly characterizes the set of voters with maximum indirect persuasion effects on other voters. Because determining the most influential group is *NP*-hard in general networks (see Kempe et al., 2003; Ballester et al., 2009), the literature has focused on obtaining other types of results. Kempe et al. (2003, 2005) and Ballester et al. (2009) construct algorithms that search for approximately influential groups and provide bounds on the goodness of the approximation. In an earlier paper, Ballester et al. (2006) are able to determine the single most influential player in their model (which is quite different from this paper’s model; in their model, individuals in the network play a one-shot game). Galeotti and Goyal (2007) build a tractable model by assuming that the targeting problem consists of deciding on what large of a fraction of individuals to target. The influencing agent is a firm that decides on the level of advertising. A continuum of consumers forms a network of which the firm knows only the degree distribution. The firm chooses an action  $x \in [0, 1]$ , thought of as the fraction of consumers it reaches through costly advertising. Caillaud and Tirole (2008), as discussed in the introduction, determine which voters are targeted by the sponsor of a proposal, but abstract from a network structure. As a result, who is targeted by the sponsor depends solely on the voters’ initial stand on the matter.<sup>62</sup> In contrast, Proposition 33 shows that the network structure matters for the question of who is influential.

### **Bringing about unanimous decisions**

How much does it take to bring about unanimous support for the proposal? And how does the answer to this question depend on the group structure and the voters’

---

<sup>61</sup>In this case, if the lobbyist can approach  $c_t$  voters in period  $t$ , it is optimal to visit the  $c_t$  best connected voters that currently oppose the proposal.

<sup>62</sup>Their model is more refined in other aspects. For example, a voter’s initial opinion is a point in the unit interval, and the voters’ influence on each other is generated through a combination of uncertainty about the proposal and correlation in payoffs.

cutoffs? This subsection addresses these questions. Since the voters' initial opinions will affect the answers and dilute comparisons across network structures, I assume that all voters initially oppose the proposal. This approach can also be interpreted as a worst-case scenario for the lobbyist. Any other set of initial opinions constitutes a more favorable environment for her. To emphasize that the lobbyist has to plant proponents of the proposal in the network, a strategy in this subsection is called a "seed." For simplicity, I examine cases where all voters have the same absolute or relative cutoff, and where the lobbyist has only one shot at seeding the network (therefore, superfluous time superscripts are dropped). Proposition 34 provides upper bounds on the minimum size of a seed that suffices to turn everybody into a supporter. Proposition 35 shows that it can only hurt the lobbyist if connections are added to the network while preserving its threshold property.

A *complete seed* is a strategy  $S \subseteq N$  that leads to unanimous support at some time  $t < \infty$ . A seed  $S$  is a *minimum complete seed* if there is no seed  $S'$  with  $|S'| < |S|$  that is also a complete seed. Consider two scenarios: (a) Every voter supports the proposal if and only if at least  $\theta_a \in \mathbb{N}_0$  individuals in his closed neighborhood support the proposal; (b) Every voter supports the proposal if and only if at least a fraction  $\theta_r$  of his closed neighborhood supports the proposal. Let  $\bar{n}_i = n_i + 1$  be the size of  $i$ 's closed neighborhood. Scenario (a) translates into the cutoffs  $\alpha_i = \min\{\theta_a - (\bar{n}_i - \theta_a), n_i - 1\} = \min\{2\theta_a - \bar{n}_i, n_i - 1\}$  for all  $i \in N$ , where the upper bound  $n_i - 1$  is a result of the above assumption). Let  $c(g, \theta_a)$  and  $c(g, \theta_r)$  denote the sizes of minimum complete seeds in, respectively, scenarios (a) and (b). Let  $\lceil \theta_r n \rceil$  denote the smallest integer at least as large as  $\theta_r n$ .

**Proposition 34** *Let  $g$  be a threshold network. Under the above assumption,  $c(g, \theta_a) \leq \theta_a$  and  $c(g, \theta_r) \leq \lceil \theta_r n \rceil$ .*

**Proof.** By Proposition 33, we can consider seeds of the form  $S = \{i, i + 1, \dots, n\}$ .



(a) Let  $S = \{n - \theta_a + 1, n - \theta_a + 2, \dots, n\}$ . For any  $i$ , we have

$$\sum_{j \in \bar{N}_i} p_j^0[S] = \min\{\theta_a - (\bar{n}_i - \theta_a), n_i - 1\} = \alpha_i.$$

Thus  $\Delta_i(p^0[S]) = +1$  for all  $i \in N$ , so  $S$  is a complete seed, showing that  $c(g, \theta_a) \leq \theta_a$ .

(b) Consider the seed  $S = \{n - \lceil \theta_r n \rceil + 1, n - \lceil \theta_r n \rceil + 2, \dots, n\}$ . If  $i \in K$ , we have  $\bar{N}_i = \{j, j + 1, \dots, n\}$  for some  $j \in N$ . Thus, at least a fraction  $\theta_r$  of  $i$ 's closed neighborhood is in  $S$ , so  $\Delta_i(p^0[S]) = +1$  for all  $i \in K$ . If  $i \in I$ , we have  $\bar{N}_i = \{i\} \cup \{j, j + 1, \dots, n\}$  for some  $j \in N$ . If  $j < n - \lceil \theta_r n \rceil + 1$  at least a fraction  $\theta_r$  of  $i$ 's closed neighborhood is in  $S$ . If not, all of  $i$ 's neighbors are in  $S$ . Either way,  $\Delta_i(p^0[S]) = +1$  for all  $i \in I$ . Therefore,  $S$  is a complete seed, showing that  $c(g, \theta_r) \leq \lceil \theta_r n \rceil$ . ■

The bounds on minimum complete seeds provide measures of the ease of diffusion in the network. Similar measures are examined by Morris (2000) and Lopez-Pintado (2008). In Morris (2000), networked players choose whether to adopt an action in a coordination game. Coordinating on taking the action yields a payoff of  $1 - q$  and coordinating on not taking the action yields a payoff of  $q \in (0, 1)$ . Players make their decisions according to whatever is the best response to their neighbors' decisions in the previous period. As mentioned in the introduction, this model generates a cutoff rule of the kind that is assumed here. The best-response dynamics of the coordination game lead players to adopt the action whenever at least a fraction  $q$  of their neighbors adopted the action in the previous period. Morris defines the "contagion threshold," to be the maximum  $q$  such that there exists a finite seed that eventually leads the entire population (which is countably infinite) to adopt the action. His main finding is that the diffusion threshold is at most  $1/2$  for any network. Lopez-Pintado (2008) analyzes diffusion in networks with a probabilistic diffusion rule (as in the cascade model by Kempe et al., 2003, discussed in the introduction). As in Morris (2000) the population is countably infinite. She derives a closed-form solution for the "diffusion

threshold," a minimum bound on the effective (local) spreading rate such that starting with an infinitely small seed ultimately leads to a positive fraction of adopters.

The bounds  $\theta_a$  and  $\lceil \theta_r n \rceil$  are attained in a complete network of size at least  $\theta_a + 1$  in (a) and of size  $n$  in (b). This suggests that it might be harder to bring about unanimous support in better connected threshold networks. Proposition 35 shows that this is the case if the goal is to generate complete support after one period. Let  $c(g, \alpha, t)$  be the size of a minimum complete seed that generates full support in period  $t$  in a threshold network with cutoffs  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ .

**Proposition 35** *Let  $g$  and  $g'$  be two threshold networks such that  $g \subseteq g'$ . Under the above assumption,  $c(g, \alpha, 1) \leq c(g', \alpha, 1)$ .*

**Proof.** Note that, when moving from network  $g'$  to  $g$ , the order of individuals in terms of their degrees might change. Let  $N' = \{1, 2, \dots, n\}$  be such that the labels are in ascending order according to degrees in  $g'$ . Let  $f : N' \rightarrow N$  be a one-to-one and onto function that produces a relabeling so that  $n_{f(1)} \leq n_{f(2)} \leq \dots \leq n_{f(n)}$  in  $g$ .

Let  $S' = \{\hat{i}, \hat{i} + 1, \dots, n\}$  be a minimum complete seed for  $g'$  that generates full support in  $t = 1$ . Consider the seed  $S = \{f(\hat{i}), f(\hat{i} + 1), \dots, f(n)\}$  for  $g$ . Let  $\bar{N}_{f(i)} = \{f(h), f(h + 1), \dots, f(n)\} \cup \{f(i)\}$  be  $i$ 's closed neighborhood in network  $g$ , and let  $\bar{N}'_i = \{h - k, h - k + 1, \dots, n\} \cup \{i\}$  be  $i$ 's closed neighborhood in network  $g'$  (because  $g \subseteq g'$ , we know that  $k \geq 0$ ). If  $\hat{i} \leq h$ , then  $S \subseteq \bar{N}_{f(i)}$  and, by the assumption,  $p_i^1 = \Delta_i(p^0[S]) = +1$  in  $g$ . If  $h < \hat{i}$ , then  $S' \subseteq \bar{N}'_i$  and  $S \subseteq \bar{N}_{f(i)}$ . Because  $g \subseteq g'$ , the degree of any  $i \in N'$  in  $g'$  is at least as large as that individual's degree in  $g$ , that is,  $n_{f(i)} \leq n'_i$  for all  $i \in N'$ . It follows that

$$\begin{aligned} \sum_{j \in \bar{N}_i} p_j^1[S] &= n - \hat{i} + 1 - (n_{f(i)} + 1 - (n - \hat{i} + 1)) \\ &\geq n - \hat{i} + 1 - (n'_i + 1 - (n - \hat{i} + 1)) = \sum_{j \in \bar{N}'_i} p_j^1[S']. \end{aligned}$$

So if  $\Delta_i(p^0[S']) = +1$  in  $g'$ , then  $\Delta_i(p^0[S]) = +1$  in  $g$ , showing that  $S$  is a complete seed for  $g$  that generates full support in  $t = 1$ . Thus  $c(g, \alpha, 1) \leq c(g', \alpha, 1)$ . ■

Proposition 35 says that more tightly connected groups are harder to convince in the short run. This might not seem intuitive at first. To see why the result holds, consider a threshold network and a complete seed that generates full support after one period. By Proposition 33, it is without loss of generality to consider a seed that contains the voters with the highest number of connections. Now remove a link between  $i$  and  $j$ . To maintain the threshold property,  $i$  and  $j$  must be the lowest labeled neighbors for each other. Consider  $i$ . If  $j$  is not in the seed, removing the link cannot hurt. If  $j$  is in the seed, so is every other of  $i$ 's neighbors and after removing the link, again all of  $i$ 's neighbors are in the seed. So, the above assumption,  $i$  supports the proposal in period 1. A symmetric argument holds for  $j$ .

The question of whether better connected networks help diffusion is also addressed by Jackson and Yariv (2007). In their model, players choose whether to adopt an action in strategic interactions with others in the network. They only know the degree distribution but not the details of the network structure. Accordingly, Jackson and Yariv examine the effects of shifts in the network's degree distribution on the fraction of adopters. They find that an upward shift in the degree distribution facilitates diffusion of the action if actions are strategic complements. This contrasts with Proposition 35. Even though voters here are not strategic, if anything, their actions are complements. Yet, better connected networks hinder diffusion.<sup>63</sup>

## Opinion Formation in General Networks

This section focuses on the process of opinion formation in arbitrary networks in the absence of lobbying. A slightly more general model is analyzed, in which voters

---

<sup>63</sup>Galeotti and Goyal (2007, 2008) also address what happens in their model as the degree distribution shifts, but their results are not directly comparable to Proposition 35 as their model is too different from this paper's model.

are allowed to be indifferent (and can abstain from voting). Voters either agree with the proposal (+1), disagree with it (-1), or are indifferent to it (0). In some settings, such as adoption of a new technology, there is no natural interpretation of abstaining, so that a more accurate model would be the one analyzed so far. This is not a problem since all results in this section extend straightforwardly when the possibility of indifference is removed. All notation is adjusted in the obvious way ( $P = \{-1, 0, +1\}$ ,  $N_0^t = \{i \in N : p_i^t = 0\}$ , etc.). In each period, a voter agrees (disagrees) with the proposal if in the previous period in  $i$ 's closed neighborhood the difference between proponents and opponents of the proposal was sufficiently large (small). Otherwise, the voter is indifferent to the proposal and intends to abstain from voting. This process is captured by the function  $\Delta_i : \{-1, 0, +1\}^n \rightarrow \{-1, 0, +1\}$ , where

$$p_i^t = \Delta_i(p^{t-1}) = \begin{cases} -1 & \text{if } \sum_{j \in \bar{N}_i} p_j^{t-1} < \alpha_i \\ 0 & \text{if } \alpha_i \leq \sum_{j \in \bar{N}_i} p_j^{t-1} \leq \beta_i \\ +1 & \text{if } \beta_i < \sum_{j \in \bar{N}_i} p_j^{t-1}, \end{cases}$$

and where  $\alpha_i \leq 0 \leq \beta_i$ . In this section, no further assumption is made on the cutoffs.

The process of opinion formation does not necessarily reach a stable state, that is, there might not exist a  $t < \infty$  such that  $\Delta^t(p^0) = \Delta^{t+1}(p^0)$ . To see this, consider the four voters in the network in Figure 22. Their initial opinions are the ones indicated in the figure.

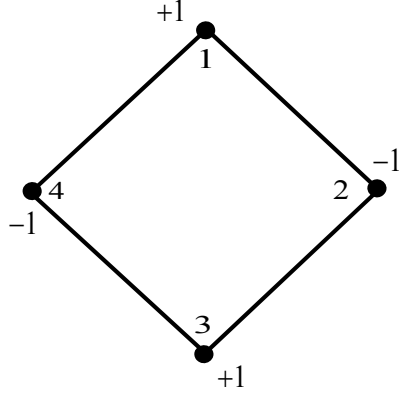


Figure 22: A stable state might not exist

Let  $\alpha_i = -\frac{1}{2}$  and  $\beta_i = \frac{1}{2}$  for all  $i \in N$ . Thus, in  $t = 1$ , voters 1 and 3 will disagree with the proposal while voters 2 and 4 will agree with the proposal, and in  $t = 2$  the situation reverses. Thus, we get

$$\begin{aligned}
 & p^0, p^1, p^2, p^3, \dots \\
 = & (+1, -1, +1, -1), (-1, +1, -1, +1), (+1, -1, +1, -1), (-1, +1, -1, +1), \dots
 \end{aligned}$$

All voters' opinions switch back and forth between agreement and disagreement, indefinitely.

The following proposition is an extension of a result on generalized threshold function in Goles and Olivos (1980). It asserts that, while a stable state might not exist, eventually opinions become periodic, with the period being either one or two. The proof of Proposition 36 appears in the appendix.

**Proposition 36** *There exists  $\hat{t} < \infty$  such that either  $\Delta^t(p^0) = \Delta^{t+1}(p^0)$  for all  $t \geq \hat{t}$  or  $\Delta^t(p^0) = \Delta^{t+2}(p^0)$  for all  $t \geq \hat{t}$ .*

Thus, from time  $\hat{t}$  onwards, there are three (possibly empty) stable groups: Supporters, opponents, and a group of indifferent voters. In addition, there might be a

fourth group, which one might label undecided or "swing voters," whose opinions keep swinging back and forth. There is a notion of swing voters being the ones sought after in election campaigns. It would be interesting to see how, depending on the network structure, targeting swing voters affects the other groups in the network.

A remark on the robustness of Proposition 36 is in order. The result carries over to settings in which each link  $ij \in g$  has a weight, measuring to which degree  $i$  and  $j$  influence each other.<sup>64</sup> It is also possible to introduce a weight on own opinion - one might imagine that an individual's own current opinion weighs more than the opinions of others - and still obtain the above result.<sup>65</sup>

However, when influence is asymmetric ( $i$ 's impact on  $j$  is not the same as  $j$ 's impact on  $i$ ), Proposition 36 does not hold. An example is given in Figure 23 below. The network consists of three individuals with influences as indicated in the figure. For simplicity, suppose that an individual's own current opinion does not have an impact on the individual's opinion in the next period. Further, suppose that  $\alpha_i = \beta_i = 0$  for  $i = 1, 2, 3$ , and that  $p^0 = (-1, +1, 0)$ . We get

$$\begin{aligned}
 p^1 &= (+1, -1, +1) \\
 p^2 &= (+1, +1, -1) \\
 p^3 &= (-1, 0, 0) \\
 p^4 &= (0, -1, +1) \\
 p^5 &= (+1, +1, -1).
 \end{aligned}$$

Hence, the system becomes periodic in period 2 with a period of length 3, which shows that the conclusion of the proposition does not hold.

---

<sup>64</sup>In the proof a link is represented by  $g_{ij} = 1$ . Since the proof does not depend on the value of  $g_{ij}$ , the result extends straightforwardly.

<sup>65</sup>Proofs of these claims are available upon request.

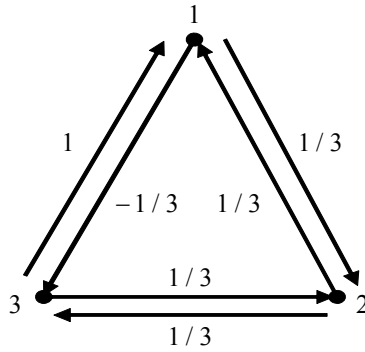


Figure 23: A network with asymmetric influences

As discussed, it is hard to say something about the outcome of opinion formation in general networks, without making any kind of concession in other parts of the model (see the discussion of the literature at the end of Section 3.1). Nonetheless, it might be possible to obtain results that improve the understanding of diffusion in networks. In the following I will informally discuss such a result.

Recall that a clique is a completely connected group. What seems to be a fruitful approach to deal with arbitrary networks is to examine a network’s clique structure (the analysis in Chwe (2000) points in that direction). Each network can be viewed as a structure of cliques that have overlaps.<sup>66</sup> Thus diffusion could be imagined as something that is passed from one group to another whenever their overlap is sufficiently big. This idea can be formalized by constructing a directed network of cliques (so that each clique is a node of the network). A directed link from clique  $a$  to clique  $b$  means that whenever clique  $a$  has adopted some behavior so does  $b$ . Notice that the link should be directed because even if  $b$  follows  $a$ , it might not be the case that  $a$  follows  $b$ . More specifically, this would happen if the intersection of  $a$  and  $b$

---

<sup>66</sup>For the reader who is interested in the formalization of this idea, here are the formal definitions of the underlying concepts. A *clique* is a complete subgraph of  $g$ . A clique  $h$  is *maximal in  $g$*  if there does not exist a clique  $h'$  such that  $h \subset h'$ . The *union* of two graphs  $g$  and  $g'$  with vertex sets  $N$  and  $N'$  is the graph  $g \cup g'$  on vertex set  $N \cup N'$ . The collection of cliques  $\mathcal{H} = \{h_1, h_2, \dots, h_K\}$  is a *clique cover* of  $g$  if it contains all maximal cliques of  $g$  and if  $\cup_{h \in \mathcal{H}} h = g$ . Every graph has a unique clique cover.

constitutes a large fraction of  $b$  but only a small fraction of  $a$  (in other words,  $a$  is a bigger group than  $b$ ).

In such a directed network of cliques one can then identify those cliques or cycles of cliques that have no arcs pointing toward them, but only paths emanating from them. In graph-theoretic terminology, those cliques are the sources of the graph. A path emanating from a source can be interpreted as a hierarchy of cliques. Individuals belonging to cliques closer towards the tail of a path are early adopters and individuals belonging to cliques closer to the head of a path are later adopters. For somebody who wants to maximize the spread of a certain behavior those cliques should be of particular interest.

## Conclusion

This paper examines networks with a core-periphery structure and shows that the group of most influential individual in the network can be explicitly determined. The story told here was one of lobbying but the model is more broadly applicable. Here, the lobbyist is shown to target those group members that have high degrees and oppose the proposal. Using this result, bounds are provided on the number of voters that have to be convinced to eventually reach unanimous support. These bounds depend on whether cutoffs are absolute or relative. In the case of absolute cutoffs, the bound depends only on the value of the cutoff, but not on the size of the network. In addition, it is shown that, for any cutoffs, more tightly connected groups are harder to convince, at least in the short-run.

The process of opinion formation in any network is shown to become periodic eventually, with individuals either holding the same opinion each period or switching back and forth between two opinions. Thus, in the periodic state there might be stable groups of supporters, opponents, and indifferent individuals, as well as a group of undecided, or swing voters.



There are two natural extensions of this paper. One is to pursue further the understanding of diffusion in general networks. The second is to examine what happens when there are two agents with opposing interests. For example, it would be interesting to see what happens when two lobbyists (one lobbying in favor of the proposal, the other lobbying against the proposal) try to persuade a group. These extensions are left for future work.

## Appendix

### A behavioral model of voters

The following is a behavioral model that generates the behavior of voters in the paper. Suppose that whether the proposal benefits a voter depends on the future state of the world  $X \in \{L, H\}$ , which is unknown to the voters. Let  $A$  stand for an outcome where the proposal is accepted and  $R$  for an outcome where the proposal is rejected. Suppose the payoffs for voter  $i$ , depending on approval or rejection of the proposal and of the state of the world, are

$$\begin{aligned} u_i(A | H) &= H_i, \\ u_i(A | L) &= L_i, \text{ and} \\ u_i(R | H) &= u_i(R | L) = 0, \end{aligned}$$

where  $L_i \leq 0 \leq H_i$ . For  $i \in N$ , let  $\bar{n}_i = n_i + 1$ , and let  $\bar{n}_i^t(A)$  and  $\bar{n}_i^t(R)$  be the numbers of voters in  $i$ 's closed neighborhood who, respectively, accept and reject the proposal in period  $t$ . Let  $i$ 's belief in period  $t + 1$  be  $P_i^{t+1}(X = H) = \frac{\bar{n}_i^t(A)}{\bar{n}_i}$ . Thus  $i$  will be in favor of the proposal in period  $t + 1$  if and only if

$$\begin{aligned} P_i^{t+1}(X = H)u_i(A | H) + P_i^{t+1}(X = L)u_i(A | L) &\geq 0 \\ \Rightarrow \bar{n}_i^t(A) &\geq \frac{-L_i}{H_i - L_i}\bar{n}_i \equiv \hat{n}_i, \end{aligned}$$

which translates into a threshold of  $\alpha_i = \hat{n}_i - (\bar{n}_i - \hat{n}_i) = 2\hat{n}_i - \bar{n}_i = \frac{H_i + L_i}{L_i - H_i} \bar{n}_i$ .

## Proofs

The following lemma is useful to prove Proposition 33.

**Lemma 37** *If  $p \geq p'$ , then  $\Delta^t(p) \geq \Delta^t(p')$  for any  $t = 1, 2, \dots$*

**Proof.** If  $p \geq p'$ , then  $\sum_{j \in \bar{N}_i} p_j \geq \sum_{j \in \bar{N}_i} p'_j$  for all  $i$ , which implies that  $\Delta_i(p) \geq \Delta_i(p')$ . Therefore, at any time  $t$ , we have  $\Delta^t(p) \geq \Delta^t(p')$ . ■

**Proof of Proposition 33.** (If  $g$  is a threshold network, then any strategy in  $\mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  is optimal for the lobbyist.)

An optimal strategy maximizes  $u(\pi)$ . The proof proceeds by showing that a strategy in  $\mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  generates a  $\pi^*$  such that  $\pi^* \geq \pi$  (and therefore  $u(\pi^*) \geq u(\pi)$ ) for any  $\pi$  generated by some other strategy. Step 1 below shows that if  $\pi^*$  is the sequence of opinions resulting from a strategy  $(S^{1*}, S^{2*}, \dots, S^{d*}) \in \mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  and  $\pi$  is the sequence of opinions from a strategy  $(S^1, S^2, \dots, S^d)$ , then  $\pi^* \geq \pi$ . Step 2 shows that if  $\pi^*$  and  $\pi^{**}$  result from two different strategies in  $\mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$ , then  $\pi^* = \pi^{**}$ .

*Step 1.* Fix a strategy  $(S^{1*}, S^{2*}, \dots, S^{d*}) \in \mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$  and a strategy  $(S^1, S^2, \dots, S^d) \notin \mathcal{S}^{1*} \times \mathcal{S}^{2*} \times \dots \times \mathcal{S}^{d*}$ , and let  $\pi^*$  and  $\pi$  be the corresponding sequences of opinions. In view of Lemma 37 and since lobbying only takes place until period  $d$ , it is sufficient to show that  $p^{t*} \geq p^t$  for  $t = 1, \dots, d$ . The proof of this part is by induction on  $t$ .

(i)  $t = 1$

To show that  $p^{1*} \geq p^1$ , is equivalent to showing that  $\Delta(p^0[S^{1*}]) \geq \Delta(p^0[S^1])$ . If  $|N_-^0| \leq c$ , it is clear that  $S^{1*} \cup N_+^0 = N$ , and by the assumption,  $\Delta_i(p^0[S^{1*}])$  for all  $i$ , so this case is trivial. For ease of exposition, I drop the superscript 1 from the strategies in this part of the proof (so  $S^{1*} = S^*$  and  $S^1 = S$ ). If  $|N_-^0| > c$ , then  $|S^*| = c$ , and, without loss of generality, I consider only alternative strategies  $S$  such that  $S \subseteq N_-^0$

and  $|S| = c$ . Thus the cardinality of both  $S^*$  and  $S$  is  $c$ , so that  $|S^* \setminus S| = |S \setminus S^*|$ . Now, let  $|S \setminus S^*| = R$  and move from  $S$  to  $S^*$  in  $R$  steps by successively replacing one individual in  $S \setminus S^*$  by an individual in  $S^* \setminus S$ . Denote the set obtained after  $r$  steps by  $S^r$  with  $S^0 = S$  and  $S^R = S^*$ . Furthermore, let  $i_r$  denote the individual removed from  $S^{r-1}$  in step  $r$ , and let  $f(i_r)$  denote the individual that is added to  $S^{r-1}$  in the same step. This can be done in a way such that at each step  $r$ ,  $i_r$  is the individual with the lowest label in  $S^{r-1} \setminus S^*$  and  $f(i_r)$  is the individual with the highest label in  $S^* \setminus S^{r-1}$ . Therefore  $f : S \setminus S^* \rightarrow S^* \setminus S$  is a one-to-one and onto mapping with  $f(i) > i$  for all  $i \in S \setminus S^*$ .<sup>67</sup> We have  $S^r = (S \setminus \{i_1, \dots, i_r\}) \cup (\{f(i_1), \dots, f(i_r)\})$ , for  $r = 1, \dots, R$ . I will show that  $p^0[S^*] \geq p^0[S]$  by showing  $p^0[S^{r+1}] \geq p^0[S^r]$  for  $r = 1, \dots, R - 1$ . So consider  $S^r$  and  $S^{r+1}$  for an arbitrary  $r = 1, \dots, R - 1$ .

Case (1)  $i_{r+1} \in K$

Because  $f(i_{r+1}) > i_{r+1}$ , it also holds that  $f(i_{r+1}) \in K$ , and therefore, by Lemma 32,  $\bar{N}_{i_{r+1}} \subseteq \bar{N}_{f(i_{r+1})}$ . Since  $S^r$  and  $S^{r+1}$  only differ in that  $S^{r+1}$  contains  $f(i_{r+1})$  instead of  $i_{r+1}$ , this implies that  $p^0[S^{r+1}] \geq p^0[S^r]$ .

Case (2)  $i_{r+1} \in I$

In this case  $N_{i_{r+1}} \subseteq \bar{N}_{f(i_{r+1})}$ , again by Lemma 32, so the only person for whom we could have  $p_i^0[S^r] = +1$  and  $p_i^0[S^{r+1}] = -1$  is  $i_{r+1}$  itself. This can happen only if  $i_{r+1} \notin N_{f(i_{r+1})}$ . So suppose that this is the case. Note that for any  $i \in I$  we have  $N_i = \{r, r + 1, \dots, n\}$  for some  $r$ . Therefore, since  $i_{r+1} \notin N_{f(i_{r+1})}$ , we have  $N_{i_{r+1}} = \{r, r + 1, \dots, n\}$  where  $r > f(i_{r+1})$ . However, for every  $j > f(i_{r+1})$  we have, by construction,  $j \in S^{r+1}$  or  $j \in N_+$ . It follows then from the assumption that  $p_i^0[S^{r+1}] = +1$ .

Thus  $p^0[S^*] \geq p^0[S]$ , implying that, reverting to the original notation,  $\Delta(p^0[S^{1*}]) \geq \Delta(p^0[S^1])$ .

---

<sup>67</sup>Here the assumption that  $S^1 \notin S^{1*}$  is needed.

(ii)  $1 < t \leq d$

Now suppose that  $p^{\hat{t}^*} \geq p^{\hat{t}}$  for  $\hat{t} = 1, \dots, t-1$ . We need to show that  $p^{t^*} \geq p^t$ . Consider the sequence of lobbying  $(S^{1^*}, \dots, S^{t-1^*}, S^t)$ , and let  $\tilde{p}^t$  be the corresponding vector of opinions in period  $t$ . Part (i) of this proof implies that  $\tilde{p}^t \leq p^{t^*}$ . Moreover, from the induction hypothesis, we know that  $p^{t-1^*} \geq p^{t-1}$ . Thus, we get  $\sum_{j \in \bar{N}_i \setminus S^t} p_j^{t-1^*} + \sum_{j \in \bar{N}_i \cap S^t} 1 \geq \sum_{j \in \bar{N}_i \setminus S^t} p_j^{t-1} + \sum_{j \in \bar{N}_i \cap S^t} 1$  for all  $i \in N$ , implying that  $\tilde{p}^t \geq p^t$ . Putting everything together, we get  $p^t \leq \tilde{p}^t \leq p^{t^*}$ .

Thus  $p^{t^*} \geq p^t$  for  $t = 1, \dots, d$ . Together with Lemma 37, this shows that  $\pi^* \geq \pi$ .

*Step 2.* Let  $(S^{1^*}, S^{2^*}, \dots, S^{d^*})$  and  $(S^{1^{**}}, S^{2^{**}}, \dots, S^{d^{**}})$  be two strategies in  $\mathcal{S}^{1^*} \times \mathcal{S}^{2^*} \times \dots \times \mathcal{S}^{d^*}$  and let  $\pi^*$  and  $\pi^{**}$  be the resulting sequences of opinions. Consider  $S^{1^*}$  and  $S^{1^{**}}$ . They can only differ on the set of  $c - (|D_-^{\hat{t}}| + \dots + |D_-^L|)$  individuals in  $D_-$ . However, for any pair  $i, j \in D_-$ , we have  $\bar{N}_i = \bar{N}_j$ . Therefore, each individual is equally affected by the strategy  $S^{1^*}$  and by the strategy  $S^{1^{**}}$ , and therefore  $p^{1^*} = p^0[S^{1^*}] = p^0[S^{1^{**}}] = p^{1^{**}}$ . This argument can be repeated an arbitrary number of times, showing that  $\pi^* = \pi^{**}$ .

Combining Step 1 and 2 together shows that  $\pi^* \geq \pi$  for all  $\pi$  generated from some strategy  $(S^1, S^2, \dots, S^d) \in \mathcal{S}^1 \times \mathcal{S}^2 \times \dots \times \mathcal{S}^d$ . ■

**Proof of Proposition 36.** (There exists  $\hat{t} < \infty$  such that either  $\Delta^t(p^0) = \Delta^{t+1}(p^0)$  for all  $t \geq \hat{t}$  or  $\Delta^t(p^0) = \Delta^{t+2}(p^0)$  for all  $t \geq \hat{t}$ .)

The proof is an extension of a proof on generalized threshold functions in Goles and Olivos (1980). Let  $P$  denote the set of all possible opinion vectors, i.e.  $P = \{-1, 0, +1\}^n$ . Consider  $(p^0, \Delta(p^0), \Delta^2(p^0), \dots) = (p^0, p^1, p^2, \dots)$ . Since  $p^j \in P$  for all  $j = 0, 1, \dots$  and  $P$  is a finite set, the sequence of opinions becomes periodic after a finite number of steps. Let  $\hat{t}$  be the first period in which the sequence becomes periodic. Let  $z > 0$  be the least period, i.e.,

$$\Delta^{t+z}(p^0) = \Delta^t(p^0) \text{ and } \Delta^{t+r}(p^0) \neq \Delta^t(p^0) \text{ for } r = 1, 2, \dots, z-1, \text{ for all } t \geq \hat{t}.$$

Define the  $n \times z$  matrix

$$\begin{aligned} X(p^0, z) &= \begin{bmatrix} x_1(0) & \cdots & x_1(z-1) \\ \vdots & & \vdots \\ x_n(0) & \cdots & x_n(z-1) \end{bmatrix} \\ &= \left( \Delta^{\hat{t}}(p^0), \Delta^{\hat{t}+1}(p^0), \dots, \Delta^{\hat{t}+z-1}(p^0) \right). \end{aligned}$$

For every  $i \in N$  and every  $k \in \{0, 1, \dots, z-1\}$ ,  $x_i(k)$  indicates the state of voter  $i$  in the  $k^{\text{th}}$  time period following  $\hat{t}$ . Note that, for  $i \in N$ ,

$$(i) \quad x_i(0) = \Delta_i(x_1(z-1), \dots, x_n(z-1))$$

and

$$(ii) \quad x_i(l+1) = \Delta_i(x_1(l), \dots, x_n(l)) \quad \text{for } l = 0, 1, \dots, z-2.$$

Let  $\zeta_i$  be the smallest period of row  $x_i$  (note that  $\zeta_i$  must be a divisor of  $z$ ). Let  $S$  be the set of rows of  $X(p^0, z)$ , and let  $g_{ij} = 1$  if  $ij \in g$ , and  $g_{ij} = 0$  otherwise. Let  $x_i(l) = x_i((l) \bmod z)$  and define

$$L : S \times S \rightarrow R$$

where

$$L(x_i, x_j) = g_{ij} \sum_{l=0}^{z-1} (x_j(l+1) - x_j(l-1)) x_i(l)$$

The following three facts are useful and are proven separately.

1.  $L(x_i, x_j) + L(x_j, x_i) = 0$  for all  $i, j \in N$ .

Straightforward algebra yields

$$\begin{aligned}
& L(x_i, x_j) + L(x_j, x_i) \\
= & g_{ij} \sum_{l=0}^{z-1} (x_j(l+1) - x_j(l-1)) x_i(l) \\
& + g_{ij} \sum_{l=0}^{z-1} (x_i(l+1) - x_i(l-1)) x_j(l) \\
= & g_{ij} \sum_{l=0}^{z-1} x_j(l) [x_i(l+1) - x_i(l+1) + x_i(l-1) - x_i(l-1)] \\
= & 0.
\end{aligned}$$

2. If  $\zeta_i \leq 2$ , then  $L(x_i, x_j) = 0$  for all  $j \in N$ .

This holds because

$$\begin{aligned}
& L(x_i, x_j) \\
= & g_{ij} \sum_{l=0}^{z-1} (x_j(l+1) - x_j(l-1)) x_i(l) \\
= & g_{ij} \sum_{l=0}^{z-1} (x_i(l-1) - x_i(l+1)) x_j(l) \\
= & 0
\end{aligned}$$

(since  $x_i(l-1) = x_i(l+1)$  for  $l = 0, 1, \dots, z-1$ ).

3. Let  $i \in N$  s.t.  $\gamma_i \geq 3$ . Then it holds that  $\sum_{j=1}^n L(x_i, x_j) < 0$ .

To see this, fix some  $i$  so that  $\zeta_i \geq 3$ . We have

$$\begin{aligned}
& \sum_{j=1}^n L(x_i, x_j) \\
&= \sum_{j=1}^n g_{ij} \sum_{l=0}^{z-1} [x_j(l+1) - x_j(l-1)] x_i(l) \\
&= \sum_{l=0}^{z-1} \sum_{j \in \bar{N}_i} x_j(l) (x_i(l-1) - x_i(l+1)) \\
&= \sum_{l=0}^{z-1} (x_i(l-1) - x_i(l+1)) \sum_{j \in \bar{N}_i} x_j(l).
\end{aligned}$$

Define  $b_l \equiv (x_i(l-1) - x_i(l+1)) \sum_{j \in \bar{N}_i} x_j(l)$  for  $l = 0, 1, \dots, z-1$  so that  $\sum_{j=1}^n L(x_i, x_j) = b_0 + b_1 + \dots + b_{z-1}$ . Because  $(x_i(l-1), x_i(l+1)) \in \{-1, 0, +1\}^2$  there are only nine different combinations of  $x_i(l-1)$  and  $x_i(l+1)$  with different implications for the value of  $b_l$ . The nine cases are summarized in Table 2 below. Notice that except for cases 4 and 6 the implied values for  $b_l$  are all nonpositive. If  $z$  is odd, we will consider the sequence  $\{x_i(0), x_i(2), \dots, x_i(z-1), x_i(1), x_i(3), \dots, x_i(z-2)\}$ , listing every other element of  $x_i$ , beginning with  $x_i(0)$  and starting over from  $x_i(1)$  after  $x_i(z-1)$  is reached. Because  $\zeta_i \geq 3$ , it cannot be the case that all elements of the sequence are the same. Therefore, there exists some  $k$  such that  $(x_i(k), x_i(k+2))$  is as in one of the cases 1, 2, 4, 6, 8, or 9. If cases 4 and 6 do not occur it follows immediately that  $\sum_{l=0}^{z-1} b_l < 0$ . Suppose that case 4 occurs so that  $(x_i(k), x_i(k+2)) = (+1, 0)$  and  $b_{k+1} \leq \beta_i$ . Search the sequence for the *next* occurring transition to  $+1$ , starting over from the beginning of the sequence if no such transition exists until the end of the sequence. Say this transition occurs at  $k'$ . This must either be a switch from 0 to  $+1$  or from  $-1$  to  $+1$ . In either case we have  $b_{k'+1} < -\beta_i$ . And so  $b_{k+1} + b_{k'+1} < 0$ . The argument if case 6 occurs is analogous. Therefore, if cases 4 or 6 occur, we can match any  $b_l$  such that  $b_l > 0$  with a  $b_{l'}$  such that  $b_{l'} < -b_l$ , implying that  $\sum_{l=0}^{z-1} b_l < 0$ . If  $z$  is even we examine the two sequences  $\{x_i(0), x_i(2), \dots, x_i(z-2)\}$

and  $\{x_i(1), x_i(3), \dots, x_i(z-1)\}$ . In either case because  $\zeta_i \geq 3$  at least one sequence must be such that not all of its elements are the same. The rest of the argument is analogous to the case when  $z$  is odd.

Finally to prove the result, for  $p \in P$  let  $X(p, z)$  be the corresponding matrix as before. If  $z \geq 3$ , then at least one  $\zeta_i \geq 3$ , and by the facts proven under points 2 and 3 above we have

$$\sum_{i=1}^n \sum_{j=1}^n L(x_i, x_j) < 0.$$

However, by the fact proven under point 1

$$\sum_{i=1}^n \sum_{j=1}^n L(x_i, x_j) = 0,$$

a contradiction. Therefore,  $z \leq 2$ . ■

Table 2: Nine cases

Case	$x_i(l-1)$	$x_i(l+1)$	implies for $\sum_{j \in \bar{N}_i} x_j(l)$	implies for $b_l$
<b>1)</b>	+1	-1	$< \alpha_i$	$< 2\alpha_i$
<b>2)</b>	0	-1	$< \alpha_i$	$< \alpha_i$
<b>3)</b>	-1	-1	$< \alpha_i$	$= 0$
<b>4)</b>	+1	0	$\in [\alpha_i, \beta_i]$	$\leq \beta_i$
<b>5)</b>	0	0	$\in [\alpha_i, \beta_i]$	$= 0$
<b>6)</b>	-1	0	$\in [\alpha_i, \beta_i]$	$\leq -\alpha_i$
<b>7)</b>	+1	+1	$> \beta_i$	$= 0$
<b>8)</b>	0	+1	$> \beta_i$	$< -\beta_i$
<b>9)</b>	-1	+1	$> \beta_i$	$< -2\beta_i$



## REFERENCES

- ANDANOVA, L.B. (2003) "Transnational Politics of the Environment: The European Union and Environmental Policy in Central and Eastern Europe," MIT Press, Cambridge, MA.
- ANTWEILER, A., COPELAND, B., and M.S. TAYLOR (2001) "Is Free Trade Good for the Environment?," *American Economic Review*, vol. 91, No. 4, pp. 877-908.
- BALA, V. and S. GOYAL (2000) "A Noncooperative Model of Network Formation," *Econometrica*, vol. 68, No. 5, pp. 1181-1229.
- BALLESTER, C., CALVO-ARMENGO, A., and Y. ZENOU (2006) "Who's Who in Networks. Wanted: The Key Player," *Econometrica*, vol. 74, No. 5, pp. 1403-1417.
- BALLESTER, C., CALVO-ARMENGOL, A., and Y. ZENOU (2009) "Delinquent networks," *Journal of the European Economic Association*, forthcoming.
- BELLEFLAMME, P. and F. BLOCH (2004) "Market Sharing Agreements and Collusive Networks," *International Economic Review*, vol. 45, No. 2, pp. 387-411.
- BLOCH, F. and M.O. JACKSON (2006) "Definitions of equilibrium in network formation games," *International Journal of Game Theory*, vol. 34, pp. 305-318.
- BRENNAN, G. and J.M. BUCHANAN (1980) "The Power To Tax: Analytical Foundations of a Fiscal Constitution," Cambridge University Press, Cambridge.
- BRIATA, F., GARCIA-JURADO, I., GONZALES-DIAZ, J., and F. PATRONE (2008) "Essentializing Equilibrium Concepts," Working paper.
- CAILLAUD, B. and J. TIROLE (2007) "Consensus Building: How to Persuade a Group," *American Economic Review*, vol. 97, No. 5, pp. 1877-1900.
- CALVO-ARMENGOL, T. (2004) "Job contact networks," *Journal of Economic Theory*, vol. 115, pp. 191-206.
- CHWE, M.S. (2000) "Communication and Coordination in a Network," *Review of Economic Studies*, vol. 67, pp. 1-16.
- COLEMAN, J., KATZ, E., and H. MENZEL (1957) "The Diffusion of an Innovation Among Physicians," *Sociometry*, vol. 20, No. 4, pp. 253-270.

D'ASPREMONT, C., GABSZEWICZ, J., and J. THISSE (1979) "On Hotelling's Stability in Competition," *Econometrica*, vol. 17, pp. 1145-1151.

DAVIES, R.B. and C.J. ELLIS (2007) "Competition in Taxes and Performance Requirements for Foreign Direct Investment," *European Economic Review*, vol. 51, No. 5, pp. 1303-1325.

DEMARZO, P.M., VAYANOS, D., and J. ZWIEBEL (2003) "Persuasion Bias," *Quarterly Journal of Economics*, vol. 118, No. 3, pp. 909-968.

EDERINGTON, J., LEVINSON, A., and J. MINIER (2004) "Trade Liberalization and Pollution Havens," in *Advances in Economic Analysis and Policy*, vol. 4, No. 2, Article 6, the Berkley Electronic Press.

FISCHEL, W.A. (1975) "Fiscal and Environmental Considerations in the Location of Firms in Suburban Communities," in E. Mills and W. Oates (eds.) *Fiscal Zoning and Land Use Controls*, D.C. Heath, Lexington, MA.

FUDENBERG, D. and D.K. LEVINE (1993) "Self-Confirming Equilibrium," *Econometrica*, vol. 61, No. 3, pp. 523-545.

GALE, D. and S. KARIV (2003) "Bayesian Learning in Networks," *Games and Economic Behavior*, vol. 45, pp. 329-346.

GALEOTTI, A., GOYAL, S., and J. KAMPHORST (2006) "Network formation with heterogenous players," *Games and Economic Behavior*, vol. 54, pp. 353-372.

GALEOTTI, A. and S. GOYAL (2007) "Games of Social Influence," University of Cambridge, Working paper.

GALEOTTI, A. and S. GOYAL (2008) "A Theory of Strategic Diffusion," University of Cambridge, Working paper.

GAVIRIA, A. and S. RAPHAEL (2001) "School-Based Peer Effects and Juvenile Behavior," *Review of Economic Studies*, vol. 83, No. 2, pp. 257-268.

GLAESER, E.L., SACERDOTE, B., and J.A. SCHEINKMAN (1996) "Crime and Social Interactions," *The Quarterly Journal of Economics*, vol. 111, No. 2, pp. 507-548.

GOLES, E. and J. OLIVOS (1980) "Periodic Behavior of Generalized Threshold Functions," *Discrete Mathematics*, vol. 30, pp. 187-189.

GOYAL, S. and S. JOSHI (2003) "Networks of collaboration in oligopoly," *Games and Economic Behavior*, vol. 43, pp. 57-85.

- GOYAL, S. and S. JOSHI (2006) "Unequal connections," *International Journal of Game Theory*, vol. 34, pp. 319-349.
- GOYAL, S. and F. VEGA-REDONDO (2005) "Network formation and social coordination," *Games and Economic Behavior*, vol. 50, pp. 178-207.
- HOJMAN, D.A. and A. SZEIDL (2008) "Core and periphery in networks," *Journal of Economic Theory*, vol. 139, pp. 295-309.
- HOTELLING, H. (1929) "Stability in Competition," *Economic Journal*, vol. 39, pp. 41-57.
- JACKSON, M.O. (2005) "A Survey of Network Formation Models: Stability and Efficiency," in *Group Formation in Economics: Networks, Clubs, and Coalitions* by Demange, G. and M. H. Wooders (eds), Cambridge University Press.
- JACKSON, M.O. and A. WATTS (2001) "The existence of pairwise stable networks," *Seoul Journal of Economics*, vol. 14, No. 3, pp. 299-321.
- JACKSON, M.O. and A. WATTS (2002) "The Evolution of Social and Economic Networks," *Journal of Economic Theory*, vol. 106, pp. 265-295.
- JACKSON, M.O. and A. WOLINSKY (1996) "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, vol. 71, pp. 44-74.
- JACKSON, M.O. and L. YARIV (2007) "Diffusion of Behavior and Equilibrium Properties in Network Games," *American Economic Review*, vol. 97, No. 2, pp. 92-98.
- KALAI, E. and A. NEME (1992) "The strength of a little perfection," *International Journal of Game Theory*, vol. 20, No. 4, pp. 335-355.
- KALAI, E. (2004) "Large robust games," *Econometrica*, vol. 72, No. 6, pp. 1631-1666.
- KALAI, E. (2005) "Partially specified large games," *Lecture notes in computer science* 3828, pp. 3-13.
- KEMPE, D., KLEINBERG, J., and E. TARDOS (2003) "Maximizing the Spread of Influence through a Social Network," *Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*.
- KEMPE, D., KLEINBERG, J., and E. TARDOS (2005) "Influential Nodes in a Diffusion Model for Social Networks," *Proc. 32nd International Colloquium on Automata, Languages and Programming (ICALP)*.

KLEINBERG, J. (2007) "Cascading Behavior in Networks: Economic and Algorithmic Issues," in *Algorithmic Game Theory*, N. Nisan, Roughgarden, T., Tardos, É., and V.V. Vazirani (eds), Cambridge University Press, New York

KOHLER, H.P., BEHRMAN, J.R., and S. WATKINS (2001) "The Density of Social Networks and Fertility Decisions: Evidence from South Nyanza District, Kenya," *Demography*, vol. 38, No. 1, pp. 43-58.

KREPS, D.M. and R. WILSON (1982) "Sequential Equilibria;" *Econometrica*, vol. 50, No. 4, pp. 863-894.

KREPS, D.M. and J.A. SCHEINKMAN (1983) "Quantity Precommitment and Bertrand Competition Yield Cournot Outcome," *The Bell Journal of Economics*, vol. 14, No. 2, pp. 326-337.

LEVITAN, R. and M. SHUBIK (1972) "Price Duopoly and Capacity Constraints," *International Economic Review*, vol. 13, pp. 111-122.

LOPEZ-PINTADO, D. (2008) "Diffusion in Complex Social Networks," *Games and Economic Behavior*, vol. 62, pp. 573-590.

MARKUSEN, J.R., MOREY, E.R., and N. OLEWILER (1995) "Competition in Regional Environmental Policies When Plant Locations are Endogenous," *Journal of Public Economics*, vol. 56, pp. 55-77.

MORRIS, S. (2000) "Contagion," *Review of Economic Studies*, vol. 67, pp. 57-78.

MYERSON, R.B. (1991) "Game Theory," Harvard University Press, Cambridge, Massachusetts.

NISKANEN, JR., W. (1977) "Bureaucracy and Representative Government," Aldine, Chicago, IL.

OATES, W.E. and R.M. SCHWAP (1988) "Economic Competition Among Jurisdictions: Efficiency Enhancing or Distortion Inducing?," *Journal of Public Economics*, vol. 35, pp. 333-354.

PAGE, F.H. and M. WOODERS (2008) "Strategic basins of attraction, the path dominance core, and network formation games," *Games and Economic Behavior*, forthcoming.

PAGE, F.H., WOODERS, M., and S. KAMAT (2005) "Networks and farsighted stability," *Journal of Economic Theory*, vol. 120, No. 2, pp. 257-269.

POST, D. (2002) "Closing the Deception Gap: Accession to the European Union and Environmental Standards in East Central Europe," in University of California International and Area Studies (UCIAS) Edited Volume 1: Dynamics of Regulatory Change: How Globalization Affects National Regulatory Policies, Berkeley Electronic Press.

ROGERS, E.M. (2003) "Diffusion of Innovations," 5th edition, New York: The Free Press.

SELTEN, R. (1975) "Reexamination of the perfectness concept for equilibrium points in extensive form games," International Journal of Game Theory, vol. 4, pp. 25-55.

SELTEN, R., MITZKEWITZ, M. and G.R. UHLICH (1997) "Duopoly Strategies Programmed by Experienced Players," Econometrica, vol. 65, No. 3, pp. 517-555.

SHUBIK, M. (1955) "A Comparison of Treatments of a Duopoly Problem (Part II)," Econometrica, vol. 53, pp. 417-431.

TAYLOR, M.S. (2004) "Unbundling the Pollution Haven Hypothesis," in Advances in Economic Analysis and Policy, the Berkley Electronic Press.

VAN DAMME, E. (1983) "Refinements of the NE concept," Berlin: Springer-Verlag.

WHITE, M.J. (1975) "Firm Location in a Zoned Metropolitan Area," in E. S. Mills and W. E. Oates (eds.) Fiscal Zoning and Land Use Controls, D.C. Heath, Lexington, MA.

WILSON, J.D. (1996) "Capital Mobility and Environmental Standards: Is There a Theoretical Case for a Race to the Bottom?," Chapter 10 in Fair Trade and Harmonization by J. N. Bhagwati and R. E. Hudec. MIT Press, Cambridge.

ZUCKERMAN, A.S., VALENTINO, A.N., and E.W. ZUCKERMAN (1994) "A Structural Theory of Vote Choice: Social and Political Networks and Electoral Flow in Britain and the United States," The Journal of Politics, vol. 56, No. 4, pp. 1008-1033.