

# AXIOMATIC EQUILIBRIUM SELECTION FOR GENERIC TWO-PLAYER GAMES

SRIHARI GOVINDAN AND ROBERT WILSON

ABSTRACT. We apply three axioms adapted from decision theory to refinements of the Nash equilibria of games with perfect recall that select connected closed subsets called solutions.

*Undominated Strategies:* No player uses a weakly dominated strategy in an equilibrium in a solution.

*Backward Induction:* Each solution contains a quasi-perfect equilibrium and thus a sequential equilibrium in strategies that provide conditionally admissible optimal continuations from information sets.

*Small Worlds:* A refinement is immune to embedding a game in a larger game with additional players provided the original players' strategies and payoffs are preserved, i.e. solutions of a game are the same as those induced by the solutions of any larger game in which it is embedded.

For games with two players and generic payoffs, we prove that these axioms characterize each solution as an essential component of equilibria in undominated strategies, and thus a stable set as defined by Mertens (1989).

---

*Date:* 28 May 2009.

*Key words and phrases.* game, equilibrium, refinement, axiom, admissibility, backward induction, small worlds, stability. *JEL subject classification:* C72.

This work was funded in part by a grant from the National Science Foundation of the United States.

## CONTENTS

List of Figures	2
1. Introduction	3
2. Notation	5
3. The Axioms	6
3.1. Undominated Strategies	6
3.2. Backward Induction	6
3.3. Small Worlds	7
3.4. Summary of the Axioms	9
4. Additional Notation and Properties	10
4.1. Payoffs	10
4.2. Notation for the Extensive Form	10
4.3. Enabling Strategies	11
4.4. Stable Sets of Equilibria	12
4.5. The Associated Pseudomanifold	13
5. Statement and Proof of the Theorem	13
5.1. Preliminaries	14
5.2. A Game with Redundant Strategies	15
5.3. Extensive Form of the Metagames	16
5.4. Outsiders' Payoffs in the Metagames	17
5.5. Outsiders' Strategies in a Quasi-Perfect Equilibrium	17
5.6. Limits of the Quasi-Perfect Equilibria of the Metagames	19
5.7. Insiders' Strategies in a Quasi-Perfect Equilibrium	20
5.8. The Induced Lexicographic Probability System	21
5.9. Limit of the Lexicographic Probability System	22
5.10. Final Step of the Proof	24
6. Concluding Remarks	25
Appendix A. Enabling Strategies	27
Appendix B. Proofs of Propositions	28
B.1. Proof of Proposition 3.3	28
B.2. Proof of Proposition 3.4	29
Appendix C. Proof of Theorem 4.1	30
Appendix D. The Genericity Assumption	45
References	46

## LIST OF FIGURES

1 A game tree and its pure reduced normal form in which each pure strategy is identified by the terminal nodes it does not exclude.	27
2 The game tree of a signaling game and its pure reduced normal form in which each pure strategy is identified by the terminal nodes it does not exclude.	28

## 1. INTRODUCTION

Nash's [22, 23] definition of equilibrium is insufficient when applied to a game in extensive form. Even for games with perfect information, some equilibria use weakly dominated strategies and yield outcomes different from the outcome predicted by backward induction; see [11, §4.2] for an example. The literature on refinements aims to sharpen Nash's definition to exclude such equilibria; recent surveys include [8, 12, 31].

Kohlberg and Mertens [14] suggest that a refinement should be characterized by axioms adapted from decision theory. They also specify properties that axioms should imply. Subsequently, Mertens [19, 20, 21] defines the set-valued refinement called *stability* and shows that it has these and other properties, including the following.<sup>1</sup>

1. *Admissibility and Perfection.* All equilibria in a stable set are perfect (hence admissible) so each player's strategy in each equilibrium is undominated.
2. *Backward Induction and Forward Induction.* A stable set includes a proper equilibrium that induces a quasi-perfect (hence sequential) equilibrium in every extensive-form game with perfect recall that has the same normal form. A subset of a stable set survives iterative elimination of weakly dominated strategies and strategies that are inferior replies at every equilibrium in the set.
3. *Invariance, Small Worlds and Decomposition.* The stable sets of a game are the projections of the stable sets of any larger game in which it is embedded. The stable sets of the product of two independent games are the products of their stable sets.

For games in extensive form with perfect recall, we propose three axioms that are versions of these properties. We assume that for each game a refinement selects a nonempty collection of nonempty connected closed subsets of equilibria, called *solutions*.<sup>2</sup> The axioms are the following:

- A. *Undominated Strategies:* No player uses a weakly dominated strategy in any equilibrium in a solution.

---

<sup>1</sup>Mertens considers the graph of equilibria over the space of perturbed games obtained by perturbing players' strategies toward mixed strategies, and then defines a connected closed set of equilibria as stable if the projection from a neighborhood in this graph is an essential map, i.e. has nonzero degree. See Mertens [19] for further topological aspects of the definition, and Govindan and Mertens [3] for an equivalent definition in terms of players' best-reply correspondences.

<sup>2</sup>Solutions are assumed to be sets because Kohlberg and Mertens [14, pp. 1015, 1019, 1029] show that there need not exist a single equilibrium satisfying weaker properties than the axioms invoked here. The technical requirement that a solution is connected excludes the trivial refinement that selects all equilibria. If only a single (possibly unconnected) subset is selected then only the trivial refinement satisfies the conditions invoked by Norde, Potters, Reijniere, and Vermeulen [24]. Kreps and Wilson [16, Theorem 2] show that if payoffs are generic then all equilibria in a connected set yield the same probability distribution over terminal nodes, and thus the same paths of equilibrium play in the extensive form.

B. Backward Induction: A solution contains a quasi-perfect equilibrium.

S. Small Worlds: A solution is immune to embedding the game within larger games that preserve players' strategies and payoffs.

In Axiom A we invoke only the implication of admissibility that no player uses a weakly dominated strategy. Axiom A is equivalent to admissibility in two-player games. Axiom B insists on inclusion of a quasi-perfect equilibrium because it induces a sequential equilibrium in strategies that provide admissible optimal continuations from players' information sets.<sup>3</sup> Axiom S ensures that a refinement is not vulnerable to framing effects depending on how the game is presented within wider contexts called metagames. A metagame can include additional players and additional pure strategies, provided these additional features do not alter optimal decisions in the original game. Our version of small worlds is stronger than Mertens' [21] version because our definition of embedding allows a more general class of metagames.

We prove that these three axioms characterize refinements that select some or all of the stable sets of a game in extensive form with perfect recall, two players, and generic payoffs. Our characterization is cast in terms of the 'enabling form' of a game in which two pure strategies of a player are considered equivalent if they exclude the same terminal nodes of the game tree, and thus enable the same sets of terminal nodes to occur as outcomes of the game. The enabling form is defined in Section 4.3 and explained further in Appendix A.

Our main theorem establishes that the axioms imply that each solution is an essential component of admissible equilibria. For the enabling form of a game this is the defining property of a stable set.<sup>4</sup>

When payoffs are generic, all equilibria in a component yield the same probability distribution over outcomes. Therefore, for economic modeling and econometric studies the axioms' chief implication is that a predicted outcome distribution should result from an essential component of the game's admissible equilibria. A secondary implication is that a solution must include all equilibria in the component. This is germane for predicting players' behaviors only after deviations from equilibrium play, but it addresses the issue of whether sequential rationality is a relevant decision-theoretic criterion after deviations (Reny [26]). We show that each equilibrium in a solution is induced by a quasi-perfect equilibrium in a

---

<sup>3</sup>We intend that a revised version will weaken Axiom B to only require inclusion of a sequential equilibrium for which each player's strategy is conditionally admissible from each of his information sets.

<sup>4</sup>A component is a maximal closed connected set, and it is essential if it has the property described in footnote 1. For the usual normal form of the game, a solution can be a subset of a component that maps to a component of the enabling form of the game.

solution of some metagame; therefore, it is sequentially rational when viewed in the wider context of this metagame. See [7, Section 2.3] for an example.

Section 2 establishes notation for Section 3, which specifies Axioms A (undominated strategies), B (backward induction), and S (small worlds), including a precise definition of embedding a game in a larger game. Appendix B verifies that Nash equilibria satisfy this definition of embedding. Section 4 establishes further notation and states a key technical property proved in Appendix C. Section 5 states and proves the main theorem, Theorem 5.1. The proof is constructive in that each equilibrium in a stable set is shown to be induced by a quasi-perfect equilibrium in a solution of a particular metagame with perfect recall that embeds the given game. Section 6 interprets this result and provides concluding remarks.

## 2. NOTATION

This section provides sufficient notation for statements of the axioms in Section 3. Section 4 introduces additional notation for the theorems in Section 5 and Appendix C.

A typical game in extensive form is denoted  $\Gamma$ . Its specification includes a set  $N$  of players, a game tree with perfect recall for each player, and an assignment of players' payoffs at each terminal node of the tree. Let  $H_n$  be player  $n$ 's collection of information sets, and let  $A_n(h)$  be his set of feasible actions at information set  $h \in H_n$ . The specification of the tree can include a completely mixed strategy of Nature.

A player's pure strategy chooses an action at each of his information sets. Denote  $n$ 's simplex of mixed strategies by  $\Sigma_n$  and interpret its vertices as his set  $S_n$  of pure strategies. The sets of profiles of players' pure and mixed strategies are  $S = \prod_n S_n$  and  $\Sigma = \prod_n \Sigma_n$ . The normal form of  $\Gamma$  assigns to each profile of players' pure strategies the profile of their expected payoffs; equivalently, it is the multilinear function  $G : \Sigma \rightarrow \mathbb{R}^N$  that assigns to each profile of their mixed strategies the profile of their expected payoffs.

A player's behavioral strategy specifies a mixture over his actions at each of his information sets. Let  $B_n$  be  $n$ 's set of behavioral strategies, and  $B = \prod_n B_n$  the set of profiles of players' behavioral strategies. Each mixed strategy  $\sigma_n$  induces a behavioral strategy  $b_n$ . Because the game has perfect recall, for each behavioral profile there are profiles of mixed strategies that induce it and yield the same distribution of outcomes (Kuhn [17]).

As defined by Nash [22, 23], an equilibrium is a profile of players' strategies such that each player's strategy is an optimal reply to other players' strategies. That is, if  $\text{BR}_n(\sigma) \equiv \arg \max_{\sigma'_n \in \Sigma_n} G_n(\sigma'_n, \sigma_{-n})$  is player  $n$ 's best-reply correspondence, then  $\sigma \in \Sigma$  is an equilibrium iff  $\sigma_n \in \text{BR}_n(\sigma)$  for every player  $n$ . The analogous definition of equilibrium in behavioral strategies is equivalent for games with perfect recall.

A refinement is a correspondence that assigns to each game a nonempty collection of nonempty connected closed subsets of its equilibria. Each selected subset is called a solution.

According to the above definitions, a mixed or behavioral strategy makes choices even at information sets that its previous choices exclude from being reached. In Section 4 we consider pure strategies to be equivalent if they make the same choices at information sets they do not exclude. And, we further simplify mixed and behavioral strategies by considering only their induced probability distributions on non-excluded terminal nodes. The definitions of an equilibrium and a refinement have equivalent statements in terms of these strategy spaces. Each equilibrium in a reduced strategy space corresponds to a set of equilibria as defined above, and analogously for solutions selected by a refinement. The axioms in Section 3 are stated in terms of mixed and behavioral strategies as defined above. One reason for this is that Axiom S implies invariance to redundant strategies, and thus justifies our later use of equivalence classes of strategies.

### 3. THE AXIOMS

**3.1. Undominated Strategies.** The first axiom requires that no player uses a weakly dominated strategy. Say that a profile of players' strategies is undominated if each player's strategy is undominated.

**Axiom A** (Undominated Strategies): Each equilibrium in a solution is undominated.

**3.2. Backward Induction.** We interpret sequential equilibrium as the generalization to games with perfect recall of backward induction in games with perfect information, and to be consistent with Axiom A we insist on conditionally admissible optimal continuations from information sets. Here we obtain these properties from a quasi-perfect equilibrium, which requires a consistency property of beliefs that is stronger than required by Kreps and Wilson's [16] definition of sequential equilibrium. Van Damme [30, p. 8] defines a quasi-perfect equilibrium in terms of players' behavioral strategies.

**Definition 3.1** (Quasi-Perfect Equilibrium). A profile  $b \in B$  of behavioral strategies is a *quasi-perfect* equilibrium if it is the limit of a sequence of profiles of completely mixed behavioral strategies for which, for each player  $n$ , from each of his information sets, continuation of his strategy  $b_n$  is an optimal reply to every profile in the sequence.

That is, if  $BR_n(\cdot|h)$  is  $n$ 's best-reply correspondence in terms of behavioral strategies that continue from his information set  $h \in H_n$ , and  $b_n(h)$  is the continuation of his behavioral

strategy  $b_n$  from this information set, then the profile  $b \in B$  is quasi-perfect if

$$(\forall k) \quad (\forall n \in N, h \in H_n) \quad b_n(h) \in BR_n(\hat{b}^k|h)$$

for some sequence  $\hat{b}^k \in B \setminus \partial B$  for which  $\hat{b}^k \rightarrow b$ .

If the mixed-strategy profile  $\sigma \in \Sigma$  induces a behavioral profile  $b \in B$  that is a quasi-perfect equilibrium then we say that  $\sigma$  too is quasi-perfect. Similarly, the justifying sequence  $\hat{b}^k$  can be represented by a sequence  $\hat{\sigma}^k$  in  $\Sigma \setminus \partial \Sigma$  for which  $\hat{\sigma}^k \rightarrow \sigma$ .

Van Damme shows that a quasi-perfect equilibrium induces a perfect equilibrium of the normal form, and a sequential equilibrium of the extensive form in which all players' beliefs are induced by the same justifying sequence. Moreover, by construction a quasi-perfect equilibrium provides for each player an admissible optimal continuation from each of his information sets. It can be shown that if payoffs are generic then every sequential equilibrium is quasi-perfect and extensive-form perfect.

The second axiom requires that some equilibrium in a solution is quasi-perfect.

**Axiom B** (Backward Induction): Each solution contains a quasi-perfect equilibrium.

If payoffs are generic then Axioms A and B imply that each solution lies in a component of undominated equilibria, each of which yields the same distribution over outcomes as a sequential equilibrium in the solution.

**3.3. Small Worlds.** The third axiom requires that a refinement is not affected by extraneous features of wider contexts in which a game is embedded, provided such contexts do not alter players' feasible strategies and payoffs. An embedding allows the presence of additional players whose actions might provide the original players with additional pure strategies equivalent to mixed strategies in the original game, and thus redundant. For simplicity, to define an embedding we use the normal form  $G : \Sigma \rightarrow \mathbb{R}^N$  of the extensive-form game  $\Gamma$ .

An embedding is described by a 'larger' game  $\tilde{G} : \tilde{\Sigma} \times \Sigma_o \rightarrow \mathbb{R}^{N \cup o}$  in which game  $G$  is 'embedded,' subject to certain restrictions specified below. The larger game  $\tilde{G}$  has outsiders in a set  $o$ , in addition to insiders who are the players in  $N$ , and there can be additional moves by Nature. An insider  $n$  can have additional pure strategies in  $\tilde{\Sigma}_n$  that are not pure strategies in  $S_n$  but are equivalent to mixed strategies in  $\Sigma_n$ . The basic requirement is that an embedding should preserve the game among insiders, conditional on actions by outsiders.

These restrictions have a technical formulation. There should exist a multilinear map  $f : \tilde{\Sigma} \times \Sigma_o \rightarrow \Sigma$  that is surjective and such that  $\tilde{G}_n = G_n \circ f$  for each insider  $n$ . Moreover, to exclude an embedding from introducing correlation among insiders' strategies,  $f$  should factor into separate multilinear maps  $(f_n)_{n \in N}$ , where each component is a map  $f_n : \tilde{\Sigma}_n \times \Sigma_o \rightarrow$

$\Sigma_n$  such that  $f_n(\cdot, \sigma_o)$  maps  $\tilde{\Sigma}_n$  surjectively onto  $\Sigma_n$  for each profile  $\sigma_o \in \Sigma_o$  of outsiders' strategies.

Admittedly, a statement of the axiom that uses this technical language could contain unsuspected implications. However, after stating the formal definition, we provide in Proposition 3.3 an equivalent formulation that is more detailed and more transparent, and that verifies the requisite properties. Also, Proposition 3.4 applies a precise test of whether the axiom is correctly stated, namely, a refinement that satisfies the axiom should be immune to the same embeddings that equilibria are.

**Definition 3.2** (Embedding). A game  $\tilde{G} : \tilde{\Sigma} \times \Sigma_o \rightarrow \mathbb{R}^{N \cup o}$  and a collection of multilinear maps  $f_n : \tilde{\Sigma}_n \times \Sigma_o \rightarrow \Sigma_n$ , one for each player  $n \in N$ , embed a game  $G : \Sigma \rightarrow \mathbb{R}^N$  if

- (a) for each  $\sigma_o \in \Sigma_o$ ,  $f_n(\cdot, \sigma_o)$  maps  $\tilde{\Sigma}_n$  surjectively onto  $\Sigma_n$ , and
- (b)  $\tilde{G}_n = G_n \circ f$ , where  $f = (f_n)_{n \in N}$ .

Condition (a) ensures that embedding has no net effect on an insider's set of mixed strategies, conditional on outsiders' strategies, and condition (b) ensures that there is no net effect on any insider's payoffs.

Hereafter, if  $\tilde{G}$  embeds  $G$  via maps  $f = (f_n)$  then we say that  $(\tilde{G}, f)$  embeds  $G$  and that  $\tilde{G}$  is a *metagame* for  $G$ . We omit description of  $f$  for metagames in extensive form that embed a game in extensive or strategic form. An elaborate example of a metagame in extensive form that embeds a game in extensive form is constructed in proving Theorem 5.1. Note that the property that  $\tilde{G}$  is a metagame for  $G$  does not depend on specifying payoffs for the outsiders; in particular, it depends only on their extensive forms and the insiders' payoffs in the two games.

A multilinear map  $f_n : \tilde{\Sigma}_n \times \Sigma_o \rightarrow \Sigma_n$  is completely specified by its values at profiles of pure strategies. Let  $\hat{f}_n$  be the restriction of  $f_n$  to the set  $\tilde{S}_n \times S_o$  of profiles of pure strategies of player  $n$  and the outsiders in  $o$ . The following Proposition, proved in Appendix B, provides an alternative definition of embedding in terms of pure strategies.

**Proposition 3.3.**  $\tilde{G}$  embeds  $G$  via a collection of multilinear maps  $f = (f_n)_{n \in N}$  if and only if for each player  $n$  there exists  $\tilde{T}_n \subseteq \tilde{S}_n$  and a bijection  $\pi_n : \tilde{T}_n \rightarrow S_n$  such that for each  $(\tilde{s}, s_o) \in \tilde{S} \times S_o$  and  $\tilde{t}_n \in \tilde{T}_n$ :

- (1)  $\hat{f}_n(\tilde{t}_n, s_o) = \pi_n(\tilde{t}_n)$ ,
- (2)  $\tilde{G}_n(\tilde{s}, s_o) = G_n(\hat{f}(\tilde{s}, s_o))$ , where  $\hat{f} = (\hat{f}_n)_{n \in N}$ .



Property (2) assures that players' payoffs from pure strategies of  $G$  are preserved by the metagame  $\tilde{G}$ . Hence property (1) assures that each pure strategy  $s_n \in S_n$  is equivalent to some pure strategy  $\tilde{t}_n = \pi^{-1}(s_n) \in \tilde{T}_n$ , independently of the outsiders' profile  $s_o$ .

Pure strategies in  $\tilde{S}_n \setminus \tilde{T}_n$  are redundant because payoffs from profiles in  $\prod_n \tilde{T}_n$  exactly replicate payoffs from corresponding profiles in  $\prod_n S_n$  for the embedded game  $G$ . In particular, if  $\hat{f}_n(\tilde{s}_n, s_o) = \sigma_n \notin S_n$  then conditional on  $s_o$  the pure strategy  $\tilde{s}_n$  is equivalent for insiders to the mixed strategy  $\sigma_n \in \Sigma_n$ .

The next Proposition, proved in Appendix B, verifies that equilibria are not affected by embedding in a metagame.

**Proposition 3.4.** *If  $(\tilde{G}, f)$  embeds  $G$  then the equilibria of  $G$  are the  $f$ -images of the equilibria of  $\tilde{G}$ .*

A corollary of Proposition 3.4 is that embedding does not introduce correlation among insiders' strategies.

Using Definition 3.2 of embedding, the small worlds axiom is the following.

**Axiom S** (Small Worlds): If  $(\tilde{G}, f)$  embeds  $G$  then the  $f$ -images of the solutions that a refinement selects for  $\tilde{G}$  are the solutions selected for  $G$ .

In view of Proposition 3.4, this axiom is an instance of the general principle that a refinement should inherit invariance properties of equilibria.

Two special cases of Axiom S are the following.

**Invariance:** Suppose  $\Sigma_o$  and  $o$  are singletons and insiders' payoffs and strategies in  $\tilde{\Sigma}$  differ from  $\Sigma$  only by treating some mixed strategies in  $\Sigma$  as additional pure strategies in  $\tilde{S}$ . Then Axiom S implies that solutions depend only on a game's reduced normal form obtained by deleting such redundant pure strategies.

**Mertens' Small Worlds Axiom** [21]: Suppose  $\tilde{\Sigma} = \Sigma$ . Then Axiom S implies that a solution does not depend on the presence of outsiders, i.e. solutions of the original game are the projections of the solutions of metagames obtained by adding dummy players.

**3.4. Summary of the Axioms.** We study refinements that are independent of embeddings in metagames that, for each profile of outsiders' strategies, preserve the strategies and payoffs of the game among insiders. And, we require that each of their solutions is a closed connected set of undominated equilibria, including one that is quasi-perfect. In particular, a solution of a metagame must contain a quasi-perfect equilibrium whose image is in the corresponding solution of the embedded game.

## 4. ADDITIONAL NOTATION AND PROPERTIES

In the sequel we consider only a game  $\Gamma$  in extensive form with perfect recall, two players, and generic payoffs. In this section we prepare for the statement and proof of the main theorem in Section 5.

**4.1. Payoffs.** Let  $Z$  be the set of terminal nodes of the game tree. Players' payoffs are given by a point  $u$  in  $U = \mathbb{R}^{N \times Z}$ , where  $u_n(z)$  is the payoff to player  $n \in N$  at terminal node  $z \in Z$ . We assume that payoffs are generic in that there exists a lower dimensional subset  $U_\circ$  of  $U$  such that our results are true for all games in  $U \setminus U_\circ$ . The set  $U_\circ$  includes the nongeneric set described in [4]. Therefore, each game outside  $U_\circ$  has finitely many equilibrium outcome distributions, and in particular all equilibria in a component yield the same distribution over outcomes. However, the proofs in Section 5 and Appendix C require some genericity properties that are not necessarily implied by the construction in [4]. To avoid disrupting the main exposition, we defer to Appendix D the description of the exact set of genericity properties required for the proofs, and an explanation of why the resulting set  $U_\circ$  of excluded payoffs has lower dimension.

**4.2. Notation for the Extensive Form.** The set of players is  $N = \{1, 2\}$ , typically represented as a player  $n$  and the other player  $m \neq n$ . Let  $X$  be the set of nodes in the game tree. Let  $X_n$  be the set of nodes where player  $n$  moves, partitioned into his information sets  $h \in H_n$ . For a node  $x \in X_n$  we write  $h(x)$  for the unique information set  $h \in H_n$  that contains  $x$ . For each  $n$  and  $h \in H_n$ , let  $A_n(h)$  be the set of actions available to player  $n$  at  $h$ . Assume that actions at all information sets are labeled differently, and let  $A_n$  be the set of all actions for player  $n$ .

Node  $x$  precedes another node  $y$ , written  $x \prec y$ , if  $x$  is on the unique path from the root of the tree to  $y$ . For a node  $x \in X_n$  and  $a \in A_n(h(x))$  write  $(x, a) \prec y$  if  $x \prec y$  and the path from the root of the tree to  $y$  requires player  $n$  to choose  $a$  at  $h(x)$ . If  $(x, a) \prec y$  and  $x$  and  $y$  belong to  $n$ 's information sets  $h$  and  $h'$ , respectively, then every node in  $h'$  follows some node in  $h$  by the choice of  $a$ , so we write  $(h, a) \prec h'$ .

The set of pure strategies of player  $n$  is the set  $S_n$  of functions  $s_n : H_n \rightarrow A_n$  such that  $s_n(h) \in A_n(h)$  for all  $h \in H_n$ . For each  $n$ ,  $s_n \in S_n$  and  $y \in X$ , let  $\beta_n(y, s_n)$  be the probability that  $s_n$  does not exclude  $y$ , i.e.  $\beta_n(y, s_n) = 1$  if for each  $(x, a) \prec y$  with  $x \in X_n$ ,  $s_n(h(x)) = a$ , and otherwise  $\beta_n(y, s_n) = 0$ . By perfect recall, if  $y \in X_n$  then  $\beta_n(y', s_n) = \beta_n(y, s_n)$  for all  $y' \in h(y)$  and we write  $\beta_n(h(y), s_n)$  for this probability. Likewise, for any node  $y$  we write  $\beta_0(y)$  for the probability that Nature does not exclude  $y$ . Then for a profile  $s \in S$  the probability that node  $y$  is reached is  $\beta(y, s) \equiv \beta_0(y)\beta_1(y, s_1)\beta_2(y, s_2)$ .

Recall that  $\Sigma_n$  is the set of mixed strategies of player  $n$ . For each node  $y$  the function  $\beta_n(y, \cdot)$  extends to a function over  $\Sigma_n$  via  $\beta_n(y, \sigma_n) = \sum_{s_n \in S_n} \beta_n(y, s_n) \sigma_n(s_n)$  for  $\sigma_n \in \Sigma_n$ . Recall also that  $B_n$  is the set of behavioral strategies of player  $n$ . For each  $b_n \in B_n$ ,  $\beta_n(y, b_n)$  is the product of  $b_n$ 's probabilities of  $n$ 's actions on the path to  $y$ .

Similarly extend  $\beta$  to profiles of mixed or behavioral strategies. Given a mixed-strategy profile  $\sigma \in \Sigma$ , the probability that outcome  $z$  results is  $\beta(z, \sigma) = \beta_0(z) \beta_1(z, \sigma_1) \beta_2(z, \sigma_2)$ .

**4.3. Enabling Strategies.** For each player  $n$  define  $\rho_n : \Sigma_n \rightarrow [0, 1]^Z$  by the formula  $\rho_n(\sigma_n) = (\beta_n(z, \sigma_n))_{z \in Z}$ , and let  $\rho = (\rho_n)_{n \in N}$ . Similarly, if  $b_n \in B_n$  is the behavioral strategy induced by  $\sigma_n$  then  $\rho_n(\sigma_n) = (\beta_n(z, b_n))_{z \in Z}$ . Let  $P_n$  be the image of  $\rho_n$ , and  $P = \prod_n P_n$  the image of  $\rho$ . Then  $P_n$  is a compact convex polyhedron, called the space of  $n$ 's enabling strategies in [5]. Each vertex of  $P_n$  corresponds to an equivalence class of  $n$ 's pure strategies that exclude the same outcomes.<sup>5</sup>

If  $\sigma \in \Sigma$  and  $p = \rho(\sigma)$  then the probability of outcome  $z$  is  $\gamma_z(p) = \beta_0(z) p_1(z) p_2(z)$ . Thus the function  $\gamma : P \rightarrow \Delta(Z)$  summarizes the extensive form. The analog of the game  $\Gamma$ 's normal form  $G : \Sigma \rightarrow \mathbb{R}^N$  is the enabling form  $\mathcal{G} : P \rightarrow \mathbb{R}^N$  that assigns to each profile of enabling strategies the profile of players' expected payoffs, where  $\mathcal{G}_n(p) = \sum_z \gamma_z(p) u_n(z)$ . Note that  $\gamma$  and  $\mathcal{G}$  are multilinear functions. From players' best-reply correspondences in terms of enabling strategies one obtains the definition of equilibrium in enabling strategies. To each equilibrium in enabling strategies there correspond families of outcome-equivalent equilibria in behavioral and mixed strategies. Axioms A and S have direct analogs in terms of enabling strategies, as shown in [11].

For games in extensive form with perfect recall, enabling strategies are minimal representations. For example, using perfect recall, by working backward in the induced tree of a player's information sets, from his enabling strategy one can construct the corresponding behavioral strategy at his information sets that his prior actions do not exclude. Because Axiom S implies Invariance, it is immaterial whether solutions are characterized in terms of mixed, behavioral, or enabling strategies. We use enabling strategies here because induced distributions over outcomes are multilinear functions of enabling strategies, like they are for mixed strategies but unlike the nonlinear dependence on behavioral strategies. Also, the dimensions of the spaces of enabling and behavioral strategies are the same, which is important for the technical property established in Theorem 4.1 below. Using these features, Section 5 derives the implications of the axioms in terms of enabling strategies.

<sup>5</sup>In [5], enabling strategies are defined in terms of terminal actions rather than terminal nodes. The present definition is equivalent because a player's probability of enabling a terminal action is the probability of enabling each terminal node that follows it. The vertices of  $P_n$  are  $n$ 's pure strategies in the 'pure reduced normal form' defined by Mailath, Samuelson, and Swinkels [18]. See Appendix A for illustrations.

**4.4. Stable Sets of Equilibria.** Now let  $\bar{\Sigma}^*$  be a component of the equilibria of  $\Gamma$  in terms of mixed strategies, and let  $\bar{\Sigma}_n^*$  be the projection of  $\bar{\Sigma}^*$  in  $\Sigma_n$ .

By genericity, all equilibria in  $\bar{\Sigma}^*$  induce the same distribution over outcomes. Therefore, for each node  $x$ ,  $\beta(x, \sigma)$  is the same for all  $\sigma \in \bar{\Sigma}^*$ ; in particular, if  $x$  belongs to information set  $h \in H_n$  and  $h$  is on an equilibrium path then  $\beta_n(h, \sigma_n)$  is the same for every equilibrium strategy  $\sigma_n$  of player  $n$  in  $\bar{\Sigma}^*$ . We therefore denote these probabilities by  $\beta_n^*(x)$  and  $\beta_n^*(h)$ . Let  $H_n^*$  be the collection of information sets  $h \in H_n$  of player  $n$  such that  $\beta_n^*(h) > 0$  and let  $A_n^*$  be the set of actions at information sets in  $H_n^*$  that are chosen with positive probability by the equilibria in  $\bar{B}^*$ , where  $\bar{B}^*$  is the set of profiles of behavioral strategies induced by equilibria in  $\bar{\Sigma}^*$ .

Let  $S_n^0 \subset S_n$  be the set of pure strategies  $s_n^0$  with the property that, at each information set  $h \in H_n^*$  that  $s_n^0$  does not exclude,  $s_n^0$  prescribes an action in  $A_n^*$ . Let  $S_n^1 = S_n \setminus S_n^0$ , i.e. each pure strategy  $s_n$  in  $S_n^1$  chooses a non-equilibrium action at some information set  $h \in H_n^*$  that it does not exclude.

For  $i = 0, 1$ , let  $\Sigma_n^i$  be the set of mixed strategies whose support is contained in  $S_n^i$ . Observe that the support of  $n$ 's strategy in every equilibrium in  $\bar{\Sigma}^*$  is contained in  $S_n^0$  and that every strategy in  $S_n^0$  is a best reply against every equilibrium in  $\bar{\Sigma}^*$ . Thus  $\bar{\Sigma}_n^*$  is contained in  $\Sigma_n^0$  and  $\bar{\Sigma}^* = \bar{\Sigma}_1^* \times \bar{\Sigma}_2^*$ .

If  $S_n^1$  is empty for each  $n$ , then each equilibrium in  $\bar{B}^*$  is completely mixed; by genericity,  $\bar{B}^*$  is a singleton and its equivalent mixed strategy is stable. The only interesting case, therefore, is one where  $S_n^1$  is nonempty for at least one of the players. In order to avoid dealing with different cases, we assume that  $S_n^1$  is nonempty for each  $n$ . (Along the way, we will indicate how the proof changes when  $S_n^1$  is empty for exactly one player.)

Let  $P_n^0$  be the image of  $\Sigma_n^0$  under  $\rho_n$ . Let  $Z_n^1 \subset Z$  be the set of terminal nodes  $z$  such that  $(h, a) \prec z$  for some  $h \in H_n^*$  and  $a \notin A_n^*$ . Let  $Z_n^0 = Z_n \setminus Z_n^1$ . Then  $P_n^0$  is the set of  $p_n \in P_n$  such that  $p_n(z) = 0$  for all  $z \in Z_n^1$  and thus  $P_n^0$  is a face of  $P_n$ . However, the image  $P_n^1$  of  $\Sigma_n^1$  under  $\rho_n$  need not be a face of  $P_n$ . For  $i = 0, 1$ , let  $P^i = P_1^i \times P_2^i$  and define  $\mathbb{P} = P^0 \times P^1$ .

For each  $p_n \in P_n$ , let  $\psi_{Z_n^1}(p_n)$  be the projection of  $p_n$  to  $\mathbb{R}_+^{Z_n^1}$ ; then  $\psi_{Z_n^1}(p_n) = 0$  iff  $p_n \in P_n^0$ . Fix a point  $\bar{p}_m$  in the interior of  $P_m$  and define  $\eta_n : P_n \rightarrow \mathbb{R}$  by  $\eta(p_n) = \sum_{z \in Z_n^1} p_0(z) \bar{p}_m(z) p_n(z)$ , where  $p_0$  is Nature's enabling strategy. Then  $\eta(p_n) = 0$  iff  $p_n \in P_n^0$ . Choose  $\varepsilon > 0$  such that  $\eta_n(p_n) > \varepsilon$  for all  $p_n \in P_n^1$ . Let  $\mathcal{H}_n$  be the hyperplane in  $\mathbb{R}_+^{Z_n^1}$  with normal  $(p_0(z) \bar{p}_m(z))_{z \in Z_n^1}$  and constant  $\varepsilon$ .  $\mathcal{H}_n$  separates the origin from  $\psi_{Z_n^1}(P_n^1)$ . Let  $\Pi_n^1$  be the intersection of  $\mathcal{H}_n$  with  $\psi_{Z_n^1}(P_n^1)$ . Then there is a function  $\bar{\pi}_n^1$  from  $P_n \setminus P_n^0$  to  $\Pi_n^1$  that maps each  $p_n \notin P_n^0$  to the point  $\varepsilon(\eta_n(p_n))^{-1} \psi_{Z_n^1}(p_n)$ .

Now let  $\Sigma^*$  be a component of the undominated equilibria of the game  $\Gamma$  that is contained in  $\bar{\Sigma}^*$ . Since  $\bar{\Sigma}^* = \bar{\Sigma}_1^* \times \bar{\Sigma}_2^*$ ,  $\Sigma^*$  is expressible as a product  $\Sigma_1^* \times \Sigma_2^*$ , where  $\Sigma_n^*$  is a component of the intersection of  $\bar{\Sigma}_n^*$  with the set of undominated strategies. For each  $n$ , let  $Q_n^*$  be the image of  $\Sigma_n^*$  under  $\rho_n$ , and  $Q^*$  the image of  $\Sigma^*$  under  $\rho$ , i.e. represented in enabling strategies.

Hereafter we say that  $Q^*$  is stable if  $\Sigma^*$  is a stable set of equilibria.<sup>6</sup>

**4.5. The Associated Pseudomanifold.** Given  $Q^*$ , let  $\mathcal{Q}$  be the set of  $(q^*, (p^0, p^1), \pi^1) \in Q^* \times \mathbb{P} \times \Pi^1$  such that there exist  $r^0, \tilde{p}_n^0 \in P_n^0$ ,  $r^1 \in P_n^1$ , and for each  $n$  scalars  $\lambda_n^0, \lambda_n^1, \mu_n^1$  in the interval  $(0, 1]$  such that, if

$$q_n^0 = \lambda_n^0 p_n^0 + (1 - \lambda_n^0) r_n^0 \quad \text{and} \quad q_n^1 = (1 - \lambda_n^1) \tilde{p}_n^0 + \lambda_n^1 (\mu_n^1 p_n^1 + (1 - \mu_n^1) r_n^1),$$

then for each  $n$ :<sup>7</sup>

- (i)  $\bar{\pi}_n^1(q_n^1) = \pi_n^1$ .
- (ii)  $q_n^0$ , and  $r_n^0$  if  $\lambda_n^0 < 1$ , are lexicographic best replies against  $(q_m^*, q_m^0, q_m^1)$ .
- (iii) If  $\mu_n^1 < 1$  then  $r_n^1$  is a best reply against  $q_n^*$  and lexicographically as good a reply against  $(q_m^*, q_m^0, q_m^1)$  as other strategies in  $P_n^1$ .

Let  $\psi : \mathcal{Q} \rightarrow \mathbb{P}$  be the natural projection, i.e.  $\psi(q^*, (p^0, p^1), \pi^1) = (p^0, p^1)$ . Let  $\partial\mathcal{Q} = \psi^{-1}(\partial\mathbb{P})$ . The following technical property is proved in Appendix C.

**Theorem 4.1.**  *$(\mathcal{Q}, \partial\mathcal{Q})$  is a pseudomanifold of the same dimension as  $(\mathbb{P}, \partial\mathbb{P})$ . Moreover, the projection map  $\psi : (\mathcal{Q}, \partial\mathcal{Q}) \rightarrow (\mathbb{P}, \partial\mathbb{P})$  is essential iff  $Q^*$  is stable.*

In case  $S_n^1$  is empty (and  $S_m^1$  is not), then the set  $\Pi_n^1$  is empty. We set  $\mathbb{P} = P^0 \times P_m^1$ . Points in  $\mathcal{Q}$  are of the form  $(q^0, p^0, \pi_m^1)$  and we drop the optimality requirement (iii) for  $n$ .

## 5. STATEMENT AND PROOF OF THE THEOREM

We now state and prove the main theorem. We assume that a solution is represented in terms of enabling strategies, i.e.  $Q^* \subset P$  is a solution iff it is the image under  $\rho$  of a solution  $\Sigma^* \subset \Sigma$ .

**Theorem 5.1.** *Axioms A, B, S imply that each solution is an essential component of the game's undominated equilibria.*

<sup>6</sup>Alternatively, as noted by Mertens [21], one can apply the definition of stability directly to  $Q^*$  as a component of equilibria, represented in terms of enabling strategies, over the space of perturbations of players' enabling strategies.

<sup>7</sup>Lexicographically optimal replies are defined by Blume, Brandenburger, and Dekel [1] and Govindan and Klumpp [2].

*Proof.* Let  $Q^* \subset P$  be a solution. Let  $\deg(\psi)$  be the degree of the projection map  $\psi : (\mathcal{Q}, \partial P) \rightarrow (\mathbb{P}, \partial \mathbb{P})$  defined in Section 4.5 above. Let  $q^{0,*}$  be an arbitrary point in  $Q^*$ . For each  $m$ , choose a point  $p_m^{0,*}$  in the interior of  $P_m^0$  against which  $q^{0,*}$  is a best reply. Such a choice is possible by genericity of payoffs: the interior of the projection of  $\bar{\Sigma}_m^*$  belongs to the interior of  $P_m^0$  and all strategies in  $P_n^1$  are inferior replies against every such point. Since  $q_n^{0,*}$  belongs to  $Q_n^*$ , which consists only of undominated (hence admissible) strategies, there exists a point  $p_m$  in the interior of  $P_m$  against which  $q_n^{0,*}$  is a best reply.  $p_m$  is equivalent to a completely mixed strategy  $\sigma_m$  in  $\Sigma_m$ . Express  $\sigma_m$  as a convex combination of  $\sigma_m^0$  and  $\sigma_m^1$ , where for  $i = 0, 1$ ,  $\sigma_m^i$  belongs to the interior of  $\Sigma_m^i$ . Let  $p_m^0$  and  $p_m^{1,*}$  be the enabling strategies that are equivalent to  $\sigma_m^0$  and  $\sigma_m^1$ , respectively. Then  $p_m^0$  and  $p_m^{1,*}$  are in the relative interiors of  $P_m^0$  and  $P_m^1$  respectively, and  $p_m$  is a convex combination of  $p_m^0$  and  $p_m^{1,*}$ . It follows that  $x^* \equiv (q^{0,*}, (p^{0,*}, p^{1,*}), \pi^{1,*})$  belongs to  $\mathcal{Q} \setminus \partial \mathcal{Q}$ , where for each  $n$ ,  $\pi_n^{1,*} = \bar{\pi}_n^1(p_n^1)$ , and in the definition of  $\mathcal{Q}$ ,  $q_n^0$  is  $p_n^{0,*}$  and  $q_n^1$  is  $p_n^1$ .

For any neighborhood  $U(q^{0,*})$  of  $q^{0,*}$  we can choose a neighborhood  $U(x^*)$  of  $x^*$  that is homeomorphic to a simplex, is contained in  $\mathcal{Q} \setminus \partial \mathcal{Q}$  and such that the projection onto the first factor is contained in the neighborhood  $U(q^{0,*})$ , i.e. if  $(q^0, (p^0, p^1), \pi^1) \in U(x^*)$  then  $q^0 \in U(q^{0,*})$ . For the proof we construct a metagame such that: (i) if  $\deg(\psi)$  is zero then there is no quasi-perfect equilibrium of the metagame projecting to an equilibrium in  $Q^*$ ; and (ii) if  $\deg(\psi)$  is nonzero then all its quasi-perfect equilibria that project to points in  $Q^*$ , project to points in  $U(q^{0,*})$ . Since  $U(q^{0,*})$  is an arbitrary neighborhood of the arbitrary point  $q^{0,*}$  in  $Q^*$ , this proves the theorem.

**5.1. Preliminaries.** Fix some  $\bar{p}$  in the interior of  $\mathbb{P}$  and define  $\iota : \partial \mathbb{P} \rightarrow \partial \mathbb{P}$  as follows:  $\iota(p)$  is the unique point in the boundary of the form  $\lambda p + (1 - \lambda)\bar{p}$  for  $\lambda < 0$ . Then  $\iota$  has no fixed point. Let  $g_{\partial \mathcal{Q}} : \partial \mathcal{Q} \rightarrow \mathbb{P}$  be the function  $\iota \circ \psi_{\partial \mathcal{Q}}$ , where  $\psi_{\partial \mathcal{Q}}$  is the restriction of  $\psi$  to  $\partial \mathcal{Q}$ . Then  $g_{\partial \mathcal{Q}}$  has no point of coincidence with  $\psi$ , i.e., for each  $x \in \partial \mathcal{Q}$ ,  $g_{\partial \mathcal{Q}}(x) \neq \psi(x)$ . Let  $d$  be the dimension of  $\mathbb{P}$  and let  $\alpha$  be the degree of  $\psi$ . If  $\alpha = 0$  then, since  $\iota$  is a homeomorphism,  $\delta^* i^* g_{\partial \mathcal{Q}}^*(1) = 0$  in  $H^\delta(\mathcal{Q}, \partial \mathcal{Q})$ , where  $i : \partial \mathcal{Q} \rightarrow \mathcal{Q}$  is the inclusion map, 1 is the generator of  $H^{d-1}(\mathbb{P}) \approx \mathbb{Z}$ , and  $\delta^*$  is the coboundary operator. By the Hopf Extension Theorem [29, Corollary 8.1.18],  $g_{\partial \mathcal{Q}}$  can be extended to a map  $g : \mathcal{Q} \rightarrow \partial \mathbb{P}$ . Obviously  $g$  has no point of coincidence with  $\psi$ . If  $\alpha \neq 0$  then construct a map  $g_{\partial U(x^*)} : \partial U(x^*) \rightarrow \partial \mathbb{P}$  of degree  $\alpha$  such that  $\delta^* i^* g_{\partial \mathcal{Q} \cup \partial U(x^*)}^*(1) = 0$  in  $H^\delta(\mathcal{Q} \setminus (U(x^*) \setminus \partial U(x^*)), \partial \mathcal{Q} \cup \partial U(x^*))$ , where now  $i : (\partial \mathcal{Q} \cup \partial U(x^*)) \rightarrow \mathcal{Q} \setminus (U(x^*) \setminus \partial U(x^*))$  is the inclusion map and  $\delta^*$  is the coboundary operator. Using again the Hopf Extension Theorem, there exists a map  $g_{\partial U(x^*)}$  such that the two maps  $g_{\partial \mathcal{Q}}$  and  $g_{\partial U(x^*)}$  can be extended to a map from  $\mathcal{Q} \setminus (U(x^*) \setminus \partial U(x^*))$  to  $\mathbb{P}$ ;

furthermore, by mapping points in  $U(x^*)$  to  $\mathbb{P}$  in a way that extends  $g_{\partial U(x^*)}$ , we obtain a map  $g : \mathcal{Q} \rightarrow \mathbb{P}$  such that all its points of coincidence with  $\psi$ , of which there is at least one, are contained in  $U(x^*) \setminus \partial U(x^*)$ .

By construction there exists  $\alpha > 0$  such that for all  $x \in \mathcal{Q}$ ,  $\|g(x) - \psi(x)\| \leq \alpha$  iff  $\alpha \neq 0$  and  $x \in U(x^*)$ . Choose a triangulation  $\mathcal{K}_n^i$  of  $P_n^i$  for each  $n$  and  $i = 0, 1$  such that the diameter of each simplex is no more than  $\alpha/2$ . For each  $n$  there exists for each  $i$  a triangulation  $\mathcal{L}_n^i$  of  $P_n^i$  and a triangulation  $\mathcal{L}_n^{\Pi}$  of  $\Pi_n^1$  such that, letting  $\mathcal{L}$  be the resulting multisimplicial subdivision of  $P^0 \times \mathbb{P} \times \Pi^1$ ,  $g$  has a multisimplicial approximation  $\tilde{g}$  [9, Appendix B] with the triangulation of the range given by  $\mathcal{K} = \prod_{n,i} \mathcal{K}_n^i$ . Observe that if for some  $x = (q^0, p^0, p^1, \pi^1)$  there exists a multisimplex  $K$  that contains  $\psi(x)$  and  $\tilde{g}(x)$  then, since  $\tilde{g}$  is a multisimplicial approximation of  $g$ , that multisimplex is a face of one that contains  $g(x)$  and thus  $\|g(x) - \psi(x)\| \leq \alpha$ ; in particular, if  $x$  also belongs to  $\mathcal{Q}$ , then  $x \in U(x^*)$ , and thus  $q^0 \in U(q^{0,*})$ . As in [9, Theorem B.2], now take a further polyhedral subdivision  $\mathcal{T}$  of  $\mathcal{L}$  and let  $\gamma$  be the convex function generated by  $\mathcal{T}$ , i.e.  $\gamma$  is piecewise linear and linear precisely on each full-dimensional polyhedron of the subdivision.

Hereafter, unless explicitly stated, we represent players' pure and mixed strategies in terms of the induced enabling strategies; thus, pure strategies in  $S_n$  are represented as vertices of  $P_n$  and mixed strategies are mixtures of these vertices.

**5.2. A Game with Redundant Strategies.** We first construct a game with redundant pure strategies that is the basis for the metagame specified later. For each fixed  $\hat{p} = (p^0, p^1) \in \mathbb{P}$  and  $\delta \in (0, 1)$ , consider the following game  $\Gamma(\delta, \hat{p})$ . Each player  $n$  chooses a strategy in the following manner (and unaware of his opponent's choices). Initially, player  $n$  provisionally chooses a pure strategy  $s_n^0 \in S_n^0$ , or he rejects all strategies in  $S_n^0$ .

- If initially he chooses a strategy  $s_n^0$  then at a subsequent second stage he can retain  $s_n^0$  or revise his choice. At this second stage the revisions available are 'duplicate' pure strategies in a set  $T_n^0(\delta, p_n^0)$ , each of which is a mixed strategy of the form  $t_n(\delta, p_n^0) \equiv (1 - \delta)t_n + \delta p_n^0$  for some  $t_n \in S_n^0$ .
- If he rejects all strategies in  $S_n^0$  then at a second stage he can choose among the strategies in  $T_n^0(\delta, p_n^0)$  or reject them all.<sup>8</sup> If he rejects them all then at a third stage he chooses among the pure strategies in  $S_n^1 \cup T_n^1(\delta, p_n^1)$ , where each strategy in  $T_n^1(\delta, p_n^1)$  is a duplicate of the form  $t_n(\delta, p_n^1) \equiv (1 - \delta)t_n + \delta p_n^1$  for some  $t_n \in S_n^0$ .

---

<sup>8</sup>It would have sufficed, at this stage, to give player  $n$  the option of playing just the strategy  $p_n^0$  instead of all the strategies in  $T_n^0(\delta, p_n^0)$ , which we do only for economy in notation.

In  $\Gamma(\delta, \hat{p})$  the set of  $n$ 's pure strategies is  $\tilde{S}_n(\delta, \hat{p}_n) \equiv S_n \cup T_n^0(\delta, p_n^0) \cup T_n^1(\delta, p_n^1)$ . Thus, game  $\Gamma(\delta, \hat{p})$  has the same reduced normal form as  $\Gamma$ .

**5.3. Extensive Form of the Metagames.** Now we specify a family of similar metagames  $\tilde{\Gamma}^\delta$ , one for each  $\delta > 0$ .

Before the insiders play, thirteen outsiders, denoted players  $o_0$  and  $o_{n,j}^i$  for  $n = 1, 2$ ,  $i = 0, 1$  and  $j = 1, 2, 3$ , move simultaneously. Outsider  $o_0$  chooses a full-dimensional polyhedron  $T$  of the polyhedral complex  $\mathcal{T}$ . Outsider  $o_{n,j}^i$ , for  $n = 1, 2$ ,  $j = 1, 3$  and  $i = 0, 1$ , chooses a point in the vertex set  $V_n^i$  of  $\mathcal{K}_n^i$ . Outsider  $o_{n,2}^0$  chooses a point in a finite subset  $S_n^{0,\delta}$  of  $P_n^0$  chosen such that every point in  $P_n^0$  is within  $\delta$  of some point in  $S_n^{0,\delta}$ ; outsider  $o_{n,2}^1$  chooses a point in a finite subset  $S_n^{1,\delta}$  of  $\Pi_n^1$  such that every point in  $\Pi_n^1$  is within  $\delta$  of some point in  $S_n^{1,\delta}$ .

For outsider  $o_{n,1}^i$ , each pure strategy  $v_n^i$  corresponds to a point in  $P_n^i$  denoted  $p_n^i(v_n^i)$ . Therefore, each mixed strategy  $\sigma_{n,1}^i$  corresponds to a point in  $P_n^i$ , denoted  $p_n^i(\sigma_{n,1}^i)$ , which is obtained by taking the appropriate average of the points induced by the pure strategies in the support of  $\sigma_{n,1}^i$ . Likewise a mixed strategy  $\sigma_{n,2}^0$  of  $o_{n,2}^0$  corresponds to a point  $q_n^0(\sigma_{n,2}^0)$  in  $P_n^0$ , and a mixed strategy  $\sigma_{n,2}^1$  of  $o_{n,2}^1$  corresponds to a point  $\pi_n^1(\sigma_{n,2}^1)$  in  $\Pi_n^1$ .

A mixed-strategy profile  $\tilde{\sigma}_o$  for the outsiders induces a point  $(q^0(\tilde{\sigma}_o), (p^0(\tilde{\sigma}_o), p^1(\tilde{\sigma}_o)), \pi^1(\tilde{\sigma}_o))$  in  $P^0 \times \mathbb{P} \times \Pi^1$ , where for each  $n$  and  $i$ ,  $p_n^i(\tilde{\sigma}_o)$  depends on the choice by  $o_{n,1}^i$ , and  $q_n^0(\tilde{\sigma}_o)$  and  $\pi_n^1(\tilde{\sigma}_o)$  depend on the choices by  $o_{n,2}^0$  and  $o_{n,2}^1$  respectively.

After each pure-strategy profile  $\tilde{s}_o$  of the outsiders there follows a copy of the game  $\Gamma(\delta, \hat{p}(\tilde{s}_o))$ . That is, if in the profile  $\tilde{s}_o$  outsiders  $o_{n,1}^i$  choose points  $v_n^i$ , then there follows a copy of  $\Gamma(\delta, \hat{p})$  after these choices, where for each  $n$ ,  $\hat{p}_n = (p_n^0(v_n^0), p_n^1(v_n^1))$ . However, the information sets in  $\tilde{\Gamma}^\delta$  are such that the insiders play without knowing which copy of  $\Gamma(\delta, \hat{p})$  they are playing. The sets of duplicate strategies available are therefore now denoted by  $T_n^0(\delta)$  and  $T_n^1(\delta)$ , omitting the reference to  $p_n^0$  and  $p_n^1$ , since the insiders are uninformed about which mixtures were implemented by outsiders. Put differently, for  $i = 0, 1$  and  $t_n \in S_n^i$ , the exact duplicate strategy implemented by choosing  $t_n^i(\delta) \in T_n^i(\delta)$  depends on the choice by the outsider  $o_{n,1}^i$  that insiders do not observe. Thus in the metagame  $\tilde{\Gamma}^\delta$ , player  $n$ 's set of pure strategies, up to duplication of pure strategies, is  $\tilde{S}_n(\delta) \equiv S_n \cup T_n^0(\delta) \cup T_n^1(\delta)$ .

That the metagame  $\tilde{\Gamma}^\delta$  embeds  $\Gamma$  follows from Proposition 3.3. The strategies in  $S_n$  are available as pure strategies in  $\tilde{S}_n(\delta)$  and the other pure strategies, which belong to  $T_n^i(\delta)$ , for  $i = 0, 1$ , implement mixtures in  $\Sigma_n$  that depend on the choices of outsiders  $o_{n,1}^i$ .



**5.4. Outsiders' Payoffs in the Metagames.** Next we describe the outsiders' payoffs in each metagame  $\tilde{\Gamma}^\delta$ .

The payoffs to  $o_0$  depend on the choices of all outsiders except outsiders  $o_{n,3}^i$  for  $i = 0, 1$  and  $n = 1, 2$ . Recall that the convex function  $\gamma$  is linear over each full-dimensional polyhedron  $T$  of  $\mathcal{T}$ . This linear function extends uniquely to a linear function  $\gamma_T$  over  $P^0 \times \mathbb{P} \times \Pi^1$ . Every mixed strategy profile of the other outsiders induces a unique point  $(\tilde{q}, \hat{p}, \pi^1) \in P^0 \times \mathbb{P} \times \Pi^1$  and  $o_0$ 's payoff from choosing  $T$  is  $\gamma_T(\tilde{q}, \hat{p}, \pi^1)$ .

Outsider  $o_{n,1}^i$  wants to mimic  $o_{n,3}^i$ . In particular, if he chooses  $v_n^i$  and  $o_{n,3}^i$  chooses  $w_n^i$  then his payoff is one if  $v_n^i = w_n^i$  and zero otherwise.

Outsider  $o_{n,2}^0$  wants to mimic the actual choice implemented by player  $n$  when this choice belongs to  $P_n^0$ . Similarly, outsider  $o_{n,2}^1$  wants to mimic the  $\pi_n^1$  implied by  $n$ 's choice when he plays a strategy in  $P_n^1$ . Specifically, for  $i = 0, 1$ , let  $\varphi_n^i : \mathbb{R}^{Z_n^i} \rightarrow \mathbb{R}$  be the function given by  $\varphi_n^i(r) = \sum_{z \in Z_n^i} r_z^2$ . For each  $r \in \mathbb{R}^{Z_n^i}$ , let  $\xi_n^i(r, \cdot)$  be the affine approximation to  $\varphi_n^i$  at  $r$ , i.e. for each  $r' \in \mathbb{R}^{Z_n^i}$ ,  $\xi_n^i(r, r') = \sum_z (r_z^2 + 2r_z(r'_z - r_z))$ . Suppose now that  $o_{n,2}^i$  chooses a pure strategy  $s_{n,2}^{i,\delta}$  and  $n$  chooses a pure strategy  $\tilde{s}_n$  in  $\tilde{S}_n(\delta)$ . If  $\tilde{s}_n$  is in  $T_n^0(\delta)$  or  $T_n^1(\delta)$  then let  $q_n$  be the actual strategy in  $P_n$  that is implemented based on  $p_n^0(v_n^0)$  or  $p_n^1(v_n^1)$  where for each  $i$ ,  $v_n^i$  is the choice of outsider  $o_{n,1}^i$ ; and otherwise let  $q_n = \tilde{s}_n$ . For outsider  $o_{n,2}^0$ , his payoff is zero if  $q_n \notin P_n^0$ ; otherwise, it is  $\xi_n^0(s_{n,2}^{0,\delta}, q_n)$ . For outsider  $o_{n,2}^1$ , his payoff is zero if  $q_n \in P_n^0$ ; otherwise it is  $\xi_n^1(s_{n,2}^{1,\delta}, \bar{\pi}_n^1(q_n))$ .

The payoff to outsider  $o_{n,3}^i$  depends on the choices of all other outsiders. If  $o_0$  chooses a polyhedron  $T$  then there exists a unique multisimplex  $L$  of  $\mathcal{L}$  that contains  $T$ . For each vertex  $w_n^i$  of  $V_n^i$ , and each vertex  $\tilde{v}$  of  $L$ , let  $u_{n,2}^i(T, \tilde{v}, w_n^i) = 1$  if  $w_n^i$  is the image of  $\tilde{v}$  under  $\tilde{g}_n^i$  and zero otherwise, where  $\tilde{g}_n^i$  is the  $(n, i)$ -th coordinate map of  $\tilde{g}$ . The function  $u_{n,2}^i$  extends multilinearly to  $L$  and, since  $L$  is full-dimensional, to the whole of  $P^0 \times \mathbb{P} \times \Pi^1$ , denoted still by  $u_{n,2}^i(T, \cdot, w_n^i)$ . Given an arbitrary mixed strategy of the other players, if  $o_0$  chooses  $T$  and  $o_{n,2}^i$  chooses  $w_n^i$  then the payoff of  $o_{n,2}^i$  is  $u_{n,2}^i(T, (p, q), w_n^i)$ , where  $(p, q)$  is the point in  $P^0 \times \mathbb{P} \times \Pi^1$  induced by the mixed strategies of the other players.

**5.5. Outsiders' Strategies in a Quasi-Perfect Equilibrium.** If  $Q^*$  is a solution of  $\Gamma$  then, by Axioms B and S, in the metagame  $\tilde{\Gamma}^\delta$  there exists a quasi-perfect equilibrium  $\tilde{b}^\delta$  whose equivalent mixed-strategy profile  $\tilde{\sigma}^\delta$  belongs to a solution and whose image under the map from the metagame  $\tilde{\Gamma}^\delta$  to  $\Gamma$  is a point in  $Q^*$ .

Each player  $n$ 's strategy in  $\tilde{b}^\delta$  necessarily has the following feature. He avoids going to his information set where his choices are among the strategies in  $S_n^1(\delta) \cup T_n^1(\delta)$ , since each of

these strategies chooses a non-equilibrium action at some information set on the equilibrium path. Let  $q_n^{0,\delta}$  be  $n$ 's actual strategy in  $P_n^0$  that is implemented by  $n$ 's strategy in the profile  $\tilde{b}^\delta$  in the metagame  $\tilde{\Gamma}^\delta$ . By construction,  $q_n^{0,\delta}$  belongs to  $Q_n^*$ .

Let  $\tilde{x}^\delta \equiv (\tilde{q}^{0,\delta}, p^{0,\delta}, \tilde{\pi}^{1,\delta})$  be the point in  $P^0 \times \mathbb{P} \times \Pi^1$  that is induced by the profile  $\tilde{\sigma}_o^\delta$  of the outsiders' strategies in the equilibrium  $\tilde{\sigma}^\delta$ .

Under  $\tilde{b}^\delta$ , after  $n$  has rejected all strategies in  $S_n^0$  and  $T_n^0(\delta)$ , consider the strategy implemented by  $n$ . Let  $(1 - \alpha_n^{1,\delta})$  be the total probability of choosing a duplicate in  $T_n^1(\delta)$  under  $\tilde{b}^\delta$ . Then  $n$ 's choice at this information set is equivalent to an enabling strategy in  $P_n$  of the form

$$\check{q}_n^\delta \equiv (1 - \alpha_n^{1,\delta})((1 - \delta)\check{p}_n^{0,\delta} + \delta p_n^{1,\delta}) + \alpha_n^{1,\delta} r_n^{1,\delta},$$

where: (i)  $\check{p}_n^0(\delta)$  is the mixture over strategies  $t_n$  such that the strategy  $t_n^1(\delta)$  is played with positive probability at this information set; (ii)  $p_n^{1,\delta} = p_n^i(\tilde{\sigma}_{n,1}^i)$  is the enabling strategy induced by the equilibrium strategy  $\tilde{\sigma}_{n,1}^i$  of outsider  $o_{n,1}^i$ ; (iii)  $r_n^{1,\delta}$  is the enabling strategy in  $P_n^1$  that is obtained from  $n$ 's actual mixture over strategies in  $S_n^1$  if  $\alpha_n^{1,\delta} > 0$ , and is arbitrary otherwise. Let

$$q_n^{1,\delta} = (\delta(1 - \alpha_n^{1,\delta}) + \alpha_n^{1,\delta})^{-1}(\delta(1 - \alpha_n^{1,\delta})p_n^{1,\delta} + \alpha_n^{1,\delta} r_n^{1,\delta}).$$

Then  $\bar{\pi}_n^1(\check{q}_n^\delta) = \bar{\pi}_n^1(q_n^{1,\delta}) \equiv \pi_n^{1,\delta}$ .

The following lemma characterizes the important aspects of the outsiders' equilibrium strategies.

**Lemma 5.2.** *The equilibrium strategies of the outsiders satisfy the following properties.*

- (1) *For each  $n$ , suppose the vertices in the support  $V_n^{i,\delta}$  of  $o_{n,3}^i$ 's equilibrium strategy span a simplex  $K_n^{i,\delta}$  of  $\mathcal{K}_n^i$ . Then  $p_n^{i,\delta}$  belongs to  $K_n^{i,\delta}$ .*
- (2) *If every polyhedron in the support of  $o_0$ 's strategy contains  $\tilde{x}^\delta$  then, for each  $n$  and  $i$ , the vertices in  $V_n^{i,\delta}$  span a simplex  $K_n^{i,\delta}$ , and  $\tilde{g}_{n,i}(\tilde{x}^\delta)$  belongs to the interior of a simplex  $\bar{K}_n^{i,\delta}$  that has  $K_n^{i,\delta}$  as a face.*
- (3) *Every polyhedron in the support of  $o_0$ 's strategy contains  $\tilde{x}^\delta$ .*
- (4)  *$\tilde{q}_n^{0,\delta}$  is within  $\delta$  of  $q_n^{0,\delta}$  and  $\tilde{\pi}_n^{1,\delta}$  is within  $\delta$  of  $\pi_n^{1,\delta}$ .*

*Proof of Lemma.* Outsider  $o_{n,1}^i$  wants to mimic outsider  $o_{n,3}^i$ . So, if the vertices of  $V_n^{i,\delta}$  span a simplex  $K_n^{i,\delta}$  then the payoff to  $o_{n,1}^i$  from choosing a vertex  $w_n^i$  is positive if it belongs to  $V_n^{i,\delta}$ , and zero otherwise. Point (1) follows.

Let  $\bar{L} = L^0 \times (\tilde{L}^0 \times L^1) \times L^\Pi$  be the unique multisimplex of  $\mathcal{L}$  that contains  $\tilde{x}^\delta$  in its interior. For each polyhedron  $T$  in the support of  $o_0$ 's strategy, there exists a full-dimensional multisimplex  $\hat{L}$  of  $\mathcal{L}$  that contains  $T$ . Obviously  $\hat{L}$  has  $\bar{L}$  as a face.  $o_{n,3}^i$ 's payoff from choosing

a strategy  $w_n^i$  if  $o_0$  chooses such a  $T$ , and given the strategies of the other outsiders, is positive if it is the image of a vertex of  $\bar{L}$  under  $\tilde{g}_n^i$  and zero otherwise. Since the image of the vertices of  $\bar{L}$  under the coordinate function  $\tilde{g}_n^i$  span a simplex  $\bar{K}_n^{i,\delta}$ ,  $\tilde{g}_n^i(\tilde{x}^\delta) \in \bar{K}_n^{i,\delta}$  and the vertices of  $V_{n,i}^\delta$  span a face of  $\bar{K}_n^{i,\delta}$ . Therefore, point (2) follows.

For each polyhedron  $T$  of  $\mathcal{T}$ ,  $o_0$ 's payoff from  $T$  is  $\gamma_T(\tilde{x}^\delta)$  and by construction,  $\gamma_T(\tilde{x}^\delta) \leq \gamma(\tilde{x}^\delta)$  with the inequality being strict iff  $\tilde{x}^\delta$  does not belong to  $T$ , which proves (3).

It remains to prove (4). The actual strategy implemented by  $n$  is  $q_n^{0,\delta}$ , which belongs to  $P_n^0$ .  $o_{n,2}^0$ 's payoff function is such that his best replies to  $q_n^{0,\delta}$  are the points in  $S_n^{0,\delta}$  that are closest to  $q_n^{0,\delta}$ . Thus  $\tilde{q}_n^{0,\delta}$  is within  $\delta$  of  $q_n^{0,\delta}$ . Since  $q_n^{0,\delta}$  belongs to  $P_n^0$ , all of  $o_{n,2}^1$ 's strategies yield a payoff of zero against  $q_n^{0,\delta}$ . However, since the behavioral strategy  $\tilde{b}^\delta$  is a quasi-perfect equilibrium, there exists a sequence of completely mixed behavioral strategies  $\tilde{b}^{\varepsilon,\delta}$  converging to  $\tilde{b}^\delta$  against which  $o_{n,2}^1$ 's equilibrium strategy  $\tilde{\sigma}_{n,2}^\delta$  is a best reply. Under the sequence  $\tilde{b}^{\varepsilon,\delta}$ , there is a positive probability that player  $n$  moves into the third stage of his choice and makes a choice among strategies in  $S_n^1 \cup T_n^1(\delta)$ . The fact that  $\tilde{\sigma}_{n,2}^\delta$  is optimal against the sequence implies that it is optimal against the limiting choice  $\tilde{q}_n^\delta$  there. Since  $\tilde{\pi}_n^1(\tilde{q}_n^\delta) = \pi_n^{1,\delta}$ ,  $o_{n,2}^1$ 's best replies are within  $\delta$  of  $\pi_n^{1,\delta}$  and thus  $\pi_n^{1,\delta}$  is within  $\delta$  of  $\pi_n^{1,\delta}$  as well.  $\square$

**5.6. Limits of the Quasi-Perfect Equilibria of the Metagames.** Consider a sequence of  $\delta$ 's converging to zero and a corresponding sequence  $\tilde{b}^\delta$  of quasi-perfect equilibria in solutions of the metagames  $\tilde{\Gamma}^\delta$ . Let  $\tilde{\sigma}^\delta$  be an equivalent sequence of mixed strategies and let  $(\tilde{q}^{0,\delta}, (p^{0,\delta}, p^{1,\delta}), \tilde{\pi}^{1,\delta})$  be the sequence in  $Q^* \times \mathbb{P} \times \Pi^1$  induced by the outsiders' strategies.

Let  $\tilde{q}^{0,0}$ ,  $(p^{0,0}, p^{1,0})$ , and  $\tilde{\pi}^{1,0}$  be the corresponding limits of  $\tilde{q}^{0,\delta}$ ,  $(p^{0,\delta}, p^{1,\delta})$ , and  $\tilde{\pi}^{1,\delta}$ . Let  $q^{0,0}$  and  $\pi^{1,0}$  be the limits of  $q^{0,\delta}$  and  $\pi^{1,\delta}$ . By properties (1)-(3) of the previous lemma, for each  $\delta, n, i$ , there exists a simplex  $\bar{K}_n^{i,\delta}$  of  $\mathcal{K}_n^i$  that contains both  $\tilde{g}_{n,i}(\tilde{q}^{0,\delta}, (p^{0,\delta}, p^{1,\delta}), \tilde{\pi}^{1,\delta})$  and  $p_n^{i,\delta}$ , with the former belonging to its interior. By property (4) of the previous lemma,  $\tilde{q}^{0,0} = q^{0,0}$  and  $\tilde{\pi}^{1,0} = \pi^{1,0}$ .

By passing to a subsequence, we can assume that there exist multisimplices  $\bar{L}$  of  $\mathcal{L}$  and  $\bar{K}$  of  $\mathcal{K}$  such that for all  $\delta$ ,  $(\tilde{q}^{0,\delta}, (p^{0,\delta}, p^{1,\delta}), \tilde{\pi}^{1,\delta})$  belongs to the interior of  $\bar{L}$  and its image under  $\tilde{g}$  belongs to the interior of  $\bar{K}$ —hence  $\bar{K}$  also contains  $(p^{0,\delta}, p^{1,\delta})$ . Going to the limit,  $(q^{0,0}, (p^{0,0}, p^{1,0}), \pi^{1,0})$  belongs to  $\bar{L}$  and its image under  $\tilde{g}$  belongs to  $\bar{K}$ ; also  $(p^{0,0}, p^{1,0})$  belongs to  $\bar{K}$ . Since  $\bar{L}$  and  $\bar{K}$  are multisimplices and  $\tilde{g}$  is a multisimplicial map, we have that for all  $\delta$ ,  $x^\delta \equiv (q^{0,0}, (p^{0,\delta}, p^{1,0}), \pi^{1,0})$  belongs to  $\bar{L}$ , and  $\tilde{g}(x^\delta), (p^{0,\delta}, p^{1,0}) \in \bar{K}$ . If we show that for small  $\delta$ ,  $x^\delta$  belongs to  $\mathcal{Q}$ , then by construction this implies that the degree of  $\psi$  is nonzero and  $x^\delta \in U(x^*)$ , and therefore  $q^{0,0} \in U(q^{0,*})$ , which proves the theorem. The remainder of the proof establishes this fact.

**5.7. Insiders' Strategies in a Quasi-Perfect Equilibrium.** Let  $\tilde{b}^{\varepsilon, \delta}$  be a sequence of completely mixed behavioral strategies converging to  $\tilde{b}^\delta$  against which for each insider  $n$  and each information set of  $n$  in  $\tilde{\Gamma}^\delta$ , his continuation strategy is optimal. We now study the important features of both the sequence and its limit. If  $n$  chooses  $s_n \in S_n^0$  in the first stage then in the second stage he has the option of revising this strategy to play something in  $T_n^0(\delta)$ . Therefore, quasi-perfection implies that player  $n$  will end up implementing  $s_n$  with positive probability in  $\tilde{b}_n^\delta$  only if this strategy is at least as good a reply against the sequence  $\tilde{b}^{\varepsilon, \delta}$  as the strategies in  $T_n^0(\delta)$ . Likewise, player  $n$  has the option of playing a strategy in  $T_n^0(\delta)$  before he decides to play a strategy in  $S_n^1$  and even when he makes a choice among these strategies, he has the option of choosing a strategy in  $T_n^1(\delta)$ . Therefore, at the information set that follows his choice of avoiding strategies in  $S_n^0$ ,  $\tilde{b}_n^\delta$  assigns a positive probability to moving on to a third stage and then choosing a strategy in  $T_n^1(\delta) \cup S_n^1$  only if one of these strategies is at least as good a reply against the sequence  $\tilde{b}^{\varepsilon, \delta}$  as all the strategies in  $T_n^0(\delta)$ . Furthermore, at the information set obtained after  $n$  avoids strategies in  $S_n^0 \cup T_n^0(\delta)$ ,  $\tilde{b}_n^\delta$  assigns a positive probability to a strategy  $s_n$  in  $S_n^1$  only if  $s_n$  is at least as good a reply against the sequence  $\tilde{b}^{\varepsilon, \delta}$  as all the strategies in  $S_n^1 \cup T_n^1(\delta)$ .

Let  $\tilde{\sigma}^{\varepsilon, \delta}$  be a sequence of mixed strategy profiles that is equivalent to the sequence  $\tilde{b}^{\varepsilon, \delta}$  of behavioral strategy profiles. For each player  $n$ , his strategy  $\tilde{\sigma}_n^{\varepsilon, \delta}$  in the sequence is a mixture over his pure strategy set  $\tilde{S}_n = S_n \cup T_n^0(\delta) \cup T_n^1(\delta)$ . However, the implications of  $n$ 's strategy (for  $m$ 's choices) depend on the choices of the outsiders through their implications for strategies in  $T_n^0(\delta)$  and  $T_n^1(\delta)$ . Each strategy  $t_n^i(\delta)$  plays  $t_n$  with probability  $(1 - \delta)$  and with probability  $\delta$  plays a strategy in  $P_n^i$  that is determined by  $o_{n,1}^i$ 's strategy. In order to fully capture the impact that  $o_{n,1}^i$  has on  $t_n^i(\delta)$ , let  $W_n^i$  be the vertex set of  $\mathcal{K}_n^i$ . Let  $\bar{T}_n^i(\delta)$  be the union over all  $w_n^i \in W_n^i$  of the sets  $T_n^i(\delta, p_n^i(w_n^i))$ . Let  $\bar{S}_n = S_n \cup \bar{T}_n^0(\delta) \cup \bar{T}_n^1(\delta)$  and let  $\bar{\Sigma}_n$  be the set of mixtures over  $\bar{S}_n$ .

Since  $W_n^i$  is the pure strategy set of outsider  $o_{n,1}^i$ , and  $\tilde{\sigma}^{\varepsilon, \delta}$  is a sequence of completely mixed strategies,  $\tilde{\sigma}_{o_{n,1}^i}^{\varepsilon, \delta}$  has  $W_n^i$  as its support for all  $\varepsilon$  in the sequence. The sequence  $\tilde{\sigma}^{\varepsilon, \delta}$  induces a mixed strategy  $\bar{\sigma}_n^{\varepsilon, \delta}$  in  $\bar{S}_n$  for each  $n$  as follows. For each  $s_n \in S_n$ , the probability  $\bar{\sigma}_n^{\varepsilon, \delta}(s_n)$  of  $s_n$  is  $\tilde{\sigma}_n^{\varepsilon, \delta}(s_n)$ . For each  $i = 0, 1$ ,  $w_n^i \in W_n^i$  and  $t_n \in S_n^0$ , the probability  $\bar{\sigma}_n^{\varepsilon, \delta}(t_n^i(\delta, p_n^i(w_n^i)))$  is  $\tilde{\sigma}_n^{\varepsilon, \delta}(t_n^i(\delta))\tilde{\sigma}_{o_{n,1}^i}^{\varepsilon, \delta}(w_n^i)$ . From player  $m$ 's perspective it is the sequence  $\bar{\sigma}_n^{\varepsilon, \delta}$ , or rather its equivalent sequence in  $P_n$ , that matters for his choice.

**5.8. The Induced Lexicographic Probability System.** By Blume, Brandenburger, and Dekel [1, Appendix Proposition 2], we can construct for each player  $n$  a lexicographic probability system (LPS)  $\bar{\Lambda}_n^\delta = (\bar{\sigma}_n^{0,\delta}, \dots, \bar{\sigma}_n^{l_n(\delta),\delta})$  over his strategies in  $\bar{S}_n$  such that for each  $\varepsilon$  in a subsequence converging to zero,

$$\bar{\sigma}_n^{\varepsilon,\delta} = (1 - \nu_0(\varepsilon))(\bar{\sigma}_n^{0,\delta} + \nu_0(\varepsilon)((1 - \nu_1(\varepsilon))\bar{\sigma}_n^{1,\delta} + \nu_1(\varepsilon)((1 - \nu_2(\varepsilon))\bar{\sigma}_n^{2,\delta} + \dots + \nu_{l_n(\delta)-1}(\varepsilon)\bar{\sigma}_n^{l_n(\delta),\delta})),$$

where  $(\nu_0(\varepsilon), \dots, \nu_{l_n(\delta)-1}(\varepsilon))$  is a sequence in  $\mathbb{R}_+^{l_n(\delta)}$  converging to the origin.

Quasi-perfection implies that  $\bar{\sigma}_n^{0,\delta}$  is equivalent to the enabling strategy  $q_n^{0,\delta}$  in  $\Gamma$  and it is a lexicographic best reply to  $\bar{\Lambda}_m^\delta$ . Let  $l_n^{0,\delta}$  be the first level in  $\bar{\Lambda}_n^\delta$  at which some strategy in  $\bar{T}_n^0(\delta)$  appears with positive probability. Let  $l_n^{1,\delta}$  be the first level of  $\bar{\Lambda}_n^\delta$  at which some strategy in  $S_n^1 \cup \bar{T}_n^1(\delta)$  appears with positive probability.

From the LPS  $\bar{\Lambda}_n^\delta$  construct an LPS  $\Lambda^\delta = (q_n^{0,\delta}, \dots, q_n^{l_n^{1,\delta},\delta})$  for  $\Gamma$  where for each  $l$ ,  $q_n^{l,\delta}$  is an enabling strategy in  $\Gamma$  that is equivalent to  $\bar{\sigma}_n^{l,\delta}$ .  $q_n^{0,\delta}$  is a lexicographic best reply against  $\Lambda_m^\delta$ . For each  $w_n^i \in W_n^i$ , the total probability under  $\bar{\sigma}_n^{l_n^{i,\delta}}$  of the strategies in  $T_n^i(\delta, p_n^i(w_n^i))$  is the limit  $\tilde{\sigma}_{\sigma_{n,1}^i}^\delta(w_n^i)$  of the sequence  $\sigma_{\sigma_{n,1}^i}^{\varepsilon,\delta}(w_n^i)$ . Since  $p_n^i(\tilde{\sigma}_{\sigma_{n,1}^i}^\delta)$  is, by definition,  $p_n^{i,\delta}$ , we have that level  $l_n^{i,\delta}$  is expressible as a convex combination  $\lambda_n^{l_n^{i,\delta},\delta} p_n^{i,\delta} + (1 - \lambda_n^{l_n^{i,\delta},\delta}) r_n^{l_n^{i,\delta},\delta}$  with  $\lambda_n^{l_n^{i,\delta},\delta} > 0$  and  $r_n^{l_n^{i,\delta},\delta} \in P_n$ . Since  $l_n^{1,\delta}$  is the first level of  $\Lambda^\delta$  that does not have its support in  $P_n^0$ ,  $\bar{\pi}_n^1(q_n^{l_n^{1,\delta},\delta})$  equals  $\pi_n^{1,\delta}$ , which recall was computed from  $n$ 's strategy under  $\tilde{b}^\delta$  in the third stage of his choice.

We claim that for each  $l \leq l_n^{0,\delta}$ ,  $q_n^{l,\delta}$  induces the equilibrium outcome against  $q_m^{0,\delta}$ . Indeed, since  $q_m^{0,\delta}$  belongs to  $Q^*$ , and the strategies in  $T_n^0(\delta)$  always implement some strategy in  $P_n^0$  regardless of the choice of outsider  $\sigma_{n,1}^0$ , all the strategies in  $T_n^0(\delta)$  are optimal against  $q_m^{0,\delta}$ . By quasi-perfection, this implies that every strategy in  $\bar{S}_n$  that appears at a level  $l \leq l_n^{0,\delta}$  of  $\bar{\Lambda}_n^\delta$  must be a best reply to  $q_m^{0,\delta}$ . Thus for all  $l \leq l_n^{0,\delta}$ ,  $q_n^{l,\delta}$  is a best reply to  $q_m^{0,\delta}$ . Since  $q_m^{0,\delta}$  is a lexicographic best reply to  $\Lambda_n^\delta$ ,  $(\tilde{q}_n(\varepsilon), q_m^{0,\delta})$  is an equilibrium of the game  $\Gamma$  for all small  $\varepsilon$ , where

$$\tilde{q}_n(\varepsilon) = (1 - \varepsilon) q_n^{0,\delta} + \varepsilon((1 - \varepsilon)q_n^{1,\delta} + \varepsilon^2((1 - \varepsilon)q_n^{2,\delta} + \varepsilon^3(\dots + \varepsilon^{l_n^{0,\delta}} q_n^{l_n^{0,\delta},\delta})).$$

By genericity of payoffs,  $\Gamma$  has finitely many equilibrium outcomes, so each of these equilibria induces the same equilibrium outcome—hence the claim follows. Three implications of this claim are: (i)  $l_n^{1,\delta} > l_n^{0,\delta}$ ; (ii) the enabling strategy  $r_n^{l_n^{1,\delta},\delta}$  in the previous paragraph belongs to  $P_n^0$ ; (iii) all levels up to  $l_n^{0,\delta}$  prescribe the same mixture at each information set on the equilibrium path and differ only at information sets excluded by (all of)  $m$ 's equilibrium strategies in  $\bar{Q}_m^*$ , which is the  $\rho_m$ -image of the component  $\bar{\Sigma}_m^*$  of equilibria.

**5.9. Limit of the Lexicographic Probability System.** Take a subsequence of the  $\delta$ 's such that the following properties hold for the associated LPSs  $\Lambda_n^\delta$  for each  $n$ : (i)  $l_n^{i,\delta}$  is independent of  $\delta$  for each  $i$ , call it  $l_n^i$ ; (ii) for each  $l \leq l_n^1$ , the face of  $P_n$  that contains  $q_n^{l,\delta}$  in its interior, as well as the strategies for  $m$  that are best replies to  $q_n^{l,\delta}$ , are independent of  $\delta$ .

Let  $\sigma_n^{l_n^1,\delta}$  be a sequence of strategies in  $\Sigma_n$  that is equivalent to the sequence  $q_n^{l_n^1,\delta}$ . Again using Blume, Brandenburger, and Dekel [1, Appendix Proposition 2], there is now an LPS  $(\sigma_n^{l_n^1,0}, \dots, \sigma_n^{l_n^1,k_n})$  and a sequence  $(\mu_0(\delta), \dots, \mu_{k_n-1}(\delta)) \in \mathbb{R}^{k_n}$  converging to zero such that for a subsequence of  $\delta$ 's,  $\sigma_n^{l_n^1,\delta}$  is expressible as the nested combination

$$\sigma_n^{l_n^1,\delta} = (1 - \mu_0(\delta))(\sigma_n^{l_n^1,0} + \mu_0(\delta)((1 - \mu_1(\delta))\sigma_n^{l_n^1,1} + \mu_1(\delta)((1 - \mu_2(\delta))\sigma_n^{l_n^1,2} + \dots + \mu_{k_n-1}(\delta)\sigma_n^{l_n^1,k_n}))$$

This LPS induces an equivalent LPS  $(q_n^{l_n^1,0}, \dots, q_n^{l_n^1,k_n})$  in enabling strategies.

Let  $k_n^1$  be the first level  $k$  of this LPS where  $q_n^{l_n^1,k}$  does not belong to  $P_n^0$ . Recall from the previous section that  $q_n^{l_n^1,\delta}$  is expressible as a convex combination  $\lambda_n^{l_n^1,\delta} p_n^{1,\delta} + (1 - \lambda_n^{l_n^1,\delta}) r_n^{l_n^1,\delta}$  and that  $\bar{\pi}_n^1(q_n^{l_n^1,\delta}) = \pi_n^{1,\delta}$ . Express  $r_n^{l_n^1,\delta}$  as a convex combination  $\alpha_n^{0,\delta} r_n^{0,\delta} + \alpha_n^{1,\delta} r_n^{1,\delta}$ , where for  $i = 0, 1$ ,  $r_n^{i,\delta} \in P_n^i$ . Then  $\lambda_n^{l_n^1,\delta} + (1 - \lambda_n^{l_n^1,\delta})\alpha_n^{1,\delta} > 0$  since  $q_n^{l_n^1,k}$  does not belong to  $P_n^0$ . Going to a subsequence, let  $\lambda_n^1$  be the limit of  $(\lambda_n^{l_n^1,\delta} + (1 - \lambda_n^{l_n^1,\delta})\alpha_n^{1,\delta})^{-1} \lambda_n^{l_n^1,\delta}$  and let  $r_n^{1,0}$  be the limit of  $r_n^{1,\delta}$ . Since the limit of  $p_n^{1,\delta}$  is  $p_n^{1,0}$ , we have that  $q_n^{l_n^1,k_n^1}$  is expressible as a convex combination  $\zeta_n(\lambda_n^1 p_n^{1,0} + (1 - \lambda_n^1) r_n^{1,0}) + (1 - \zeta_n) \check{p}_n^0$  for some  $\zeta_n > 0$  and  $\check{p}_n^0 \in P_n^0$ . Moreover, since  $\pi_n^{1,0}$  is the limit of  $\pi_n^{1,\delta}$ ,  $\bar{\pi}_n^1(q_n^{l_n^1,k_n^1}) = \pi_n^{1,0}$ .

For each  $\delta$  in the subsequence used above, define now an LPS  $\hat{\Lambda}_n^\delta = (\hat{q}_n^{0,\delta}, \dots, \hat{q}_n^{l_n^1+k_n^1+1,\delta})$  as follows:  $\hat{q}_n^{0,\delta} = q_n^{0,0}$ ,  $\hat{q}_n^{l,\delta} = q_n^{l-1,\delta}$  if  $0 < l \leq l_n^1$ , and  $\hat{q}_n^{l,\delta} = q_n^{l_n^1, l-l_n^1-1}$  otherwise. The strategy  $\hat{q}_n^{l,\delta}$  is independent of  $\delta$  for  $l = 0$  and  $l \geq l_n^1$ . Each level  $l < l_n^1 + k_n^1 + 1$  is a strategy in  $P_n^0$ .  $\hat{q}_n^{l_n^1+1,\delta}$  is a convex combination  $\lambda_n^{l_n^1,\delta} p_n^{0,\delta} + (1 - \lambda_n^{l_n^1,\delta}) r_n^{l_n^1,\delta}$  while  $\hat{q}_n^{l_n^1+k_n^1+1,\delta}$  is a convex combination  $\zeta_n(\lambda_n^1 p_n^{1,0} + (1 - \lambda_n^1) r_n^{1,0}) + (1 - \zeta_n) \check{p}_n^0$  where  $\zeta_n > 0$  and  $\bar{\pi}_n^1(\hat{q}_n^{l_n^1+k_n^1+1,\delta}) = \pi_n^{1,0}$ . The next lemma sets out the key properties of  $\hat{\Lambda}_n^\delta$  that lead to a conclusion of our proof.

**Lemma 5.3.** *For all small  $\delta$ , the LPS  $\hat{\Lambda}_n^\delta$  satisfies the following properties.*

- (1) *A strategy is at least as good a reply against  $\Lambda_n^\delta$  as another only if it is at least as good a reply against  $\hat{\Lambda}_n^\delta$ .*
- (2) *If  $\lambda_n^1 < 1$ , then the strategy  $r_n^{1,0}$  is a best reply to  $q_m^{0,0}$  and is at least as good a reply lexicographically against  $\Lambda_m^\delta$  as every strategy in  $S_n^1$ .*
- (3)  *$\lambda_n^1 > 0$  and every level  $l < l_n^1 + k_n^1 + 1$  induces the equilibrium outcome against  $q_m^{0,0}$ .*

- (4) The strategy  $\hat{q}_n^{l,\delta}$  for  $l < l_n^0$  and the strategy  $r_n^{l_n^0,\delta}$  if  $\lambda_n^{l_n^0,\delta} < 1$  are lexicographic best replies against  $\hat{q}_m^\delta$ .

*Proof of Lemma.* Suppose  $s_m$  is a better reply against  $\hat{\Lambda}_n^\delta$  than another strategy  $t_m$ . We show that  $s_m$  is also a better reply against  $\Lambda_n^\delta$ . Let  $l$  be the first level of  $\hat{\Lambda}_n^\delta$  such that  $s_m$  is a better reply against  $\hat{q}_m^{l,\delta}$  than  $t_m$ . If  $l = 0$  then for all small  $\delta$ ,  $s_m$  is a better reply against  $q_n^{0,\delta}$  since  $\bar{q}_n^{0,\delta}$  equals the limit  $q_n^{0,0}$  of  $q_n^{0,\delta}$ ; thus against  $\Lambda_n^\delta$ ,  $t_m$  is a worse reply against the very first level. If  $0 < l < l_n^1 + 1$  then obviously  $s_m$  is better reply against  $\Lambda_n^\delta$  than  $t_m$  since level  $l$  of  $\bar{\Lambda}_n^\delta$  corresponds to level  $l - 1$  of  $\Lambda_n^\delta$ . Suppose then that  $l_n^1 + 1 \leq l \leq l_n^1 + k_n^1 + 1$ . Then for all small  $\delta$ ,  $s_m$  is a better reply against  $q_n^{l,\delta}$  since  $q_n^{l,\delta}$  is a nested combination of  $(q_n^{l_n^1,0}, \dots, q_n^{l_n^1,k_n^1})$ . Thus,  $s_m$  is a better reply against  $\Lambda_n^\delta$ . This proves (1).

In the game  $\tilde{\Gamma}^\delta$  player  $n$ , when finally making a choice among the strategies in  $S_n^1 \cup T_n^1(\delta)$ , would choose a strategy  $s_n$  in  $S_n^1$  with positive probability only if it is at least as good a reply against  $\Lambda_n^\delta$  as the other strategies in  $S_n^1$ . Therefore, such a strategy would show up with positive probability under level  $l_n^1$  of  $\tilde{\Lambda}_n^\delta$  (and hence in  $\bar{\Lambda}_n^\delta$ ) only if this is the case. This implies that if  $(1 - \mu_n^{l_n^1,\delta})\alpha_n^1$  is positive, then the strategy  $r_n^{1,\delta}$  is at least as good a reply against  $\Lambda_n^\delta$  as the strategies in  $S_n^1$ . Recall that  $r_n^{1,0}$  is the limit of  $r_n^{1,\delta}$  and  $\lambda_n^1$  is the limit of  $(\mu_n^{l_n^1,\delta} + (1 - \mu_n^{l_n^1,\delta})\alpha_n^1)^{-1} \mu_n^{l_n^1,\delta}$ . Therefore, by point (1) of this lemma,  $r_n^{1,0}$  is at least as good a reply against  $\hat{\Lambda}_n^\delta$  as strategies in  $S_n^1$  if  $\lambda_n^1 < 1$ . To show that it is also a best reply against  $q_m^{0,0}$ , suppose to the contrary that it is not. Then every strategy in  $T_n^1(\delta)$ , regardless of outsider  $o_{n,1}^1$ 's choice, is a better reply against  $q_m^{0,\delta}$  for all small  $\delta$ , since for  $\delta = 0$  these strategies are in  $P_n^0$ , which are best replies to  $q_m^{0,0}$ . Therefore, quasi-perfection implies that  $n$  would prefer to play the strategies in  $T_n^1(\delta)$  rather than implementing  $r_n^{1,0}$  (or  $r_n^{1,\delta}$  when  $\delta$  is small), which shows that  $(1 - \mu_n^{l_n^1,\delta})\alpha_n^{l_n^1,\delta} = 0$  for all small  $\delta$  and hence that  $\lambda_n^1 = 1$ . This proves (2).

We turn now to (3). Every strategy  $\hat{q}_n^{l,\delta}$  for  $l < l_n^1 + k_n^1 + 1$  belongs to  $P_n^0$  and is thus optimal against  $q_m^{0,0}$ , which belongs to  $Q_m^*$ . The strategy  $\check{p}_n^0$  (which recall is part of the expression defining  $q_n^{l_n^1,k_n^1}$ ) which is chosen with positive probability is also in  $P_n^0$  and hence optimal against  $q_m^{0,0}$ . As we saw in the previous paragraph, if  $\lambda_n^1 < 1$ , the strategy  $r_n^1$  must also be optimal against  $q_m^{0,0}$ . Obviously  $q_m^{0,0}$  is optimal against  $\hat{\Lambda}_n^\delta$  since it is the limit of  $q_m^{0,\delta}$ , which by point (1) is optimal against  $\hat{\Lambda}_n^\delta$ . Therefore, for all small  $\varepsilon$ ,  $(q_m^{0,0}, q_n(\varepsilon))$  is an equilibrium

of  $\Gamma$ , where

$$\tilde{q}_n(\varepsilon) = \left( \sum_{l=0}^{l_n^1+k_n^1} \varepsilon^l + (1 - 1_{\lambda_n^1>0})\varepsilon^{l_n^1+k_n^1+1} \right)^{-1} \left( \sum_{l=0}^{l_n^1+k_n^1} \varepsilon^l \hat{q}_n^{l,\delta} + (1 - 1_{\lambda_n^1>0})\varepsilon^{l_n^1+k_n^1+1} \hat{q}_n^{l_n^1+k_n^1+1} \right)$$

where  $1_{\lambda_n^1>0}$  is the indicator function. This is impossible if  $\lambda_n^1 = 0$ , since  $\hat{q}_n^{l_n^1+k_n^1+1}$  is a convex combination of a strategy in  $P_n^0$  and one in  $P_n^1$ . Thus  $\lambda_n^1 > 0$ . Moreover, since this is a continuum of equilibria, genericity implies that all of them induce the same outcome. Therefore, all strategies at levels preceding  $l_n^1+k_n^1+1$  induce the equilibrium outcome against  $q_m^{0,0}$ .

Lastly we prove (4). By the previous paragraph, all strategies in  $P_n^0$  are optimal against  $\hat{q}_m^{k,\delta}$  for  $k \leq l_m^1+k_m^1$ . Thus the optimality of a strategy in  $P_n^0$  depends on how it fares against  $\hat{q}_m^{l_m^1+k_m^1+1}$ . If a strategy  $s_n \in S_n^0$  is not optimal against this strategy, then for all small  $\delta$ , some strategy in  $T_n^0(\delta)$  is a superior reply to  $\hat{q}_m^{l_m^1,k_m^1}$  regardless of what outsider  $o_{n,1}^0$ 's choice is. Thus, such a duplicate strategy is a better reply against  $\Lambda_m^\delta$  than  $s_n$ . Therefore, in the quasi-perfect equilibrium of  $\Gamma^\delta$ , player  $n$  when given a choice of actually implementing  $s_n$  will prefer this duplicate strategy in  $T_n^0(\delta)$ , i.e. the probability of  $s_n$  is zero for all  $l \leq l_n^0$  in  $\Lambda_n^\delta$ . Point (4) now follows.  $\square$

**5.10. Final Step of the Proof.** Fix a small  $\delta$  that satisfies the properties enumerated in Lemma 5.3 of the previous subsection. Let

$$\begin{aligned} \bar{q}_n^0(\varepsilon) &= \left( \sum_{l \leq l_n^0+1} \varepsilon^l \right)^{-1} \left( \sum_{l \leq l_n^0+1} \varepsilon^l \hat{q}_n^{l,\delta} \right) \\ \bar{q}_n^1(\varepsilon) &= \left( \sum_{l=l_n^0+2}^{l_n^1+k_n^1+1} \varepsilon^l \right)^{-1} \left( \sum_{l=l_n^0+2}^{l_n^1+k_n^1+1} \varepsilon^l \hat{q}_n^{l,\delta} \right). \end{aligned}$$

Observe that  $\bar{q}_n^0(\varepsilon)$  belongs to  $P_n^0$  for all  $\varepsilon$  and it is a convex combination of  $p_n^{0,\delta}$  and a subset  $R_n^0$  of strategies that are at least as good replies against  $\hat{\Lambda}_m^\delta$  as other strategies in  $S_n^0$ . Likewise  $\bar{q}_n^1(\varepsilon)$  is a convex combination of  $p_n^1$ ,  $r_n^1$  and a point in  $P_n^0$  such that the strategy  $r_n^1$  if it has a positive weight is at least as good a reply against  $\hat{\Lambda}_m^\delta$  as other strategies in  $P_n^1$ . There exists a small  $\varepsilon$  such that a strategy  $s_m$  is at least as good as a strategy  $t_m$  against  $\hat{\Lambda}_n^\delta$  iff it is lexicographically at least as good a reply against the LPS  $(q^0, \bar{q}_n^0(\varepsilon), \bar{q}_n^1(\varepsilon))$ . Since  $\bar{\pi}_n^1(\bar{q}_n^1(\varepsilon)) = \bar{\pi}_n^1(q_n^{1,0}) = \pi^{1,0}$  for all  $\varepsilon$ , it follows that  $(q^0, (p^{0,\delta}, p^{1,0}), \pi^{1,0})$  belongs to  $\mathcal{Q}$ . As we noted earlier, proving that this point belongs to  $\mathcal{Q}$  shows that in fact it belongs to  $U(x^*)$  and hence that  $q^0$  in  $U(q^{0,*})$ , which completes the proof.  $\square$



**Remark 5.4.** In the case where  $S_n^1$  is empty for exactly one player  $n$ , as we said initially in the description of  $\mathbb{P}$  and  $\mathcal{Q}$ , we do not have the factor  $P_n^1$  or  $\Pi_n^1$ . In the family of games  $\Gamma(\delta, \hat{p})$ , player  $n$  decides provisionally in the first stage on the strategy in  $S_n$  to play and in the second stage gets to execute it or switch to playing a strategy in  $T_n(\delta, \hat{p})$ . In the metagame, we do not have outsiders  $o_{n,j}^1$  for  $j = 1, 2, 3$ . The rest of the analysis is essentially the same modulo these provisions.

## 6. CONCLUDING REMARKS

Like our previous paper [10] on forward induction, the characterization in this paper is a step toward a theory of equilibrium refinement using axioms adapted from decision theory. Theorem 5.1 is confined to games in extensive form with perfect recall, two players, and generic payoffs, but it suggests that an extension to more general games might be possible. We hope also that Axiom B can be weakened to require only that a solution contains a sequential equilibrium for which each player’s strategy is conditionally admissible from each of his information sets—thus keeping the axioms entirely within standard decision theory.

Previously, some proposed refinements selected equilibria with one or more desirable properties, like admissibility, subgame perfection, or sequential rationality. Other proposed refinements derived some properties from limits of equilibria of perturbed games, such as perfect, quasi-perfect, and proper equilibria. In the latter approach, a key step forward was Kohlberg and Mertens’ [14] argument that an axiomatic development requires set-valued refinements. Their program achieved remarkable success with Mertens’ [19] definition of a stable set, which has the desirable properties 1, 2, 3 listed in Section 1 and others too, such as ordinality and immunity to splitting players into agents.

However, an axiomatic theory of refinement should be based on basic principles of rational behavior in the game at hand, as in decision theory. This precludes reliance on perturbed games obtained by perturbing players’ sets of feasible strategies. The challenge, therefore, has been to establish why consideration of perturbed games yields the requisite decision-theoretic properties.

Our answer here begins with Axiom S, which generalizes the invariance criterion of Kohlberg and Mertens’ [14] and the small worlds criterion of Mertens’ [21], as explained in Section 3.3. Absent a strong invariance property like Axiom S, a refinement is vulnerable to ‘framing effects’ depending on wider contexts in which the given game might be embedded. In decision theory, such effects were examined by Savage [28], and in cognitive psychology they play a prominent role in interpreting decisions by subjects in experiments, as for instance in Kahneman and Tversky [13]. For a theory of thoroughly rational behavior, however, an

axiom should exclude framing effects. Axiom S does this by requiring a solution of a game to be consistent with the solution of any metagame in which it is embedded. As shown in Proposition 3.4, it is already true of any equilibrium that it is consistent with an equilibrium of any metagame in which the game is embedded. Axiom S merely extends to refinements this fundamental invariance property of equilibria.

Our answer continues with the proof of Theorem 4.1 in Appendix B. There it is shown that a set  $Q^*$  of equilibria in enabling strategies is stable iff the corresponding projection map from the pseudomanifold  $\mathcal{Q}$  to the space  $\mathbb{P}$  of enabling strategies is essential. Using this key property, the proof of Theorem 5.1 shows that for each equilibrium in a component of undominated equilibria there exists a corresponding metagame for which the equilibrium is the image of a quasi-perfect equilibrium in the metagame if and only if the projection map is essential. Hence Axioms A, B, S imply that a solution is a stable set, and conversely due to Mertens' previous proofs.

The answer to the 'why' question above is thus that, given Axioms A and B, stability with respect to perturbed games is equivalent to an analogous 'stability' with respect to embeddings in metagames, as required by Axiom S. Because in practice every game is embedded in some wider context, we view Axiom S's requirement that a refinement is immune to presentation effects as the relevant criterion from the perspective of decision theory. This view is reinforced by the facts that Nash equilibria satisfy Axiom S, and that together with Axioms A and B, the implied refinement agrees with stability based on perturbed games.

For a refinement satisfying the axioms, Theorem 5.1 establishes that a solution of a game must be a component of its undominated equilibria, and that the component must be essential. Because payoffs are assumed to be generic, all equilibria in the component have the same paths of equilibrium play and thus the same distribution of outcomes. Therefore the main implication for equilibrium refinements is that a predicted outcome distribution should result from equilibria in an essential component of undominated equilibria, and in particular, from the sequential equilibria it necessarily contains.

A secondary implication is that after deviations from equilibrium play, the continuations of all equilibria in the component remain admissible and sequentially rational, where those that are not sequential equilibria of the original game are justified by beliefs induced by quasi-perfect equilibria of corresponding metagames that embed the given game. This provides one resolution of the conundrum posed by Reny [25, 26, 27].

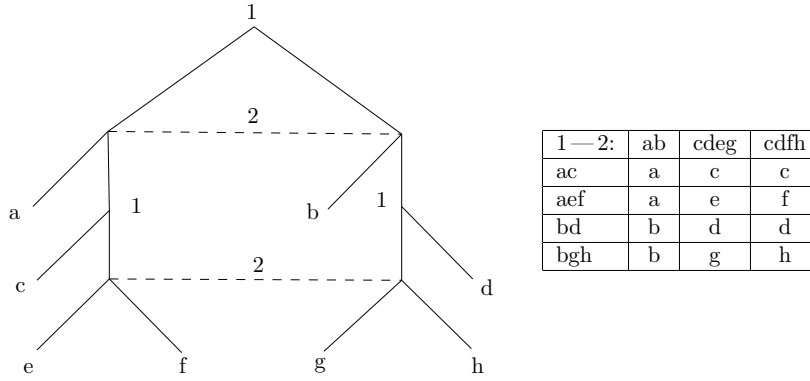


FIGURE 1. A game tree and its pure reduced normal form in which each pure strategy is identified by the terminal nodes it does not exclude.

APPENDIX A. ENABLING STRATEGIES

In the normal-form representation of a game in extensive form, a player’s pure strategy specifies the action chosen at each of his information sets in the game tree. However, outcomes are not affected by a strategy’s actions at information sets excluded by his previous actions. One therefore considers equivalence classes of pure strategies. Say that two pure strategies are outcome equivalent if the sets of terminal nodes they do not exclude are the same. For instance, the game in Figure 1 is shown on the left side in extensive form and on the right side in the ‘pure reduced normal form’ (PRNF) introduced by Mailath, Samuelson, and Swinkels [18]. In the PRNF, each outcome-equivalent class of player 1’s pure strategies is identified by the terminal nodes it does not exclude, as indicated by labels of rows along the left side; and each equivalence class of player 2’s pure strategies is identified by the terminal nodes it does not exclude, as indicated by labels of columns along the top. Because this game has no moves by Nature, each row and column determine a unique outcome that is the intersection of the row and column labels, shown as the corresponding entry in the matrix.

A similar example is shown in Figure 2 for the game tree of a signaling game. In this case, each profile of pure strategies determines a pair of outcomes such that the first or second outcome occurs depending on whether Nature’s initial move is up or down. For instance, the outcome of 1’s strategy *abcd* and 2’s strategy *aceg* is *a* with probability *p* and *c* with probability  $1 - p$ .

Say that a terminal node that is not excluded is an enabled outcome. A pure strategy of a player enables outcome *z* if it chooses all his actions on the path to *z*. A player’s mixed strategy randomizes over his pure strategies, whereas a behavioral strategy randomizes over

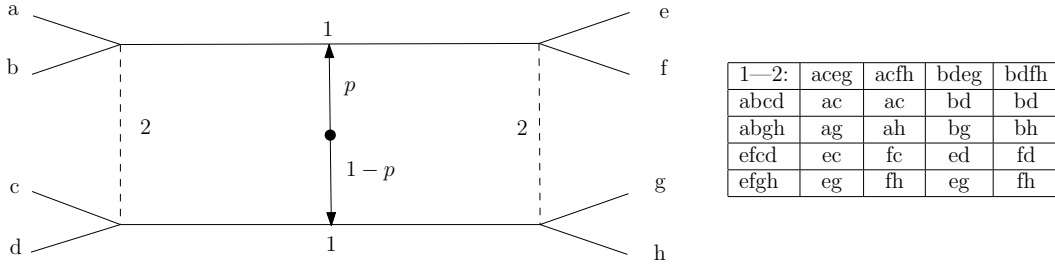


FIGURE 2. The game tree of a signaling game and its pure reduced normal form in which each pure strategy is identified by the terminal nodes it does not exclude.

actions at each of his information sets. A strategy of either kind induces a probability distribution over outcome-equivalent classes of his pure strategies, and thus a distribution over enabled outcomes. Such a distribution is called an *enabling strategy*. A point  $p_n \in [0, 1]^Z$  is an enabling strategy for player  $n$  if it is the distribution over enabled outcomes induced by some mixed or behavioral strategy, i.e.  $p_n(z)$  is the mixed strategy's probability of those pure strategies that enable outcome  $z$ . The vertices of the polyhedron  $P_n$  of  $n$ 's enabling strategies correspond to outcome-equivalent classes of  $n$ 's pure strategies in the PRNF, as in Figures 1 and 2. Enabling strategies are minimal representations of strategic behavior in games with perfect recall.<sup>9</sup>

Let  $p_*(z)$  be the probability that Nature's strategy enables outcome  $z$ , which is 1 if Nature has no moves. Then for each profile  $p \in P = \prod_n P_n$  of players' enabling strategies, the probability that outcome  $z$  results is  $\gamma_z(p) = p_*(z) \prod_n p_n(z)$ , because Nature and the players randomize independently. The extensive form is therefore summarized by the multilinear function  $\gamma : P \rightarrow \Delta(Z) \subset \mathbb{R}^Z$  that assigns to each profile of players' enabling strategies a distribution over terminal nodes, including the effect of Nature's enabling strategy. Player  $n$ 's expected payoff is  $\mathcal{G}_n(p) = \sum_z \gamma_z(p) u_n(z)$ . The game  $\Gamma$  is therefore summarized by the multilinear function  $\mathcal{G} : P \rightarrow \mathbb{R}^N$  that assigns to each profile of players' enabling strategies their expected payoffs. This summary specification is called the *enabling form* of the game.

## APPENDIX B. PROOFS OF PROPOSITIONS

**B.1. Proof of Proposition 3.3.** A multilinear map  $f_n : \tilde{\Sigma}_n \times \Sigma_o \rightarrow \Sigma_n$  is completely specified by its values at profiles of pure strategies. We use  $\hat{f}_n$  to denote the restriction of  $f_n$  to the set  $\tilde{S}_n \times S_o$  of profiles of pure strategies.

<sup>9</sup>Mertens [21, p. 554] introduces the mapping of mixed strategies to induced distributions on terminal nodes. Koller and Megiddo [15] call them realization plans.

**Proposition B.1.**  $\tilde{G}$  embeds  $G$  via a collection of multilinear maps  $f = (f_n)_{n \in N}$  if and only if for each player  $n$  there exists  $\tilde{T}_n \subseteq \tilde{S}_n$  and a bijection  $\pi_n : \tilde{T}_n \rightarrow S_n$  such that for each  $(\tilde{s}, s_o) \in \tilde{S} \times S_o$  and  $\tilde{t}_n \in \tilde{T}_n$ :

- (1)  $\hat{f}_n(\tilde{t}_n, s_o) = \pi_n(\tilde{t}_n)$ ,
- (2)  $\tilde{G}_n(\tilde{s}, s_o) = G_n(\hat{f}(\tilde{s}, s_o))$ , where  $\hat{f} = (\hat{f}_n)_{n \in N}$ .

*Proof.* Suppose we have a game  $\tilde{G} : \tilde{\Sigma} \times \Sigma_o \rightarrow \mathbb{R}^{N \cup o}$  and a collection of multilinear maps  $f_n : \tilde{\Sigma}_n \times \Sigma_o \rightarrow \Sigma_n$ , one for each  $n \in N$ , such that conditions (1) and (2) of the proposition are satisfied. Then, by condition (1) and multilinearity of  $f_n$  for each  $n$ , for each fixed  $\sigma_o$ ,  $f_n(\cdot, \sigma_o)$  is surjective because it maps the face spanned by  $\tilde{T}_n$  homeomorphically onto  $\Sigma_n$ . Also, condition (2) and multilinearity of each  $f_n$  imply that  $\tilde{G} = G \circ f$ . According to Definition 3.2, therefore,  $(\tilde{G}, f)$  embeds  $G$ .

Now suppose that  $(\tilde{G}, f)$  embeds  $G$ . Let  $\sigma_o$  be a profile of completely mixed strategies for outsiders. Because  $f_n$  is multilinear it induces a linear mapping  $f_n(\cdot, \sigma_o)$  from  $\tilde{\Sigma}_n$  to  $\Sigma_n$  that, by the definition of an embedding, is surjective. Hence, for each  $s_n \in S_n$  there exists a pure strategy  $\tilde{t}_n(s_n)$  in  $\tilde{S}_n$  that is mapped to  $s_n$  by this linear map. We claim that  $f_n(\tilde{t}_n(s_n), s_o) = s_n$  for all  $s_o \in S_o$ . Indeed, observe that  $f_n(\tilde{t}_n(s_n), \sigma_o) = \sum_{s_o} f_n(\tilde{t}_n(s_n), s_o) \sigma_o(s_o)$ , where for each  $s_o$ ,  $\sigma_o(s_o)$  is the probability of  $s_o$  under  $\sigma_o$ . Therefore, since  $\sigma_o$  is completely mixed, if  $f_n(\tilde{t}_n(s_n), s_o) \neq s_n$  for some  $s_o$  then  $f_n(\tilde{t}_n(s_n), \sigma_o)$ , which is an average of values at vertices of  $S_o$ , cannot be  $s_n$ . Thus,  $f_n(\tilde{t}_n(s_n), s_o) = s_n$  for all  $s_o$ . Let  $\tilde{T}_n \subset \tilde{S}_n$  be a collection comprising a different pure strategy  $\tilde{t}_n(s_n)$  for each  $s_n \in S_n$  and let  $\pi_n$  be the associated bijection. Define  $\hat{f}_n : \tilde{S}_n \times S_o \rightarrow \Sigma_n$  by  $\hat{f}_n(\tilde{s}_n, s_o) = f_n(\tilde{s}_n, s_o)$ . Then conditions (1) and (2) of the proposition are satisfied.  $\square$

## B.2. Proof of Proposition 3.4.

**Proposition B.2.** If  $(\tilde{G}, f)$  embeds  $G$  then the equilibria of  $G$  are the  $f$ -images of the equilibria of  $\tilde{G}$ .

*Proof.* Suppose  $(\tilde{\sigma}, \sigma_o)$  is an equilibrium of  $\tilde{G}$  and let  $\sigma = f(\tilde{\sigma}, \sigma_o)$ . For any insider  $n$  and his strategy  $\tau_n \in \Sigma_n$  there exists  $\tilde{\tau}_n \in \tilde{\Sigma}_n$  such that  $f_n(\tilde{\tau}_n, \sigma_o) = \tau_n$  because  $f_n(\cdot, \sigma_o)$  is surjective by condition (a) of Definition 3.2 an embedding. Using condition (b),

$$G_n(\tau_n, \sigma_{-n}) = G_n(f(\tilde{\tau}_n, \tilde{\sigma}_{-n}, \sigma_o)) = \tilde{G}_n(\tilde{\tau}_n, \tilde{\sigma}_{-n}, \sigma_o) \leq \tilde{G}_n(\tilde{\sigma}, \sigma_o) = G_n(f(\tilde{\sigma}, \sigma_o)) = G_n(\sigma),$$

where the inequality obtains because  $(\tilde{\sigma}, \sigma_o)$  is an equilibrium of  $\tilde{G}$ . Hence  $\sigma$  is an equilibrium of  $G$ .

Conversely, suppose  $\sigma$  is an equilibrium of  $G$ . For each  $n$ , let  $\pi_n$  be the bijection given by Proposition 3.3. Let  $\tilde{\sigma}_n$  be the strategy for insider  $n$  in  $\tilde{G}$  defined by  $\tilde{\sigma}_n(\tilde{t}_n) = \sigma_n(\pi_n(\tilde{t}_n))$  for  $\tilde{t}_n \in \tilde{T}_n$  and  $\tilde{\sigma}_n(\tilde{s}_n) = 0$  for  $\tilde{s}_n \notin \tilde{T}_n$ . Since  $f_n$  is multilinear, by condition (1) of Proposition 3.3,  $f_n(\tilde{\sigma}_n, \cdot) = \sigma_n$  and thus  $f(\tilde{\sigma}, \cdot) = \sigma$ . Hence, it suffices to show that there exists a strategy profile  $\sigma_o$  for outsiders such that  $(\tilde{\sigma}, \sigma_o)$  is an equilibrium of  $\tilde{G}$ . By fixing the profile of insiders' strategies to be  $\tilde{\sigma}$  one induces a game among outsiders. Let  $\sigma_o$  be an equilibrium of this induced game among outsiders. To see that  $(\tilde{\sigma}, \sigma_o)$  is an equilibrium of  $\tilde{G}$ , observe that for each pure strategy  $\tilde{s}_n$  of an insider  $n$ :

$$\tilde{G}_n(\tilde{s}_n, \tilde{\sigma}_{-n}, \sigma_o) = G_n(f_n(\tilde{s}_n, \sigma_o), \sigma_{-n}) \leq G_n(\sigma) = G_n(f(\tilde{\sigma}, \sigma_o)) = \tilde{G}_n(\tilde{\sigma}, \sigma_o),$$

where the first and second equalities use the property  $f(\tilde{\sigma}, \cdot) = \sigma$  established above, and the inequality obtains because  $\sigma$  is an equilibrium of  $G$ .  $\square$

#### APPENDIX C. PROOF OF THEOREM 4.1

**Theorem C.1.**  *$(\mathcal{Q}, \partial\mathcal{Q})$  is a pseudomanifold of the same dimension as  $(\mathbb{P}, \partial\mathbb{P})$ . Moreover,  $\psi : (\mathcal{Q}, \partial\mathcal{Q}) \rightarrow (\mathbb{P}, \partial\mathbb{P})$  is an essential map iff  $Q^*$  is stable.*

*Proof.* The proof invokes genericity of payoffs by assuming that certain points and polyhedra, identified as they arise during the proof, are in general position. See Appendix D for elaboration of the character of these genericity requirements.

For any set  $X$ , we write  $d(X)$  for its dimension. For any subset  $T_n$  of  $S_n$ , let  $P_n(T_n)$  be the convex hull (in  $P_n$ ) of the strategies in  $T_n$ . For simplicity, we write  $d(T_n)$  for  $d(P_n(T_n))$ . See Sections 4.4 and 4.5 for additional notation used below.

For any vertex  $\pi_n^1$  of  $\Pi_n^1$ , let  $S_n^1(\pi_n^1)$  be the set of pure strategies that map to  $\pi_n^1$ . For any face  $\Psi_n^1$  of  $\Pi_n^1$ , let  $S_n^1(\Psi_n^1)$  be the set of strategies  $s_n^1$  such that  $\bar{\pi}_n^1(s_n^1) \in \Psi_n^1$ .

Let  $H_n^0$  be the set of information sets  $h_n \in H_n \setminus H_n^*$  of player  $n$  such that at the last information set  $h'_n \in H_n^*$  that precedes  $h_n$ , the action there leading to  $h_n$  belongs to  $A_n^*$ , which is the set of his equilibrium actions. If a subset  $T_n^0$  of  $S_n^0$  is such that  $P_n(T_n^0)$  contains an equilibrium in  $\bar{Q}_n^*$ , then for each first information set  $h_n \in H_n^0$  there exists a strategy in  $T_n^0$  that enables  $h_n$ . Let  $H_n^1 = H_n \setminus (H_n^* \cup H_n^0)$ .

**Lemma C.2.** *Suppose  $T_n^0$  is a subset of strategies in  $S_n^0$  such that  $P_n(T_n^0)$  contains an equilibrium  $q_n^*$  in  $\bar{Q}_n^*$ . If the strategies in  $T_n^0$  are at least as good replies against  $p_m^0 \in P_m^0$  as other strategies in  $S_n^0$ , then all the strategies in  $S_n^0$  are equally good replies against  $p_m^0$ .*

*Proof.* Let  $\Pi_n^*$  be the projection of  $P_n^0$  to  $\mathbb{R}^{Z^*}$ , where  $Z^*$  is the set of terminal nodes reached with positive probability under the equilibria in  $\bar{Q}_n^*$ . Then  $Z^* = Z_n^0 \cap Z_m^0$ . Consequently, the

payoff to a strategy  $s_n^0$  against  $p_m^0$  depends on  $s_n^0$  only through its projection to  $\Pi_n^*$ . Since the projection of  $q_n^*$ —which is at least as good a reply as the other strategies in  $S_n^0$  against  $p_m^0$ —belongs to the relative interior of  $\Pi_n^*$ , every strategy in  $S_n^0$  is a best reply against  $p_m^0$ .  $\square$

**Lemma C.3.** *Suppose that  $T_n^0$  is a subset of  $S_n^0$  such that  $P_n(T_n^0)$  is a face of  $P_n^0$  containing an equilibrium strategy in  $\bar{Q}_n^*$ . For each  $t_n \notin T_n^0$  there exists an information set  $h_n \in H_n^0$  that  $t_n$  enables and where the action chosen by it is avoided by all  $t_n^0 \in T_n^0$  that enable  $h_n$ .*

*Proof.* Since  $P_n(T_n^0)$  is a face of  $P_n^0$ , which is itself a face of  $P_n$ , there exists a linear function  $f : \mathbb{R}^Z \rightarrow \mathbb{R}$  that is zero on  $P_n(T_n^0)$  and negative everywhere else on  $P_n$ . Fix  $p_m$  in the interior of  $P_m$  and define a payoff function  $\tilde{u}_n$  for  $n$  by the equation:  $p_0(z)p_m(z)\tilde{u}_n(z) = f(e_z)$  where  $e_z$  is the  $z$ -th unit vector in  $\mathbb{R}^Z$ . Then when  $n$ 's payoff function is  $\tilde{u}_n$ , the strategies in  $P_n(T_n^0)$  are the best replies against  $p_m$ . Take  $t_n \notin T_n^0$ . Since it is suboptimal against  $p_m$ , there exists an information set  $h_n$  where the action  $a$  chosen by  $t_n$  is suboptimal. Then  $h_n$  does not belong to  $H_n^*$ : indeed, since  $t_n$  belongs to  $S_n^0$ ,  $a$  is an equilibrium action if  $h_n \in H_n^*$ ; and since  $P_n(T_n^0)$  contains an equilibrium strategy, there exists a strategy in  $T_n^0$  that enables such an  $h_n$  and chooses  $a$ . Let  $h'_n$  be the last information set preceding  $h_n$  that belongs to  $H_n^*$ . Then obviously the action chosen by  $t_n$  at  $h'_n$  is an equilibrium action and thus  $h_n$  belongs to  $H_n^0$ . Any strategy in  $T_n^0$  that enables  $h_n$  avoids  $a$ , which proves the lemma.  $\square$

For the next lemma, it is worth recapitulating the exact definition of the set  $\Pi_n^1$ . Recall from Section 4.5 that we fix a completely mixed enabling strategy  $\bar{p}_m$  for player  $m$  and compute for each  $p_n$  the total probability  $\eta(p_n)$  of reaching a terminal node in  $Z_n^1$  under  $(\bar{p}_m, p_n)$ .  $\mathcal{H}_n$  is a hyperplane in  $\mathbb{R}^{Z_n^1}$  that separates the projection  $\psi_{Z_n^1}(P_n^1)$  of  $P_n^1$  to  $\mathbb{R}^{Z_n^1}$  from the origin of  $\mathbb{R}^{Z_n^1}$  and that has  $(p_0(z)\bar{p}_m(z))_{z \in Z_n^1}$  as its normal and some  $\varepsilon > 0$  as its constant. The function  $\bar{\pi}_n^1$  maps each point  $p_n \in P_n \setminus P_n^0$  to  $\varepsilon(\eta(p_n))^{-1}\psi_{Z_n^1}(p_n) \in \mathcal{H}_n$ .

**Lemma C.4.** *For a strategy  $s_n$  of  $S_n^1$ ,  $\bar{\pi}_n^1(s_n)$  is a vertex of  $\Pi_n^1$  iff there exists a unique information set  $h_n \in H_n^*$  with the property that  $s_n$  enables  $h_n$  and chooses a non-equilibrium action there.*

*Proof.* Let  $s_n \in S_n^1$  be a pure strategy satisfying the condition of the lemma. We prove by contradiction that  $\bar{\pi}_n^1(s_n)$  is a vertex of  $\Pi_n^1$ . Therefore suppose to the contrary that  $\bar{\pi}_n^1(s_n)$ , which equals  $\varepsilon(\eta(p_n))^{-1}\psi_{Z_n^1}(p_n)$ , is not a vertex of  $\Pi_n^1$ . We can express  $\bar{\pi}_n^1(s_n)$  as a unique convex combination  $\sum_{j=1}^J \lambda^j \pi_n^{1,j}$  where  $J > 1$  and for each  $1 \leq j \leq J$ ,  $\pi_n^{1,j}$  is a vertex of  $\Pi_n^1$ . For each  $j$ , since  $\pi_n^{1,j}$  is a vertex of  $\Pi_n^1$ , we can express  $\pi_n^{1,j}$  as  $\varepsilon(\eta(s_n^{1,j}))^{-1}\psi_{Z_n^1}(s_n^{1,j})$  for some  $s_n^{1,j}$  in  $S_n^1$ . Since  $\lambda^j > 0$ ,  $s_n^{1,j}$  cannot choose a non-equilibrium action at any  $h'_n \neq h_n$

in  $H_n^*$  that it enables; since it belongs to  $S_n^{1,j}$  it must therefore enable  $h_n$ ; and it cannot choose a different non-equilibrium action from  $s_n$  at  $h_n$ . Observe now that the probability  $\eta(p_n)$  and  $\eta(s_n^{1,j})$  for all  $j$  are equal and exactly the probability that Nature and the strategy  $\bar{p}_m$  do not exclude  $h_n$ . Therefore,  $\psi_{Z_n^1}(p_n) = \sum_j \lambda^j \psi_{Z_n^1}(s_n^{1,j})$ . Modify  $s_n^{1,j}$  to a strategy  $t_n^{1,j}$  so that at every information set other than the successors to  $h_n$ ,  $t_n^{1,j}$  agrees with  $s_n$ , and at the successors to  $h_n$  it agrees with  $s_n^{1,j}$ . It is now clear that when viewed as enabling strategies,  $s_n = \sum_j \lambda^j t_n^{1,j}$  and thus  $s_n$  is a convex combination of the strategies  $t_n^{1,j}$ . But that is a contradiction since  $s_n$  is a vertex of  $P_n$  and all the strategies  $t_n^{1,j}$  are different from one another and from  $s_n$  as they induce the points  $\pi_n^{1,j}$  that are different from one another and from  $\bar{\pi}_n^1(s_n)$  in  $\mathcal{H}_n$ .

To prove the other way around, suppose  $s_n$  is a strategy that, at a collection  $h_n^k$  for  $k = 1, \dots, K$  of at least two information sets in  $H_n^*$ , chooses a non-equilibrium action. For each  $k$ , choose a strategy  $s_n^{1,k}$  in  $S_n^1$  that enables  $h_n^k$ , agrees with  $s_n$  there and at all its successors, but at other  $h_n \in H_n^*$ , chooses an equilibrium action. Then  $\psi_{Z_n^1}(s_n) = \sum_k \psi_{Z_n^1}(s_n^{1,k})$ . Therefore,  $\bar{\pi}_n^1(s_n)$  cannot be a vertex of  $\Pi_n^1$ .  $\square$

For each  $T_n^0 \subseteq S_n^0$  such that  $P_n(T_n^0)$  is a face of  $P_n^0$  and contains an equilibrium strategy for  $n$ , let  $S_n^1(T_n^0)$  be the subset of  $S_n^1$  consisting of strategies  $s_n^1$  such that for each first information set  $h_n^0$  in  $H_n^0$  that  $s_n^1$  does not exclude,  $s_n^1$  agrees with some  $t_n^0 \in T_n^0$  at  $h_n^0$  and all its successors, where  $t_n^0$  has the property that it does not exclude  $h_n^0$ . For a face  $\Psi_n^1$  of  $\Pi_n^1$ , let  $S_n^1(T_n^0; \Psi_n^1) = S_n^1(\Psi_n^1) \cap S_n^1(T_n^0)$  and  $T_n \equiv T_n^0 \cup S_n^1(T_n^0; \Psi_n^1)$ . For notational simplicity, we refer to  $S_n^1(T_n^0; \Psi_n^1)$  as  $T_n^1$ . The following lemma provides an important feature of the set  $T_n$ .

**Lemma C.5.** *The strategies in  $T_n$  are the vertices of a face of  $P_n$  whose dimension is  $d(T_n) \equiv d(T_n^0) + d(\Psi_n^1) + 1$ .*

*Proof.* Let  $\hat{T}_n^1$  be the set of strategies  $t_n^1$  in  $T_n^1$  such that  $\bar{\pi}_n^1(t_n^1)$  is a vertex of  $\Psi_n^1$ . We will first show that every  $t_n \in T_n \setminus (T_n^0 \cup \hat{T}_n^1)$  is affinely dependent on the strategies in  $T_n^0 \cup \hat{T}_n^1$ . Let  $t_n^1 \in T_n \setminus (T_n^0 \cup \hat{T}_n^1)$ . By Lemma C.4, there exist information sets  $h_n^1, \dots, h_n^K$ ,  $K > 1$ , in  $H_n^*$  such that for each  $k$ ,  $t_n^1$  chooses a non-equilibrium action  $a^k$  at  $h_n^k$ , and at each other  $h_n \in H_n^*$  it chooses an equilibrium action. Fix  $t_n^0 \in T_n^0$  that agrees with  $t_n^1$  everywhere except at the information sets  $h_n^k$  for each  $k$ . For each  $k$  let  $t_n^{1,k}$  be the strategy that agrees with  $t_n^1$  at  $h_n^k$  and its successors, but everywhere else agrees with  $t_n^0$ . Each  $t_n^{1,k}$  belongs to  $\hat{T}_n^1$  and also,  $t_n^1 = \sum_k t_n^{1,k} - (K-1)t_n^0$ . Thus,  $t_n^1$  is an affine combination of the strategies in  $T_n^0 \cup \hat{T}_n^1$ .



For each  $j = 0, \dots, d(\Psi_n^1)$ , pick a strategy  $t_n^{1,j} \in S_n^1(T_n^0)$  such that  $\bar{\pi}_n^1(t_n^{1,j})$  is a vertex of  $\Psi_n^1$ . Let  $\tilde{T}_n^1$  be the collection of these strategies. We show that strategies in  $\hat{T}_n^1 \setminus \tilde{T}_n^1$  are now affinely dependent on the strategies in  $T_n^0 \cup \tilde{T}_n^1$ . Fix  $t_n^1 \in \hat{T}_n^1 \setminus \tilde{T}_n^1$ . By Lemma C.4 there exists a unique information set  $h_n \in H_n^*$  enabled by  $t_n^1$  and where it chooses a non-equilibrium action. By construction of  $\tilde{T}_n^1$ , there exists a subset  $(t_n^{1,j})_{j=1}^J$  of  $\tilde{T}_n^1$  such that  $\bar{\pi}_n^1(t_n^1)$  is expressible as an affine combination  $\sum_j \lambda^j \bar{\pi}_n^1(t_n^{1,j})$  with  $\lambda^j \neq 0$  for all  $j$ . For each  $j$ ,  $s_n^{1,j}$  enables  $h_n$  and chooses  $a$  at  $h_n$ ; at all other information sets in  $H_n^*$  it chooses an equilibrium action. Let  $t_n^0$  be a strategy that agrees with  $t_n^1$  everywhere except at  $h_n$  and its successors. For each  $j$ , let  $t_n^{0,j}$  be a strategy that agrees with  $t_n^{1,j}$  everywhere except at  $h_n$  and its successors, where it agrees with  $t_n^0$ . Then  $t_n^1 = t_n^0 + \sum_j \lambda^j (t_n^{1,j} - t_n^{0,j})$  and is affinely dependent on the strategies in  $T_n^0 \cup \tilde{T}_n^1$ .

It follows now from the above arguments that the affine space  $A$  spanned by  $T_n^0 \cup \tilde{T}_n^1$  contains  $P_n(T_n)$  and that the dimension of  $P_n(T_n)$  is as stated. To finish the proof of the lemma, we show that  $P_n(T_n)$  is a face of  $P_n$ . Let  $Q_n$  be the smallest face of  $P_n$  that contains  $P_n(T_n)$ . Suppose  $Q_n \neq P_n(T_n)$ . Then there exists a point  $p_n$  in the relative interior of  $P_n(T_n)$  and  $Q_n$ . Therefore  $p_n$  can be expressed as a convex combination of the vertices of  $Q_n$  in two different ways: (a)  $\sum_i \lambda^i t_n^{0,i} + \sum_j \lambda^j t_n^{1,j}$ , where the  $t_n^{0,i}$ 's are in  $T_n^0$  and the  $t_n^{1,j}$ 's are in  $T_n^1$ ; (b)  $\sum_i \mu^i t_n^{0,i} + \sum_j \mu^j t_n^{1,j} + \sum_k \mu^k t_n^{2,k}$ , where now the  $t_n^{2,k}$ 's are the vertices of  $Q_n$  that are not in  $T_n$ . Consider one of the  $t_n^{2,k}$ 's. If it belongs to  $S_n^0 \setminus T_n^0$  then by Lemma C.3 there is an information set  $h_n$  in  $H_n^0$  that is enabled by  $t_n^{2,k}$  where the action chosen by  $h_n$  is avoided all strategies in  $T_n^0$  that enable it: that implies under the expression in (b) that the nodes following this action are assigned a positive probability, but not under (a), which is impossible. If  $t_n^{2,k}$  belongs to  $S_n^1 \setminus T_n^1$  then it must belong to  $S_n^1(\Psi_n^1)$  since otherwise under (b)  $\bar{\pi}_n^1(p_n) \notin \Psi_n^1$ . Since  $s_n \notin T_n^1$  there exists an information set  $h_n \in H_n^0$  enabled by  $t_n^{2,k}$  where the continuation strategy of  $t_n^{2,k}$  coincides with that of some  $t_n \in S_n^0 \setminus T_n^0$  but not for any  $s_n \in T_n^0$ . As in the previous case, this too is impossible.  $\square$

One corollary of the above result obtains when we take  $T_n^1$  to be  $S_n^1$  and  $\Psi_n^1$  to be  $\Pi_n^1$ . The dimension of  $P_n$  is  $d(P_n^0) + d(\Pi_n^1) + 1$ . Observe that  $P_n^1$  is a face of  $P_n$  iff  $\psi_{Z_n^1}(P_n^1)$  is a face of  $\psi_{Z_n^1}(P_n)$ , which is equivalent to saying that  $\Pi_n^1$  is homeomorphic to  $\psi_{Z_n^1}(P_n^1)$ . Thus, the dimension of  $P_n^1$  equals  $d(\Pi_n^1)$  if  $P_n^1$  is a face of  $P_n$  and otherwise it equals the dimension of  $P_n$ .

Fix  $T \equiv ((T_1^0, \Psi_1^1), (T_2^0, \Psi_2^1))$ , where for each  $n$ ,  $P_n(T_n) \cap Q_n^*$  nonempty and  $\Psi_n^1$  is a (possibly empty) face of  $\Pi_n^1$ . Let  $A_n(T)$  be the set of points in  $P_n(T_n^0)$  such that the strategies in  $T_m$

are all best replies. Let  $\tilde{T}_n^0$  be the unique subset of  $T_n^0$  such that the interior of  $A_n(T)$  is contained in the interior of  $P_n(\tilde{T}_n^0)$ . Let  $\tilde{d}_n^*(T)$  be the dimension of  $A_n(T)$ .

**Lemma C.6.** *If  $T_m^1$  is nonempty then  $A_n(T)$  is a proper face of  $\bar{Q}_n^* \cap P_n(\tilde{T}_n^0)$ .*

*Proof.* Since  $Q_n^* \cap P_n(T_n)$  is nonempty, the strategies in  $T_n^0$  are undominated. Hence,  $P_n(\tilde{T}_n^0) \cap \bar{Q}_n^* = P_n(\tilde{T}_n^0) \cap Q_n^* \equiv A_n^*(T)$ .  $A_n(T)$  is thus a face of  $A_n^*(T)$ . There remains to show that it is a proper face. This follows from the genericity of payoffs. Fix  $t_m^1 \in T_m^1$ . There exists an information set  $h_m \in H_m^*$  where it chooses a non-equilibrium action  $a$ . If the path from each  $(x, a)$ , for  $x \in h_m$  that is reached under the equilibrium outcome, does not pass through an information set  $h_n \in H_n^0$ , then  $a$  would be suboptimal against every equilibrium  $\bar{Q}_n^*$ . Thus, there exists a first information set  $h_n \in H_n^0$  and nodes  $x \in H_m^*$  and  $y \in h_n$  such that  $(x, a) \prec y$  and  $x$  is reached under the equilibrium outcome. Because  $A_n^*(T)$  is nonempty, there is a strategy  $t_n^0 \in \tilde{T}_n^0$  that enables  $h_n$ . Clearly, there must be multiple such strategies, again by genericity. Perturbing the probabilities of the terminal nodes following  $y$  does not affect the payoffs to strategies in  $T_m^0$  but they affect the payoff to  $t_m^1$ . Hence  $A_n(T)$  is a proper face of  $A_n^*(T)$ .  $\square$

Let  $\hat{T}_m^1$  be the set of strategies  $s_m$  in  $S_m^1(T_m^0) \setminus T_m^1$  such that: (i)  $s_m$  is an equally good reply against every point in  $P_n^0$  to which the strategies in  $T_m$  are equally good replies. Let  $\bar{T}_m^1$  be the set of strategies  $s_m$  in  $S_m^1(T_m^0) \setminus (\hat{T}_m^1 \cup T_m^1)$  that are best replies against every point in  $A_n(T)$ .

Let  $B_n^*(T)$  be the closure of the points in the interior of  $P_n^0$  against which the strategies in  $T_m$  are equally good replies and at least as good replies as strategies in  $\bar{T}_m^1$ . Let  $d_n^*(T)$  be the dimension of  $B_n^*(T)$ . If  $T_m^1$  is empty, let  $B_n(T) = P_n^0$ . Otherwise, let  $B_n(T)$  be the set of points in  $P_n^0$  that are of the form  $\lambda q_n^0 + (1 - \lambda)r_n^0$  such that  $\lambda \geq 1$ ,  $q_n^0 \in B_n^*(T)$ , and  $r_n^0 \in P_n(T_n^0)$ .

**Lemma C.7.** *Suppose  $T_m^1$  is nonempty.  $B_n(T)$  is a polyhedron of dimension  $d_n^*(T) + d_n(T_n^0) - \tilde{d}_n^*(T)$ . Each maximal face  $B'_n(T)$  of  $B_n(T)$  satisfies exactly one of the following:*

- (1) *The relative interior of  $B'_n(T)$  is contained in the relative interior of a maximal proper face of  $P_n^0$ .*
- (2) *There exists a strategy  $r_m \in \bar{T}_m^1$  such that for each  $p_n^0 \in B'_n(T)$ ,  $r_m$  is an equally good reply against every point of the form  $\lambda p_n^0 + (1 - \lambda)r_n^0$  in  $B_n^*(T)$ ; moreover, in this case, letting  $\check{R}_m^1$  be the set of such  $r_m$ , for any  $r_m \in \check{R}_m^1$ , if  $r_m$  is a best reply against a point in  $B_n^*(T)$  then every point in  $\check{R}_m^1$  is also a best reply against this point.*

- (3) *There exists a maximal proper face of  $P_n(T_n^0)$ , say  $P_n(R_n^0)$ , such that for each  $q_n \in B_n^*(T)$ ,  $r_n \in P_n(T_n^0)$  and  $\lambda > 1$ , if  $\lambda q_n + (1 - \lambda)r_n$  belongs to  $\hat{B}'_n(T)$ , then  $r_n$  belongs to  $P_n(R_n^0)$ ; moreover,  $A_n(T)$  is contained in  $P_n(R_n^0)$ .*

*Proof.* Let  $\tilde{P}_n^0$ ,  $\tilde{P}_n(T_n^0)$ , and  $\tilde{B}_n^*(T)$  be the convex cones spanned by  $P_n^0$ ,  $P_n(T_n^0)$ , and  $B_n^*(T)$  respectively. Let  $\xi : P_n^0 \times \tilde{P}_n(T_n^0) \rightarrow \tilde{P}_n^0$  be the function  $\xi(p_n^0, \tilde{r}_n^0) = p_n^0 + \tilde{r}_n^0$ . Then for each  $\tilde{p}_n^0$ ,  $\xi^{-1}(\tilde{p}_n^0)$  is a set of dimension  $d(T_n^0)$ . Hence the dimension of  $\hat{B}(T) \equiv \xi^{-1}(\tilde{B}_n^*(T))$  is  $d(T_n^0) + d_n^*(T) + 1$ . Obviously  $\hat{B}(T)$  is a polyhedron. For each face  $\hat{B}'_n(T)$  of  $\hat{B}_n(T)$  and for all points  $(p_n^0, \tilde{r}_n^0) \in \hat{B}'_n(T)$  at least one of the following holds: (i)  $p_n^0$  belongs to the boundary of  $P_n^0$ ; (ii) there exists a strategy  $t_m \in \bar{T}_m^1$  such that  $\xi(p_n^0, \tilde{r}_n^0)$  belongs to the convex cone spanned by the face of  $B_n^*(T)$  where this strategy  $t_m$  is an equally good reply; (iii)  $\tilde{r}_n^0$  belongs to the boundary of  $\tilde{P}_n(T_n^0)$ .

Observe now that  $B_n(T)$  is the projection of  $\hat{B}_n(T)$  onto the first factor. Obviously it is a polyhedron. For each  $p_n^0 \in B_n(T)$  and each  $\tilde{r}_n^0$  such that  $(p_n^0, \tilde{r}_n^0) \in \hat{B}_n(T)$ ,  $(p_n^0, \tilde{r}_n^0 + \lambda r_n^0) \in \hat{B}_n(T)$  for all  $r_n^0 \in B_n^*(T) \cap P_n(T_n^0)$  and  $\lambda \geq 0$ . If the set  $B_n^*(T)$  is in generic position (i.e. if the payoffs are in generic position), then all points in  $\hat{B}_n(T)$  that project to  $p_n^0$  are of this form. Since the set  $B_n^*(T) \cap P_n(T_n^0)$  is the intersection of  $P_n(T_n^0)$  with the affine space spanned by  $A_n(T)$ , the dimension of  $B_n(T)$  is as asserted. The enumerated properties of  $B_n(T)$  now follow directly from the corresponding points above; only the last part of property (iii) needs a proof. Suppose  $r_n^0$  belongs to a proper face  $P_n(R_n^0)$  of  $P_n(T_n^0)$  and  $A_n(T)$  is not contained in  $P_n(R_n^0)$ . If  $(p_n^0, \lambda r_n^0)$  belongs to  $\hat{B}_n(T)$ , then so does  $(p_n^0, \lambda r_n^0 + \tilde{r}_n^0)$  for  $\tilde{r}_n^0 \in A_n(T) \setminus P_n(R_n^0)$  and  $\lambda r_n^0 + \tilde{r}_n^0$  does not belong to the convex cone generated by  $P_n(R_n^0)$ .  $\square$

Let  $C_n^*(T)$  be the closure of the set of  $q_n$  in the interior of  $P_n$  such that the strategies in  $T_m$  are all equally good replies and at least as good as strategies in  $\hat{T}_m^1$  and  $S_n^0 \setminus T_n^0$ . By Lemma C.5, the dimension of the face spanned by  $T_m$  is  $d(T_m^0) + d(\Psi_m^1) + 1$ . By genericity of payoffs, the dimension of  $C_n^*(T)$  is therefore  $d(P_n) - d(T_m^0) - d(\Psi_m^1) - 1$ .

Let  $C_n(T)$  be the set of  $(p_n^1, \pi_n^1) \in P_n^1 \times \Pi_n^1$  such that there exist  $p_n^0 \in P_n^0$ ,  $p_n^2 \in P_n(T_n^1)$ ,  $q_n \in C_n^*(T)$ , and  $\mu \in \mathbb{R}_+^3$ , such that  $\sum_i \mu^i p_n^i \in C_n^*(T)$ ,  $\sum_i \mu^i = 1$ ,  $\mu^1 > 0$ , and  $\bar{\pi}_n^1(\sum_i \mu^i p_n^i) = \pi_n^1$ .

**Lemma C.8.** *The set  $C_n(T)$  is a polyhedron of dimension  $d(P_n^0) + d(P_n^1) + d_n(\Psi_n^1) - d(T_m^0) - d(\Psi_m^1) - d_n^*(T)$ . On each maximal proper face  $C'$  of  $C_n(T)$ , exactly one of the following holds for all  $(p_n^1, \pi_n^1)$  in  $C'$ . If  $q_n \in C^*(T)$  is of the form  $\sum_i \mu^i p_n^i$  for some  $p_n^0 \in P_n^0$  and  $p_n^2 \in P_n(T_n^1)$ , and  $\bar{\pi}_n^1(\sum_i \mu^i p_n^i) = \pi_n^1$ , then:*

- (1)  $p_n^1$  belongs to a maximal proper face of  $P_n^1$ ;

- (2) for  $i = 0$  or  $i = 1$ , but not both, there exists  $s_m \in S_m^i \setminus T_m^i$ , which actually belongs to  $\hat{T}_m^i$  if  $i = 1$ , that is as good a reply as points in  $T_m$  against  $q_n$ ; moreover, in this case, if  $i = 0$ ,  $T_m^0$  and  $s_m$  span a face of  $P_m^0$  of which  $P_m(T_m^0)$  is a maximal proper face; and if  $i = 1$ ,  $\Psi_m^1$  and  $\bar{\pi}_n^1(s_m)$  span a face of  $\Pi_m^1$  of which  $\Psi_m^1$  is a maximal proper face.
- (3) there exists a maximal proper face  $\Psi'$  of  $\Psi_n$  such that  $\bar{\pi}_n^1(p_n^2)$  belongs to  $\Psi'$ .

*Proof.* We show that  $C_n(T)$  is a polyhedron of the stated dimension. Since the construction is similar to that in the previous lemma, the enumerated properties can be proved just as before. Let  $P_n^2$  be the convex hull of  $P_n^0$  and  $P_n(T_n^1)$ . Using Lemma C.5, the dimension of  $P_n^2$  is  $d(P_n^0) + d(\Psi_n^1) + 1$ . Let  $\tilde{P}_n^2$  and  $\tilde{P}_n$  be the convex cones spanned by  $P_n^2$  and  $P_n$ , respectively. Define  $\xi : P_n^1 \times \tilde{P}_n^2 \rightarrow \tilde{P}_n$  by  $\xi(p_n^1, \tilde{p}_n^2) = p_n^1 + \tilde{p}_n^2$ . Then for each  $\tilde{p}_n$  in the interior of  $\tilde{P}_n$ ,  $\xi^{-1}(\tilde{p}_n)$  is a set of dimension  $d(P_n^1) + d(P_n^0) + d(\Psi_n^1) + 1$ . Letting  $\tilde{C}_n^*(T)$  be the convex cone spanned by  $C_n^*(T)$ , the dimension of  $\hat{C}_n(T) \equiv \xi^{-1}(\tilde{C}_n^*(T))$  is  $d(P_n^0) + d(P_n^1) + d(\Psi_n^1) + 1 - d(T_m^0) - d(\Psi_m^1)$ . The function  $\bar{\pi}_n^1$  extends to  $\tilde{P} \setminus \{0\}$ .  $C_n(T)$  is the image of  $\hat{C}_n(T)$  under the function  $\chi : P_n^1 \times \tilde{P}_n^2$  given by  $\chi(p_n^1, \tilde{p}_n^2) = (p_n^1, \bar{\pi}_n^1(p_n^1 + \tilde{p}_n^2))$  and is thus a polyhedron. As will be shown in the course of the proof of the next lemma,  $C_n^*(T) \cap P_n(T_n^1) \subset P_n^0$ . Therefore,  $C_n^*(T) \cap P_n^2$  is the intersection of  $P_n^0$  with the affine space spanned by  $B_n^*(T)$ . For each  $(p_n^1, \tilde{p}_n^2) \in \hat{C}_n(T)$ , the point  $(p_n^1, \tilde{p}_n^2 + \mu q_n^0)$  belongs to  $\hat{C}_n(T)$  for all  $\mu > 0$  and  $q_n^0 \in C_n^*(T) \cap P_n^2$ , and has the same image under  $\chi$  as  $(p_n^1, \tilde{p}_n^2)$ . Moreover, if  $C_n^*(T)$  is in general position then for each  $(p_n^1, \pi_n^1)$  in  $C_n(T)$  every point in its inverse image under  $\chi$  is expressible in this form. Therefore, the dimension of  $C_n(T)$  is as given.  $\square$

Let  $\mathcal{T}$  be the collection of  $T$ 's such that  $A_n(T)$ ,  $B_n(T)$  and  $C_n(T)$  are nonempty for each  $n$ . For each  $T \in \mathcal{T}$ , let  $\mathcal{Q}_n(T) = A_n(T) \times B_n(T) \times C_n(T)$  for each  $n$  and let  $\mathcal{Q}(T) = \mathcal{Q}_1(T) \times \mathcal{Q}_2(T)$ .

**Lemma C.9.**  $(q^*, p^0, p^1, \pi^1)$  belongs to  $\mathcal{Q}$  iff it belongs to  $\mathcal{Q}(T)$  for some  $T \in \mathcal{T}$ .

*Proof.* Suppose for each  $n$  that  $q_n^* \in A_n(T)$ ,  $p_n^0 \in B_n(T)$ ,  $(p_n^1, \pi_n^1) \in C_n(T)$  for some  $T$ . Choose  $r_n^0 \in P_n(T_n^0)$  and  $\lambda_n^0$  such that  $q_n^0 \equiv (1 - \lambda_n^0)p_n^0 + \lambda_n^0 r_n^0 \in B_n^*(T)$ . Also, fix  $\tilde{p}_n^0, r_n^2 \in P_n(T_n^1)$ ,  $\mu_n^0, \mu_n^1, \mu_n^2$  such that  $q_n^1 \equiv \mu_n^0 \tilde{p}_n^0 + \mu_n^1 p_n^1 + \mu_n^2 r_n^2$  belongs to  $C_n^*(T)$ . Fix points  $\tilde{q}_n^0$  and  $\tilde{q}_n^1$  in the interior of  $A_n(T)$  and  $B_n(T)$  for each  $n$  and consider for each  $0 < \varepsilon < 1$ , the LPS  $(q^*, \tilde{q}^0(\varepsilon), \tilde{q}^1(\varepsilon))$  where for each  $n$ ,  $\tilde{q}_n^0(\varepsilon) = (1 - \varepsilon)\tilde{q}_n^0 + \varepsilon q_n^0$ ; and  $\tilde{q}_n^1(\varepsilon) = (1 - \varepsilon)\tilde{q}_n^1 + \varepsilon \tilde{q}_n^1(\varepsilon) + \varepsilon^2 q_n^1$ . The strategies in  $T_n$  are equally good replies to  $q_m^*, \tilde{q}_m^0(\varepsilon), \tilde{q}_m^1(\varepsilon)$  for all  $\varepsilon$ . We show that these strategies are lexicographic best replies to  $(q^*, \tilde{q}^0(\varepsilon), \tilde{q}^1(\varepsilon))$  for all small  $\varepsilon$ , which proves that  $(q^*, p^0, p^1, \pi^1)$  belongs to  $\mathcal{Q}$ .

Observe first that for all  $\varepsilon$ , a strategy in  $S_n^0 \setminus T_n^0$  is an equally good reply against  $q_m^*$  and  $\tilde{q}_m^0(\varepsilon)$  as strategies in  $T_n^0$ , by Lemma C.2, and no better a reply against  $\tilde{q}_m^1(\varepsilon)$  by construction of  $C_n^*(T)$ . Now for a strategy  $s_n^1$  in  $S_n^1$ , consider the strategy  $t_n^1$  that agrees with  $s_n^1$  everywhere except that starting at each first information set  $h_n \in H_n^0$  that  $s_n^1$  enables,  $t_n^1$  agrees with some  $t_n^0$  in  $T_n^0$ . Since strategies in  $T_n^0$  are at least as good as the other strategies in  $S_n^0$ , clearly  $t_n^1$  is at least good a reply against  $(q_m^*, \tilde{q}_m^0(\varepsilon), \tilde{q}_m^1(\varepsilon))$ . Observe now that  $t_n^1$  belongs to  $S_n^1(T_n^0)$ . If it belongs to  $\hat{T}_n^1$ , then it is an equally good reply as strategies in  $T_n$  against  $q_m^*$  and  $\tilde{q}_m^0(\varepsilon)$  and no better reply a reply against  $\tilde{q}_m^1(\varepsilon)$  by definition. If it belongs to  $\bar{T}_n^1$  then it is an equally good reply to  $q_m^*$ , no better reply against  $\tilde{q}_m^0(\varepsilon)$  for all  $\varepsilon$ , and a strictly worse reply against  $\tilde{q}_m^1(\varepsilon)$  for all small  $\varepsilon$ , again by construction, since it is an inferior reply against  $\bar{q}_n^1$  which belongs to the interior of  $B_n^*(T)$ . Finally, if it belongs to  $S_n^1(T_n^1) \setminus (\hat{T}_n^1 \cup \bar{T}_n^1)$  then it is no better a reply against  $q_m^*$  and an inferior reply to  $\tilde{q}_m^0(\varepsilon)$  for all small  $\varepsilon$ , since it is an inferior reply to  $\bar{q}_n^0$  by construction of  $A_n(T)$ . Thus the strategies in  $T$  are lexicographic best replies to  $(q^*, \tilde{q}^0(\varepsilon), \tilde{q}^1(\varepsilon))$  for all small  $\varepsilon$ .

Before proceeding to prove the converse, we use the above argument to show that the intersection of  $C_n^*(T)$  with the convex hull  $P_n^2$  of  $P_n^0$  is in fact the intersection  $F$  of  $P_n^0$  with the affine space spanned by  $B_n^*(T)$ . Take a point  $q_n^1$  in  $C_n^*(T) \cap P_n^2$ . If it belongs to  $P_n^0$ , then in fact it belongs to  $F$  by the definitions of  $B_n^*(T)$  and  $\hat{T}_n^1$ . If it does not belong to  $P_n^0$ , then it assigns a positive weight to some strategy  $s_n^1 \in T_n^1$ . The above argument applied when using this  $q_n^1$  shows that  $q_m^*$  is a best reply to  $(q_n^*, \tilde{q}_n^0(\varepsilon), \tilde{q}_n^1(\varepsilon))$  and the strategies in  $T_n$  and  $S_n^0$  are best replies to  $q_m^*$ . Observe now that  $\tilde{q}_n^1(\varepsilon)$  is a convex combination of strategies in  $S_n^0$  and  $T_n^1$ . Therefore, for all small  $\delta$ ,  $((1 - \delta - \delta^2)q_n^* + \delta\tilde{q}_n^0(\varepsilon) + \delta^2\tilde{q}_n^1(\varepsilon), q_m^*)$  is an equilibrium if  $\varepsilon$  is small as well. But these points induce different outcomes because  $\tilde{q}_n^1(\varepsilon)$  has a non-equilibrium strategy, namely one in  $T_n^1$ , in its support, which is impossible. Thus,  $C_n^*(T) \cap P_n^2 = F$  as claimed.

Returning to the proof of this lemma, suppose  $(q^*, (p^0, p^1), \pi^1)$  belongs to  $\mathcal{Q}$ . Let  $q_n^0 = (1 - \lambda_n^0)p_n^0 + \lambda_n^0r_n^0$  and let  $q_n^1 = \mu_n^0\tilde{p}_n^0 + \mu_n^1p_n^1 + \mu_n^2r_n^2$  where  $(1 - \lambda_n^0)q_n^* + \lambda_n^0r_n^0$  is a best reply against  $(q^*, q^0, q^1)$ ; and  $r_n^2$ , if  $\mu_n^2 > 0$ , is a best reply against  $q_n^*$  and a weakly better reply against  $(q^*, q^0, q^1)$  than all the strategies in  $P_n^1$ . Let  $Q_n^0$  be the face of  $P_n^0$  that contains  $(1 - \lambda_n^0)q_n^* + \lambda_n^0r_n^0$ . Let  $Q_n^1$  be the face of  $P_n^1$  that contains  $r_n^2$  in its interior if  $\mu_n^2 > 0$ . Let  $T_n^0$  be the set of strategies  $t_n$  in  $S_n^0$  such that if  $t_n$  enables a first information set  $h_n \in H_n^0$  then the choices from there on prescribed by  $t_n^0$  coincide with the choices dictated by some vertex of  $Q_n^0$  or  $Q_n^1$  that enables  $h_n$ . Observe that each  $t_n^0 \in T_n^0$  is optimal against  $(q^*, q^0, q^1)$ . If  $\mu_n^2 > 0$ , let  $\Psi_n^1$  be the face of  $\Pi_n^1$  that contains  $\bar{\pi}_n^1(r_n^2)$  in its interior; otherwise let  $\Psi_n^1$  be the empty set.

We can now assume without loss of generality that the strategies in  $Q_n^0$  and  $Q_n^1$  are equally good replies against  $(q_m^*, q_m^0, q_m^1)$  and hence best replies. Indeed, if for  $i$  either 0 or 1, if the strategies in  $Q_n^1$  do not yield the same payoff against  $q_m^i$  as those in  $Q_n^0$ , modify  $q_m^i$  as follows: pick a point  $\tilde{r}_m^0$  in the face  $\tilde{Q}_n^0$  of  $Q_n^0$  containing  $q_m^*$  in its interior such that the strategies in  $Q_n^0$  are equally good replies, the strategies in  $Q_n^1$  do strictly better than the strategies in  $Q_n^0$  and at least as well as the other strategies in  $S_n^1$ . There exists a unique  $\nu_m^i \in [0, 1]$  such that the strategies in  $Q_n^0$  and  $Q_n^1$  are now equally good replies against  $(1 - \nu_m^i)q_m^i + \nu_m^i \tilde{r}_m^i$ . Thus, our assumption is without loss of generality.

Since the strategies in  $Q_n^0$  and  $Q_n^1$  are best replies. There remains to show that every strategy in  $S_n(T_n^0; \Psi_n^1)$  is a best reply against  $(q_m^*, q_m^0, q_m^1)$ . Fix  $s_n \in S_n(T_n^0; \Psi_n^1)$ . To show that it is a best reply it is sufficient to show that an information set  $h_n$  that is enabled by  $s_n$  is enabled by some vertex of either  $Q_n^0$  or  $Q_n^1$  and that this vertex agrees with  $s_n$ 's choice  $a_n$  there. Suppose that this  $h_n$  is in  $H_n^*$  and  $a_n \in A_n^*$ , or  $h_n$  belongs to  $H_n^0$ ; then obviously some strategy in  $Q_n^0$  or  $Q_n^1$  enables  $h_n$  and chooses  $a_n$ , by the definition of  $T_n^0$ . If  $h_n \in H_n^*$  and  $a_n \notin A_n^*$  or  $h_n$  follows some information set in  $H_n^*$  by the choice of a non-equilibrium action, then some strategy in  $Q_n^1$  enables it and chooses this action, since otherwise  $s_n$  enables a terminal node that is excluded by all strategies in  $Q_n^1$ , contradicting the assumption that  $\bar{\pi}_n^1(s_n) \in \Psi_n^1$ . Thus  $s_n$  is a best reply and  $(q^*, q^0, q^1)$  belongs to  $\mathcal{Q}(T)$ .  $\square$

**Lemma C.10.**  $\mathcal{Q}$  is a pseudomanifold of dimension  $\hat{d} \equiv d(\mathbb{P})$ .

*Proof.* Each  $\mathcal{Q}(T)$  is a polyhedron of dimension  $\hat{d}$ . By the previous lemma  $\mathcal{Q} = \cup_{T \in \mathcal{T}} \mathcal{Q}(T)$ . Therefore,  $\mathcal{Q}$  has dimension  $\hat{d}$ . To show that  $\mathcal{Q}$  is a pseudomanifold, we establish three facts for each  $(q^*, (p^0, p^1), \pi^1)$  that belongs to some  $\mathcal{Q}(T)$ : (1) if  $(q^*, (p^0, p^1), \pi^1)$  belongs to the interior of  $\mathcal{Q}(T)$ , then it does not belong to the interior of  $\mathcal{Q}(R)$  for  $R \neq T$ ; (2) if  $(q^*, (p^0, p^1), \pi^1)$  is a generic point in a maximal proper face  $\mathcal{Q}'$  of  $\mathcal{Q}(T)$ , then it does not belong to  $\mathcal{Q}(R)$  for any  $R \neq T$  if  $(p^0, p^1) \in \partial \mathbb{P}$ , and it belongs to the boundary of  $\mathcal{Q}(R)$  for exactly one other  $R \neq T$  if  $(p^0, p^1) \notin \partial \mathbb{P}$ ; moreover in the latter case it belongs to the interior of a maximal proper face of this  $\mathcal{Q}(R)$  as well; (3) given  $T, R \in \mathcal{T}$ , there exists a finite chain  $T = T(0), \dots, T(k) = R$  such that for each  $0 \leq j \leq k-1$ ,  $\mathcal{Q}(T(j)) \cap \mathcal{Q}(T(j+1))$  is a subset of a maximal proper face of each and has a nonempty interior in this face.

Fix  $T = (T^0, \Psi^1)$  and  $x = (q^*, (p^0, p^1), \pi^1) \in \mathcal{Q}(T)$ . For each  $n$ , choose  $q_n^0 \equiv (1 - \lambda_n^0)p_n^0 + \lambda_n^0 r_n^0$  in  $B_n^*(T)$  and  $q_n^1 \equiv \mu_n^0 \tilde{p}_n^0 + \mu_n^1 p_n^1 + \mu_n^2 r_n^1 \in C_n^*(T)$ .

We start with (1). Suppose now that  $x$  belongs to the interior of  $\mathcal{Q}(R)$  for some  $R$ . We show that  $R = T$ . Since  $x$  belongs to the interior of  $\mathcal{Q}(T)$ , we can assume that every strategy in  $S_m^0 \setminus T_m^0$  is inferior to  $q_m^1$ . Let  $s_m$  be a strategy in  $S_m^0 \setminus T_m^0$ . Since  $s_m$  is an inferior reply

against  $q_m^1$  compared to the strategies in  $T_n^0$ , by Lemma C.3 there exists an information set  $h_n \in H_n^0$  that is enabled by  $s_n$  where the action chosen by  $s_n$  is suboptimal and different from the action chosen by every  $t_n \in T_n^0$  that enables  $h_n$ . But the posterior belief over the terminal nodes following  $h_n$  computed from  $q_m^1$  can be computed from  $\pi_m^1$ . This implies that for any  $q'_m$  such that  $\bar{\pi}_m^1(q'_m) = \pi_m^1$ ,  $s_n$  is an inferior strategy. Therefore,  $(p_m^1, \pi_m^1)$  cannot belong to  $C_m(R)$  unless  $R_m^0 \subseteq T_m^0$ . Moreover, if  $R_m^0 \subsetneq T_m^0$ , then it cannot belong to the interior of  $C_m(R)$ , since strategies in  $T_m^0 \setminus R_m^0$  are also optimal. Thus, if  $x$  belongs to the interior of  $\mathcal{Q}(R)$ ,  $R_m^0 = T_m^0$  for each  $m$ . If  $\Psi_m^1$  is empty, this implies that  $R_m = T_m$ . Suppose now that  $\Psi_m^1$  is nonempty. Since  $x$  is in the interior we can assume that  $\bar{\pi}_m^1(p_m^1) \neq \pi_m^1$  and that  $\bar{\pi}_n(r_m^1)$  is in the interior of  $\Psi_m^1$ . Observe that  $\bar{\pi}_n(r_m^1)$  can be computed uniquely from  $p_m^1$  and  $\pi_m^1$  by taking the line segment from  $\bar{\pi}_m^1(p_m^1)$  through  $\pi_m^1$  and computing the boundary point of this line. This implies that if  $x$  belongs to the interior of  $\mathcal{Q}(R)$ , then  $R_m = (T_m^0, \Psi_m^1)$ . Thus  $R = T$ .

We turn to point (2). Suppose now  $x$  belongs to the relative interior of a maximal proper face of  $\mathcal{Q}(T)$  and that  $(p^0, p^1)$  belongs to  $\partial P$ . Then if it belongs to another  $\mathcal{Q}(R)$  it cannot be in the interior and must belong to the boundary. The arguments of the previous paragraph apply to show that  $x$  does not belong to the interior of a maximal face of  $\mathcal{Q}(R)$ . Indeed, it relied on strategies in  $S_m^0 \setminus T_m^0$  being inferior to  $q_n^1$ , and  $q_n^0$  (resp.  $q_n^1$ ) not belonging to the boundary of  $B_n(T)$  (resp.  $C_n(T)$ ). Thus  $x$  must belong to a face of dimension at most  $\hat{d} - 2$ . The set of such points in this maximal proper face of  $\mathcal{Q}(T)$  then has dimension at most  $\hat{d} - 2$ , i.e. it is nongeneric.

Suppose that  $x$  belongs to the relative interior of a maximal proper face of  $\mathcal{Q}(T)$  but that  $(p^0, p^1)$  is in the interior in  $\mathbb{P}$ . Then for exactly one  $n$ , just one of the following hold: (2a)  $q_n^*$  belongs to the boundary of  $A_n^*(T)$ ; (2b)  $p_n^0$  belongs to the boundary of  $B_n(T)$ ; (2c)  $(p_n^1, \pi_n^1)$  belongs to the boundary of  $C_n(T)$ . We start with (2c). By the properties we proved for  $C_n(T)$  in Lemma C.8, and since  $(p^0, p^1) \notin \partial \mathbb{P}$ , either property (ii) or property (iii) of that lemma holds. Under property (ii)  $x$  belongs to the boundary of  $\mathcal{Q}(R)$  where  $R = ((R_m^0, \Phi_m^1), (T_n, \Psi_n^1))$  is defined as follows. If the strategy  $r_m^i$  identified there belongs to  $S_m^0$ , then  $R_m^0$  is the vertex set of the face spanned by  $r_m^i$  and  $T_m^0$ , while  $\Phi_m^1 = \Psi_n^1$ ; if the strategy  $r_m^i$  belongs to  $S_n^1(T_n^0; \Psi_n^1)$ , then  $R_m^0 = T_m^0$  and  $\Phi_m^1$  is a face of  $\Pi_m^1$  that has  $\Psi_m^1$  as a maximal face with  $\pi_m^1(r_m^1) \in \Phi_m^1 \setminus \Psi_m^1$ . Under property (iii)  $x$  belongs to the boundary of  $\mathcal{Q}(R)$  where  $R = ((T_n^0, \Phi_n^1), (T_m^0, \Psi_m^1))$  where  $\Phi_n^1$  is the maximal proper face of  $\Psi_n^1$  identified there.

Suppose  $x$  satisfies (2b). Then by the properties we proved for  $B_n(T)$  in Lemma C.7, either property (ii) or property (iii) of that lemma holds. Under property (ii) let  $\tilde{R}_m^1$  be the

set of strategies in  $\tilde{T}_m^1$  that are now best replies against  $q_n^0$ . Let  $\tilde{\Phi}_m^1$  be the smallest face of  $\Pi_m^1$  that contains  $\Psi_m^1$  and the vectors  $\pi_m^1(\tilde{r}_m^1)$  for  $\tilde{r}_m^1 \in \tilde{R}_m^1$ . Then the strategies in  $T_m^0$  and  $S_m^1(\Phi_m^1)$  are equally good replies against  $q_n^0$ . Moreover by the genericity of  $x$ , if one of these strategies is a best reply against a point in  $B_n^*(T)$  then all these points are best replies as well. For each face  $\Phi_m^1$  of  $\tilde{\Phi}_m^1$  that has  $\Psi_m^1$  a maximal proper face, choose a strategy  $r_m(\Phi_m^1)$  that maps to a vertex of  $\Phi_m^1$  that is not contained in  $\Psi_m^1$ . The set of points in  $B_n^*(T)$  against which the strategies in  $\tilde{R}_m$  are as good replies as  $T_m$  has dimension  $d_n^*(T) - 1$ . However, the set of points in  $C_n^*(T)$  where two or more of these strategies  $r_m(\Phi_m^1)$  are also best replies has dimension  $d(P_n) - d(T_m^0) - d(\Psi_m^1) - 3$  or less. Therefore, for  $R$  and  $R'$  of the form  $((T_n^0, \Psi_n^1), (T_m^0, \Phi_m^1))$  the set of  $(p_n^1, \pi_n^1)$  that lies in the intersection  $C_n(T) \cap C_n(R) \cap C_n(R')$  is at most  $d_n^0 + d_n^1 + d_n(\Psi_n^1) - d_n^*(T) - 1$  or less. This implies that generic  $(p_n^1, \pi_n^1)$  in  $C_n(T)$  belongs to at most one of these sets. Moreover, if  $x$  belongs to  $\mathcal{Q}(R)$ , then it belongs to the boundary of  $\mathcal{Q}(R)$ : indeed the point  $\phi_m^1$  in  $\Phi_m^1$  such that  $\pi_m^1$  is a convex combination of  $\bar{\pi}_m^1(p_m^1)$  and  $\phi_m^1$  is uniquely determined, as we argued above; since this point belongs to  $\Psi_m^1$ , which is a face of  $\Phi_m^1$ ,  $x$  indeed belongs to the boundary of  $\mathcal{Q}(R)$  if it belongs to  $\mathcal{Q}(R)$ . To finish the proof of this case, we now show that  $x$  belongs to at least one  $\mathcal{Q}(R)$ . Take an  $\tilde{r}_m^1$  that yields the highest payoff against  $q_n^1$  among the strategies in  $\tilde{R}_m^1$ . If this payoff is higher than the payoff to the strategies in  $T_m^0$ , pick a point  $\bar{q}_n^0$  in the interior of  $B_n^*(T)$  and replace  $q_n^1$  with  $\tilde{q}_n^1(\varepsilon) \equiv (1 - \varepsilon)q_n^1 + \varepsilon\bar{q}_n^0$  where  $\varepsilon$  is the unique number where the strategies  $T_m^0$  and  $\tilde{r}_m^1$  are equally good replies; then  $x$  belongs to some  $\mathcal{Q}(R)$  that has  $\pi_m^1(\tilde{r}_m^1)$  as an extra vertex. If the payoff to  $\tilde{r}_m^1$  is lower, take a point  $\bar{q}_n^0$  in  $P_n^0$  against which the strategies in  $T_m^0$  are equally good and worse than the strategies in  $\tilde{R}_m^1$  and repeat the argument to show that  $x$  belongs to  $\mathcal{Q}(R)$ .

Finally, suppose that  $x$  satisfies (2a). Let  $A'_n$  be a maximal proper face of  $A_n(T)$  that contains  $q_n^*$  in its interior. Let  $\check{R}_m^1$  be the set of strategies  $\check{r}_m$  in  $S_m^1(T_m^0) \setminus (T_m^1 \cup \hat{T}_m^1 \cup \bar{T}_m^1)$  that are best replies against all the points in  $A'_n$ . Observe that  $\check{R}_m^1$  is nonempty if  $q_n^*$  belongs to the interior of  $P_n(\tilde{T}_n^0)$ . Indeed, in this case, the interior of  $A'_n$  which is a face of  $A_n(T)$  is contained in the interior of  $P_n(\tilde{T}_n^0)$ , which implies that some strategy in  $S_m^1$  is now optimal against every point in this face. Let  $\mathcal{R}_n^0$  be the set of subsets  $R_n^0$  of  $T_n^0$  such  $P_n(R_n^0)$  is a maximal proper face of  $P_n(T_n^0)$  and  $P_n(R_n^0) \cap Q_n^* = A'_n$ . Observe that  $\mathcal{R}_n^0$  is nonempty if  $q_n^*$  belongs to the boundary of  $P_n(\tilde{T}_n^0)$ .

Let  $\check{\Phi}_m^1$  be the smallest face of  $\Pi_m^1$  that contains  $\Psi_m^1$  and the vectors  $\pi_m^1(\check{r}_m^1)$  for  $\check{r}_m^1 \in \check{R}_m^1$ . For each face  $\Phi_m^1$  of  $\check{\Phi}_m^1$  that has  $\Psi_m^1$  as a maximal proper face, choose a strategy  $\check{r}_m^1(\Phi_m^1)$  that maps to a vertex of  $\Phi_m^1$  that is not contained in  $\Psi_m^1$ . Take  $R$  and  $R'$  of



the form  $((T_n^0, \Psi_n^1), (T_m^0, \Phi_m^1))$ . Since  $A'_n$  is a face of  $A_n(T)$ ,  $\tilde{d}_n^*(R) = \tilde{d}_n^*(R') = \tilde{d}_n^*(T) - 1$ . Since the strategies in  $\tilde{R}_m^1$  are inferior replies to points in the interior of  $A_n(T)$ ,  $B_n^*(R)$  if nonempty has dimension  $\tilde{d}_n^*(T) - 1$ . Therefore, if  $B_n(R) \neq B_n(R')$ , their intersection with  $B_n(T)$  has codimension 1 in  $B_n(T)$  and  $x$  cannot belong to two of these sets at once. On the other hand, if  $B_n(R) = B_n(R')$  then an argument similar to that in the previous paragraph shows that generic  $(p_n^1, \pi_n^1) \in C_n(T)$  can belong to at most one of these sets,  $C_n(R)$  and  $C_n(R')$ . Hence a generic  $x$  belongs to at most one of these sets. Likewise, for  $R$  of the form  $((R_n^0, \Psi_n^1), (T_m^0, \Psi_m^1))$  with  $R_n^0 \in \mathcal{R}_n^0$  and  $R'$  of the form  $((R_n'^0, \Psi_n^1), (T_m^0, \Psi_m^1))$  or  $((T_n^0, \Psi_n^1), (T_m^0, \Phi_m^1))$ , the intersection of  $B_n(T)$  with  $B_n(R)$  and  $B_n(R')$  has codimension at least one. Thus  $x$  belongs to at most one set  $\mathcal{Q}(R)$ . To finish the proof of this part, we show that it belongs to at least one such set.

Let  $\tilde{r}_m^1(\Phi_m^1)$  be a strategy that is a lexicographic best reply to  $(q_n^0, q_n^1)$  among the strategies in this class. If  $\tilde{r}_m^1(\Phi_m^1)$  is a lexicographic (weakly) better reply against  $(q_n^0, q_n^1)$  (or it is a worse reply and  $q_n^*$  belongs to the interior of  $P_n(\tilde{T}_n^0)$ ) we choose a point  $\bar{q}_n^0$  in the interior of  $P_n(\tilde{T}_n^0)$  against which the strategy  $\tilde{r}_m^1(\Phi_m^1)$  is at least as good as the other points in this class and inferior (resp. superior) to strategies in  $T_m^0$ . The strategies in  $T_m^0$  and  $S_m(T_m^0; \Phi_m^1)$  are now equally good replies against some average of  $q_n^0$  and  $\bar{q}_n^0$ , as well as some average of  $q_n^1$  and  $\bar{q}_n^1$ . The point  $x$  then belongs to  $((T_m^0, \Phi_m^1), (T_n^0, \Psi_n^1))$ . If  $q_n^*$  belongs to the boundary of  $P_n(\tilde{T}_n^0)$ , and  $\tilde{r}_m^1(\Phi_m^1)$  is an inferior reply, then  $\tilde{R}_n^0$  is nonempty. There exists  $R_n^0$  such that  $q_n^0$  is a convex combination of a point in  $P_n(\tilde{T}_n^0)$  and  $p_n^0$ . Then  $x$  belongs to  $((\tilde{T}_n^0, \Psi_n^1), (T_m^0, \Psi_m^1))$ . This concludes the proof of (2).

We turn now to (3). Given  $\mathcal{Q}(T)$  and  $\mathcal{Q}(R)$  for  $T = (T^0, \Psi^0) \neq R = (R^0, \Phi^1)$ , we will first construct a sequence  $T = T(1), \dots, T(k) = \tilde{T}$  where  $\tilde{T} = ((\tilde{T}_n^0, \emptyset), (\tilde{T}_m^0, \emptyset))$ . And, likewise one from  $R$  to  $\tilde{R}$ . Then we will show how to construct a sequence from  $\tilde{T}$  to  $\tilde{R}$ .

In case  $\tilde{T}_n^0 \neq T_n^0$  for some  $n$ , let  $T = T(0), \dots, T(k)$  be a sequence where for each  $j > 0$ ,  $T(j) = ((T_n^0(j), \Psi_n^1), (T_m^0(j), \Psi_m^1))$  with  $P_n(T_n^0(j)) \times P_m(T_m^0(j))$  being a maximal proper face of  $P_n(T_n^0(j-1)) \times P_m(T_m^0(j-1))$ , and  $T_n(k) = \tilde{T}_n^0$  for each  $n$ . This sequence generates a sequence of polyhedra  $\mathcal{Q}(T) = \mathcal{Q}(T(0)), \dots, \mathcal{Q}(T(k))$  where for each  $j > 0$ , the intersection of  $\mathcal{Q}(T(j))$  with  $\mathcal{Q}(T(j-1))$  is contained in a maximal proper face of each and has a nonempty interior. After this operation we have  $\mathcal{Q}(\tilde{T}(k))$ , where  $\tilde{T}(k) = ((\tilde{T}_n^0, \Psi_n^1), (\tilde{T}_m^0, \Psi_m^1))$ . In case  $\Psi_n^1$  is nonempty for some  $n$ , let  $\Psi^1 = (\Psi_n^1(0), \Psi_m^1(0)), \dots, (\Psi_n^1(l), \Psi_m^1(l)) = (\emptyset, \emptyset)$  be a sequence such that for each  $1 \leq j \leq l$ ,  $\Psi_n^1(j) \times \Psi_m^1(j)$  is a maximal proper face of  $\Psi_n^1(j-1) \times \Psi_m^1(j-1)$ , and  $\Psi^1(l) = \emptyset$ . This way we can connect  $\mathcal{Q}(\tilde{T}(k))$  with  $\mathcal{Q}(\tilde{T})$  where  $\tilde{T} = ((\tilde{T}_n^0, \emptyset), (\tilde{T}_m^0, \emptyset))$ . Now we show how to connect  $\tilde{T}$  with  $\tilde{R}$  for two sets  $T, R$  in  $\mathcal{T}$ . Because  $Q_n^*$  is connected for

each  $n$ , there exists a sequence  $\tilde{T}^0 = (S_n^0(0), S_m^0(0)), \dots, (S_n^0(l), S_m^0(l)) = \tilde{R}_n^0$  where for each  $1 \leq j \leq l$ , either  $P_n(S_n^0(j)) \times P_n(S_m^0(j))$  is either a maximal proper face of  $P_n(S_n^0(j+1))$  or vice versa and for each  $n$ ,  $Q_n^*$  intersects the interior of the set  $P_n(S_n^0(j))$ . This generates a sequence  $\mathcal{Q}^0, \dots, \mathcal{Q}^j$ , where  $\mathcal{Q}^j = ((S_n^0(j), \emptyset), (S_m^0(j), \emptyset))$ . Thus we have constructed a sequence of sets in  $\mathcal{T}$  that connect  $T$  and  $R$ .  $\square$

This concludes the proof of the first statement in the theorem. Now we prove the second part.

**Lemma C.11.**  $\psi : (\mathcal{Q}, \partial\mathcal{Q}) \rightarrow (\mathbb{P}, \partial\mathbb{P})$  is essential iff  $Q^*$  is stable.

*Proof.* Let  $Y = [0, 1] \times P$ . For each  $0 < \varepsilon \leq 1$ , let  $Y_\varepsilon = [0, \varepsilon] \times P$  and let  $\partial Y_\varepsilon$  be the boundary of  $Y_\varepsilon$ . Each  $(\varepsilon, p) \in Y$  defines a strategic game  $G(\varepsilon, p)$  where the strategy set is  $P$  but where the payoff from an enabling strategy profile  $q$  is the payoff in  $G$  from the profile  $(1 - \varepsilon)q + \varepsilon p$ . If  $q$  is an equilibrium of  $G(\varepsilon, p)$ , we say that  $(1 - \varepsilon)q + \varepsilon p$  is a perturbed equilibrium of  $G(\varepsilon, p)$ . Let  $\mathcal{E}$  be the closure of the set of  $(\varepsilon, p, q)$  such that  $(\varepsilon, p) \in Y_1 \setminus \partial Y_1$  and  $q$  is a perturbed equilibrium of  $G(\varepsilon, p)$ . Let  $\theta$  be the projection map from  $\mathcal{E}$ . For each subset  $E$  of  $\mathcal{E}$  and each  $0 < \varepsilon$ , let  $(E_\varepsilon, \partial E_\varepsilon)$  be  $E \cap \theta^{-1}(P_\varepsilon, \partial P_\varepsilon)$ .

In [9] we show that there exists  $0 < \bar{\varepsilon} \leq 1$  and a finite number of subsets  $E^1, \dots, E^K$  of  $\mathcal{E}$  such that for each  $0 < \varepsilon \leq \bar{\varepsilon}$ : (i)  $(E_\varepsilon^k, \partial E_\varepsilon^k)$  is a pseudomanifold (in fact a homology manifold) of dimension  $d(P) + 1$  for each  $k$ ; (ii)  $E_\varepsilon^k \cap E_\varepsilon^j \subset \partial\phi^{-1}(\partial P_1)$  for  $k \neq j$ ; (iii)  $\cup_k E_\varepsilon^k = \mathcal{E}_\varepsilon$ .

In [6] we show that there exists a neighborhood of  $\Sigma^*$  and an  $\varepsilon > 0$  such that the set of  $\varepsilon$ -perfect equilibria in this neighborhood is connected. The image of this set of  $\varepsilon$ -perfect equilibria under  $\rho$  is therefore connected. Thus there exists some  $k$  such that  $E_0^k = \{0\} \times Q^*$ .  $Q^*$  is stable iff the projection  $\theta$  from  $E^k$  to  $Y_\varepsilon$  is essential (Mertens [19]). For simplicity in notation we refer to this  $E^k$  as simply  $E$ .

It is now sufficient to prove that  $\psi$  is essential iff the projection  $\theta$  from  $E$  is essential. For each  $n$ ,  $\tilde{P}_n \equiv [0, 1] \times \mathbb{P}_n$  and define  $\chi_n : \tilde{P}_n \rightarrow P_n$  by  $\chi_n(\lambda, p_n^0, p_n^1) = (1 - \lambda)p_n^0 + \lambda p_n^1$ . Then we have that  $((\tilde{P}_n, \partial\tilde{P}_n), (P_n, \partial P_n), \chi_n)$  is a ball-bundle. Let  $\chi$  be the product map  $\chi_1 \times \chi_2$ ; then  $((\tilde{P}, \partial\tilde{P}), (P, \partial P), \chi)$  is a ball-bundle too. Let  $\hat{Y}_\varepsilon = [0, \varepsilon] \times \tilde{P}$ . Then  $\chi$  induces a map  $h_Y : \hat{Y}_1 \rightarrow Y_1$  by  $h(\varepsilon, \lambda, p^0, p^1) = (\varepsilon, \chi(\lambda, p^0, p^1))$ . Now  $((\hat{Y}_\varepsilon, \partial\hat{Y}_\varepsilon), (Y_\varepsilon, \partial Y_\varepsilon), h_Y)$  is a ball-bundle. Let  $\tilde{E}$  be the set of all  $((\varepsilon, \lambda, p^0, p^1), q) \in \hat{Y}_1 \times P$  such that  $h_E(\varepsilon, \lambda, p^0, p^1, q) \equiv (\varepsilon, f(\lambda, p^0, p^1), q) \in E$ . Then also  $((\tilde{E}_\varepsilon, \partial\tilde{E}_\varepsilon), (E_\varepsilon, \partial E_\varepsilon), h_E)$  is a ball-bundle. Moreover, letting  $\tilde{\theta}$  be the projection from  $\tilde{E}$  to  $\mathbb{P}$ , we have that  $h_Y \circ \tilde{\theta} = \theta \circ h_E$ . Therefore, by the Thom Isomorphism Theorem,  $\theta$  is essential iff  $\tilde{\theta}$  is; cf. [20, Appendix IV.3].

Let  $\hat{E}$  be the closure of the set of  $(\varepsilon, \lambda, p^0, p^1, q, \pi^1(q))$  such that  $(\varepsilon, \lambda, p^0, p^1, q) \in \tilde{E}$  and  $\lambda \neq 0$ . By the strong excision property, the natural projection  $\hat{\phi}$  from  $\hat{E}$  to  $\tilde{E}$  induces an isomorphism of their cohomology groups. Let  $\hat{\theta}$  be the projection from  $\hat{E}$  to  $\hat{Y}_\varepsilon$ . Then  $\hat{\theta} = \tilde{\theta} \circ \hat{\phi}$ . Therefore,  $\tilde{\theta}$  is essential iff  $\hat{\theta}$  is.

Let  $\eta : \hat{E} \rightarrow \mathbb{R}_+^3$  be the projection map  $\eta(\varepsilon, \lambda, (p^0, p^1), q, \pi^1) = (\varepsilon, \lambda)$ . Let  $D = \eta(\hat{E})$ . By the generic local triviality theorem, there exists a partition of  $D$  into a finite number of connected subsets  $D_1^0, \dots, D_l^0$ , and for each  $D_i$  a semi-algebraic fibre pair  $(F_i, \partial F_i)$ , a homeomorphism  $h_i : D_i^0 \times (F_i, \partial F_i) \rightarrow (\eta^{-1}(D_i^0), \eta^{-1}(D_i^0) \cap \partial \hat{E})$  such that  $\eta \circ h_i$  is the projection from  $D_i^0 \times F_i$  to  $D_i^0$ . There now exists an  $i$ , say 1, such that the closure  $D_1$  of a subset of  $D_i^0$  is homeomorphic to a 3-simplex and contains  $[0, \tilde{\varepsilon}] \times \{0\}$  for some  $\tilde{\varepsilon} < \bar{\varepsilon}$ .

Let  $(\hat{Y}(D_1), \partial \hat{Y}(D_1)) \equiv (D_1, \partial D_1) \times (\mathbb{P}, \partial \mathbb{P})$  and let  $\hat{E}(D_1)$  be the closure of the inverse image of  $D_1 \times F_1$  under  $h_1$ . Let  $\hat{\theta}(D_1) : (\hat{E}(D_1), \partial \hat{E}(D_1)) \rightarrow (\hat{Y}(D_1), \partial \hat{Y}(D_1))$  be the projection. We claim that  $(\hat{E}(D_1), \partial \hat{E}(D_1))$  is a pseudomanifold. Since  $\hat{E}$  is a pseudomanifold, this claim is proved if we show that  $\hat{E}(D_1) \setminus \partial \hat{E}(D_1)$  is path-connected. There exist  $0 < \hat{\varepsilon} < \tilde{\varepsilon}$  and integers  $r_n > 1$  for each  $n$  such that  $(\varepsilon, \lambda) \in D_1$  if  $0 < \varepsilon < \hat{\varepsilon}$  and  $0 \leq \lambda_n \leq \varepsilon^{r_n-1}$ . Now given two points  $x(0)$  and  $x(1)$  in  $\hat{E}_{\hat{\varepsilon}}(D_1) \setminus \partial \hat{E}_{\hat{\varepsilon}}(D_1)$ , connect them by a curve  $x(t) = (\varepsilon(t), \lambda(t), p^0(t), p^1(t), q(t), \pi^1(q(t)))$  in  $\hat{E}_{\hat{\varepsilon}} \setminus \partial \hat{E}_{\hat{\varepsilon}}$  as  $t$  goes from 0 to 1. For each  $t$ , express  $q(t)$  as  $\varepsilon(t)(\lambda(t)p^0(t) + (1 - \lambda(t))p^1(t)) + (1 - \varepsilon(t))r(t)$  where  $r(t)$  is a best reply to  $q(t)$ . For each  $t$  we can now find a point  $q^*(t) \in Q^*$  such that  $r(t)$  is a best reply against  $q^*(t)$ . Choose a positive  $\varepsilon$  such that  $\varepsilon + \varepsilon^{r_n} < \hat{\varepsilon}$ . For each  $n$ , modify  $x_n(t)$  to the vector

$$\tilde{x}_n(t) = (\varepsilon + \varepsilon^{r_n}\varepsilon(t), \tilde{\lambda}(t), \tilde{p}^0(t), p^1(t), \tilde{q}(t), \pi^1(t))$$

where

$$\begin{aligned} \tilde{\lambda}_n(t) &= \frac{\varepsilon^{r_n}\varepsilon(t)}{\varepsilon + \varepsilon^{r_n}}\lambda_n(t), \\ \tilde{p}_n^0(t) &= \frac{\varepsilon q^*(t) + \varepsilon^{r_n}\varepsilon(t)(1 - \lambda(t))p^0(t)}{\varepsilon + \varepsilon^{r_n}\varepsilon(t)(1 - \lambda(t))}, \end{aligned}$$

$$\tilde{q}_n(t) = (1 - \varepsilon^{r_n})q^*(t) + \varepsilon^{r_n}(1 - \varepsilon(t))r(t) + \varepsilon^{r_n}\varepsilon(t)((1 - \lambda(t))p^0(t) + \lambda(t)p^1(t)).$$

Then  $\tilde{x}(t)$  belongs to  $\hat{E}(D_1) \setminus \partial \hat{E}(D_1)$  for all  $t$ . Moreover, for  $t = 0, 1$ ,  $x(t)$  and  $\tilde{x}(t)$  can now be connected by a path  $\hat{x}(t; s)$  defined as follows. For  $s \in [0, 1]$ , let  $k_n(s) = \min(1, 2s)r_n$  and then:

$$\hat{x}_n(t; s) = ((2s - 1)^+ \varepsilon + \varepsilon^{k_n(s)}\varepsilon(t), \hat{\lambda}(t; s), p^0(t; s), p^1(t), \hat{q}(t; s), \pi^1(t))$$

where

$$\hat{\lambda}_n(t) = \frac{\varepsilon^{k_n(s)}\varepsilon(t)\lambda_n(t)}{(2s-1)^+\varepsilon + \varepsilon^{k_n(s)}\varepsilon(t)},$$

$$\tilde{p}_n^0(t; s) = \frac{(2s-1)^+\varepsilon q_n^*(t) + \varepsilon^{r_n}\varepsilon(t)(1-\lambda_n(t))p_n^0(t)}{(2s-1)^+\varepsilon + \varepsilon^{k_n(s)}\varepsilon(t)(1-\lambda_n(t))},$$

$$\tilde{q}_n(t; s) = (1 - \varepsilon^{k_n(s)})q_n^*(t) + \varepsilon^{k_n(s)}(1 - \varepsilon(t))r(t) + \varepsilon^{k_n(s)}\varepsilon(t)((1 - \lambda_n(t))p_n^0(t) + \lambda_n(t)p^1(t)).$$

Thus  $\hat{E}(D_1) \setminus \partial\hat{E}(D_1)$  is connected and hence a pseudomanifold of dimension  $d(\mathbb{P})+3$ .  $\hat{Y}(D_1)$  is a full-dimensional subset of  $Y_\varepsilon$ . Therefore  $\hat{\theta}$  is essential iff  $\hat{\theta}(D_1)$  is essential.

Since  $(\hat{E}(D_1), \partial\hat{E}(D_1))$  is a pseudomanifold,  $(F_1, \partial F_1)$  is a pseudomanifold of dimension  $d(\mathbb{P})$ . Moreover, for each  $(\varepsilon, \lambda) \in D_1^0$ , letting  $(\hat{E}_{\varepsilon, \lambda}, \partial\hat{E}_{\varepsilon, \lambda}) \equiv h_1^{-1}(\{\varepsilon, \lambda\} \times (F_1, \partial F_1))$  we have that  $\hat{\theta}$  is essential iff  $\hat{\theta}_{\varepsilon, \lambda}$ , the projection map  $(\hat{E}_{\varepsilon, \lambda}, \partial\hat{E}_{\varepsilon, \lambda}) \rightarrow (\mathbb{P}, \partial P)$ , is essential.

Let  $L$  be the set of  $(\varepsilon, \lambda) \in D_1$  such that  $\lambda_n = \varepsilon^r$  for some  $r \geq r_n$  for each  $n$ . Let  $\hat{E}(L)$  be the closure of the inverse image of  $L \setminus \{(0, 0)\}$  under  $h^1$ . Let  $\partial\hat{E}(L)$  be the inverse image of  $\partial\mathbb{P}$  under the projection map  $\theta(L)$  from  $\hat{E}(L)$ . For each  $(\varepsilon, \lambda) \in L$ ,  $(E_{\varepsilon, \lambda}, \partial E_{\varepsilon, \lambda})$  is  $(\theta(L))^{-1}(\{\varepsilon, \lambda\} \times (\mathbb{P}, \partial\mathbb{P}))$ . Our next objective is to show that  $(\mathcal{Q}, \partial\mathcal{Q})$  equals the set of  $(q^*, (p^0, p^1), \pi^1)$  such that  $(0, 0, (p^0, p^1), q^*, \pi^1)$  belongs to  $E_{0,0}(L)$ . Given  $(0, 0, (p^0, p^1), q^*, \pi^1)$  in  $\hat{E}(L)$  there exists a sequence  $(\varepsilon(k), \lambda(k), (p^0, p^1)(k), q(k), \pi^1(k))$  in  $\hat{E}(L) \setminus E_{0,0}$  converging to it. For each  $n$  and  $k$ , we can express  $q_n(k)$  as  $(1 - \varepsilon(k))((1 - \mu_n^1)q_n^0(k) + \mu_n^1 r_n^1(k)) + \varepsilon(k)(\lambda(k)p_n^0(k) + (1 - \lambda(k))p_n^1(k))$ , where  $\mu_n^1$  and  $\lambda(k)$  converge to zero,  $q_n^0(k)$  belongs to  $P_n^0$  and converges to  $q_n^*$ , and  $r_n^1(k)$  belongs to  $P_n^1$ . Also  $q_n^0(k)$  and  $r_n^1(k)$  if  $\mu_n^1 > 0$  are best replies to  $q(k)$  for all  $k$ . By going to a subsequence, the faces to which  $q_n^0(k)$  and  $r_n^1(k)$  belong are constant for all  $k$ . The sequence generates for each  $n$  an LPS  $\Lambda_n = (\bar{q}_n^0, \dots, \bar{q}_n^l)$  where  $\bar{q}_n^0 = q_n^*$ . As in the proof of Theorem 5.1, there exists a level  $l_n^i$  for each  $i = 0, 1$  that is expressible as a convex combination of  $p_n^0$  and another strategy. Since  $\lambda(k)$  converges to zero,  $l_n^0 < l_n^1$ . As in the proof of Theorem 5.1, we can prove that  $\bar{q}_n^l$  belongs to  $\bar{Q}_n^*$  for each  $l < l_n^0$  and if we express  $\bar{q}_n^0 = \nu_n^0 p_n^0 + (1 - \nu_n^0)r_n^0$ , then the strategies  $\bar{q}_n^l$  for  $l < l_n^0$  and  $r_n^0$  if  $\nu^0 < 1$  are lexicographic best replies against  $\Lambda_m$ . Likewise, if we express  $\bar{q}_n^1$  as  $\bar{\nu}_n^0 \tilde{p}_n^0 + \bar{\nu}_n^1 p_n^1 + \bar{\nu}_n^2 r_n^2$ , then the strategy  $r_n^2$  if  $\bar{\nu}_n^2 > 0$  is a lexicographic best reply against  $\Lambda_m$ . As in the proof of Theorem 5.1, Section 5.6, we can now write down an LPS  $(q^*, \bar{q}^0(\varepsilon), \bar{q}^1(\varepsilon))$  to show that  $(q^*, (p^0, p^1), \pi^1)$  belongs to  $\mathcal{Q}$ .

Given  $(q^*, (p^0, p^1), \pi^1)$  in  $\mathcal{Q}$ , it belongs to  $\mathcal{Q}(T)$  for some  $T$ . There exist  $q_n^0 = (1 - \lambda_n^0)p_n^0 + \lambda_n^0 r_n^0$  and  $q_n^1 = \mu_n^0 \tilde{p}_n^0 + \mu_n^1 p_n^1 + \mu_n^2 r_n^2$  such that the strategies in  $T_m$  are best replies against the LPS  $(q^*, q_n^0, q_n^1)$ . For all small  $\varepsilon$ , choose  $\alpha(\varepsilon)$  such that  $\varepsilon^r \mu_1 = (\varepsilon^r \mu_1 + \varepsilon^r \mu_0 + \alpha(\varepsilon)(1 - \lambda_n^0)p_n^0)^r$ ,

where  $r$  satisfies the property in the first line of the previous paragraph. Then for all small  $\varepsilon$ ,  $(\varepsilon, \lambda(\varepsilon), (p^0, p^1)(\varepsilon), q(\varepsilon))$  belongs to  $\hat{E}(L)$ , where:

$$\lambda(\varepsilon) = \frac{\varepsilon^r \mu_1}{\varepsilon^r \mu_1 + \varepsilon^r \mu_0 + \alpha(\varepsilon)(1 - \lambda_n^0)},$$

and  $(p^0, p^1)(\varepsilon) = (p^0(\varepsilon), p^1)$  with for each  $n$ ,

$$p_n^0(\varepsilon) = \frac{\varepsilon^r \mu_0 \tilde{p}_n^0 + \alpha(\varepsilon)(1 - \lambda_n^0)p_n^0}{\varepsilon^r \mu^0 + \alpha(\varepsilon)}$$

and  $\hat{q}(\varepsilon) = (1 - \alpha(\varepsilon) - \varepsilon^r)q^* + \alpha(\varepsilon)q^0 + \varepsilon^r q^1$ .

For each  $(\varepsilon, \lambda)$ ,  $(E_{\varepsilon, \lambda}, \partial E_{\varepsilon, \lambda})$  is a pseudomanifold. Indeed for  $(\varepsilon, \lambda) \neq (0, 0)$  this follows from the fact that this pair is homeomorphic to  $(F_1, \partial F_1)$ , which is a pseudomanifold; for  $(0, 0)$ , this follows from the fact that  $(E_{0,0}, \partial E_{0,0})$  is homeomorphic to  $(\mathcal{Q}, \partial \mathcal{Q})$ . The inclusion map  $(E_{\varepsilon, \lambda}, \partial E_{\varepsilon, \lambda})$  induces an isomorphism of the  $d(\mathbb{P})$ -th cohomology groups. Thus  $\hat{\theta}_{0,0}$  is essential iff  $\hat{\theta}_{\varepsilon, \lambda}$  is essential for some (and then all)  $(\varepsilon, \lambda) \in L$ . The projection map  $\hat{\theta}_{0,0}$  is just the map  $\psi$ . Hence  $Q^*$  is stable iff  $\psi$  is essential.  $\square$

In case  $S_n^1$  is empty, the construction is modified as follows. We can omit the sets  $C_n(T)$  and  $B_m(T)$  from the description of  $\mathcal{Q}(T)$ . In the last lemma above, the vector  $\lambda$  is now just a number, one for player  $m$ . The simplex  $D$  constructed there is 2-dimensional and contains a curve  $L$  of the form  $\lambda = \varepsilon^r$ . The rest of the proof is essentially the same.

This concludes the proof of the Theorem.  $\square$

#### APPENDIX D. THE GENERICITY ASSUMPTION

The conclusions of this paper necessarily hold only when, fixing the game tree, the payoffs lie in a generic set. Here we outline the nature of the genericity that is invoked. First, we require that the game have finitely equilibrium outcomes: in [4] we show that outside a lower-dimensional set every game has finitely many outcomes. Second, the constructions in Appendix C rely on certain polyhedra being in general position. Each of these polyhedra—there are finitely many of them—is a set of enabling strategies for a player  $n$  against which, in a certain class of strategies for player  $m$ , a subclass is optimal. Since these are defined by linear equations and inequalities in the payoffs of player  $m$ , the set of games where the arguments fail is a lower-dimensional set. Third, Lemma C.11 requires a characterization of stable sets that in [9] we show holds for all games outside a lower-dimensional set.

## REFERENCES

- [1] Blume, L., A. Brandenburger, and E. Dekel (1991): Lexicographic Probabilities and Equilibrium Refinements, *Econometrica*, 59, 81-98.
- [2] Govindan, S., and T. Klumpp (2002): Perfect Equilibrium and Lexicographic Beliefs, *International Journal of Game Theory*, 31, 229-243.
- [3] Govindan, S., and J.-F. Mertens (2003): An Equivalent Definition of Stable Equilibria, *International Journal of Game Theory*, 3, 339-357.
- [4] Govindan, S., and R. Wilson (2001): Direct Proofs of Generic Finiteness of Nash Equilibrium Outcomes, *Econometrica*, 69, 765-769.
- [5] Govindan, S., and R. Wilson (2002): Structure Theorems for Game Trees, *Proceedings of the National Academy of Sciences USA*, 99, 9077-9080. URL: [www.pnas.org/cgi/reprint/99/13/9077.pdf](http://www.pnas.org/cgi/reprint/99/13/9077.pdf)
- [6] Govindan, S., and R. Wilson (2002): Maximal Stable Sets of Two-Player Games, *International Journal of Game Theory*, 30, 557-566.
- [7] Govindan, S., and R. Wilson (2006): Sufficient Conditions for Stable Equilibria, *Theoretical Economics*, 1, 167-206.
- [8] Govindan, S., and R. Wilson (2008): "Nash Equilibrium, refinements of," in S. Durlauf and L. Blume (eds.), *The New Palgrave Dictionary of Economics*, 2nd Edition. New York: Palgrave Macmillan.
- [9] Govindan, S., and R. Wilson (2008): Metastable Equilibria, *Mathematics of Operations Research*, 33, 787-820.
- [10] Govindan, S., and R. Wilson (2009): On Forward Induction, *Econometrica*, 77, 1-28.
- [11] Govindan, S., and R. Wilson (2009): Axiomatic Theory of Equilibrium Selection in Games with Games with Two Players, Perfect Information, and Generic Payoffs, Research Paper 2008, Stanford Business School, Stanford CA.
- [12] Hillas, J., and E. Kohlberg (2002): Conceptual Foundations of Strategic Equilibrium, in R. Aumann and S. Hart (eds.), *Handbook of Game Theory*, III, 1597-1663. New York: Elsevier.
- [13] Kahneman, D., and A. Tversky, editors (2000): *Choices, Values, and Frames*. Cambridge, UK: Cambridge University Press.
- [14] Kohlberg, E., and J.-F. Mertens (1986): On the Strategic Stability of Equilibria, *Econometrica*, 54, 1003-1037.
- [15] Koller, D., and N. Megiddo (1992): The Complexity of Two-Person Zero-Sum Games in Extensive Form, *Games and Economic Behavior*, 4, 528-552.
- [16] Kreps, D., and R. Wilson (1982): Sequential Equilibria, *Econometrica*, 50, 863-894.
- [17] Kuhn, H. (1953): Extensive Games and the Problem of Information, in H. Kuhn and A. Tucker (eds.), *Contributions to the Theory of Games*, II, 193-216. Princeton: Princeton University Press. Reprinted in H. Kuhn (ed.), *Classics in Game Theory*, Princeton University Press, Princeton, New Jersey, 1997.
- [18] Mailath, G., L. Samuelson, and J. Swinkels (1993): Extensive Form Reasoning in Normal Form Games, *Econometrica*, 61, 273-302.
- [19] Mertens, J.-F. (1989): Stable Equilibria—A Reformulation, Part I: Definition and Basic Properties, *Mathematics of Operations Research*, 14, 575-625.
- [20] Mertens, J.-F. (1991): Stable Equilibria—A Reformulation, Part II: Discussion of the Definition and Further Results, *Mathematics of Operations Research*, 16, 694-753.
- [21] Mertens, J.-F. (1992): The Small Worlds Axiom for Stable Equilibria, *Games and Economic Behavior*, 4, 553-564.

- [22] Nash, J. (1950): Equilibrium Points in N-Person Games, Proceedings of the National Academy of Sciences USA, 36, 48-49.
- [23] Nash, J. (1951): Non-Cooperative Games, Annals of Mathematics, 54, 286-295.
- [24] Norde, H., J. Potters, H. Reijniere, D. Vermeulen (1996): Equilibrium Selection and Consistency, Games and Economic Behavior, 12, 219-225.
- [25] Reny, P. (1992): Backward Induction, Normal Form Perfection and Explicable Equilibria, Econometrica, 60, 627-649.
- [26] Reny, P. (1992): Rationality in Extensive-Form Games, Journal of Economic Perspectives, 6, 103-118.
- [27] Reny, P. (1993): Common Belief and the Theory of Games with Perfect Information, Journal of Economic Theory, 59, 257-274.
- [28] Savage, L.J. (1954): Foundations of Statistics, New York: John Wiley & Sons.
- [29] Spanier, E.H. (1966): Algebraic Topology, New York: McGraw-Hill.
- [30] van Damme, E. (1984): A Relation between Perfect Equilibria in Extensive Form Games and Proper Equilibria in Normal Form Games, International Journal of Game Theory, 13, 1-13.
- [31] van Damme, E. (2002): Strategic Equilibrium, in R. Aumann and S. Hart (eds.), Handbook of Game Theory, III, 1523-1596. New York: Elsevier.

DEPARTMENT OF ECONOMICS, UNIVERSITY OF IOWA, IOWA CITY IA 52242, USA.

*E-mail address:* srihari-govindan@uiowa.edu

STANFORD BUSINESS SCHOOL, STANFORD, CA 94305-5015, USA.

*E-mail address:* rwilson@stanford.edu