

RESEARCH PAPER NO. 2000

**AXIOMATIC THEORY OF EQUILIBRIUM SELECTION
IN SIGNALING GAMES WITH GENERIC PAYOFFS**

Srihari Govindan

Robert Wilson

September 2008

This work was partially funded by a National Science Foundation grant.

ABSTRACT

Three axioms from decision theory select sets of Nash equilibria of signaling games in extensive form with generic payoffs. The axioms require undominated strategies (admissibility), inclusion of a sequential equilibrium (backward induction), and dependence only on the game's normal form even when embedded in a larger game with redundant strategies or irrelevant players (small worlds). The axioms are satisfied by a set that is stable (Mertens, 1989) and conversely the axioms imply that each selected set is stable and thus an essential component of admissible equilibria with the same outcome.

AXIOMATIC THEORY OF EQUILIBRIUM SELECTION IN SIGNALING GAMES WITH GENERIC PAYOFFS

SRIHARI GOVINDAN AND ROBERT WILSON

ABSTRACT. Three axioms from decision theory select sets of Nash equilibria of signaling games in extensive form with generic payoffs. The axioms require undominated strategies (admissibility), inclusion of a sequential equilibrium (backward induction), and dependence only on the game's normal form even when embedded in a larger game with redundant strategies or irrelevant players (small worlds). The axioms are satisfied by a set that is stable (Mertens, 1989) and conversely the axioms imply that each selected set is stable and thus an essential component of admissible equilibria with the same outcome.

CONTENTS

1. Introduction	2
2. Notation	2
3. The Axioms	3
3.1. Axiom A: Admissibility	3
3.2. Axiom B: Backward Induction	4
3.3. Axiom S: Small Worlds	5
4. Signaling Games	9
4.1. Definitions and the Theorem	9
4.2. Preliminaries	11
4.3. The Essentiality Property	12
4.4. The Link Between the Axioms and Stability	13
5. Concluding Remarks	20
5.1. Set-Valuedness and Connectedness	20
5.2. Alternative Axiom S	21
5.3. Extensive-Form Analysis	21
Appendix A. The Pseudo-Manifold Property	22
Appendix B. Proof of Proposition 4.3	25
References	28

Date: 25 September 2008.

Key words and phrases. equilibrium, refinement, admissibility, backward induction, small worlds, essentiality, stability.

JEL subject classification: C72.

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

1. INTRODUCTION

Kohlberg and Mertens [19] argue that Nash's [27, 28] criterion of equilibrium in a non-cooperative game should be refined by applying principles from decision theory.¹ As in the single-person case, the goal is to obtain sharp predictions from axioms that specify minimal requirements for rational behavior in multi-person contexts.

We prove that three axioms adapted from decision theory imply that a refinement selects stable sets (Mertens [24]) for signaling games with generic payoffs.² The methods developed here enable a similar conclusion for general games in extensive form with perfect recall, two players, and generic payoffs that will be published in a later working paper—our intent here is to present the main ideas for the simpler case of signaling games.

Section 2 establishes notation for Section 3, which specifies Axioms A (admissibility), B (backward induction), and S (small worlds). Except for Axiom S's more general version of small worlds, these axioms are among those proposed by Kohlberg and Mertens [19]. The axioms are stated for general games in extensive form with perfect recall. Section 4 proves the main theorem for signaling games with generic payoffs. Appendices A and B establish technical properties used in the proof. Section 5 provides concluding remarks.

2. NOTATION

A typical game in extensive form is denoted Γ . Its specification includes a set N of players, a game tree that has perfect recall for each player, and an assignment of real-valued utility payoffs to all players at each terminal node of the tree. The tree can include a specified mixed strategy of Nature. We assume throughout the standard epistemic conditions that the game is common knowledge and that players' rationality is common knowledge.

Denote player n 's simplex of mixed strategies by Σ_n and interpret its vertices as his set S_n of pure strategies. A pure strategy assigns an action at each of the player's information sets. 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 a game 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 mixed strategies the players' expected payoffs from that profile. A player's pure strategy is *redundant* if its payoffs for all players are replicated by a mixture of his other pure strategies. The *reduced normal form* $G_\circ : \Sigma_\circ \rightarrow \mathbb{R}^N$ is the normal form obtained from G by deleting redundant pure strategies to obtain the reduced set S_\circ of profiles of pure strategies,

¹Also see Kohlberg [18]. Hillas and Kohlberg [17] survey subsequent developments.

²This class of generic games is studied by Banks and Sobel [1], Cho and Kreps [4], Cho and Sobel [5], Fudenberg and Tirole [8, Chap. 11], and Kreps and Sobel [21], among others.

which is unique up to labeling of pure strategies. Each pure strategy in the normal form has an equivalent representation as a pure or mixed strategy in its reduced normal form.

As defined by Nash [27, 28], 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 $BR_n(\sigma) \equiv \arg \max_{\sigma'_n \in \Sigma_n} G_n(\sigma'_n, \sigma_{-n})$ is player n 's best-reply correspondence, then $\sigma \in \Sigma$ is an equilibrium iff $\sigma_n \in BR_n(\sigma)$ for every player n . We represent an equilibrium by a profile of mixed strategies but for a game in extensive form with perfect recall there exists an equivalent representation in behavioral strategies (Kuhn [23]).

A *refinement* is a correspondence that assigns to each game a nonempty collection of nonempty connected closed subsets of its equilibria. This restriction on its range reflects the premise that equilibrium is a basic implication of common knowledge of players' rationality when there is no coordination mechanism to correlate their strategies. Each selected subset is called a *solution*. A typical solution is denoted \mathbf{S}^* .

For the axioms in Section 3 we assume straightaway that solutions are sets because there need not exist a single equilibrium that satisfies the axioms (Kohlberg and Mertens [19, pp. 1015, 1019, 1029]). The technical requirement that a solution is connected excludes the trivial refinement that always selects the set of all equilibria. Recall too that for a game with generic payoffs, all equilibria in a connected subset (and thus its closure) have the same outcome and thus the same paths of equilibrium play.³ In Section 5.1 we develop further justification for refinements that select connected sets.

3. THE AXIOMS

3.1. Axiom A: Admissibility. The first axiom requires that each player uses only admissible strategies, defined as follows.

Definition 3.1 (Admissible Strategy). A player's strategy is *admissible* if it is an optimal reply to some profile of completely mixed strategies.

That is, $\sigma_n \in \Sigma_n$ is admissible if $\sigma_n \in BR_n(\hat{\sigma})$ for some $\hat{\sigma} \in \Sigma \setminus \partial\Sigma$.

For a game with two players a strategy is admissible iff it is not weakly dominated by another strategy. Thus in this case admissibility is the same as in decision theory.

Say that a profile of players' strategies is admissible if each player's strategy is admissible.

Axiom A [Admissibility]: Each equilibrium in a solution is admissible.

³Kreps and Wilson [22, Theorem 2]. S. Elmes' corrections to the alternative proof in Kohlberg and Mertens [19, Appendix C] are noted in [12, p. 9079]. We use here the stronger characterization in [11] for which nongeneric payoffs lie in a lower dimensional subset.

We invoke admissibility as the first axiom because we anticipate the case of two players—a stronger version is required for games with more than two players. In Section 5 we remark that admissibility can be replaced by an axiom requiring that a solution is a minimal set satisfying the two other axioms.

3.2. Axiom B: Backward Induction. The second axiom invokes consistent beliefs and sequential equilibrium as defined by Kreps and Wilson [22, p. 872].

Definition 3.2 (Consistent Beliefs). A player's *belief* assigns to each of his information sets a conditional probability distribution over profiles of players' and Nature's strategies that satisfies Bayes' rule where defined. Players' beliefs are *consistent* with an equilibrium if they are limits of conditional probabilities induced by a sequence of profiles of completely mixed strategies converging to the equilibrium.

This definition appears to depart from standard decision theory because it invokes perturbed strategies, but Kohlberg and Reny [20] show that consistency of beliefs can be derived from primitive axioms appropriate for a frequency interpretation of probabilities. Kreps and Wilson interpret a belief at an information set as only the conditional distribution over its nodes induced by a belief as defined above, but we omit this unnecessary restriction. In [15] we provide motivation for the general specification used here.

Definition 3.3 (Sequential Equilibrium). An equilibrium is *sequential* if there exists a profile of consistent beliefs such that, conditional on a player's belief at an information set, the continuation of his strategy is an optimal reply to the profile of players' and Nature's strategies.

This property is called *sequential rationality*. Sequential equilibria are usually specified by behavioral strategies but we omit this aspect here.

The second axiom ensures that some equilibrium in a solution is sequentially rational:

Axiom B [Backward Induction]: Each solution contains a sequential equilibrium.

The two axioms of admissibility and backward induction account for differences in the definition of stability in Kohlberg and Mertens [19] and the definition in Mertens [24]. The former reject refinements called hyperstability and full stability that violate admissibility, and they judge inadequate their tentative definition of stability (called here *KM-stability*) because it violates backward induction. Unlike Mertens, Kohlberg and Mertens do not require that a selected set is connected. Mertens' revised definition of stability satisfies both admissibility and backward induction, as well as the third axiom that we describe next.

3.3. Axiom S: Small Worlds. The first two axioms invoke principles of rational decisions by individuals. The third axiom requires that a refinement is not affected by extraneous features. It implies two criteria called invariance and small-worlds by Kohlberg and Mertens [19] and Mertens [26], respectively. To motivate the third axiom we define these criteria separately before stating their generalization.

Invariance. As in decision theory, invariance requires that it is irrelevant whether a mixed strategy is treated as a pure strategy, i.e. it excludes dependence on a presentation effect. Because a given normal form is just its reduced normal form augmented by redundant pure strategies, invariance is defined as follows:

Definition 3.4 (Invariance). A refinement satisfies *invariance* if it depends only on each game's reduced normal form.

That is, the solutions for any two games with the same reduced normal forms have the same equivalent representations as a solution of the reduced normal form.

Invariance requires a refinement to inherit the property of equilibria that they depend only on the reduced normal form.⁴ This enforces the decision-theoretic principle that rational decisions are not affected by presentation effects, that is, not affected by which among many equivalent extensive forms represent the same essential strategic situation (Dalkey [7], Thompson [31], Kohlberg and Mertens [19]).

Say that two games in extensive form with perfect recall are *equivalent* if they have the same normal form and hence the same reduced normal form. Then backward induction and invariance together require that a solution contains an *invariant* sequential equilibrium that is sequential in every equivalent game. Examples are analyzed in detail in [14, §2.3] and [15, §2.7] to show how the sequential equilibria of an extensive-form game are refined by applying these two criteria to expanded games obtained by appending redundant strategies. More generally, [15] uses the properties of invariant sequential equilibria to prove that these criteria imply forward induction for two-player games with generic payoffs.

Small-worlds. In decision theory, the small-worlds criterion posits that rational decisions are not affected by presentations in expanded contexts with additional features irrelevant for optimal choices (Savage [29, §5.5]). Equilibria satisfy the following analog of the small-worlds criterion.

⁴Equilibria also depend only on players' best-reply correspondences, as in the decision-theoretic formulation of revealed preferences. We do not invoke this property explicitly, but it is implied by Theorem 4.1 below due to the demonstration in Govindan and Mertens [10] that the definition of stable sets can be modified to depend only on players' best-reply correspondences.

Using the normal form for simplicity, say that a game $G : \Sigma \rightarrow \mathbb{R}^N$ is *trivially embedded* in a game $\tilde{G} : \Sigma \times \Sigma^o \rightarrow \mathbb{R}^{N \cup o}$ with additional players in a set o (the ‘outsiders’) if the feasible pure strategies and their payoffs for the players in N (the ‘insiders’) remain the same as in G . That is, outsiders are *dummy* players in the game among insiders. Because the best-reply correspondences of insiders are independent of outsiders’ strategies, each equilibrium of the larger game \tilde{G} projects to an equilibrium of G . The small-worlds criterion requires that a refinement inherits this property of equilibria.

Definition 3.5 (Small-Worlds). A refinement satisfies *small-worlds* if the solutions for each game are projections of the solutions for any larger game in which the game is trivially embedded.

This is the version of small-worlds proposed by Mertens [26].

The small-worlds criterion is distinct from invariance because the latter considers a non-trivial embedding in which players have additional redundant strategies. However, the two share common features that we exploit in Definition 3.6 below. In the case of invariance one considers a game \tilde{G} with sets \tilde{S}_n of pure strategies for players $n \in N$, and its reduced normal form $\tilde{G}_o = G$ with players’ sets S_n of pure strategies such that pure strategies in $\tilde{S}_n \setminus S_n$ are payoff-equivalent (for all players) to mixed strategies in Σ_n . A mathematical statement of this situation is that for each player n there exists an affine surjective map $\tilde{f}_n : \tilde{S}_n \rightarrow \Sigma_n$ that maps pure strategies in \tilde{S}_n to payoff-equivalent mixed strategies in Σ_n and maps other mixed strategies by linear interpolation. A general statement of the converse is that there exist affine surjective maps $\tilde{f} = (\tilde{f}_n)$ that preserve payoffs, i.e. each $\tilde{G}_n = G_n \circ \tilde{f}$. Small-worlds considers the special case that \tilde{f} is the projection map from strategies in $\Sigma \times \Sigma^o$ to insiders’ strategies in Σ . The third axiom, called *small worlds* without a hyphen, considers the general case.

The small worlds axiom is more general than invariance and small-worlds, but it is motivated by similar considerations. It excludes dependence on another presentation effect when the embedding in a larger game is not trivial, and not due solely to redundant strategies. One usually interprets a player’s mixed or behavioral strategy as implemented by a private randomization. But if the game is embedded in a larger game with additional players then the player might use private observations of outsiders’ actions as a source of his randomization. But conditioning on outsiders’ actions enlarges the player’s set of pure strategies, so the embedding cannot be trivial. Nevertheless, among insiders the larger game has the same reduced normal form as the original game, so any dependence on the embedding is again a presentation effect.

Embedding. We define the general form of embedding as follows, using again a formulation in terms of the normal form.

Definition 3.6 (Embedding). A pair (\tilde{G}, f) *embeds* a game $G : \Sigma \rightarrow \mathbb{R}^N$ if $\tilde{G} : \tilde{\Sigma} \times \Sigma^o \rightarrow \mathbb{R}^{N \cup o}$ is a game with additional players in o , and maps $f = (f_n)_{n \in N}$ satisfy: (a) each $f_n : \tilde{\Sigma}_n \times \Sigma^o \rightarrow \Sigma_n$ is a multilinear map; (b) for each fixed $\sigma^o \in \Sigma^o$, $f_n(\cdot, \sigma^o)$ is surjective; and (c) $\tilde{G}_N = G \circ f$.

Hereafter we say that \tilde{G} is a *metagame* for G if there exists f such that (\tilde{G}, f) embeds G . We omit description of f for the analogous metagame in extensive form that embeds a game in extensive or normal form.

Invariance uses the special case in which G is the reduced normal form of \tilde{G} , viz., Σ^o is a single point and each f_n maps strategies in $\tilde{\Sigma}_n$ to payoff-equivalent strategies in Σ_n . The small-worlds criterion uses the trivial embedding in which $\tilde{\Sigma} = \Sigma$ and f is the projection map.

More generally, embedding allows that insiders can have more pure strategies in $\tilde{\Sigma}$ than in Σ , provided there is no net effect on their sets of strategies (conditions (a) and (b)), and no net effect on their payoffs (condition (c)).

The proof in Section 4 uses the special case mentioned above in which an insider's private observation of outsiders' actions (or additional moves of Nature in the larger game) substitutes for his private randomization. In this case, a player n obtains a larger set of strategies in $\tilde{\Sigma}_n$ by conditioning his choice among his pure strategies in Σ_n on an outsider's action, and f_n maps such a strategy in $\tilde{\Sigma}_n$ into the strategy in Σ_n that is chosen conditional on observations of outsiders' strategies. From the viewpoint of other insiders, player n 's choice of a strategy in $\tilde{\Sigma}_n$ is equivalent to the strategy in Σ_n that is its image under f .

Note that Nash equilibria are not affected by embedding in a metagame:

Proposition 3.7. *If (\tilde{G}, f) embeds a game 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 (b). Using condition (c) and the best-reply property of the equilibrium,

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

where the inequality obtains because $(\tilde{\sigma}, \sigma^o)$ is an equilibrium of \tilde{G} . Hence σ is an equilibrium of G . Conversely, suppose σ is an equilibrium of G . Define the correspondence $\phi : \tilde{\Sigma} \times \Sigma^o \rightarrow$

$\tilde{\Sigma} \times \Sigma^o$ from the strategy space of \tilde{G} to itself by letting $\phi(\tilde{\tau}, \tau^o)$ be the set of $(\tilde{\sigma}, \sigma^o)$ such that $f_n(\tilde{\sigma}_n, \tau^o) = \sigma_n$ for each insider n , and σ^o is a best reply for the outsiders to $(\tilde{\tau}, \tau^o)$. Then ϕ has nonempty values because each $f_n(\cdot, \tau^o)$ is surjective by condition (b), and values are closed and convex because f is multilinear by condition (a). The correspondence is upper-semi-continuous because the multilinearity of f implies that it is continuous. The assumptions of Kakutani's fixed point theorem are thus satisfied, so there exists a fixed point $(\tilde{\sigma}, \sigma^o)$ of ϕ . By construction, σ^o is a best reply to $(\tilde{\sigma}, \sigma^o)$. That $\tilde{\sigma}$ is also a best reply to $(\tilde{\sigma}, \sigma^o)$ is established as follows. For each insider n , $f_n(\tilde{\sigma}_n, \sigma^o) = \sigma_n$. For each insider n and $\tilde{\tau}_n \in \tilde{\Sigma}_n$, let $\tau_n = f_n(\tilde{\tau}_n, \sigma^o)$. Then

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

where the inequality obtains because σ is an equilibrium of G . Therefore $(\tilde{\sigma}, \sigma^o)$ is an equilibrium of \tilde{G} . \square

The Small Worlds Axiom. Because embedding preserves the reduced normal form among insiders, the third axiom is the following generalization of invariance and small-worlds.

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

In particular, for each solution \mathbf{S}^* of G , Axioms B and S together require that for any extensive-form game $\tilde{\Gamma}$ with perfect recall and the normal form \tilde{G} there exists a sequential equilibrium $\tilde{\sigma}$ in a solution $\tilde{\mathbf{S}}^*$ of $\tilde{\Gamma}$ and \tilde{G} such that $f(\tilde{\sigma}) \in \mathbf{S}^* = f(\tilde{\mathbf{S}}^*)$. This property is the fulcrum of the proof in Section 4.

The small worlds axiom strengthens the exclusion of presentation effects by requiring that a refinement depend only on each game's reduced normal form even when the embedding is nontrivial.⁵ Even so, Axiom S does not exclude the possibility that embedding in a metagame accounts for a 'focal point,' i.e. the context in which a game presented to the insiders might account for which solution, or which equilibrium in that solution is used, explains the outcome of the insiders' strategic interaction. Thus the refinement may be incomplete because it does not necessarily identify a unique solution nor a unique equilibrium in a solution.

⁵The proof in Section 4 requires only that the refinement selects subsets for G that are images under f of subsets selected for \tilde{G} .

To summarize: we study refinements whose solutions are subsets of equilibria that use only admissible strategies and that contain an equilibrium that is sequential, and for which solutions are independent of embedding in metagames with additional strategies and additional players.

4. SIGNALING GAMES

This section proves for signaling games with generic payoffs that Axioms A, B, S characterize a significant property of a refinement. We show that each solution is a stable set as defined by Mertens [24]. Thus, for such games any further refinement is restricted to selecting among stable sets.

The proof shows in particular that each solution is a component of admissible strategies whose neighborhood in the space of pairs of a belief and a strategy has an essential projection into the space of beliefs. There are two remarkable aspects. One is that it is necessary to select an entire component of admissible strategies. Even though admissibility and backward induction can be satisfied by a single equilibrium, the small worlds axiom implies that *all* admissible equilibria in a component must be included in a solution to account for all the metagames in which the signaling game can be embedded. The second is that stable sets are identified by properties of nearby games with perturbed strategy sets but the axioms invoke only properties of a given game and its embeddings in metagames.⁶

We begin by defining signaling games and stating the main theorem, and then subsequent subsections present its proof.

4.1. Definitions and the Theorem. A signaling game is a game in extensive-form with perfect recall, two players in $N = \{1, 2\}$, and a tree obtained from the three finite sets T, M, R . The tree has three stages:

- (1) Nature chooses player 1's type $t \in T$ with a specified probability $\pi_t > 0$.
- (2) After observing his type t , player 1 sends a message $m \in M$ to player 2.
- (3) After observing 1's message m , but not 1's type t , player 2 chooses a response $r \in R$.

Thus, player 1 has perfect information but player 2 observes only 1's message. In the normal form, player 1's set of pure strategies is $S_1 = M^T$ and 2's is $S_2 = R^M$. In the extensive form, the set of terminal nodes is $Z = T \times M \times R$. Player 1's set of behavioral strategies is the set of $T \times M$ stochastic matrices in $B_1 = \Delta(M)^T$ with typical member $\mu = (\mu_t)_{t \in T}$, and 2's set of behavioral strategies is the set of $M \times R$ stochastic matrices in $B_2 = \Delta(R)^M$ with

⁶This indicates that analyses of perturbed games can be construed as shortcuts to analyses of metagames. This view is important for the foundations of game theory because ultimately a theory of rational play must pertain to the actual game, which in Axiom S is the game among insiders.

typical member $\rho = (\rho_m)_{m \in M}$. That is, $\mu_t(m)$ is 1's probability of sending message m after observing his type t , and $\rho_m(r)$ is 2's probability of her reply r after observing message m .

The specification of a signaling game Γ is completed by assigning a payoff $u_n(t, m, r)$ to each player n at each terminal node $(t, m, r) \in Z$. The payoff assignment u is a point in $U = \mathbb{R}^{N \times Z}$. A well-known example of a signaling game is the Beer-Quiche game studied by Cho and Kreps [4, §II] and analyzed further in [15, §2.2].

We assume that payoffs are generic, that is, we specify a lower dimensional subset U_\circ of U such that the theorem is true if $u \in U \setminus U_\circ$. The conditions for genericity are specified cumulatively during the proof, using the property that the intersection of two generic classes is generic, or equivalently U_\circ is the union of lower dimensional sets of payoffs excluded in successive steps of the proof. Prominent examples of nongeneric payoffs occur in 'cheap talk' signaling games in which player 1's payoffs do not depend on his message (Crawford and Sobel [6], Chen, Kartik and Sobel [3]).

The Main Theorem. We prove the following:

Theorem 4.1. *For signaling games with generic payoffs, Axioms A, B, S imply that solutions are stable sets.*

Since admissibility is basically a minimality requirement, the gist is this. Suppose one accepts sequential rationality but prefers an alternative solution that is not stable. The theorem establishes that the game can be embedded in a larger game for which no sequential equilibrium is equivalent for insiders to any equilibrium (sequential or not) in the alternative solution. The refinement that selects this alternative solution therefore depends on how insiders are influenced by alternative embeddings, i.e. on presentation effects.

Mertens [24, 25] proves for general games that stable sets satisfy Axiom A, Axiom B, invariance, and small-worlds. A modification of his proof extends this conclusion to Axiom S. We prove a converse here: if a refinement satisfies Axioms A, B, S then there is a lower dimensional set U_\circ of payoff assignments such that if Γ is a signaling game with payoffs $u \in U \setminus U_\circ$ then each solution \mathbf{S}^* is a stable set of the equilibria of Γ .

The proof is broken into several steps, and a technical construction is consigned to Appendix A. Subsection 4.2 uses special features of signaling games to reduce the problem to minimal considerations. Subsection 4.3 defines three objects invoked in proving stability. Mertens' general definition requires essentiality of the map that is the projection from a neighborhood in the graph of equilibria to a space of perturbed games obtained from perturbed strategies. For signaling games we show that it suffices to prove essentiality of the projection from the space of player 2's pairs of beliefs and optimal strategies after observing

a message from player 1 that is not sent in any equilibrium in the solution. Subsection 4.3 proves essentiality of this projection, and also proves that a solution is an entire component of admissible strategies. The proof shows that if a solution \mathbf{S}^* is not an entire component of admissible strategies, or if for some unsent message the projection from belief-strategy pairs to beliefs is an inessential map, then there exists a metagame $(\tilde{\Gamma}, f)$ that embeds Γ for which the image $\sigma = f(\tilde{\sigma})$ of each sequential equilibrium $\tilde{\sigma}$ of $\tilde{\Gamma}$ is not in \mathbf{S}^* — thus, either Axiom B or Axiom S is violated. The final step of the proof is relegated to Appendix B, which proves that these properties imply that a solution is a stable set as defined by Mertens [24].

4.2. Preliminaries. We start with preliminaries concerning the structure of the problem.

Because Γ has perfect recall it is sufficient to represent equilibria in terms of players' behavioral strategies (Kuhn [23]). Moreover, due to the simple structure of a signaling game, to prove stability of a solution represented in normal-form strategies it is sufficient to prove the requisite properties for the game's agent-normal form or for its representation in behavioral strategies. Therefore, we assume henceforth that a solution \mathbf{S}^* is represented in behavioral strategies.

Assume initially that $U \setminus U_o$ is included in the generic set for which all equilibria in a connected set have the same outcome, i.e. the same play along paths of equilibrium play and thus the same probability distribution on terminal nodes and the same equilibrium payoffs. Because solutions are connected sets, all equilibria in a solution \mathbf{S}^* induce the same outcome. Therefore, player 1's strategy is the same, say μ^* , for all equilibria in the solution \mathbf{S}^* . For each type t , let $M_t^* = \text{supp}(\mu_t^*)$ be the set of messages sent with positive probability by 1's type t in equilibrium, and let $M^* = \cup_t M_t^*$ be the set of all messages sent by equilibria in \mathbf{S}^* . For each message $m \in M^*$, player 2's strategy after seeing m is also constant across the equilibria in \mathbf{S}^* , say ρ_m^* . Thus, the only variation among equilibria in \mathbf{S}^* is among strategies of player 2 after receiving messages in the set $M_o = M \setminus M^*$ of *unsent* messages when play adheres to an equilibrium in \mathbf{S}^* .

For each type t of player 1, let v_t^* be his equilibrium payoff, and let v^* be the vector of these payoffs. For player 2, let $B_{2o} = \prod_{m \in M_o} B_m = \{(\rho_m)_{m \in M_o} \mid \rho \in B_2\}$ where $B_m = \Delta(R)$ is her set of behavioral strategies after receiving message m . Denote the projection of \mathbf{S}^* to B_{2o} by \mathbf{S}_o^* . Also for each $m \in M_o$, let \mathbf{S}_m^* be the projection of \mathbf{S}^* to B_m . Then $\prod_{m \in M_o} \mathbf{S}_m^*$ contains \mathbf{S}_o^* .

Along an equilibrium path (i.e. after a message $m \in M^*$), player 2 can compute the unique conditional distribution over 1's types and pure strategies given the message she observed and 1's equilibrium strategy. Necessarily this posterior distribution is her belief at that information set if the equilibrium is sequential. After an unsent message, she need

not have a uniquely implied belief, even among sequential equilibria. Even so, an important discovery by Cho and Kreps [4] is that KM-stability as defined by Kohlberg and Mertens [19] imposes significant restrictions on the range of beliefs for which her strategy is optimal, and thus on the equilibria in a KM-stable set. Moreover, for a class of games that includes the ones studied here, we show in [15] that backward induction and invariance imply forward induction, which is also a restriction on beliefs at information sets off the paths of equilibrium play. Similarly, the gist of the proof here is to show that Axioms A, B, S impose restrictions on beliefs that ultimately imply stability. Thus in Subsection 4.3 below, Proposition 4.2 and its proof are cast in terms of restrictions on 2's beliefs implied by the axioms.

Thus for each unsent message $m \in M_o$ let $P_m \equiv \Delta(T)$ be the set of 2's possible beliefs about 1's types after observing m . Her belief about 1's types is the only payoff-relevant part of her belief since she observes his message and player 1 has no further moves.

4.3. The Essentiality Property. Cho and Kreps [4] establish for generic payoffs that \mathbf{S}^* contains a KM-stable set iff for each unsent message $m \in M_o$ and belief $p_m \in P_m$ there exist $\rho_m \in \mathbf{S}_m^*$, $\lambda \in (0, 1]$ and $\phi_m \in P_m$ such that ρ_m is an optimal reply for 2 when her belief is $\lambda p_m + (1 - \lambda)\phi_m$, and if $\lambda < 1$ then sending message m yields player 1 his equilibrium payoff v_t^* for every type t in the support of ϕ_m . Appendix B proves that stability as defined by Mertens [24] is implied by the stronger property that the projection from the set of all such pairs (p_m, ρ_m) of 2's beliefs and strategies—those for which there exist λ and ϕ_m satisfying the above conditions—to the set P_m of beliefs is an essential map. Therefore, the proof here proceeds by first proving that \mathbf{S}^* satisfies this essentiality property, and then proving that this essentiality property implies stability. The next three paragraphs define the three objects needed for the first part.

1. Components of admissible strategies. Neither player moves twice in a signaling game, so the admissibility required by Axiom A is equivalent to requiring conditional admissibility of 2's behavioral strategy after each unsent message. A response r after message m is conditionally admissible for player 2 iff it is an optimal reply for some belief with full support. Axiom A implies that \mathbf{S}_m^* for an unsent message $m \in M_o$ is contained in the subset of conditionally admissible strategies in $\Delta(R)$ such that no type t gets a payoff exceeding v_t^* by choosing m ; moreover, this latter set has finitely many connected components. Let \mathbf{S}_m^\bullet be the unique connected component of this set that contains \mathbf{S}_m^* and define $\mathbf{S}_o^\bullet = \prod_{m \in M_o} \mathbf{S}_m^\bullet$. Then \mathbf{S}_o^\bullet contains \mathbf{S}_o^* .

2. The graph of 2's beliefs and optimal responses. Define X_m as the set of $(p_m, \rho_m) \in P_m \times \Delta(R)$ such that there exist $\lambda \in [0, 1]$ and $\phi_m \in P_m$ such that:
 (i) ρ_m is a best reply for 2 to the belief $\lambda p_m + (1 - \lambda)\phi_m$, and

(ii) if $\lambda < 1$ then sending m yields v_t^* for each type t of 1 in the support of ϕ_m .

A further restriction on the set of generic payoffs ensures that the weight λ cannot actually be zero. Indeed, suppose ρ_m is a strategy against which 1's types in a subset $T' \subset T$ get their equilibrium payoffs by sending m . Then for a generic set of player 1's payoffs, $|T'| < |\text{supp}(\rho_m)|$. But for a generic set of 2's payoffs this implies that there does not exist a belief ϕ_m with support contained in T' against which the strategies in the support of ρ_m are equally good replies. Thus any belief for which ρ_m is a best reply must have strictly larger support than the set of types for whom sending message m is not inferior.

3. The projection map. Let q_m be the projection map from X_m to P_m . Define $\partial X_m = q_m^{-1}(\partial P_m)$, where ∂P_m is the boundary of P_m . Appendix A shows that further restrictions on the generic set of payoffs imply that $(X_m, \partial X_m)$ is a $(|T| - 1)$ -dimensional pseudo-manifold with boundary. That is, X_m is connected and can be expressed as the union of simplices of dimension $(|T| - 1)$ such that every maximal proper face of a simplex is a face of at most one other simplex, and ∂X_m is the union of those maximal proper faces of simplices that are not faces of a second simplex. Since $(X_m, \partial X_m)$ is a pseudo-manifold of the same dimension as P_m , the topological degree of the projection map q_m is well-defined. This degree, denoted d_m , is a non-negative integer that provides an algebraic count of the number of points in $q_m^{-1}(p_m)$ that is the same for every belief p_m in a generic subset of P_m . The map q_m is essential iff d_m is nonzero. This characterization of essentiality is equivalent to the following one used here: q_m is essential if every map $h_m : X_m \rightarrow P_m$ has a point of coincidence with q_m , i.e. there exists $(p_m, \rho_m) \in X_m$ such that $h_m(p_m, \rho_m) = q_m(p_m, \rho_m) \equiv p_m$.

Essentiality of q_m for each $m \in M_o$ is equivalent to essentiality of the composite map $q \equiv \times_{m \in M_o} q_m$. Appendix B shows further that this implies essentiality of the projection map from a neighborhood of \mathbf{S}^* in the graph of equilibria to the space of perturbed strategies, as used in Mertens' general definition of stability.

4.4. The Link Between the Axioms and Stability. This subsection proves the crucial link from the axioms to stability.

Proposition 4.2. $\mathbf{S}_o^\bullet = \mathbf{S}_o^*$ and q_m is essential for each $m \in M_o$.

That is, for unsent messages the solution \mathbf{S}_o^* is the entire component \mathbf{S}_o^\bullet of admissible strategies, and the projection maps from pairs of 2's beliefs and strategies conditional on unsent messages are essential. These imply that the solution \mathbf{S}^* is an entire component of admissible strategies, and the projection maps from pairs of 2's beliefs and strategies conditional on all messages are essential.

Proof of Proposition. We show that if the proposition is false then there is a metagame that embeds Γ but has no sequential equilibrium equivalent for the insiders to one in \mathbf{S}^* , which violates Axiom B or S. The proof has two long parts: the first constructs the metagame and the second shows that it cannot have the requisite sequential equilibrium.

I.1. Construction of the metagame: preliminaries.

Suppose that either $\mathbf{S}_\circ^\bullet \supsetneq \mathbf{S}_\circ^*$ or q_m is inessential for some unsent message $m \in M_\circ$. For each $m \in M$ choose a strategy $\rho_m \in \mathbf{S}_m^\bullet$ such that if $\mathbf{S}_\circ^\bullet \supsetneq \mathbf{S}_\circ^*$ then $(\rho_m)_{m \in M_\circ} \notin \mathbf{S}_\circ^*$. Since ρ_m is admissible, there exists $p_m \in P_m \setminus \partial P_m$ such that ρ_m is a conditional best reply for player 2 against the posterior p_m . Therefore, $(p_m, \rho_m) \in X_m \setminus \partial X_m$. There now exists a neighborhood A_m of (p_m, ρ_m) such that A_m is homeomorphic to a $(|T| - 1)$ -dimensional simplex, and if $d_{m'} \neq 0$ for all $m' \in M_\circ$ then the projection from $\prod_{m'} A_{m'}$ to \mathbf{S}_\circ^\bullet is disjoint from \mathbf{S}_\circ^* .

The Hopf extension theorem (Spanier [30, §8.1.18]) assures existence of a map $\tilde{f}_m : X_m \rightarrow P_m$ such that the restrictions of \tilde{f}_m and q_m to ∂X_m agree and such that: (i) if $d_m = 0$ then $\tilde{f}_m(X_m) \subseteq \partial P_m$; and (ii) if $d_m \neq 0$ then $\tilde{f}_m(X_m \setminus (A_m \setminus \partial A_m)) \subseteq \partial P_m$ and $\tilde{f}_m(A_m) = P_m$. Fix a point \tilde{p}_m in the interior of P_m and construct a map φ from P_m to P_m that has \tilde{p}_m as its only fixed point, as follows: $\varphi(\tilde{p}_m) = \tilde{p}_m$; for $p_m \in \partial P_m$, $\varphi(p_m)$ is the unique point in ∂P_m that is on the line from p_m through \tilde{p}_m ; map all other p_m by linear interpolation, i.e. $\varphi(\lambda p_m + (1 - \lambda)\tilde{p}_m) = \lambda\varphi(p_m) + (1 - \lambda)\tilde{p}_m$ for all $p_m \in \partial P_m$ and $\lambda \in [0, 1]$. By construction, \tilde{p}_m is the only fixed point of φ . Let $f_m = \varphi \circ \tilde{f}_m$. Then f_m has no point of coincidence with q_m unless q_m is essential, in which case the only points of coincidence are those in $q^{-1}(\tilde{p}_m)$ and therefore are contained in $A_m \setminus \partial A_m$. Extend each f_m to a map from $P_m \times \Delta(R)$ to P_m , denoting it still by f_m . Choose $\alpha > 0$ such that for each m and $(p_m, \rho_m) \in X_m$, $\|f_m(p_m, \rho_m) - p_m\| < \alpha$ only if $d_m \neq 0$ and $(p_m, \rho_m) \in A_m \setminus \partial A_m$.

For each $m \in M_\circ$ take a simplicial subdivision \mathcal{K}_m of P_m such that the diameter of each simplex K_m of \mathcal{K}_m is strictly smaller than $\alpha/2$. Also take a further simplicial subdivision \mathcal{L}_m of \mathcal{K}_m and one \mathcal{T}_m of $\Delta(R)$ such that f_m has a multisimplicial approximation g_m from $\mathcal{L}_m \times \mathcal{T}_m$ to \mathcal{K}_m .⁷ Suppose for $(p_m, \rho_m) \in X_m$, $g_m(p_m, \rho_m)$ and p_m belong to a simplex of \mathcal{K}_m . Then, $\|f_m(p_m, \rho_m) - p_m\| \leq \|f_m(p_m, \rho_m) - g_m(p_m, \rho_m)\| + \|g_m(p_m, \rho_m) - p_m\| < \alpha$. Thus, in this case $d_m \neq 0$ and $(p_m, \rho_m) \in A_m$. Let \mathcal{P}_m be a polyhedral subdivision of $\mathcal{L}_m \times \mathcal{T}_m$ and let \mathcal{Q}_m be the set of full-dimensional polyhedra of \mathcal{P}_m . For each $m \in M_\circ$ let V_m be the vertex set of \mathcal{K}_m . Each vertex $v_m \in V_m$ corresponds to a belief in P_m , and we denote it by p_{v_m} .

I.2. Construction of the metagame: the game tree.

⁷See [13, Appendix B] for the theory of multisimplices and multisimplicial approximation.

For each small but positive δ , embed the normal form G of Γ in a metagame \tilde{G}^δ derived from an extensive form $\tilde{\Gamma}^\delta$ specified as follows. Let S_1^* be the set of pure strategies for player 1 in Γ that choose some $m \in M^*$ for each type t . There is nothing to prove if $S_1^* = S_1$ so assume that $S_1^* \neq S_1$. The game begins with player 1 deciding whether to play a strategy in S_1^* or not. If he decides to play one of these, then next he chooses which one of these to play. Following each of these choices there is a copy of the signaling game Γ in which player 1's chosen strategy is automatically implemented. If he decides not to play a strategy in S_1^* then outsiders $o_{m,1}$, one for each $m \in M_o$, move simultaneously: $o_{m,1}$ picks a vertex v_m in the vertex set V_m of \mathcal{K}_m . Player 1 observes privately the choices made by these outsiders. Then he decides which among his pure strategies in $S_1 \setminus S_1^*$ to play, or he can choose an additional one depending on the profile of observed choices $v = (v_m)$. This pure strategy, denoted $s^{\delta,v}$, is a mixed strategy in Γ that is equivalent to the following behavioral strategy, which exists if δ is small: each type t chooses $m \in M_o$ with probability $(\delta/\pi_t)p_{v_m}(t)$ and chooses each $m \in M^* \setminus M_t^*$ with zero probability. After each of these pure strategies follows a copy of the signaling game Γ with the chosen pure strategy for player 1 implemented automatically. Player 2 moves after all these choices by player 1 and the first group of outsiders ($o_{m,1}$) are made, and knowing only the message sent by player 1. After player 2's choice of a response, if some message in M_o was chosen by player 1 then all outsiders $o_{m,2}$ and $o_{m,3}$ for $m \in M_o$ move simultaneously. The only information they have is that some message in M_o was sent. An outsider $o_{m,2}$ chooses a vertex w_m in V_m , and an outsider $o_{m,3}$ chooses a polyhedron Q_m in \mathcal{Q}_m .

Observe that player 2's information and strategies are exactly as in Γ . Player 1's strategy set includes S_1^* plus choices of other pure strategies for each possible choice of the first group of outsiders. The game $\tilde{\Gamma}^\delta$ is easily seen to embed Γ regardless of the payoffs to the outsiders, since the outsiders are dummy players in the game between players 1 and 2, and player 1 conditions his strategy only on private observations of choices by the first outsiders that are not observed by player 2.

I.3. Construction of the metagame: outsiders' payoffs.

We now describe the payoffs to the outsiders. Each outsider $o_{m,1}$ wants to mimic his counterpart $o_{m,2}$. In particular, for each pure strategy choice v_m of $o_{m,1}$ and w_m of $o_{m,2}$, define $u_{m,1}(v_m, w_m)$ to be 1 if $v_m = w_m$ and zero otherwise. For any terminal node \tilde{z} of $\tilde{\Gamma}^\delta$, if the path from the root of the tree passes through the information set of $o_{m,2}$, hence also the information set of $o_{m,1}$, then $o_{m,1}$'s payoff at \tilde{z} is $u_{m,1}(v_m, w_m)$ where v_m and w_m are the unique choices of the two players on the path to \tilde{z} ; otherwise his payoff is zero. Observe that

if player 1 eschews strategies in S_1^* then play will pass through the information set of each $o_{m,2}$.

We need more notation to define the payoffs of the other players. Given a pure strategy Q_m of $o_{m,3}$, there exists a unique multisimplex $L_m \times T_m$ of $\mathcal{L}_m \times \mathcal{T}_m$ that contains it. For each pure strategy w_m of $o_{m,2}$, and each vertex (v_m, t_m) of $L_m \times T_m$, define $u_{w_m}(Q_m, (v_m, t_m)) = 1$ if $g_m(v_m, t_m) = w_m$ and 0 otherwise. Extend u_{w_m} to a multilinear map over $L_m \times T_m$. Since $L_m \times T_m$ is full-dimensional, it extends to a multilinear map over $P_m \times \Delta(R)$ that we also call u_{w_m} .

Let γ_m be the piecewise-affine function associated with the polyhedral complex \mathcal{P}_m — see [13, Appendix C] for details. For each polyhedron Q_m in \mathcal{Q}_m , the restriction of γ_m is affine and hence has a unique extension u_{Q_m} to the whole of $P_m \times \Delta(R)$.

The payoffs to $o_{m,2}$ and $o_{m,3}$ at a terminal node \tilde{z} are defined as follows. Their payoffs from \tilde{z} are zero unless the path from the root of the tree to \tilde{z} has the following features: player 1 chooses $s^{\delta,v}$ at some node v on the path to \tilde{z} , the implementation of $s^{\delta,v}$ then has type t playing m on the path to \tilde{z} , player 2 plays some response r after seeing m , player $o_{m,2}$ plays some vertex w_m , and $o_{m,3}$ plays Q_m . In these exceptional cases, $o_{m,2}$'s payoff is $u_{w_m}(Q_m, (t, r))$ and $o_{m,3}$'s payoff is $u_{Q_m}(t, r)$, where t is the belief that assigns probability 1 to t . By the construction of $s^{\delta,v}$, if after a pure strategy choice $v = (v_m)$ of the first set of outsiders, player 1 plays $s^{\delta,v}$, player 2 chooses ρ_m after m , player $o_{m,2}$ chooses w_m , and player $o_{m,3}$ chooses Q_m , then player $o_{m,2}$'s expected payoff is $\delta u_{w_m}(Q, (p_{v_m}, \rho_m))$ and $o_{m,3}$'s is $\delta u_{Q_m}(p_{v_m}, \rho_m)$.

II. Proof that no sequential equilibrium satisfies Axioms B and S.

Recall that the metagames $\tilde{\Gamma}^\delta$ are constructed on the supposition that for some solution \mathbf{S}^* of Γ either $\mathbf{S}_\circ^* \supsetneq \mathbf{S}_\circ^*$ or q_m is inessential for some unsent message $m \in M_\circ$. For each sufficiently small $\delta > 0$, Axioms B and S require that some sequential equilibrium $\tilde{\sigma}^\delta$ of the metagame $\tilde{\Gamma}^\delta$ is equivalent for the insiders to an equilibrium of Γ in the solution \mathbf{S}^* . Also, genericity of payoffs implies that in $\tilde{\Gamma}^\delta$ the outcome of σ^δ for insiders is the same as the outcome in Γ of all equilibria in the solution \mathbf{S}^* . The following proof shows that the supposed existence of these sequential equilibria leads to a contradiction, thus verifying that the supposition is false.

For player 2, let $\tilde{\phi}^\delta$ be a consistent belief for which her strategy $\tilde{\rho}^\delta$ in the equilibrium $\tilde{\sigma}^\delta$ of the metagame $\tilde{\Gamma}^\delta$ is sequentially rational. In the metagame she observes only 1's message, so after a message in $m \in M^*$ her strategy must be ρ_m^* , and after seeing an unsent message $m \in M_\circ$, her strategy must be some $\tilde{\rho}_m^\delta \in \mathbf{S}_m^*$ and be an optimal reply to her belief $\tilde{\phi}_m^\delta \in P_m$.

Consider a sequence of such metagames $\tilde{\Gamma}^\delta$ and their equilibria $\tilde{\sigma}^\delta$ for which there is a convergent subsequence $\tilde{\sigma}^\delta \rightarrow \tilde{\sigma}^0$ as δ declines to zero. In particular, $\tilde{\rho}_m^\delta \rightarrow \tilde{\rho}_m^0$ and since \mathbf{S}^* and hence \mathbf{S}_\circ^* are closed, $(\tilde{\rho}_m^*)_{m \in M_\circ} \in \mathbf{S}_\circ^*$. For a further subsequence the support of the strategy at each information set of each player is constant, and 2's beliefs converge $\tilde{\phi}_m^\delta \rightarrow \tilde{\phi}_m^0$ for each m . For outsiders $o_{m,1}$ and $o_{m,2}$, let V_m^δ and W_m^δ be the supports of their strategies along the subsequence, let V_m^0 and W_m^0 be the supports of the limit strategies, and let $V^\delta, W^\delta, V^0, W^0$ be the corresponding product sets.

Step 1: Characterization of player 2's beliefs

For each δ in the subsequence and each $v \in V^\delta$, let $\varepsilon_{m,t}^{\delta,v}$ be the probability that player 1 at his node v sends message m for type t by choosing a strategy other than $s^{\delta,v}$, i.e. $\varepsilon_{m,t}^{\delta,v}$ is the probability at 1's information set after observing $v = (v_{m'})_{m' \in M_\circ}$ of playing a strategy $s \neq s^{\delta,v}$ that for type t sends message m . Let $\varepsilon_m^{\delta,v} = \sum_t \varepsilon_{m,t}^{\delta,v}$. For each collection $v = (v_m)$ of choices by the first outsiders denote its probability by $\tilde{\pi}^\delta(v) = \prod_{m \in M_\circ} \tilde{\sigma}_{o_{m,1}}^\delta(v_m)$.

We claim that along the subsequence, if $\varepsilon_{m,t}^{\delta,v}$ is positive for some $v \in V^\delta$ with probability $\tilde{\pi}^\delta(v) > 0$ then for 1's type t , sending message m yields the payoff v_t^* when anticipating 2's strategy $\tilde{\rho}_m^0$. Indeed, if $\varepsilon_{m,t}^{\delta,v}$ is positive then there is some strategy $s \neq s^{\delta,v}$ that for type t sends message m and that has positive probability of being chosen by 1 at his information set following v . Thus this strategy must do at least as well as the strategy $s^{\delta,v}$ along the subsequence. In the limit $s^{0,v}$ is an equilibrium strategy yielding the equilibrium payoff $\sum_{v'} \pi_{v'} v_{v'}^*$. Thus s yields this payoff in the limit. In particular this implies that for type t , sending message m yields v_t^* when 2's strategy is $\tilde{\rho}_m^0$. Suppose for some $m \in M_\circ$, and some $v \in V^\delta$, either $\tilde{\sigma}_1^\delta(s^{\delta,v})$ or $\varepsilon_{m,t}^{\delta,v}$ is positive for some t , then Bayes' rule is well-defined and so player 2's belief after observing message m is:

$$\begin{aligned} \tilde{\phi}_m^\delta(t) &= \frac{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) \delta p_{v_m}(t) + \varepsilon_{m,t}^{\delta,v}]}{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) \delta + \varepsilon_m^{\delta,v}]} \\ &= \frac{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) p_{v_m}(t) + \varepsilon_{m,t}^{\delta,v} / \delta]}{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) + \varepsilon_m^{\delta,v} / \delta]}. \end{aligned}$$

For each $m \in M_\circ$, $t \in T$ and $v \in V^\delta$, we claim that $\tilde{\pi}^\delta(v) \varepsilon_{m,t}^{\delta,v} / \delta \not\rightarrow \infty$ as $\delta \downarrow 0$. Indeed, were this not true for some message m , then the above formula for $\tilde{\phi}_m^\delta(t)$ is valid and the limiting belief after m is given by:

$$\begin{aligned}\tilde{\phi}_m^0(t) &= \lim_{\delta \downarrow 0} \frac{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) p_{v_m}(t) + \varepsilon_{m,t}^{\delta,v} / \delta]}{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) + \varepsilon_m^{\delta,v} / \delta]} \\ &= \lim_{\delta \downarrow 0} \frac{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\varepsilon_{m,t}^{\delta,v} / \delta]}{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\varepsilon_m^{\delta,v} / \delta]}.\end{aligned}$$

As we argued above, for each $v \in V^\delta$ and each t such that $\varepsilon_{m,t}^{\delta,v}$ is positive along the sequence, message m for type v is optimal against $\tilde{\rho}_m^0$. Thus $\tilde{\phi}_m^0$, for which $\tilde{\rho}_m^0$ is an optimal reply, gives positive probability only to types t' for whom sending message m yields their equilibrium payoffs v_t^* against $\tilde{\rho}_m^0$. But we proved previously in Subsection 4.3.2 that this is impossible for the assumed set of generic payoffs. Hence, for a further subsequence, $\lim_{\delta \downarrow 0} \tilde{\pi}^\delta(v) \varepsilon_{m,t}^{\delta,v} / \delta$ exists and is finite as claimed. Thus define $\eta_{m,t}^v \equiv \lim_{\delta \downarrow 0} \tilde{\pi}^\delta(v) \varepsilon_{m,t}^{\delta,v} / \delta$ and $\eta_m^v \equiv \sum_t \eta_{m,t}^v$.

An implication of the above conclusion is that in the limit, at each vertex $v \in V^0$ the probability of player 1 choosing $s^{0,v}$ is $\lim_{\delta \downarrow 0} \tilde{\sigma}_1^\delta(s^{\delta,v}) = 1$. Therefore, Bayes rule can be used to compute $\tilde{\phi}_m^\delta$ for all sufficiently small δ and we obtain

$$\begin{aligned}\tilde{\phi}_m^0(t) &= \lim_{\delta \downarrow 0} \frac{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) p_{v_m}(t) + \varepsilon_{m,t}^{\delta,v} / \delta]}{\sum_{v \in V^\delta} \tilde{\pi}^\delta(v) [\tilde{\sigma}_1^\delta(s^{\delta,v}) + \varepsilon_m^{\delta,v} / \delta]} \\ &= \frac{\sum_{v \in V^0} \tilde{\pi}^0(v) p_{v_m}(t) + \sum_{v \in V^\delta} \eta_{m,t}^v}{1 + \sum_{v \in V^\delta} \eta_m^v},\end{aligned}$$

where $V^\delta = V^0$ for δ sufficiently small. If $\eta_{m,t}^v > 0$ for some v and t then it is optimal for type t to send m when anticipating $\tilde{\rho}_m^0$. Hence the above formula implies that $\tilde{\phi}_m^0$ is an average of

$$\tilde{p}_m^0 \equiv \sum_{v \in V^0} \tilde{\pi}^0(v) p_{v_m}$$

and a conditional distribution $\tilde{\varphi}_m^0$ in P_m derived from $(\eta_{m,t}^v)$ whose support is within the nonempty set of types t who get payoff v_t^* by sending message m . Therefore $(\tilde{p}_m^0, \tilde{\rho}_m^0) \in X_m$.

As argued above, in the limit player 1 plays $s^{0,v}$ with probability 1 after $v \in V^0$. So along the subsequence he plays this strategy with positive probability at each such v . Therefore, all along the subsequence 2's belief

$$\tilde{p}_m^\delta \equiv \sum_{v \in V^\delta} \left[\frac{\tilde{\pi}^\delta(v) \tilde{\sigma}_1^\delta(s^{\delta,v})}{\sum_{v'} \tilde{\pi}^\delta(v') \tilde{\sigma}_1^\delta(s^{\delta,v'})} \right] p_{v_m}$$

is well-defined.

Step 2: Characterization of outsiders' strategies

Next we establish implications of sequential rationality for the outsiders. We claim that for each unsent message $m \in M_o$:

- (1) Suppose that the support W_m^δ of $\tilde{\sigma}_{o_{m,2}}^\delta$ is a simplex of \mathcal{K} . Then the support V_m^δ of $\tilde{\sigma}_{o_{m,1}}$ is face of this simplex.
- (2) Suppose that every polyhedron Q_m in the support of $o_{m,3}$'s strategy $\tilde{\sigma}_{o_{m,3}}^\delta$ contains $(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$, where according to $\tilde{\sigma}^\delta$, $\tilde{\rho}_m^\delta$ is the response by player 2 after seeing m . Then w_M^δ is a face of the simplex that contains $g_m(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$ in its interior.
- (3) The support of every polyhedron Q_m in the support of $o_{m,3}$'s strategy contains $(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$.

When player $o_{m,1}$ chooses his action he knows that play will pass through $o_{m,2}$'s information set. Since he wants to mimic $o_{m,2}$, sequential rationality at his information set implies (1).

Let $L'_m \times T'_m$ be the multisimplex that contains $(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$ in its interior, and let K^δ be the simplex that contains $g_m(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$ in its interior. Fix Q_m in the support of $o_{m,3}$'s strategy. Conditional on player 1 avoiding strategies in S_1^* , player $o_{m,2}$'s payoff when he chooses a vertex w_m , $o_{m,3}$ plays Q_m , and the others play according to $\tilde{\sigma}^\delta$ is, by construction:

$$\sum_v \tilde{\pi}^\delta(v) \tilde{\sigma}_1^\delta(s^{\delta,v}) [\delta u_{w_m}(Q_m, p_{v_m}, \tilde{\rho}_m^\delta)] = \sum_v \tilde{\pi}^\delta(v) \tilde{\sigma}_1^\delta(s^{\delta,v}) [\delta u_{w_m}(Q_m, \tilde{p}_m^\delta, \tilde{\rho}_m^\delta)].$$

Let $L_m \times T_m$ be the multisimplex that contains Q_m . Then $L'_m \times T'_m$ is a face of $L_m \times T_m$. And, the above payoff to w_m is positive iff it is the image of one of the vertices of $L'_m \times T'_m$ under g_m . Clearly, these vertices w_m of \mathcal{K}_m are the vertices of the simplex K^δ that contains $g_m(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$ in its interior. Since Q_m was an arbitrary element in the support of $o_{m,3}$'s strategy, the best replies indeed form a face of this simplex, which verifies (2).

As with player $o_{m,2}$, when player $o_{m,3}$ plays a polyhedron Q_m , his payoff is

$$\begin{aligned} \sum_v \tilde{\pi}^\delta(v) \tilde{\sigma}_1^\delta(s^{\delta,v}) \delta u_{Q_m}(p_{v_m}, \tilde{\rho}_m^\delta) &= \sum_v \tilde{\pi}^\delta(v) \tilde{\sigma}_1^\delta(s^{\delta,v}) \delta u_{Q_m}(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta) \\ &\leq \sum_v \tilde{\pi}^\delta(v) \tilde{\sigma}_1^\delta(s^{\delta,v}) \delta \gamma_m(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta). \end{aligned}$$

From the properties of γ_m , the last inequality is an equality iff Q_m contains $(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$, which verifies (3).

Thus the support of $o_{m,3}$'s strategy is contained in the set of polyhedra that contain the point $(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$. Hence the support W^δ of $o_{m,2}$'s strategy is a face of a simplex K_m^δ that contains $g_m(\tilde{p}_m^\delta, \tilde{\rho}_m^\delta)$ in its interior. And, the vertices in V^δ span a face of K_m^δ . Hence \tilde{p}_m^δ ,

which is an average of p_{v_m} for $v_m \in V_m^\delta$, belongs to K_m^δ . Invoking a further subsequence if necessary, we can assume that this simplex K_m^δ is the same for all $\delta > 0$, say K_m .

Step 3: Establishing the contradiction

At the limit, \tilde{p}_m^0 and $g_m(\tilde{p}_m^0, \tilde{\rho}_m^0)$ both belong to K_m . Since $(\tilde{p}_m^0, \tilde{\rho}_m^0)$ belongs to X_m , by construction this is impossible if the degree of the map q_m is $d_m = 0$. If the degree is $d_m \neq 0$ for all $m \in M_o$ then $(\tilde{p}_m^0, \tilde{\rho}_m^0)$ belongs to A_m for all such m . This too is impossible since then $(\tilde{\rho}_m^0)_{m \in M_o} \notin \mathbf{S}_o^*$. Therefore, for all small $\delta > 0$ in a subsequence there is no sequential equilibrium $\tilde{\sigma}^\delta$ of $\tilde{\Gamma}^\delta$ that is equivalent for the insiders to an equilibrium in \mathbf{S}^* . This contradicts Axiom B and/or Axiom S. \square

Finally, Proposition 4.2 implies:

Proposition 4.3. \mathbf{S}^* is a stable set of Γ .

The proof of Proposition 4.3 is provided in Appendix B.

This concludes the proof of Theorem 4.1.

5. CONCLUDING REMARKS

5.1. Set-Valuedness and Connectedness. We allow a refinement to select sets of equilibria as solutions, but restrict solutions to be connected and closed. Here we provide a justification for these conditions in terms of what we learn from signaling games.

Let \mathbf{S}^* be a set of strategy profiles in a generic signaling game such that for every embedding there exists a sequential equilibrium whose image is contained in \mathbf{S}^* . We are not requiring here that \mathbf{S}^* consist only of equilibria. The techniques of Section 4 show that \mathbf{S}^* contains a component of admissible equilibria that is stable and hence satisfies our axioms. To see this, observe first that the set of equilibria of a generic signaling game has finitely many components of admissible equilibria and all equilibria in each component induce the same outcome. For each component \mathbf{S} of admissible equilibria, and for each message m that is unsent in the equilibria of \mathbf{S} , one can construct the set X_m as in Section 4. And \mathbf{S} is stable iff the projection from X_m to $\Delta(T)$ is essential for each unsent message m . Now if \mathbf{S}^* does not contain a component of admissible equilibria that is stable, then for each \mathbf{S} that is stable we can choose a point $\rho(\mathbf{S})$ such that the projection of \mathbf{S}^* to player 2's strategies does not contain $\rho(\mathbf{S})$. This yields an embedding (\tilde{G}, f) such that the f -image of every sequential equilibrium of the larger game has player 2 playing a strategy that is arbitrarily close to some $\rho(\mathbf{S})$ for some stable set \mathbf{S} .

Thus Axioms B and S require that the solution contains a component of admissible equilibria that is stable, and the proof also shows that all stable sets must be components of

admissible equilibria. Axioms B and S and a minimality axiom therefore imply that solutions must be connected sets of admissible equilibria. That they must be set-valued follows trivially since the only possible component of admissible equilibria that is a singleton is one in which *all* messages are sent in equilibrium. We do not develop this alternative axiomatization here because, as an axiom, minimality lacks a decision-theoretic justification. But this reasoning shows that imposing connectedness makes solutions as ‘minimal’ as possible.

5.2. Alternative Axiom S. Axiom S has two parts. One part requires that if (\tilde{G}, f) embeds G then the f -image of a solution of (\tilde{G}, f) is a solution of G . The second part requires each solution of G to be the f -image of a solution of (\tilde{G}, f) . In [16] we study a refinement called metastability that implies a weaker version, say Axiom SW, in which the second part is: each solution of G *contains* the image of a solution of (\tilde{G}, f) . Metastability invokes perturbations of players’ best-reply correspondences rather than perturbations of their strategies; alternatively, it can be obtained using homotopic essentiality rather than homological essentiality.

A possible advantage of metastability is that its formulation adheres to a revealed preference perspective. In addition, Axiom SW and metastability allow embeddings to refine solutions, i.e. the solution of a general game contains many subsets that are projections of solutions of metagames that embed it. This is one case in which presentations of a game as embedded in metagames justifies further refinements of solutions via ‘focal points’ induced by particular classes of embeddings. However, Theorem 4.1 is valid with Axiom SW and metastability replacing Axiom S and stability because, as shown in [16, Appendix E], metastability coincides with stability for extensive-form games with generic payoffs. Hence if payoffs are generic then no further refinement is possible.

5.3. Extensive-Form Analysis. Axiom S ensures that a refinement depends only on the reduced normal form, yet the proof of Theorem 4.1 exploits a signaling game’s extensive form. A resulting advantage is that it reveals how stability translates into restrictions on beliefs at information sets that are off paths of equilibrium play. Moreover, the extensive-form analysis translates readily into normal-form analysis of equilibria represented as lexicographic probability systems (Blume, Brandenberger, and Dekel [2]). That is, player 2’s strategy after a message that is not sent in equilibrium must be an optimal reply to her belief induced by the second strategy in the LPS representation of 1’s behavior, which in effect is what is characterized in the proof here.

A similar method of proof suffices in the case of general games in extensive form with perfect recall, two players, and generic payoffs, as will be reported in a later paper. There

too the requirements for stability translate into restrictions on the beliefs of one player after observing an initial deviation from the equilibrium by the other player. The proof is more complicated only because one must account for the possibility of subsequent deviations.

For general games with more than two players it seems clear that, besides strengthening Axiom A to require that a solution contains only perfect equilibria, it will be necessary in the proof to use LPS representations, and in particular to exploit the special form of LPS representations in Govindan and Klumpp [9] to ensure explicitly that correlations among players' strategies are excluded, which is only implicit in [2].

APPENDIX A. THE PSEUDO-MANIFOLD PROPERTY

This appendix establishes that $(X_m, \partial X_m)$ is a $(|T| - 1)$ -dimensional pseudo-manifold with boundary. It uses the notation and definitions in the text. In addition, recall that $B_m = \Delta(R)$, which is 2's set of behavioral strategies after observing message m , and let $B_\circ = \prod_{m \in M_\circ} B_m$.

Fix an unsent message $m \in M_\circ$. We first study the nature of strategic interaction when some type t sends m and player 2 responds. Recall that $\rho_m \in B_m$ is admissible iff there exists a belief $p_m \in P_m$ with full support against which ρ_m is optimal. Therefore, the set of admissible responses following m is expressible as a union of faces of the simplex B_m . Since each face of B_m is described by the set of responses that have zero probability, there exists a collection \mathcal{R}_m of subsets of R such that the set of admissible best replies is the union over $R_0 \in \mathcal{R}_m$ of $\Delta(R \setminus R_0)$.

Let \bar{B}_m be the polyhedron that is obtained as the set of $\rho_m \in \mathbb{R}^R$ that satisfy the following system of inequalities and equation:

$$\begin{aligned} u_1(t, m, \rho_m) &\leq v_t^* & \forall t \in T \\ \rho_{m,r} &\geq 0 & \forall r \in R \\ \sum_{r \in R} \rho_{m,r} &= 1 \end{aligned}$$

These are the responses following m that support the equilibrium outcome corresponding to \mathbf{S}^* . Moreover, for every face there exist unique subsets T_1 of T and R_0 of R such that the types in T_1 are the ones who get a payoff of v_t^* from playing m against any point in this face and the responses in R_0 are the ones used with zero probability in every point in this face. We write such a face as $G_m(T_1, R_0)$. Observe that by our genericity assumption, the dimension of $G_m(T_1, R_0)$ is $|R| - |T_1| - |R_0| - 1$.

Now the set of admissible strategies in \bar{B}_m is the union over faces $G_m(T_1, R_0)$ of \bar{B}_m such that $R_0 \in \mathcal{R}_m$. Let \mathbf{S}_m^\bullet be the connected component of admissible strategies in \bar{B}_m that

contains \mathbf{S}_m^* . \mathbf{S}_m^\bullet is a union of faces of \bar{B}_m . Let \mathcal{P} be the collection of all pairs (T_1, R_0) such that $G_m(T_1, R_0)$ is contained in \mathbf{S}_m^\bullet .

Fix now a face $G_m(T_1, R_0)$ in \bar{B}_m that is nonempty. Let $F_m(T_1, R_0)$ be the set of $p_m \in P_m$ such that there exists $\lambda \in [0, 1]$ and $\phi_m \in \Delta(T_1)$ such that the responses in $R \setminus R_0$ are all optimal replies against $\lambda p_m + (1 - \lambda)\phi_m$.

Claim A.1. *$F_m(T_1, R_0)$ is a nonempty polyhedron of dimension $|T_1| + |T| - |R| + |R_0|$. Moreover, for each point $p_m \in F_m(T_1, R_0)$ there exists a unique $\lambda_m \in (0, 1]$ and a unique $\phi_m \in \Delta(T_1)$ if $\lambda_m < 1$ such that the responses in $R \setminus R_0$ are all best replies against $\lambda_m p_m + (1 - \lambda_m)\phi_m$. Moreover λ_m is a continuous function of p_m .*

Proof of Claim. Since $G_m(T_1, R_0)$ is nonempty and has dimension $|R| - |T_1| - |R_0| - 1$, $|R \setminus R_0| > |T_1|$. By genericity of player 2's payoffs, there does not exist a belief in $\Delta(T_1)$ against which the strategies in $R \setminus R_0$ are all equally good replies. Therefore, if these strategies are all best replies against a strategy of the form $\lambda_m p_m + (1 - \lambda_m)\phi_m$ where $\phi_m \in \Delta(T_1)$, then $\lambda_m > 0$. Thus, $G_m(T_1, R_0)$ is the set of $p_m \in P_m$ such that there exist $\eta \in \mathbb{R}^{T_1}$ and $v_m \in \mathbb{R}$ such that:

$$\begin{aligned} \sum_{t \in T} u_2(t, m, r)p_m(t) + \sum_{t \in T_1} u_2(t, m, r)\eta_t - v_m &= 0 \quad \forall r \in R \setminus R_0 \\ \sum_{t \in T} u_2(t, m, r)p_m(t) + \sum_{t \in T_1} u_2(t, m, r)\eta_t - v_m &\leq 0 \quad \forall r \in R_0 \\ \eta &\geq 0 \\ p_m &\geq 0 \\ \sum_{t \in T} p_m(t) &= 1 \end{aligned}$$

By definition there exists a belief p_m with full support against which the strategies in $\Delta(R \setminus R_0)$ are all best replies. By genericity of player 2's payoffs, therefore, the set of such p_m against which these strategies are the only best replies is a nonempty set of dimension $|T| - |R \setminus R_0|$. Since these solutions correspond to points of the form $(p_m, 0, v_m)$ that solve the above system, the coefficient matrix for the equations in the system has full row-rank. Moreover, by genericity, if there exists a solution to the system then there exists one where all the inequalities are strict. Hence, the set of solutions to this system has dimension $|T| + |T_1| - |R \setminus R_0|$. The projection of the solutions to P_m is $F_m(T_1, R_0)$.

Next we show that for each $p_m \in F_m(T_1, R_0)$ there exists a unique (η, v_m) that solves the system. Fix $p_m \in F_m(T_1, R_0)$ for which there exists (η, v_m) that solves the system. Consider the following subsystem, which has $|R \setminus R_0|$ equations in $|T_1| + 1$ variables.

$$\begin{aligned} \sum_{t \in T_1} u_2(t, m, r)\eta_t - v_m &= -\sum_{t \in T} u_2(t, m, r)p_m(t) \quad \forall r \in R \setminus R_0 \\ \sum_{t \in T_1} u_2(t, m, r)\eta_t - v_m &\leq -\sum_{t \in T} u_2(t, m, r)p_m(t) \quad \forall r \in R_0 \\ \eta &\geq 0 \end{aligned}$$

The result is proved if we show that the matrix of coefficients for equations in the system has row-rank $|T_1| + 1$. This last point follows from the fact that if the row-rank were less then there would exist a point (η', v'_m) that solves the following system:

$$\begin{aligned} \sum_{t \in T_1} u_2(t, m, r) \eta_t - v_m &= 0 \quad \forall r \in R \setminus R_0 \\ \sum_t \eta_t &= 1 \end{aligned}$$

which is impossible by our genericity assumption that implies $|R \setminus R_0| > |T_1|$. Hence, for each $p_m \in F_m(T_1, R_0)$ there exists a unique (η, v_m) that solves this system. The fact that λ_m is continuous now follows, since for each $p_m \in F_m(T_1, R_0)$, if (p, η, v_m) solves the above system, then $\lambda_m = (1 + \sum_t \eta_t)^{-1}$. \square

The set X_m defined in the text is now the union over all faces $F_m(T_1, R_0)$ in \mathbf{S}_m^\bullet of the product $H_m(T_1, R_0) \equiv F_m(T_1, R_0) \times G_m(T_1, R_0)$. The following states and proves the pseudo-manifold property invoked in the text in subsection 4.3.

Claim A.2. $(X_m, \partial X_m)$ is a $(|T| - 1)$ -dimensional pseudo-manifold with boundary $\partial X_m \equiv \{(p_m, \rho_m) \in X_m \mid p_m \in \partial P_m\}$.

Proof of Claim. Since \mathbf{S}_m^\bullet is connected, X_m is connected too. X_m is the union of the $(|T| - 1)$ -dimensional polyhedra given by $H_m(T_1, R_0)$ for $(T_1, R_0) \in \mathcal{P}$. Hence to finish the proof it is sufficient to show that every maximal proper face of one of these polyhedra $H_m(T_1, R_0)$ is a face of at most one other $H_m(T'_1, R'_0)$ and that it is not a face of another polyhedron iff for every point (p_m, ρ_m) in this face, $p_m \in \partial P_m$. Consider now a polyhedron H'_m of dimension $|T| - 2$ that is a face of some $H_m(T_1, R_0)$. It is of the form $F_m(T_1, R_0) \times G'_m$ for a maximal face G'_m of $G_m(T_1, R_0)$ or it is of the form $F'_m \times G(T_1, R_0)$ for a maximal face F'_m of $F_m(T_1, R_0)$. Suppose it is the former. Then exactly one of the inequalities defining $G_m(T_1, R_0)$ holds with equality on G'_m . There are two possibilities. (1) $\rho_{m,r} = 0$ for some $r \notin R_0$. In this case, $(T_1, R \cup \{r\})$ belongs to \mathcal{P} and $F_m(T_1, R_0) \times G'_m$ is also a face of the polyhedron $H_m(T_1, R_0 \cup \{r\})$. (2) The payoff to some $t \notin T_1$ is v_t^* for all points on this face F'_m . In this case, $(T \cup \{t\}, R_0) \in \mathcal{P}$ and $F_m(T_1, R_0) \times G'_m$ is also a face of $H_m(T_1 \cup \{t\}, R_0)$. Consider now the case where H'_m is of the form $F'_m \times G_m(T_1, R_0)$. There are now three possibilities. (1) One of the inequalities for $r \in R_0$ holds with equality in the linear system constructed in the proof of the previous claim. Since in the interior of F'_m the other inequalities are strict, p_m in the interior of F'_m have full support, i.e. $R_0 \setminus \{r\} \in \mathcal{R}$, $(T_1, R_0 \setminus \{r\}) \in \mathcal{P}$ and $F_m(T_1, R_0) \times G'_m$ is a face of $H_m(T_1, R_0 \setminus \{r\})$. (2) One of the inequalities $\eta_t \geq 0$ is zero for $t \in T_1$; in this case H' is a face of $H_m(T_1 \setminus \{t\}, R_0)$. (3) One of the $p_m(t) \geq 0$ is zero for some $t \in T$; in this case H'_m is not a face of another full-dimensional polyhedron and $p_m \in \partial P_m$ for each (p_m, ρ_m) on this face. \square

APPENDIX B. PROOF OF PROPOSITION 4.3

This appendix proves Proposition 4.3 in the text.

Proposition B.1. \mathbf{S}^* is stable if $\mathbf{S}^\circ = \mathbf{S}_\circ^*$ and the projection $q_m : (X_m, \partial X_m) \rightarrow (P_m, \partial P_m)$ is essential for each $m \in M_\circ$.

In outline, the proof constructs (1) an essential map f from an ε -neighborhood in the graph $(X, \partial X)$ over the space $(P, \partial P)$ of beliefs to a corresponding neighborhood $(Y, \partial Y)$ of \mathbf{S}^* in the graph E of equilibria over the space $(A, \partial A)$ of perturbed strategies, and (2) a locally essential map g from the projection $(Q, \partial Q) \subset (P, \partial P)$ of the neighborhood in $(X_m, \partial X_m)$ to the corresponding neighborhood in $(A, \partial A)$, such that in the commutative diagram

$$\begin{array}{ccc} (X, \partial X) & \xrightarrow{f} & (Y, \partial Y) \subset E \\ \hat{p} \downarrow & & p \downarrow \\ (Q, \partial Q) & \xrightarrow{g} & (A, \partial A) \end{array}$$

the assumed essentiality of the projection map \hat{p} implies the essentiality of the projection map p that is required to establish that \mathbf{S}^* is stable.

Proof. First we define the space $(A, \partial A)$ of perturbed strategies. For each t , let $A_t = \{ \delta \mu_t \mid 0 \leq \delta \leq 1, \mu_t \in \Delta(M) \}$. A typical element of A_t is denoted η_t , and $\bar{\eta}_t \equiv \sum_m \eta_{t,m}$. Likewise for each m , let $A_m = \{ \delta \rho_m \mid 0 \leq \delta \leq 1, \rho_m \in \Delta(R) \}$. A typical element of A_m is denoted θ_m , and $\bar{\theta}_m \equiv \sum_{r \in R} \theta_{m,r}$. Let $A_1 = \prod_t A_t$, $A_2 = \prod_m A_m$, and $A = A_1 \times A_2$. For each $\eta \in A_1$ let $\zeta(\eta) = (\zeta_m(\eta))_{m \in M}$ where $\zeta_m(\eta) = \sum_{t \in T} \eta_{t,m}$. For each ε , let $A_1^\varepsilon = \{ \eta \mid \zeta_m(\eta) \leq \varepsilon \forall m \}$, $A_2^\varepsilon = \{ \theta \mid \bar{\theta}_m \leq \varepsilon \forall m \}$ and $A^\varepsilon = A_1^\varepsilon \times A_2^\varepsilon$. Denote their topological boundaries by ∂A_1^ε , ∂A_2^ε and ∂A^ε .

For each $(\eta, \theta) \in A$ define the perturbed game $\Gamma(\eta, \theta)$ in which the payoff to player t when he sends message m and player 2 responds with ρ_m is the payoff he would get in Γ if player 2 responded with $(1 - \bar{\theta}_m)\rho_m + \theta_m$, and similarly for player 2, if she plays response r after message m and player 1 chose μ , her payoff is the same as in Γ if each type t chose $(1 - \bar{\eta}_t)\mu_t + \eta_t$. Note that we are not allowing a player's perturbation to affect his own payoff. If (μ, ρ) is an equilibrium of $\Gamma(\eta, \theta)$ then say that (μ', ρ') is a perturbed equilibrium, where $\mu'_t = (1 - \bar{\eta}_t)\mu_t + \eta_t$ and $\rho'_m = (1 - \bar{\theta}_m)\rho_m + \theta_m$ for all t and m . Let E be the graph of the perturbed equilibrium correspondence over A , i.e. E is the set of $((\eta, \theta), (\mu, \rho))$ such that (μ, ρ) is a perturbed equilibrium of $\Gamma(\eta, \theta)$. Let p be the natural projection from E to A . For each ε and each subset $Y \subset E$, let $(Y^\varepsilon, \partial Y^\varepsilon) = Y \cap p^{-1}(A^\varepsilon, \partial A^\varepsilon)$. Because the A^ε 's form a basis of polyhedral neighborhoods of $0 \in A$, to show that \mathbf{S}^* is stable it is sufficient to show

that there exists a subset Y of E such that $\mathbf{S}^* = \{(\mu, \rho) \mid (0, \mu, \rho) \in Y\}$ and (i) Connexity: $(Y^\varepsilon \setminus \partial Y^1)$ is connected and dense in Y^ε for all small ε ; (ii) Essentiality: for some small ε the projection map from $(Y^\varepsilon, \partial Y^\varepsilon) \rightarrow (A^\varepsilon, \partial A^\varepsilon)$ is essential in Čech cohomology with integer coefficients. This is the requirement for $*$ -stability in Mertens [24].

For each $m \in M^*$, let R_m^* be the support of ρ_m^* , player 2's equilibrium response after m . For a fixed perturbation vector θ for player 2, consider the following system $L_1(\theta)$ of equations in the variables $\rho_m \in \Delta_m$ for $m \in M^*$, and $v_t \in \mathbb{R}^T$:

$$\begin{aligned} u_1(t, m, \rho_m) \rho_{m,r} - v_t &= 0 & \forall t \in T, m \in M_t^* \\ u_1(t, m, \rho_m) \rho_{m,r} - v_t &< 0 & \forall t \in T, m \in M^* \setminus M_t^* \\ \rho_{m,r} &= \theta_{m,r} & \forall m \in M^*, r \notin R_m^* \\ \rho_{m,r} &> \theta_{m,r} & \forall m \in M^*, r \in R_m^* \\ \sum_r \rho_{m,r} &= 1 & \forall m \in M^*. \end{aligned}$$

When $\theta = 0$ this system has a unique solution that is the equilibrium strategies ρ_m^* for $m \in M^*$ and the equilibrium payoff vector v^* . By genericity of player 1's payoffs, for each $m \in M^* \setminus M_t^*$, type t 's payoff against ρ_m is strictly less than v_t^* . Therefore, it still has a unique solution $(\rho^*(\theta), v^*(\theta))$ for all θ close to zero.

For a fixed vector $\delta = (\delta_t)_{t \in T}$ of positive numbers, and a perturbation $\eta \in A_1$, consider the following system $L_2(\eta, \delta)$ of equations in the variables $(\mu_{t,m})_{t \in T, m \in M^*}$ and $(v_m)_{m \in M^*}$:

$$\begin{aligned} \sum_t \pi_t u_2(t, m, r) \mu_{t,m} - v_m &= 0 & \forall m \in M^*, r \in R_m^* \\ \sum_t \pi_t u_2(t, m, r) \mu_{t,m} - v_m &< 0 & \forall m \in M^*, r \notin R_m^* \\ \mu_{t,m} &= \eta_{t,m} & \forall t \in T, m \in M^* \setminus M_t^* \\ \mu_{t,m} &> \eta_{t,m} & \forall t \in T, m \in M_t^* \\ \sum_{m \in M_t^*} \mu_{t,m} &= 1 - \delta_t & \forall t \in T. \end{aligned}$$

When the η_t 's and δ_t 's are zero, this system has a unique solution: equilibrium strategies for player 1 as well as the unconditional payoffs for player 2 after seeing message $m \in M^*$. Since this solution is unique, it continues to have a unique solution when δ and η are close to zero. The strategy component of the solution is denoted $\mu^*(\eta, \delta)$ for small but positive δ and ε .

For $m \in M_\circ$ and each face $F_m(T_1, R_0)$ of \mathbf{S}_\circ^* (which equals \mathbf{S}_\circ^\bullet by assumption here and according to Proposition 4.2 in the text) and each fixed $\theta \in A_2$ and $v \in R^T$, consider the set of solutions $G_m(T_1, R_0; \theta, v)$ to the following system:

$$\begin{aligned} u_1(t, m, \rho_m) &= v_t & \forall t \in T_1 \\ u_1(t, m, \rho_m) &\leq v_t & \forall t \notin T_1 \\ \rho_{m,r} &= \theta_{m,r} & \forall r \in R_0 \\ \rho_{m,r} &\geq \theta_{m,r} & \forall r \notin R_0 \\ \sum_r \rho_{m,r} &= 1. \end{aligned}$$

The solution when $\theta = 0$ and $v = v^*$ is exactly $G_m(T_1, R_0)$. So for θ and v close to zero and v^* respectively, $G_m(T_1, R_0; \theta, v)$ is a polyhedron of dimension $|R| - |T_1| - |R_0| - 1$. Moreover, there is a homeomorphism $h^{T_1, R_0, \theta, v}$ from $G_m(T_1, R_0)$ to $G_m(T_1, R_0; \theta, v)$ that maps vertices to vertices and everything else by linear interpolation. The vertex map for this homeomorphism has the property that the same set of inequalities that hold with equality at a vertex in $G_m(T_1, R_0)$ hold under its image in $G_m(T_1, R_0)$, albeit with different constants. As in Appendix A, let \mathcal{P} be the collection of all pairs (T_1, R_0) such that $G_m(T_1, R_0)$ is contained in \mathbf{S}_m^\bullet . Let $\mathbf{S}_m^*(\theta, v)$ be the union over $(T_1, R_0) \in \mathcal{P}$ of $G_m(T_1, R_0; \theta, v)$. Then these homeomorphisms $h_m^{T_1, R_0, \theta, v}$ induce a homeomorphism $h_m^{\theta, v}$ from \mathbf{S}_m^* to $\mathbf{S}_m^*(\theta, v)$.

For each m , each $(p_m, \rho_m) \in X_m$, and each type t , by Claim A.1, the map $\beta_{t,m}$ is well-defined and continuous, where $\beta_{t,m}(p_m, \rho_m) \equiv p_m(t) + \lambda_m^{-1}(1 - \lambda_m)\phi_m(t)$ for $\lambda_m \in (0, 1]$ and $\phi_m \in P_m$ such that ρ_m is a best reply against $\lambda_m p_m + (1 - \lambda_m)\phi_m$ and the support of ϕ_m is contained in the set of types for whom sending m yields v_t^* against ρ_m . Let c be the maximum of $\beta_{t,m}(p_m, \rho_m)$ over all m, t , and $(p_m, \rho_m) \in X_m$.

Fix $\bar{\varepsilon}$ such that the following are well-defined:

- (1) the solution $(\rho^*(\theta), v^*(\theta))$ to the system $L_1(\theta)$ for $\theta \in A_2^{\bar{\varepsilon}}$.
- (2) $\mu^*(\eta, \delta)$ for all $\eta \in A_1^{\bar{\varepsilon}}$ and $\delta \leq \bar{\varepsilon}|M^*| + \bar{\varepsilon}|M_\circ|c$.
- (3) The homeomorphism $h_m^{\theta, v^*(\theta)}$ for all $\theta \in A_2^{\bar{\varepsilon}}$ and $m \in M_\circ$.

Let $P = \prod_{m \in M} P_m$ and $Q = [0, 1]^M \times P \times A_2$. For each $0 < \varepsilon \leq 1$, let Q^ε be the set of $(\zeta, p, \theta) \in Q$ such that $\zeta_m \leq \varepsilon$ and $\bar{\theta}_m \leq \varepsilon$ for all m . Denote the topological boundary of Q^ε by ∂Q^ε . Let \hat{X} be the set of $((\zeta, p, \theta), \rho_\circ) \in Q \times B_\circ$ such that for each $m \in M_\circ$, $(p_m, \rho_m) \in X_m$. And let \hat{p} be the projection map from \hat{X} to Q . For each ε let $(\hat{X}^\varepsilon, \partial \hat{X}^\varepsilon) = \hat{p}^{-1}(Q^\varepsilon, \partial Q^\varepsilon)$.

We now define the map $f : (\hat{X}^{\bar{\varepsilon}}, \partial \hat{X}^{\bar{\varepsilon}}) \rightarrow (E^{\bar{\varepsilon}}, \partial E^{\bar{\varepsilon}})$ as follows. Let $f((\zeta, p, \theta), \rho_\circ) = (\tilde{\eta}, \theta, \tilde{\mu}, \tilde{\rho})$ where:

$$\begin{aligned} \tilde{\eta}_{t,m} &= \zeta_m p_m(t) & \forall t, m \\ \tilde{\mu}_{t,m} &= \zeta_m \beta_m(p_m, \rho_m) & \forall t, m \in M_\circ \\ \tilde{\delta}_t &= \sum_{m \notin M_t^*} \tilde{\mu}_{t,m} & \forall t \\ \tilde{\mu}_{t,m} &= \mu_{t,m}^*(\tilde{\eta}, \tilde{\delta}_t) \\ \tilde{\rho}_m &= \rho_m^*(\theta) & \forall m \in M^* \\ \tilde{\rho}_m &= h^{\theta, v^*(\theta)}(\rho_m) & \forall m \in M_\circ \end{aligned}$$

Let $Y^{\bar{\varepsilon}}$ be the image of $\hat{X}^{\bar{\varepsilon}}$ under f . We claim that for each $0 < \varepsilon \leq \bar{\varepsilon}$, f induces a homeomorphism between $\hat{X}^\varepsilon \setminus \partial \hat{X}^\varepsilon$ and $Y^\varepsilon \setminus \partial Y^\varepsilon$. Indeed, the inverse image (ζ, p, θ, ρ) of a

point $(\tilde{\eta}, \theta, \tilde{\mu}, \tilde{\rho}) \in Y^\varepsilon \setminus \partial Y^1$ under f is computed as follows.

$$\begin{aligned}\zeta_m &= \sum_{t \in T} \pi_t \tilde{\eta}_{t,m} & \forall m \in M_o \\ p_m(t) &= (\zeta_m)^{-1} \tilde{\eta}_{t,m} & \forall t, m \\ \rho_m &= (h_m^{\theta, v^*(\theta)})^{-1}(\tilde{\rho}_m) & \forall m \in M_o\end{aligned}$$

We first show that $\mathbf{S}^* = \{(\tilde{\mu}, \tilde{\rho}) \mid (0, \tilde{\mu}, \tilde{\rho}) \in Y\}$. If $(0, \mu, \rho) \in Y$, then for each $(\zeta, p, \theta, \rho_o) \in f^{-1}(0, \tilde{\mu}, \tilde{\rho})$, $\zeta = 0$ and $\theta = 0$. Hence, from the equations above defining f , $\tilde{\mu}_{t,m} = \mu_{t,m}^*$ for each t , $\tilde{\rho}_m = \rho_m^*$ for each $m \in M^*$, and $(\tilde{\rho}_m)_{m \in M_o} = \rho_o$, which belongs to \mathbf{S}_o^* . Therefore $(\tilde{\mu}, \tilde{\rho})$ belongs to \mathbf{S}^* . Going the other way, if $(\tilde{\mu}, \tilde{\rho})$ belongs to \mathbf{S}_o^* then letting $\tilde{\rho}_o$ be the projection of $\tilde{\rho}$ to \mathbf{S}_o , $f((0, p, 0), \tilde{\rho}_o) = (0, \tilde{\mu}, \tilde{\rho})$ where for each $m \in M_o$, p_m is such that $(p_m, \tilde{\rho}_m) \in X_m$ and p_m is arbitrary for $m \notin M_o$. Thus, $\mathbf{S}^* = \{(\tilde{\mu}, \tilde{\rho}) \mid (0, \tilde{\mu}, \tilde{\rho}) \in Y\}$.

We turn now to the connectedness condition for Y^ε . For each $m \in M_o$, since $(X_m, \partial X_m)$ is a pseudo-manifold, $X_m \setminus \partial X_m$ is connected and dense in X_m . Therefore, for each $0 < \varepsilon$, $\hat{X}^\varepsilon \setminus \partial \hat{X}^1$ is connected and dense in \hat{X}^ε . For each $0 < \varepsilon \leq \bar{\varepsilon}$, f maps $(\hat{X}^\varepsilon, \partial \hat{X}^\varepsilon)$ onto $(Y^\varepsilon, \partial Y^\varepsilon)$; and it maps $\hat{X}^\varepsilon \setminus \partial \hat{X}^1$ homeomorphically onto $Y^\varepsilon \setminus \partial Y^1$. Since the former set is connected and dense in \hat{X}^ε , the latter is as well. Therefore, Y^ε satisfies the connectedness condition for stability.

Finally, we show the essentiality of the projection $p : (Y^\varepsilon, \partial Y^\varepsilon) \rightarrow (A^\varepsilon, \partial A^\varepsilon)$ for $0 < \varepsilon \leq \bar{\varepsilon}$. Since \hat{X}^ε is, up to reordering of coordinates, the set $\prod_{m \in M_o} X_m \times [0, \varepsilon]^M \times \prod_{m \in M^*} P_m$, and the projection map \hat{p} is (up to reordering of coordinates) the product of the q_m 's on X_m 's with just the identity function on the remaining factors. Therefore, $(\hat{X}^\varepsilon, \partial \hat{X}^\varepsilon)$ is a pseudo-manifold whose projection \hat{p} has degree $\prod_{m \in M_o} d_m$, which is nonzero by assumption. In particular, \hat{p} is a cohomologically essential map. Let $g : Q^\varepsilon \rightarrow A^\varepsilon$ be the map $g(\zeta, p, \theta) = (\eta, \theta)$ where for each (t, m) , $\eta_{t,m} = \zeta_m p_m(t)$. Then g is a surjective map that maps $Q^\varepsilon \setminus \partial Q^1$ homeomorphically onto $A^\varepsilon \setminus A^1$. Observe that $g \circ \hat{p} = p \circ f$ in the commutative diagram

$$\begin{array}{ccc}(\hat{X}^\varepsilon, \partial \hat{X}^\varepsilon) & \xrightarrow{f} & (Y^\varepsilon, \partial Y^\varepsilon) \\ \hat{p} \downarrow & & p \downarrow \\ (Q^\varepsilon, \partial Q^\varepsilon) & \xrightarrow{g} & (A^\varepsilon, \partial A^\varepsilon)\end{array}$$

By the strong excision property [30, Theorem 6.5.5], f and g induce isomorphisms in cohomology. The essentiality of \hat{p} thus implies that the projection map $p : (Y^\varepsilon, \partial Y^\varepsilon) \rightarrow (A^\varepsilon, \partial A^\varepsilon)$ is essential. Therefore \mathbf{S}^* is a stable set. \square

REFERENCES

- [1] Banks, J., and J. Sobel (1987): Equilibrium Selection in Signaling Games, *Econometrica*, 55, 647-661.

- [2] Blume, L., A. Brandenburger, and E. Dekel (1991): Lexicographic Probabilities and Equilibrium Refinements, *Econometrica*, 59, 81-98.
- [3] Chen, Y., N. Kartik, and J. Sobel (2008): Selecting Cheap-Talk Equilibria, *Econometrica*, 76, 1171-136.
- [4] Cho, I., and D. Kreps (1987): Signalling Games and Stable Equilibria, *Quarterly Journal of Economics*, 102, 179-221.
- [5] Cho, I., and J. Sobel (1990): Strategic Stability and Uniqueness in Signaling Games, *Journal of Economic Theory*, 50, 381-413.
- [6] Crawford, V., and J. Sobel (1982): Strategic Information Transmission, *Econometrica*, 50, 1431-1451.
- [7] Dalkey, N. (1953): Equivalence of Information Patterns and Essentially Determinate Games, *Contributions to the Theory of Games*, 2, 217-245. Princeton NJ: Princeton University Press.
- [8] Fudenberg, D., and J. Tirole (1993): *Game Theory*. Cambridge MA: MIT Press.
- [9] Govindan, S., and T. Klumpp (2000): Perfect Equilibrium and Lexicographic Beliefs, *International Journal of Game Theory*, 31, 229-243.
- [10] Govindan, S., and J.-F. Mertens (2003): An Equivalent Definition of Stable Equilibria, *International Journal of Game Theory*, 3, 339-357.
- [11] Govindan, S., and R. Wilson (2001): Direct Proofs of Generic Finiteness of Nash Equilibrium Outcomes, *Econometrica*, 69, 765-769.
- [12] 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
- [13] Govindan, S., and R. Wilson (2005): Essential Equilibria, *Proceedings of the National Academy of Sciences USA*, 102, 15706-15711. URL: www.pnas.org/cgi/reprint/102/43/15706.pdf.
- [14] Govindan, S., and R. Wilson (2006): Sufficient Conditions for Stable Equilibria, *Theoretical Economics*, 1, 167-206.
- [15] Govindan, S., and R. Wilson (2009a): On Forward Induction, *Econometrica*, to appear. Available as Research Report 1986, Stanford Business School, Stanford CA.
- [16] Govindan, S., and R. Wilson (2009b): Metastable Equilibria, *Mathematics of Operations Research*, to appear. Available as Research Report 1934R, Stanford Business School, Stanford CA.
- [17] 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.
- [18] Kohlberg, E. (1990): Refinement of Nash Equilibrium: The Main Ideas, in T. Ichiishi, A. Neyman, and Y. Tauman (eds.), *Game Theory and Applications*. San Diego: Academic Press.
- [19] Kohlberg, E., and J.-F. Mertens (1986): On the Strategic Stability of Equilibria, *Econometrica*, 54, 1003-1037.
- [20] Kohlberg, E., and P. Reny (1997): Independence on Relative Probability Spaces and Consistent Assessments in Game Trees, *Journal of Economic Theory*, 75, 280-313.
- [21] Kreps, D., and J. Sobel (1994): Signalling, in R. Aumann and S. Hart, (eds.), *Handbook of Game Theory*, II, 849-868. New York: Elsevier.
- [22] Kreps, D., and R. Wilson (1982): Sequential Equilibria, *Econometrica*, 50, 863-894.
- [23] 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.
- [24] Mertens, J.-F. (1989): Stable Equilibria—A Reformulation, Part I: Definition and Basic Properties, *Mathematics of Operations Research*, 14, 575-625.

- [25] Mertens, J.-F. (1991): Stable Equilibria—A Reformulation, Part II: Discussion of the Definition and Further Results, *Mathematics of Operations Research*, 16, 694-753.
- [26] Mertens, J.-F. (1992): The Small Worlds Axiom for Stable Equilibria, *Games and Economic Behavior*, 4, 553-564.
- [27] Nash, J. (1950): Equilibrium Points in N-Person Games, *Proceedings of the National Academy of Sciences USA*, 36, 48-49.
- [28] Nash, J. (1951): Non-Cooperative Games, *Annals of Mathematics*, 54, 286-295.
- [29] Savage, L.J. (1954): *Foundations of Statistics*, New York: John Wiley & Sons.
- [30] Spanier, E. (1966): *Algebraic Topology*, New York: McGraw-Hill. Reprinted New York: Springer-Verlag, 1989.
- [31] Thompson, F. (1952): *Equivalence of Games in Extensive Form & Behavior Strategies in Finite Games*, RM 759 & 769. Santa Monica CA: The Rand Corporation.

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