Coverpage for Submission to Econometrica

Date: August 29, 2004

Title: Axiomatic Justification of Stable Equilibria

Authors: Srihari Govindan and Robert Wilson

Addresses:
(1) Professor Srihari Govindan, Department of Economics, University of Iowa, Iowa City IA 52242, USA.
Email: srihari-govindan@uiowa.edu.
(2) Professor Robert Wilson, Stanford Business School, Stanford, CA 94305-5015, USA.
Email: rwilson@stanford.edu.

Corresponding Author: Professor Robert Wilson, Stanford Business School, Stanford, CA 94305-5015, USA. Email: rwilson@stanford.edu.

Abstract: A solution concept that satisfies the axioms of weak invariance and strong backward induction selects a stable set of a game's equilibria.

Keywords: game theory, equilibrium selection, stability.

JEL subject classification: C72.

AXIOMATIC JUSTIFICATION OF STABLE EQUILIBRIA

SRIHARI GOVINDAN AND ROBERT WILSON

Abstract. A solution concept that satisfies the axioms of weak invariance and strong backward induction selects a stable set of a game's equilibria.

1. Introduction

The theory of games is inherently incomplete because a typical game has multiple equilibria. Nevertheless, because some equilibria seem more plausible than others, several authors suggest that Nash's (1950) initial definition of equilibrium should be strengthened. Hillas and Kohlberg (2002) survey proposals for solution concepts that select equilibria with additional properties. Among those derived from a game's normal form are perfect, proper, and lexicographic equilibria, represented as mixtures of pure strategies. Among those derived from an extensive form with perfect recall are refinements of sequential equilibria represented as behavior strategies that allow mixtures of actions at each information set.[1] These proposals reflect basic tensions. The implausibility of some equilibria is most evident in the extensive form where Bayes' Rule and backward induction are explicit; hence one can examine the beliefs used to justify behavior at information sets off the path of equilibrium play. In the normal form, analogous tests of plausibility can be applied to the hierarchy of beliefs implied by a lexicographic equilibrium.

The crux of the problem is that extensive-form analyses suggest criteria for equilibrium refinements, yet the theory of individual decisions suggests that a theory of rational behavior in games should depend only on the normal form. The latter, the orthodox view, is implicit in Nash's definition. Similarly, Kohlberg and Mertens (1986)[KM hereafter] show that a proper equilibrium of a normal form induces a sequential equilibrium in every extensive form having that normal form. Indeed, because the set of sequential equilibria varies depending on which among many equivalent extensive forms is used, they argue that a refinement should depend only on the normal form—or better, only on the reduced normal form obtained by deleting pure strategies that are redundant because their payoffs can be duplicated by mixtures of other pure strategies.[2] The orthodox view implements the general principle that a theory of rational behavior should be immune to presentation effects, such as which extensive form is envisioned, or whether redundant strategies are deleted. But this principle does not suffice to ensure that beliefs are plausible in an extensive-form description of

Date: February 1, 2004; this version August 29, 2004.
Web file: http://faculty-gsb.stanford.edu/wilson/pdf%20files/AxiomaticStability040829.pdf.
Key words and phrases. game theory, equilibrium selection, stability. *JEL subject classification:* C72.
This work was funded in part by a grant from the National Science Foundation of the United States.

[1] Perfect, proper, lexicographic and sequential equilibria are defined by Selten (1975), Myerson (1978), Blume, Brandenberger, and Dekel (1991), and Kreps and Wilson (1982), respectively.

[2] Thompson (1952) shows that four elementary transformations enable one of two games in extensive form having the same normal form and no moves of nature to be transformed into the other. KM add the transformation that coalesces final moves by nature by replacing them with a terminal node at which the payoffs are the players' expected payoffs. Thompson uses a weaker definition of the reduced normal form.

the game. KM therefore argue that additional criteria should be invoked, such as admissibility, backward induction, and forward induction.

Kohlberg and Mertens argue further that solution concepts have been too restrictive. Criteria that apply to a single equilibrium have inherent limitations; therefore, they propose that selection should apply criteria to sets of equilibria. This is immaterial in an extensive-form game with generic payoffs since a connected set of equilibria induce the same outcome; that is, these equilibria agree along paths of equilibrium play (Kreps and Wilson, 1982; Govindan and Wilson, 2001). With this amendment, KM propose a more ambitious program. Recall that an equilibrium of an extensive form that is perfect with respect to a sequence of perturbations of behavior strategies induces a sequential equilibrium, and conversely if the game is generic. KM argue that ideally the goal would be to select equilibria that are "truly perfect" (i.e., essential) in that they are perturbed slightly by *any* small perturbation of strategies. Although this ideal is impossible with a solution concept that selects singletons, it is feasible with one that selects sets of equilibria. KM's main result shows the existence of a component of equilibria that is stable. Stability requires a kind of continuity in that every nearby game has an equilibrium near the component. Although this general result is stated for payoff perturbations, KM argue that admissibility suggests that the relevant payoff perturbations are those induced by strategy perturbations, and further, they focus on minimal stable sets. Here, sets that are minimal among those stable against strategy perturbations are called KM-stable sets.

In this paper we advance KM's program using a dual approach.

1. **Weak Invariance**. The orthodox view requires a solution concept to be invariant in that it depends only on the reduced normal form of a game; i.e., selection and reduction commute. Our first axiom is weaker. We require that a selected set is equivalent to a superset of a set selected for the inflated game obtained by adding redundant pure strategies. Here, equivalent means they have the same probability distributions on the original set of pure strategies.
2. **Strong Backward Induction**. Accepting the relevance of extensive-form analysis, we require that behavior strategies in an extensive-form game are quasi-perfect (and therefore sequential). However, our second axiom enforces KM's more stringent truly-perfect criterion by requiring further that *every* strategy perturbation refines the selected set of equilibria.

Quasi-perfection is a refinement of sequential equilibrium proposed by van Damme (1984). It implements backward induction by requiring that at each information set a player's continuation strategy is optimal against perturbed strategies of other players. This ensures conditional admissibility; that is, a sequential equilibrium that is quasi-perfect uses a continuation strategy that is not dominated in the remainder of the game following an information set.

These axioms say that a solution should select sets of equilibria such that, for each extensive form with an inflated normal form, each perturbation of behaviors should refine a selected set by selecting quasi-perfect equilibria. The conjunction of these two axioms implies tight restrictions on a solution. Our main theorem is:

Theorem 1.1. *If a solution satisfies Weak Invariance and Strong Backward Induction then each selected set includes a KM-stable set.*

Thus the axioms imply that a selected set is affected only slightly by any perturbation of mixed strategies of the normal form of the game.

In their Appendix D, KM establish a comparable result in the special case of an equilibrium that assigns positive probability to every optimal strategy: if such an equilibrium is perfect with respect to every perturbation of behavior strategies in every extensive form with the same reduced normal form then it is essential. But such an equilibrium need not exist. Our result differs because we consider sets of equilibria, and we replace perfect by quasi-perfect. The existence of solutions that satisfy the axioms is established by the fact that they are satisfied by the stronger concept of stability proposed by Mertens (1989).

Section 2 fixes notation and specifies the axioms. Section 3 proves Theorem 1.1 for 2-player games, where a simple proof is possible. Section 4 uses an alternative proof for games with N players. Section 5 provides concluding remarks.

2. Formulation

We consider games with finite sets of players and pure strategies. The normal form of a game is specified by a payoff function $G : \prod_{n\in N} S_n \to \mathbb{R}^N$ where N is the set of players and S_n is player n's set of pure strategies. Interpret a pure strategy s_n as a vertex of player n's simplex $\Sigma_n = \Delta(S_n)$ of mixed strategies. The sets of profiles of pure and mixed strategies are $S = \prod_n S_n$ and $\Sigma = \prod_n \Sigma_n$. Say that two mixed strategies of a player are equivalent if for every profile of the other players' strategies they yield the same expected payoff for every player. A pure strategy s_n of player n is redundant if n has an equivalent mixed strategy $\sigma_n \neq s_n$. The normal form is reduced if no pure strategy is redundant. Say that two games are equivalent if their reduced normal forms are the same except for labeling of pure strategies.

In general, a solution assigns to each game a collection of nonempty sets of its equilibria, called the selected sets. However, each equilibrium induces a family of equilibria in equivalent strategies for each inflation of the game obtained by adding redundant strategies. Therefore, we assume:

Axiom 2.1. Weak Invariance. A selected set is equivalent to a superset of one selected for an inflated game. Specifically, if Σ° is a selected set for the game G° and G^* is an inflation of G° then there exists a selected set Σ^* for G^* such that the set of strategies in G° that are equivalent to those in Σ^* is included in Σ°.

To each game in normal form we associate those games in extensive form with perfect recall that have that normal form. Each extensive form specifies a disjoint collection $H = \{H_n \mid n \in N\}$ of the players' information sets, and for each information set $h \in H_n$ it specifies a set $A_n(h)$ of possible actions by n at h. In its normal form the set of pure strategies of player n is $S_n = \{s_n : H_n \to \cup_{h\in H_n} A_n(h) \mid s_n(h) \in A_n(h)\}$. The projection of S_n onto h and n's information sets that follow h is denoted $S_{n|h}$; that is, $S_{n|h}$ is the set of n's continuation strategies from h. Let $S_n(h)$ be the set of n's pure strategies that choose all of n's actions necessary to reach $h \in H_n$, and let $S_n(a|h)$ be the subset that choose $a \in A_n(h)$. Then a completely mixed strategy $\sigma_n \gg 0$ induces the conditional probability $\sigma_n(a|h) = \sum_{s_n\in S_n(a|h)} \sigma_n(s_n)/\sum_{s_n\in S_n(h)} \sigma_n(s_n)$ of choosing a at h. More generally, a behavior strategy $\beta_n \in \prod_{h\in H_n} \Delta(A_n(h))$ assigns to each information set h a probability $\beta_n(a|h)$ of action $a \in A_n(h)$ if h is reached. Kuhn (1953) shows that mixed and behavior strategies are payoff-equivalent in extensive-form games with perfect recall.

Given a game in extensive form, an action perturbation $\varepsilon : H \to (0,1)^2$ assigns to each information set a pair $(\underline{\varepsilon}(h), \bar{\varepsilon}(h))$ of small positive numbers, where $0 < \underline{\varepsilon}(h) \leqslant \bar{\varepsilon}(h)$. Use $\{\varepsilon\}$ to denote a sequence of action perturbations that converges to 0.

Definition 2.2. Quasi-Perfect.[3] A sequence $\{\sigma^\varepsilon\}$ of profiles is $\{\varepsilon\}$-quasi-perfect if for each $a \in A_n, h \in H_n, n \in N$ and each action perturbation ε:

1. $\sigma_n^\varepsilon(a|h) \geqslant \underline{\varepsilon}(h)$, and
2. $\sigma_n^\varepsilon(a|h) > \bar{\varepsilon}(h)$ only if a is an optimal action at h in reply to σ^ε; i.e., only if $s_n(h) = a$ for some continuation strategy $s_n \in \arg\max_{s \in S_{n|h}} E[G_n \mid h, s, \sigma_{-n}^\varepsilon]$.

Suppose that $\sigma_n(\cdot|h) = \lim_{\varepsilon \downarrow 0} \sigma_n^\varepsilon(\cdot|h)$. Then this definition says that at h player n's continuation strategy at h assigns a positive conditional probability $\sigma_n(a|h) > 0$ to action a only if a is chosen by a continuation strategy that is an optimal reply to perturbations $(\sigma_{n'}^\varepsilon)_{n' \neq n}$ of other players' strategies. Thus when solving his dynamic programming problem, player n takes account of vanishingly small trembles by other players but ignores his own trembles later in the game. In particular, this enforces admissibility of continuation strategies conditional on having reached h. Van Damme (1984) shows that the pair $(\mu, \beta) = \lim_{\varepsilon \downarrow 0}(\mu^\varepsilon, \beta^\varepsilon)$ of belief and behavior profiles is a sequential equilibrium, where σ^ε induces at $h \in H_n$ the conditional probability $\mu_n^\varepsilon(t|h)$ of node $t \in h$ and the behavior $\beta_n^\varepsilon(a|h) = \sigma_n^\varepsilon(a|h)$ is player n's conditional probability of choosing a at h.

Our second axiom requires that each sequence of action perturbations induces a further selection among the profiles in a selected set.

Axiom 2.3. Strong Backward Induction. For a game in extensive form with perfect recall for which the solution selects a set Σ° of equilibria, for each sequence $\{\varepsilon\}$ of action perturbations there exists a profile $\sigma \in \Sigma^\circ$ that is the limit of a convergent subsequence $\{\sigma^\varepsilon\}$ of $\{\varepsilon\}$-quasi-perfect profiles.

As the proofs in Sections 3 and 4 show, Theorem 1.1 remains true if Axiom 2.3 is weakened by requiring action perturbations to satisfy the additional restriction that $\underline{\varepsilon}(h) = \bar{\varepsilon}(h)$ for all h. The reason we do not do so is conceptual. The lower bound $\underline{\varepsilon}(\cdot)$ reflects the requirement that every action of a player is chosen with positive probability, while $\bar{\varepsilon}(\cdot)$ provides the upper bound on the "error probability" of suboptimal actions at an information set.

We conclude this section by defining stability. In general, a set of equilibria of a game in normal form is stable if, for any neighborhood of the set, every game obtained from a sufficiently small perturbation of payoffs has an equilibrium in the neighborhood. However, KM focus on minimal closed sets that are stable only against those payoff perturbations induced by strategy perturbations. For $0 \leqslant \delta \leqslant 1$, let $P_\delta = \{(\lambda_n \tau_n)_n \mid (\forall\, n)\ 0 \leqslant \lambda_n \leqslant \delta, \tau_n \in \Sigma_n\}$ and let ∂P_δ be the topological boundary of P_δ. For each $\eta \in P_1$, and $n \in N$, let $\overline{\eta}_n = \sum_{s \in S_n} \eta_n(s)$. Given any $\eta \in P_1$, a perturbed game $G(\eta)$ is obtained by replacing each pure strategy s_n of player n with $\eta_n + (1 - \overline{\eta}_n)s_n$. Thus $G(\eta)$ is the perturbed game in which the strategy sets of the players are restricted so that the probability that n plays a strategy $s \in S_n$ must be at least $\eta_n(s)$. For a vector (λ, τ), we sometimes write $G(\lambda, \tau)$ to denote the perturbed game $G((\lambda_n \tau_n)_n)$.

Definition 2.4. KM-Stability. A set of equilibria of the game G is KM-stable if it is minimal with respect to the following property: Σ° is a closed set of equilibria of G such that for each $\epsilon > 0$ there exists $\delta > 0$ such that for each $\eta \in P_\delta \backslash \partial P_\delta$ the perturbed game $G(\eta)$ has an equilibrium within ϵ of Σ°.

[3] This definition differs from van Damme (1984) in that the upper bound $\bar{\varepsilon}(\cdot)$ of the error probability can differ across information sets. However, it is easily shown that the set of quasi-perfect equilibria as defined by van Damme is the set of all behavioral-strategy profiles that are limits of sequences of behavioral-strategy profiles induced by sequences of $\{\varepsilon\}$-quasi-perfect equilibria as defined here.

KM show that every game has a KM-stable set of equilibria.

3. 2-Player Games

This section provides a direct proof of Theorem 1.1 for the special case of two players. It is simpler than the proof of the general case in Section 4 because 2-player games have a linear structure. This structure enables a generalization—Statement 3 in the following Theorem—of the characterization of KM-stability obtained by Cho and Kreps (1987) and Banks and Sobel (1987) for the special case of sender-receiver signaling games with generic payoffs.

Theorem 3.1. *[Characterization of Stability] Let G be a 2-player game, and let Σ^0 be a closed subset of Σ. The following statements are equivalent.*

1. *Σ° contains a KM-stable set of the game G.*
2. *For each $\tau \in \Sigma \backslash \partial\Sigma$ there exists sequence σ^k in Σ converging to $\sigma^\circ \in \Sigma^\circ$ and a corresponding sequence λ^k in $(0,1)^2$ converging to the origin, such that σ^k is an equilibrium of $G(\lambda^k, \tau)$.*
3. *For each $\tau \in \Sigma \backslash \partial\Sigma$ there exists $\sigma^\circ \in \Sigma^\circ$, a profile $\tilde{\sigma} \in \Sigma$ and $0 < \mu \leqslant 1$ such that, for each player n and each pure strategy $s \in S_n$ for which $\mu\sigma^\circ_n(s) + [1-\mu]\tilde{\sigma}_n(s) > 0$, s is an optimal reply for player n against both σ° and the profile $\sigma^* = \mu\tau + [1-\mu]\tilde{\sigma}$.*

Proof. We prove first that statement 1 implies statement 2. Suppose Σ° contains a KM-stable set. Fix $\tau \in \Sigma \backslash \partial\Sigma$. Then for each positive integer k one can choose a vector $\lambda^k \in (0, 1/k)^2$ and an equilibrium σ^k of $G(\lambda^k, \tau)$ whose distance from Σ° is less than $1/k$. Let σ° be the limit of a convergent subsequence of σ^k as $k \uparrow \infty$. Then $\sigma^\circ \in \Sigma^\circ$ satisfies statement 2 for τ.

Next we prove that statement 2 implies statement 3. Fix $\tau \in \Sigma \backslash \partial\Sigma$. Statement 2 assures that there exists a sequence λ^k in $(0,1)^2$ converging to zero and a sequence σ^k of equilibria of $G(\lambda^k, \tau)$ converging to an equilibrium σ° in Σ°. By passing to a subsequence if necessary, we can assume that the set of optimal replies in G to the strategies σ^k is the same for all k. Define ς by $\varsigma_n = [\sigma^1_n - \lambda^1_n \tau_n]/(1 - \lambda^1_n)$ where σ^1 is the first element of the sequence σ^k. Since σ^1 is an equilibrium of $G(\lambda^1, \tau)$, ς is an optimal reply to σ^1 and hence an optimal reply to all elements of the sequence, as well as to the limit σ°. Also, σ° must be an optimal reply to σ^1 and σ°, since the optimal replies are constant along the sequence σ^k of equilibria of perturbed games, which converges to σ°. Let $\mu = \min(\lambda^1_1, \lambda^1_2)$. Define $\tilde{\sigma}$ by $\tilde{\sigma}_n = [\mu(1-\lambda^1_n)\varsigma_n + (\lambda^1_n - \mu)\sigma^\circ_n]/\lambda^1_n(1-\mu)$ and let $\sigma^* = \mu\tau + (1-\mu)\tilde{\sigma}$. Then $\sigma^*_n = [\mu\sigma^1_n + (\lambda^1_n - \mu)\sigma^\circ_n]/\lambda^1_n$. Therefore, ς and σ° are both optimal against σ^*. For each player n, $\tilde{\sigma}_n$ is an average of ς_n and σ°_n, so it too is optimal against σ^0 and σ^*, which completes the proof.

Lastly we prove that statement 3 implies statement 1 by showing that Σ° satisfies the property in Definition 2.4. Fix an ϵ-neighborhood of Σ°. Take a sufficiently fine simplicial subdivision of Σ such that: (i) the union U of the simplices of this complex that intersect Σ° is contained in its ϵ-neighborhood; and (ii) the best-reply correspondence is constant over the interior of each simplex. Because G is a two-player game, this simplicial subdivision can be done such that each simplex is actually a convex polytope. Observe that U is itself a closed neighborhood of Σ°. Let Q be the set of all pairs $(\eta, \sigma) \in P_1 \times \Sigma$ such that $\sigma \in U$ and σ is an equilibrium of $G(\eta)$; and let Q_0 be the set of $(0, \sigma) \in Q$, i.e., the set of equilibria of the game G that are contained in U. By property (ii) of the triangulation and because the simplices are convex polytopes, Q and Q_0 are finite unions of polytopes. Triangulate Q such that Q_0 is a subcomplex, and take a barycentric subdivision so that Q_0 becomes a full subcomplex. Since Q is a union of polytopes, both the triangulation and the projection map $p : Q \to P_1$ can be made piecewise-linear. Let X be the union of simplices of Q that intersect Q_0. Since Q_0 is a full subcomplex, the intersection of each simplex of Q with Q_0 is a face of the simplex. Let $X^0 = X \cap Q_0$ and let X^1 be the union of simplices of X that do not intersect Q_0. Given $x \in X$, there exists a unique simplex K of X that contains x in its interior. Let K^0 be the face of K that belongs to X^0; and let K^1 be the face of K spanned by the vertices of K that do not belong to K^0. K^1 is then contained in X^1. We therefore have that x is expressible as a convex combination $[1-\alpha]x^0 + \alpha x^1$, where $x^i \in K^i$ for $i = 0, 1$; moreover, this combination is unique if $x \notin X^0 \cup X^1$. Finally, since the projection map p is piecewise affine, we have that $p(x) = [1-\alpha]p(x^0) + \alpha p(x^1) = \alpha p(x^1)$.

Choose $\delta^* > 0$ such that for each $(\eta, \sigma) \in X^1$, $\max_n \overline{\eta}_n > \delta^*$. Such a choice is possible since X^1 is a compact subset of Q that is disjoint from Q_0. Fix now $\delta_1, \delta_2 < \delta^*$ and $\tau \in \Sigma$. The proof is complete if we can show that the game $G(\delta_1\tau_1, \delta_2\tau_2)$ has an equilibrium in U. By statement 3, there exists $\sigma^\circ \in \Sigma^\circ$, $\tilde{\sigma} \in \Sigma$ and $0 < \mu \leqslant 1$ such that $\sigma(\gamma) = ((1-\gamma\delta_n)\sigma_n^\circ + \gamma\delta_n((1-\mu)\tilde{\sigma}_n + \mu\tau_n))_{n=1,2}$ is an equilibrium of $G(\gamma\mu(\delta_1\tau_1, \delta_2\tau_2))$ for all $0 \leqslant \gamma \leqslant 1$. Because $\sigma(0) = \sigma^\circ \in \Sigma^\circ$ we can choose γ sufficiently small that the point $x = (\gamma\mu(\delta_1\tau_1, \delta_2\tau_2), \sigma(\gamma))$ belongs to $X \backslash (X^0 \cup X^1)$; hence there exists a unique $\alpha \in (0, 1)$ and $x^i \in X^i$ for $i = 0, 1$ such that x is an α-combination of x^0 and x^1. As remarked before, $p(x) = \alpha p(x^1)$. Therefore, there exists $\sigma \in \Sigma$ such that $x^1 = (\gamma^*\mu(\delta_1\tau_1, \delta_2\tau_2), \sigma)$, where $\gamma^* = \gamma/\alpha$. Since points in X^1 project to $P_1 \backslash P_{\delta^*}$, $\gamma^*\mu\delta_n > \delta^*$ for some n; i.e., $\gamma^*\mu > 1$ since $\delta_n < \delta^*$ for each n by assumption. Therefore, the point $[1 - 1/\gamma^*\mu]x^0 + [1/\gamma^*\mu]x^1$ corresponds to an equilibrium of the game $G(\delta_1\tau_1, \delta_2\tau_2)$ that lies in U. This proves statement 1. □

The characterization in statement 3 can be stated equivalently in terms of a lexicographic probability system [LPS] as in Blume, Brandenberger, and Dekel (1991).

Corollary 3.2. *[Lexicographic Characterization] A closed set $\Sigma^\circ \subset \Sigma$ contains a KM-stable set if and only if for each $\tau \in \Sigma \backslash \partial\Sigma$ there exists $\sigma^0 \in \Sigma^\circ$, a profile $\tilde{\sigma} \in \Sigma$, and for each player n, an LPS $\mathcal{L}_n = (\sigma_n^0, \dots, \sigma_n^{K_n})$ for which $\sigma_n^{K_n} = [1-\lambda_n]\tilde{\sigma}_n + \lambda_n\tau_n$ for some $\lambda_n \in (0, 1]$, such that for each player n every strategy that is either: (i) in the support of σ^k with $k < K_n$ or (ii) in the support of $\tilde{\sigma}_n$ if $\lambda_n < 1$, is a lexicographic best reply to the LPS of the other player.*

Proof. The necessity of the condition follows from statement 3. To prove sufficiency, observe that for each sufficiently small $\alpha > 0$ the strategy profile $\sigma(\alpha)$ defined by $\sigma_n(\alpha) = \sum_{k=0}^{K_n} \alpha^k \sigma_n^k$ is an equilibrium for the perturbed game $G(\eta)$ where player n's perturbation vector is $\eta_n = \alpha^{K_n}\lambda_n\tau_n$. Since $\sigma(\alpha)$ converges to σ^0 as α goes to zero, the condition of the Corollary implies statement 2 of the Theorem. □

Theorem 3.3. *[Sufficiency of the Axioms] If a solution satisfies Weak Invariance and Strong Backward Induction then for any 2-player game a selected set includes a KM-stable subset of its normal form.*

Proof. Let G be the normal form of a 2-player game. Suppose that $\Sigma^\circ \subset \Sigma$ is a set selected by a solution that satisfies Weak Invariance and Strong Backward Induction. Let $\tau = (\tau_1, \tau_2)$ be any profile in the interior of Σ. We show that Σ° satisfies the condition of Corollary 3.2 for τ. Construct as follows the extensive-form game Γ with perfect recall that has a normal form that is an inflation of G. In Γ each player n first chooses whether or not to use the mixed strategy τ_n, and if not, then which pure strategy in S_n to use. Denote the two information sets at which n makes these choices by h'_n and h''_n. At neither of these does n have any information about the other player's analogous choices. In Γ the set of pure strategies for player n is $S^*_n = \{\tau_n\} \cup S_n$ (after identifying all strategies where n chooses to play τ_n at his first information set h'_n) and the corresponding simplex of mixed strategies is Σ^*_n. For each $\delta > 0$ in a sequence converging to zero, let $\{\varepsilon\}$ be a sequence of action perturbations that require the minimum probability of each action at h'_n to be $\underline{\varepsilon}(h'_n) = \delta$, and the maximum probability of suboptimal actions at h''_n to be $\bar{\varepsilon}(h''_n) = \delta^2$. By Weak Invariance, the solution selects a set $\tilde{\Sigma}^\circ$ for Γ that is a subset of those strategies equivalent to ones in Σ°. By Strong Backward Induction there exists a sequence $\{\tilde{\sigma}^\varepsilon\}$ of $\{\varepsilon\}$-quasi-perfect profiles converging to some point $\tilde{\sigma}^0 \in \tilde{\Sigma}^\circ$. By Blume, Brandenberger, and Dekel (1991) there exists for each player n: (i) an LPS $\tilde{\mathcal{L}}_n = (\tilde{\sigma}^0_n, \tilde{\sigma}^1_n, \dots, \tilde{\sigma}^{K_n}_n)$, with members $\tilde{\sigma}^k_n \in \Sigma^*_n$; and (ii) for each $0 \leqslant k < K_n$ a sequence of positive numbers $\lambda^k_n(\varepsilon)$ converging to zero such that each $\tilde{\sigma}^\varepsilon_n$ in the sequence is expressible as the nested combination $((1-\lambda^0_n(\varepsilon))\tilde{\sigma}^0_n + \lambda^0_n((1-\lambda^1_n(\varepsilon))\tilde{\sigma}^2_n + \lambda^1_n(\cdots + \lambda^{K_n-1}_n(\varepsilon)\tilde{\sigma}^{K_n}_n)))$. Let k^*_n be the smallest k for which $\tilde{\sigma}^k_n$ assigns positive probability to the "pure" strategy τ_n of the inflated game.

Claim 3.4. *If $s_n \in S_n$ is assigned a positive probability by some $\tilde{\sigma}^k_n$ for $k \leqslant k^*_n$ then s_n is a lexicographic best reply to the LPS of the other player.*

Proof of Claim. If s_n is not a lexicographic best reply to the LPS of the other player then sufficiently far along the sequence s_n is not a best reply against $\tilde{\sigma}^\varepsilon$. Quasi-perfection requires that $\tilde{\sigma}^\varepsilon_n(\tau_n|h'_n) \geqslant \underline{\varepsilon}(h'_n) = \delta$ and $\tilde{\sigma}^\varepsilon_n(s_n|h''_n) \leqslant \bar{\varepsilon}(h''_n) = \delta^2$. Hence $\lim_{\varepsilon\downarrow 0} \tilde{\sigma}^\varepsilon_n(s_n|h''_n)/\tilde{\sigma}^\varepsilon_n(\tau_n|h'_n) = 0$. Therefore $\tilde{\sigma}^k_n(s_n) = 0$ for all $k \leqslant k^*_n$, which proves the Claim. □

From the LPS $\tilde{\mathcal{L}}$ construct for each player n an LPS $\mathcal{L}_n = (\sigma^0_n, \sigma^1_n, \dots, \sigma^{k^*_n}_n)$ for the game G by letting σ^k_n be the mixed strategy in Σ_n that is equivalent to $\tilde{\sigma}^k_n$. Since $\tilde{\sigma}^0 \in \tilde{\Sigma}^0$, σ^0 belongs to Σ^0. By the definition of k^*_n and $\mathcal{L}_n$, there exists $\sigma'_n \in \Sigma_n$ such that $\sigma^{k^*_n}_n = \lambda_n\tau_n + [1-\lambda_n]\sigma'_n$, where λ_n is the probability of the strategy τ_n in $\tilde{\sigma}^{k^*_n}_n$. By the previous Claim, if a pure strategy is in the support of σ^k for $k < k^*_n$, or in the support of σ'_n when $\lambda_n \neq 1$, then it is a lexicographic best reply to the LPS $\mathcal{L} \equiv (\mathcal{L}_1, \mathcal{L}_2)$. Thus $\mathcal{L}$ satisfies the condition of Corollary 3.2 for τ. Hence Σ° contains a KM-stable set. □

4. N-Player Games

This section provides the proof of Theorem 1.1 for the general case with N players. We begin with some definitions. For a real-valued analytic function (or more generally a power series) $f(t) = \sum_{i=0}^\infty a_i t^i$ in a single variable t, the order of f, denoted $o(f)$, is the smallest integer i such that $a_i \neq 0$. The order of the zero function is $+\infty$. It follows that for any two power series f and g, $o(fg) = o(f) + o(g)$ and $o(f+g) \geqslant \min(o(f), o(g))$. We say that a power series f is positive if $a_{o(f)} > 0$; thus if f is an analytic function then f is positive if and only if $f(t)$ is positive for all sufficiently small $t > 0$. For two analytic functions $f(t)$ and $g(t)$, say that $f > g$ iff $f - g$ is positive.

By a slight abuse of terminology, we call a function $F : [0, \bar{t}] \to X$, where X is a subset of a Euclidean space $\mathbb{R}^l$, analytic if there exists an analytic function $F' : (-\delta, \delta) \to \mathbb{R}^l$, $\delta > \bar{t}$, such that F' agrees with F on $[0, \bar{t}]$. For an analytic function $F : [0, \bar{t}] \to \mathbb{R}^k$, the order $o(F)$ of F is $\min_i o(F_i)$. If $\sigma : [0, \bar{t}] \to \Sigma$ is an analytic function then for each pure strategy s_n of player n his payoff $G_n(\sigma_{-n}(t), s_n)$ in the game G is an analytic function as well, since payoff functions are multilinear in mixed strategies. We say that s_n is a best reply of order k for player n against an analytic function σ if for all $s'_n \in S_n$, $G_n(\sigma_{-n}(t), s_n) - G_n(\sigma_{-n}(t), s'_n)$ is either nonnegative or has order at least $k+1$; s_n is a best reply to σ if it is a best reply of order ∞.

Lemma 4.1. *Suppose $\sigma, \tau : [0, \bar{t}] \to \Sigma$ are two analytic functions such that $o(\sigma - \tau) > k$. If s_n is not a best reply of order k against σ then it is not a best reply of order k against τ.*

Proof. Let s'_n be a pure strategy such that $G_n(\sigma_{-n}(t), s_n) - G_n(\sigma_{-n}(t), s'_n)$ is negative and has order, say, $l \leqslant k$. Let $\tau' = \tau - \sigma$. We can then write $G_n(\tau_{-n}(t), s_n) - G(\tau_{-n}(t), s'_n)$ as

$$G_n(\sigma_{-n}(t), s_n) - G(\sigma_{-n}(t), s'_n) + \sum_{s_{-n}} \sum_{N' \subsetneqq N \backslash \{n\}} \left(\prod_{n' \in N'} \sigma_{n', s_{n'}}(t) \prod_{n'' \in N \backslash (N' \cup \{n\})} \tau'_{n'', s_{n''}}(t) \right) [G_n(s_{-n}, s_n) - G_n(s_{-n}, s'_n)].$$

The first term in the above expression is negative and has order l by assumption. Therefore, to prove the result it is enough to show that the order of the double summation is at least $k+1$: it then follows the whole expression is negative and has order l. To prove this last statement, using the above mentioned property of the order of sums of power series, it is sufficient to show that each of the summands in the second term has order at least k+1. Consider now a summand for a fixed s_{-n} and $N' \subsetneqq N \backslash \{n\}$. If both s_n and s'_n give the same payoff against s_{-n} then the order of this term is ∞. Otherwise, using the property of the order of products of functions, the order of this term is

$$\sum_{n' \in N'} o(\sigma_{n', s_{n'}}) + \sum_{n'' \notin (N' \cup \{n\})} o(\tau'_{n'', s_{n''}}) > k,$$

where the inequality follows from the following two facts: (i) the order of each $\sigma_{n', s_{n'}}$ is at least zero; and (ii) there exists at least one $n'' \notin (N' \cup \{n\})$ and for any such n'' the order of $\tau'_{n'', s_{n''}}$ is greater than k by assumption. □

We use the following version of a result of Blume, Brandenberger, and Dekel (1991).

Lemma 4.2. *If the map $\tau_n : [0, \bar{t}] \to \Sigma_n$ is analytic then $\tau_n(t) = \sum_{k=0}^{K} f_n^k(t) \tau_n^k$, where $K \leqslant |S_n|$, each τ_n^k is in Σ_n, and each map $f_n^k : [0, \bar{t}] \to \mathbb{R}_+$ is analytic.*

Proof. Let $\tau_n^0 = \tau_n(0)$ and $S_n^0 = \operatorname{supp} \tau_n^0$. Define $f_n^0(t)$ to be $\min_{s \in S_n^0} \tau_{n,s}(t)/\tau_{n,s}(0)$; and let $\tau_n^1(t) = [1 - f_n^0(t)]^{-1}[\tau_n(t) - f_n^0(t)\tau_n^0]$. It follows from the definitions of f_n^0 and $\tau_n^1(t)$ that the latter is an analytic function from $[0, \bar{t}]$ into Σ_n for which there exists $s \in S_n$ such that $\tau_{n,s}(t) > 0$ while $\tau_{n,s}^1(t) = 0$. Moreover, $\tau_n(t) = f_n^0(t)\tau_n^0 + [1 - f_n^0(t)]\tau_n^1(t)$. Now let $\tau_n^1 = \tau_n^1(0)$ and $S_n^1 = \operatorname{supp} \tau_n^1$. Define $\hat{f}_n^1(t)$, as before, to be $\min_{s \in S_n^1} \tau_{n,s}^1(t)/\tau_{n,s}^1(0)$; $\tau_n^2(t) = [1 - \hat{f}_n^1(t)]^{-1}[\tau_n^1(t) - \hat{f}_n^1(t)\tau_n^1]$; and $f_n^1(t) = [1 - f_n^0(t)]\hat{f}_n^1(t)$. Then $\tau_n(t) = f_n^0(t)\tau_n^0 + f_n^1(t)\tau_n^1 + [1 - f_n^0(t)][1 - f_n^1(t)]\tau_n^2(t)$. Likewise, we can obtain mixed strategies τ_n^3, etc., and corresponding coefficients $f_n^3(t)$, etc. This process must terminate in a finite number of steps since for each k there exists an $s \in S_n$ for which $\tau_{n,s}^k(t)$ is positive but $\tau_{n,s}^l(t)$ is zero for all $l > k$. □

Theorem 4.3. *If a solution satisfies Weak Invariance and Strong Backward Induction then for any game a selected set includes a KM-stable subset of its normal form.*

Proof. We show that if a solution selects a set $\Sigma^\circ \subset \Sigma$ of profiles that does not contain a KM-stable set for the normal-form game G then it satisfies Weak Invariance only if it violates Strong Backward Induction.

Suppose Σ° does not contain a KM-stable set. Then there exists $\epsilon > 0$ such that for each $\delta \in (0,1)$ there exists $\eta \in P_\delta \backslash \partial P_\delta$ such that the perturbed game $G(\eta)$ does not have an equilibrium in the ϵ-neighborhood U of Σ°. Take a sufficiently fine simplicial subdivision of Σ such that the union X of those simplices intersecting Σ° is contained in U. X is then a neighborhood of Σ^0. Let $A = \{(\lambda,\tau) \in (0,1)^N \times (\Sigma\backslash\partial\Sigma) \mid G(\lambda,\tau)$ has no equilibrium in $X\}$; then A is nonempty and there exists $\tau^\circ \in \Sigma$ such that $(0,\tau^\circ)$ is in the closure of A. Further, since X is semi-algebraic, A too is semi-algebraic. Therefore, by the Nash Curve Selection Lemma (cf. Bochnak, Coste, and Roy, 1998, Proposition 8.1.13), there exists $\bar{t} > 0$ and a semialgebraic, analytic map $t \mapsto (\lambda(t),\tau(t))$ from $[0,\bar{t}]$ to $[0,1]^N \times \Sigma$ such that $(\lambda(0),\tau(0)) = (0,\tau^\circ)$ and $(\lambda(t),\tau(t)) \in A$ for all $t \in (0,\bar{t}]$. Define the compact semi-algebraic set

$$Y = \{(t,\sigma) \in [0,\bar{t}] \times X \mid (\forall\ s_n \in S_n)\ \sigma_{n,s_n} \geqslant \lambda_n(t)\tau_{n,s_n}(t)\}\,.$$

Claim 4.4. *There exists a positive integer p such that for every analytic function $\zeta \mapsto (t(\zeta),\sigma(\zeta))$ from an interval $[0,\bar{\zeta}]$ to Y, where $t(\zeta)$ is positive, there exists a player n and a pure strategy $s_n \in S_n$ such that $\sigma(\zeta) > \lambda_n(t(\zeta))\tau_{n,s_n}(t(\zeta))$ and s_n is not a best reply of order $o(t(\zeta))p$ against $\sigma(\zeta)$.*

Proof of Claim. Define the maps $\alpha,\beta : Y \to \mathbb{R}$ via

$$\alpha(t,\sigma) = \max_{n,s_n \in S_n} \left\{[\sigma_n(s_n) - \lambda_n(t)\tau_{n,s_n}(t)] \times \max_{s'_n \in S_n}[G_n(s'_n,\sigma_{-n}) - G_n(s_n,\sigma_{-n})]\right\}$$

and $\beta(t,\sigma) = t$. By construction, $\alpha,\beta \geqslant 0$ and $\alpha^{-1}(0) \subseteq \beta^{-1}(0)$. By Lojasiewicz's inequality (see Bochnak et al., 1998, Corollary 2.6.7) there exist a positive scalar c and a positive integer p such that $c\alpha \geqslant \beta^p$. Given an analytic map $\zeta \mapsto (t(\zeta),\sigma(\zeta))$ as in the statement of the theorem, observe that for each n, s_n, s'_n, $\sigma_{n,s_n}(\zeta) - \lambda(t(\zeta))\tau_{n,s_n}(t(\zeta))$ and $G_n(s'_n,\sigma_{-n}(\zeta)) - G_n(s_n,\sigma_{-n}(\zeta))$ are also analytic in ζ. Therefore there exists a pair n, s_n that achieves the maximum in the definition of α for all small ζ. Then

$$\max_{s'_n}[G_n(s'_n,\sigma_{-n}(\zeta)) - G_n(s_n,\sigma_{-n}(\zeta))] \geqslant \alpha(t(\zeta),\sigma(\zeta)) \geqslant (t(\zeta))^p/c\,,$$

where the first inequality follows from the fact that $\sigma_{n,s_n}(\zeta) - \lambda_n(t(\zeta))\tau_{n,s_n}(t(\zeta)) \leqslant 1$. By assumption, $t(\zeta)$ is positive. Therefore, $\max_{n,s_n,s'_n}[G_n(s'_n,\sigma_{-n}(\zeta)) - G_n(s_n,\sigma_{-n}(\zeta))]$ is also a positive analytic function and, being greater than $c^{-1}(t(\zeta))^p$, has order at most $o(t(\zeta))p$. □

Using Lemma 4.2, express each $\tau_n(t)$ as the sum $\sum_{k=0}^{K_n} f_n^k(t)\tau_n^k$, where each τ_n^k is a mixed strategy in Σ_n and $f_n^k : [0,\bar{t}] \to \mathbb{R}_+$ is analytic. Construct the game Γ in extensive form in which each player n chooses among the following, while remaining uninformed of the others' choices. Player n first chooses whether to commit to the mixed strategy τ_n^0 or not; if not then n chooses between τ_n^1 or not, and so on for $k = 2,\ldots,K_n$; and if n does not commit to any strategy τ_n^k then n chooses among the pure strategies in S_n. Since the normal form of Γ is an inflation of G, Weak Invariance implies that for the game Γ the solution selects a subset of those strategies equivalent to Σ°. For perturbations of the game Γ use the following action perturbation: for the information set where n chooses between τ_n^k or not, use $\underline{\varepsilon}_n^k(t) = \bar{\varepsilon}_n^k(t) = \lambda_n(t)f_n^k(t)$; and at the information set where n chooses among the strategies in S_n, use $\underline{\varepsilon}_n^{K_n+1}(t) = \bar{\varepsilon}_n^{K_n+1}(t) = t^{p+1}$.

Let $\tilde{S}$ and $\tilde{\Sigma}$ be the sets of pure and mixed-strategy profiles in Γ. (As in the two-person case, for each player n and each $0 \leqslant k \leqslant K_n$ we identify all strategies of n that choose, at the relevant information set, to play the strategy τ_n^k.) Let E be the set of $(t,\sigma) \in (0,\bar{t}] \times \tilde{\Sigma}$ such that σ is an $\varepsilon(t)$-quasi-perfect equilibrium of Γ (i.e., satisfying conditions 1 and 2 of Definition 2.2) whose reduced-form strategy profile in Σ lies in

X. Since the minimum error probabilities are analytic functions of t, E is a semi-analytic set.[4] Strong Backward Induction requires that there exists $\tilde{\sigma}^0 \in \Sigma^*$ such that the reduced form of $\tilde{\sigma}^0$ belongs to Σ^0 and $(0, \tilde{\sigma}^0)$ belongs to the closure of E. By the Curve Selection Lemma (cf. Lojasiewicz, 1993, II.3), there exists an analytic function $\zeta \mapsto (t(\zeta), \tilde{\sigma}(\zeta))$ from $[0, \bar{\zeta}]$ to $[0, \bar{t}] \times \tilde{\Sigma}$ such that $(t(\zeta), \tilde{\sigma}(\zeta)) \in E$ for all $\zeta > 0$ and $(t(0), \tilde{\sigma}(0)) = (0, \tilde{\sigma}^0)$. By construction, $t(\zeta)$ is nonconstant, i.e., $0 < o(t(\zeta)) < \infty$.

From $\tilde{\sigma}(\zeta)$ construct the analytic function $\hat{\sigma}(\zeta)$ as follows: for each player n, choose a strategy s_n^* in Γ such that $o(\tilde{\sigma}_{n,s_n^*})$ is zero—i.e., a strategy in the support of $\tilde{\sigma}_n(0)$. Let S_n' be the set of all pure strategies s_n of the original game G that are chosen with the minimum probability in $\tilde{\sigma}(t)$ (i.e., with probability $(t(\zeta))^{p+1}$); let $\hat{\sigma}_{n,s_n}(\zeta) = 0$ for each $s_n \in S_n'$; define $\hat{\sigma}_{n,s_n^*}(\zeta) = \tilde{\sigma}_{n,s_n^*}(\zeta) + |S_n'|(t(\zeta))^{p+1}$; and finally, let the probabilities of the other strategies in $\hat{\sigma}$ be the same as in $\tilde{\sigma}$. Obviously, $o(\tilde{\sigma} - \hat{\sigma}) \geqslant o(t(\zeta))(p+1) > o(t(\zeta))p$, where the second inequality follows from the fact that $0 < o((t\zeta)) < \infty$.

If $\hat{\sigma}_{n,s_n}(\zeta) > 0$ for some $s_n \in S_n$ then s_n is a best reply against $\tilde{\sigma}(\zeta)$; hence by Lemma 4.1, s_n is a best reply of order $o(t(\zeta))p$ against $\hat{\sigma}(\zeta)$. Likewise, for each k the strategy s_n that plays τ_n^k at the appropriate information set is optimal of order $o(t(\zeta))p$ against $\hat{\sigma}_n(\zeta)$ if $\hat{\sigma}_{n,s_n}(\zeta) > \lambda_n(t(\zeta))f_n^k(t(\zeta))$.

Let $\sigma(\zeta)$ be the reduced form of $\hat{\sigma}(\zeta)$ in the game G. Then we have a well-defined analytic function $\varphi : [0, \bar{\zeta}] \to Y$, given by $\varphi(\zeta) = (t(\zeta), \sigma(\zeta))$: indeed, by definition, $\sigma(\zeta)$ is contained in X; also, for each n and $s_n \in S_n$, $\sigma_{n,s_n}(\zeta) \geqslant \lambda_n(t(\zeta))\tau_{n,s_n}(t(\zeta))$, since in $\tilde{\sigma}(\zeta)$ (and therefore in $\hat{\sigma}(\zeta)$) the "pure" strategy τ_n^k is chosen with probability at least $\lambda_n(t(\zeta))f_n^k(t(\zeta))$. Therefore, by the above Claim, there exist n, s_n such that $\sigma_n(\zeta)$ assigns s_n more than the minimum probability even though it is not a best reply of order $o(t(\zeta))p$ against $\sigma_n(\zeta)$ (and $\hat{\sigma}(\zeta)$). By the definition of $\sigma(\zeta)$ and $\hat{\sigma}(\zeta)$, either (i) s_n is assigned a positive probability by $\hat{\sigma}(\zeta)$ or (ii) a strategy τ_n^k—containing s_n in its support, when viewed as a mixed strategy in Σ_n—is assigned a probability greater than $\lambda_n(\tau(\zeta))f_n^k(t(\zeta))$, even though it is not a best reply of order $o(t(\zeta))p$ against $\hat{\sigma}(\zeta)$, which contradicts the conclusion of the previous paragraph. In the game Γ, therefore, for any sequence of sufficiently small t there cannot be a sequence of $\{\varepsilon(t)\}$-quasi-perfect profiles whose reduced forms are in X. Thus Strong Backward Induction is violated. □

5. Concluding Remarks

We accept the arguments for invariance, and by implication Weak Invariance, adduced by KM as entirely convincing—to do otherwise would reject a cornerstone of decision theory. Our results differ from KM primarily in using quasi-perfection to specify a strong form of backward induction. In spite of its awkward name, quasi-perfection seems to be an appropriate refinement of weaker forms of backward induction such as sequential equilibrium. Some strengthening is evidently necessary since a sequential equilibrium can use inadmissible strategies and strategies that are dominated in the continuation from an information set, and compared to perfection, quasi-perfection avoids pathologies from a player's anticipation of his own trembles at subsequent information sets. However, one might conjecture that similar conclusions could be derived from a formulation in which Strong Backward Induction requires only that a selected set include for each extensive form an equilibrium of the agent-normal form that excludes inadmissible strategies and that evaluates each agent's conditional payoff at its information set as the player's continuation payoff. Alternatively, the reader may have noticed that Strong Backward Induction is used in the proofs mainly to establish existence of

[4] $A \subseteq \mathbb{R}^k$ is semi-analytic if for all $x \in \mathbb{R}^k$, there exists a neighborhood U of x, such that $A \cap U$ is a finite union of sets of the form $\{y \in U \mid f_1(y) = \cdots = f_m(y) = 0, g_1(y) > 0, \ldots, g_l(y) > 0\}$ where $f_1, \ldots f_m$, $g_1, \ldots, g_l$ are analytic on U.

lexicographic probability systems that "respect preferences" as defined by Blume, Brandenberger, and Dekel (1991). Thus Axiom 2.3 might state directly that each sequence of perturbations of an extensive form should refine the selected set by selecting a lexicographic equilibrium that respects preferences, as in statement 3 of Theorem 3.1 and Corollary 3.2 for two-player games. It seems plausible that quasi-perfection can be characterized in terms of a lexicographic equilibrium with the requisite properties.

Theorems 3.3 and 4.3 remain true if Axiom 2.3 is replaced by the requirement that a selected set must include, for each information set h of each player, an equilibrium that provides a quasi-proper continuation from h. That is, it is the limit of ε-quasi-proper equilibria for which conditions 1 and 2 of Definition 2.2 are replaced by the requirement that if the expected continuation payoff from h for $a \in A_n(h)$ is less than it is for a' then $\varepsilon_h^{|A_n(h)|} \leqslant \sigma_n^\varepsilon(a|h) < \varepsilon_h \sigma_n^\varepsilon(a'|h)$. In Govindan and Wilson (2002) we establish for a generic class of sender-receiver signaling games an analog of Theorem 3.3, but rather than Axiom 2.3 we assume that the selected set contains an equilibrium that is quasi-proper in continuation from each information set of the receiver. Actually, for generic signaling, outside-option, and perfect information games it suffices to replace Axiom 2.3 by the requirement that a selected set contains a proper equilibrium.

In a companion paper (Govindan and Wilson, 2004) we show that the axioms invoked here imply a version of Hillas' (1996) conjecture that invariance and backward induction imply forward induction; that is, a selection is not affected by deleting a strategy that is inferior at every equilibrium in the selected set. This is also the gist of the "intuitive criterion" proposed by Cho and Kreps (1987) for signaling games, and its extension to the solution concept of "divinity" proposed by Banks and Sobel (1987).

References

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

Blume, L., A. Brandenberger, and E. Dekel (1991), "Lexicographic Probabilities and Choice Under Uncertainty" and "Lexicographic Probabilities and Equilibrium Refinements," *Econometrica*, 59: 61-98.

Bochnak, J., M. Coste, and M-F. Roy (1998), *Real Algebraic Geometry.* Berlin: Springer-Verlag.

Cho, I., and D. Kreps (1987), "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102: 179-202.

van Damme, E. (1984), "A Relation between Perfect Equilibria in Extensive Form Games and Proper Equilibria in Normal Form Games," *International Journal of Game Theory*, 13: 1-13.

Govindan, S., and R. Wilson (2001), "Direct Proofs of Generic Finiteness of Nash Equilibrium Outcomes", *Econometrica*, 69(3): 765-769.

Govindan, S., and R. Wilson (2002), "Invariance of Stable Equilibria of Signaling Games," *Festshrift in Honor of Steinar Ekern.* Bergen, Norway: Norwegian School of Economics and Business Administration.

Govindan, S., and R. Wilson (2004), "The Principle of Forward Induction," to appear.

Hillas, J. (1996), "How Much of Forward Induction is Implied by Backward Induction and Ordinality," University of Aukland, New Zealand, mimeo.

Hillas, J., and E. Kohlberg (2002), "Foundations of Strategic Equilibrium," in R. Aumann and S. Hart (eds.), *Handbook of Game Theory*, Volume III, Chapter 42, 1597-1663. Amsterdam: Elsevier.

Kohlberg, E., and J. Mertens (1986), "On the Strategic Stability of Equilibria," *Econometrica*, 54: 1003-1038.

Kreps, D., and R. Wilson (1982), "Sequential Equilibria," *Econometrica*, 50: 863-894.

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.

Lojasiewicz, S. (1993), "Sur la géométrie semi- et sous- analytique," *Annales de l'institut Fourier*, 43: 1575-1595.

Mertens, J. (1989), "Stable Equilibria—A Reformulation, Part I: Definition and Basic Properties," *Mathematics of Operations Research*, 14, 575-625.

Myerson, R. (1978), "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7: 73-80.

Nash, J. (1950), "Equilibrium Points in n-Person Games," *Proceedings of the National Academy of Sciences USA*, 36: 48-49. Reprinted in H. Kuhn (ed.), *Classics in Game Theory*, Princeton University Press, Princeton, New Jersey, 1997.

Selten, R. (1975), "Reëxamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4: 25-55. Reprinted in H. Kuhn (ed.), *Classics in Game Theory*, Princeton University Press, Princeton, New Jersey, 1997.

Thompson, F. (1952), "Equivalence of Games in Extensive Form," RM-759, RAND Corporation, Santa Monica, California. Reprinted in H. Kuhn (ed.), *Classics in Game Theory*, Princeton University Press, Princeton, New Jersey, 1997.

Economics Department, 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