A STRUCTURE THEOREM FOR RATIONALIZABILITY IN INFINITE-HORIZON GAMES

JONATHAN WEINSTEIN AND MUHAMET YILDIZ

Abstract. We show that in any game that is continuous at infinity, if a plan of action \( a_i \) is rationalizable for a type \( t_i \), then there are perturbations of \( t_i \) for which following \( a_i \) for an arbitrarily long future is the only rationalizable plan. One can pick the perturbation from a finite type space with common prior. As an application we prove an unusual folk theorem: Any individually rational and feasible payoff is the unique rationalizable payoff vector for some perturbed type profile.

\( JEL \) Numbers: C72, C73.

1. Introduction

In economic applications involving infinite-horizon dynamic games, the sets of equilibrium strategies and rationalizable strategies are often very large. For example, in the literature on repeated games there are folk theorems concluding that every individually rational payoff can be supported by a subgame-perfect equilibrium. For a less transparent example, in Rubinstein’s (1982) bargaining game, although there is a unique subgame-perfect equilibrium, any outcome can be supported in Nash equilibrium. Consequently, economists focus on strong refinements of equilibrium and ignore other rationalizable strategies and equilibria. This is so common that we rarely think about rationalizable strategies in extensively-analyzed dynamic games. Of course, all of these applications make strong common-knowledge assumptions.

In this paper, building on existing theorems for finite games, we prove a structure theorem for rationalizability in infinite-horizon dynamic games that characterizes the robust predictions of any refinement. The attraction of our new result is that it is readily applicable to most economic applications. We discuss two immediate applications, one to repeated games.

\footnotesize
\textit{Date}: First Version: March 2009; This Version: January 2010.

This paper was written when the authors were members at the Institute for Advanced Study. We thank them for their generous financial support and hospitality.
(with sufficiently patient players) and one to bargaining, showing that no refinement can robustly rule out any individually rational outcome in these games.

We consider an arbitrary dynamic game that is continuous at infinity, has finitely many moves at each information set, and has a finite type space. Note that virtually all games analyzed in economics, such as repeated games with discounting and bargaining games, are continuous at infinity. For any type $t_i$ in this game, consider a rationalizable plan of action $a_i$, which is a complete contingent plan that determines which move the type $t_i$ will take at any given information set of $i$. Fix any integer $L$. We show that, by perturbing the interim beliefs of type $t_i$, we can find a new type $\hat{t}_i$ who plays according to $a_i$ in the first $L$ information sets in any rationalizable plan. By perturbation, we mean that types $t_i$ and $\hat{t}_i$ have similar beliefs about the payoff functions, similar beliefs about the other players’ beliefs about the payoff functions, similar beliefs about the other players’ beliefs about the players’ beliefs about the payoff functions, and so on, up to an arbitrarily chosen finite order. Moreover, we can pick $\hat{t}_i$ from a finite model with a common prior, so our perturbations do not rely on some esoteric large type space or the failure of the common-prior assumption.

In Weinstein and Yildiz (2007) we showed this result for finite-action games in normal form, under the assumption that the space of payoffs is rich enough so that any action can be dominant under some payoff specification. While this richness assumption holds when one relaxes all common-knowledge assumptions on payoff functions in a static game, it fails if one fixes a non-trivial dynamic game tree. This is because a plan of action cannot be strictly dominant when some information sets may not be reached. Chen (2008) has nonetheless extended the structure theorem to finite dynamic games, showing that the same result holds under the weaker assumption that all payoff functions on the terminal histories are possible. This is an important extension, but the finite-horizon assumption rules out many major dynamic applications of game theory, such as repeated games and sequential bargaining.

\[\text{1The usual notation in dynamic games and games of incomplete information clash. In a dynamic game } a_i \text{ would usually be a single move while } s_i \text{ would be a complete plan; but the fact that this plan is contingent on information suggests using } a_i \text{ for the plan and } s_i \text{ for the function from type to plan. Also, } t \text{ stands for time in dynamic games but type profile in incomplete-information games; } h_i \text{ stands for history in dynamic games but hierarchy in incomplete-information games, etc. Following Chen (2008), we will use the notation customary in incomplete-information games, so } a_i \text{ is a complete contingent plan of action, which we will sometimes call a “plan”. We will sometimes say “move” to distinguish an action at a single node.} \]
Since the equilibrium strategies can discontinuously expand when one switches from finite- to infinite-horizon, as in the repeated prisoners’ dilemma game, it is not clear what the structure theorem for finite-horizon game implies in those applications. Here, we extend Chen’s results further by allowing infinite-horizon games that are continuous at infinity, as are nearly all standard applications. There is a challenge in this extension, because the construction employed by Weinstein and Yildiz (2007) and Chen (2008) relies on the assumption that there are finitely many actions. The finiteness (or countability) of the action space is used in a technical but crucial step of ensuring that the constructed type is well-defined, and there are counterexamples to that step when the action space is uncountable. Unfortunately, in infinite-horizon games, such as the infinitely-repeated prisoners dilemma, there are uncountably many strategies, even in reduced form. However, continuity at infinity turns out to be enough to make infinite-horizon games behave well enough for the result to carry over.

We now briefly explain the implications of our structure theorem to robustness. Imagine a researcher who subscribes to an arbitrary refinement of rationalizability, such as sequential equilibrium or proper equilibrium. Applying his refinement, he can make many predictions about the outcome of the game, describing which histories we may observe. Let us confine ourselves to predictions about finite-length (but arbitrarily long) outcome paths. For example, in the repeated prisoners’ dilemma game, “players cooperate in the first round” and “player 1 plays tit-for-tat in the first 1,000,000 periods” are such predictions, but “players always cooperate” and “players eventually defect” are not. Our result implies that any such prediction that can be obtained by a refinement, but not by mere rationalizability, relies crucially on assumptions about the infinite hierarchies of beliefs embedded in the model. Therefore, refinements cannot lead to any new prediction about finite-length outcome paths that is robust to misspecification of interim beliefs.

One can reformulate the main result of this paper in terms of predictions by following the formulation in Weinstein and Yildiz (2007). Here, we will informally illustrate the basic intuition. Suppose that the above researcher observes a “noisy signal” about the players’
first-order beliefs (which are about the payoff functions), the players’ second-order beliefs (which are about the first-order beliefs), \ldots, up to a finite order \( k \), and does not have any information about the beliefs at order higher than \( k \). Here, the researcher’s information may be arbitrarily precise, in the sense that the noise in his signal may be arbitrarily small and \( k \) may be arbitrarily large. Suppose that he concludes that a particular type profile \( t = (t_1, \ldots, t_n) \) is consistent with his information, in that the interim beliefs of each type \( t_i \) could lead to a hierarchy of beliefs that is consistent with his information. Suppose that for this particular specification, his refinement leads to a sharper prediction about the finite-length outcome paths than rationalizability. That is, for type profile \( t \), a particular path (or history) \( h \) of length \( L \) is possible under rationalizability but not possible under his refinement. But there are many other type profiles that are consistent with his information. In order to verify his prediction that \( h \) will not be observed under his refinement, he has to make sure that \( h \) is not possible under his refinement for any such type profile. Otherwise, his prediction would not follow from his information or solution concept; it would rather be based on his modeling choice of considering \( t \) but not the alternatives. Our result establishes that he cannot verify his prediction, and his prediction is indeed based on his choice of modeling: there exists a type profile \( \hat{t} \) that is also consistent with his information and, for \( \hat{t} \), \( h \) is the only rationalizable outcome for the first \( L \) moves. We can then also conclude that \( h \) is the only outcome for the first \( L \) moves according to his refinement.

Our structure theorem has two limitations. First, it only applies to (arbitrarily long) finite-length outcomes. Second, as in one example we will discuss, while the players’ ex ante beliefs are close to those in the original game, their updated beliefs along the unique rationalizable path may be very different, calling into question our notion of perturbation. For this to happen, the perturbed types must find the unique rationalizable outcome unlikely at the beginning of play. By narrowing our focus to Nash equilibria of complete information games, we can prove a stronger structure theorem that does not have these limitations. For any Nash equilibrium of any complete-information game that is continuous at infinity, we show that there exists a profile of perturbed types for which the equilibrium is the unique rationalizable action profile and the perturbed types assign nearly probability one to the equilibrium path.
As an application of this stronger result and the usual folk theorems, we show an unusual folk theorem. We show that every payoff $v$ in the interior of the set of individually rational and feasible payoffs can be the unique rationalizable outcome of some perturbation for sufficiently patient players. Moreover, in the actual situation described by the perturbation, all players anticipate that the payoffs are within $\varepsilon$ neighborhood of $v$. That is, the complete-information game is surrounded by types with a unique solution, but the unique solution varies in such a way that it traces all individually rational and feasible payoffs. While the multiplicity in usual folk theorems may suggest a need for a refinement, the multiplicity in our folk theorem emphasizes the impossibility of a robust refinement. In the same vein, in Rubinstein’s bargaining model, we show that any bargaining outcome can be supported as a unique rationalizable outcome for some perturbation. Once again, no refinement can rule out these outcomes without imposing a common knowledge assumption.

After laying out the model in the next section, we present our general results in Section 3. We present applications to repeated games and bargaining in Sections 4 and 5, respectively. We discuss the relation of our general results to broader literature on robustness in Section 6. The proofs of our general results are in the appendix.

2. Basic Definitions

This section will need to introduce notation for both dynamic Bayesian games and hierarchies of beliefs, and will be a bit tedious as a result. We suggest that the reader skim the section quickly and refer back as necessary. The main text is not very notation-heavy.

**Extensive-form games.** We consider standard $n$-player extensive-form games with possibly infinite horizon, as modeled in Osborne and Rubinstein (1994). In particular, we fix an extensive game form $\Gamma = (N, H, (I_i)_{i \in N})$ with perfect recall where $N = \{1, 2, \ldots, n\}$ is a finite set of players, $H$ is a set of histories, and $I_i$ is the set of information sets at which player $i \in N$ moves. We use $i \in N$ and $h \in H$ to denote a generic player and history, respectively. We write $I_i(h)$ for the information set that contains history $h$, at which player $i$ moves, i.e. the set of histories $i$ finds possible when he moves. The set of available moves at $I_i(h)$ is denoted by $B_i(h)$. We have $B_i(h) = \{b_i : (h, b_i) \in H\}$, where $(h, b_i)$ denotes the history in which $h$ is followed by $b_i$. We assume that $B_i(h)$ is finite for each $h$. An action (or plan) $a_i$ of $i$ is defined as a function that maps the information sets of $i$ to the moves available at
those information sets; i.e., \( a_i : I_i(h) \mapsto a_i(h) \in B_i(h) \). We write \( A = A_1 \times \cdots \times A_n \) for the set of action profiles \( a = (a_1, \ldots, a_n) \).\(^3\) We write \( Z \) for the set of terminal nodes, at which no player moves. We write \( z(a) \) for the terminal history that is reached by profile \( a \). We say that actions \( a_i \) and \( a'_i \) are equivalent if \( z(a_i, a_{-i}) = z(a'_i, a_{-i}) \) for all \( a_{-i} \in A_{-i} \).

**Type spaces.** Given an extensive game form, a Bayesian game is defined by specifying the belief structure about the payoffs. To this end, we write \( \theta(z) = (\theta_1(z), \ldots, \theta_n(z)) \in [0, 1]^n \) for the payoff vector at the terminal node \( z \in Z \) and write \( \Theta^* \) for the set of all payoff functions \( \theta : Z \rightarrow [0, 1]^n \). The payoff of \( i \) from an action profile \( a \) is denoted by \( u_i(\theta, a) \).

Note that \( u_i(\theta, a) = \theta_i(z(a)) \). We endow \( \Theta^* \) with the product topology (i.e. the topology of pointwise convergence). Note that \( \Theta^* \) is compact and \( u_i \) is continuous in \( \theta \). Note, however, that \( \Theta^* \) is not a metric space. We will use only finite type spaces, so by a model, we mean a finite set \( \Theta \times T_1 \times \cdots \times T_n \) associated with beliefs \( \kappa_{t_i} \in \Delta(\Theta \times T_{-i}) \) for each \( t_i \in T_i \), where \( \Theta \subseteq \Theta^* \). Here, \( t_i \) is called a type and \( T = T_1 \times \cdots \times T_n \) is called a type space. A model \((\Theta, T, \kappa)\) is said to be a common-prior model (with full support) if and only if there exists a probability distribution \( p \in \Delta(\Theta \times T) \) with support \( \Theta \times T \) and such that \( \kappa_{t_i} = p(\cdot|t_i) \) for each \( t_i \in T_i \). Note that \((\Gamma, \Theta, T, \kappa)\) defines a Bayesian game. In this paper, we consider games that vary by their type spaces, for a fixed game-form \( \Gamma \).

**Hierarchies of Beliefs.** Given any type \( t_i \) in a type space \( T \), we can compute the first-order belief \( h_t^1(t_i) \in \Delta(\Theta^*) \) of \( t_i \) (about \( \theta \)), second-order belief \( h_t^2(t_i) \in \Delta(\Theta^* \times \Delta(\Theta^*)^n) \) of \( t_i \) (about \( \theta \) and the first-order beliefs), etc., using the joint distribution of the types and \( \theta \). Using the mapping \( h_i : t_i \mapsto (h_t^1(t_i), h_t^2(t_i), \ldots) \), we can embed all such models in the universal type space, denoted by \( T^* = T_1^* \times \cdots \times T_n^* \) (Mertens and Zamir (1985) and Brandenburger and Dekel (1993)). We endow the universal type space with the product topology of the usual weak convergence. We say that a sequence of types \( t_i(m) \) converges to a type \( t_i \), denoted by \( t_i(m) \rightarrow t_i \), if and only if \( h_t^k(t_i(m)) \rightarrow h_t^k(t_i) \) for each \( k \), where the latter convergence is in the weak topology, i.e., “convergence in distribution.”

\(^3\)Notation: Given any list \( X_1, \ldots, X_n \) of sets, write \( X = X_1 \times \cdots \times X_n \) with typical element \( x \), \( X_{-i} = \prod_{j \neq i} X_j \) with typical element \( x_{-i} \), and \( (x'_i, x_{-i}) = (x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n) \). Likewise, for any family of functions \( f_j : X_j \rightarrow Y_j \), we define \( f_{-i} : X_{-i} \rightarrow X_{-i} \) by \( f_{-i}(x_{-i}) = (f_j(x_j))_{j \neq i} \). This is with the exception that \( h \) is a history as in dynamic games, rather than a profile of hierarchies \((h_1, \ldots, h_n)\). Given any topological space \( X \), we write \( \Delta(X) \) for the space of probability distributions on \( X \), endowed with Borel \( \sigma \)-algebra and the weak topology.
For each $i \in N$ and for each belief $\pi \in \Delta (\Theta \times A_{-i})$, we write $BR_i(\pi)$ for the set of actions $a_i \in A_i$ that maximize the expected value of $u_i(\theta, a_i, a_{-i})$ under the probability distribution $\pi$.

**Interim Correlated Rationalizability.** For each $i$ and $t_i$, set $S^0_i[t_i] = A_i$, and define sets $S^k_i[t_i]$ for $k > 0$ iteratively, by letting $a_i \in S^k_i[t_i]$ if and only if $a_i \in BR_i(\text{Marg}_{\Theta \times A_{-i}} \pi)$ for some $\pi \in \Delta (\Theta \times T_{-i} \times A_{-i})$ such that $\text{Marg}_{\Theta \times T_{-i}} \pi = \kappa_{t_i}$ and $\pi (a_{-i} \in S^{k-1}_{-i}[t_{-i}]) = 1$. That is, $a_i$ is a best response to a belief of $t_i$ that puts positive probability only on the actions that survive the elimination in round $k - 1$. We write $S^{k-1}_{-i}[t_{-i}] = \prod_{j \neq i} S^{k-1}_j[t_j]$ and $S^k[t] = S^k_1[t_1] \times \cdots \times S^k_n[t_n]$. The set of all rationalizable actions for player $i$ with type $t_i$ is

$$S_i^\infty[t_i] = \bigcap_{k=0}^\infty S^k_i[t_i].$$

This definition of interim correlated rationalizability (ICR) is due to Dekel, Fudenberg, and Morris (2007) (see also Battigalli and Siniscalchi (2003) for a related concept). They show that the ICR set for a given type is completely determined by its hierarchy of beliefs, so we will sometimes refer to the ICR set of a hierarchy or “universal type.” ICR is the weakest rationalizability concept, and hence our results remain true under other notions of rationalizability.

**Continuity at Infinity.** We now turn to the details of the extensive game form. If a history $h = (b')_L^{l=1}$ is formed by $L$ moves for some finite $L$, then $h$ is said to be finite and have length $L$. Otherwise, $h$ is said to be infinite. A game form is said to have finite horizon if for some $L < \infty$ all histories have length at most $L$; the game form is said to have infinite horizon otherwise. For any history $h = (b')_L^{l=1}$ and any $L'$, we write $h^{L'}$ for the subhistory of $h$ that is truncated at length $L'$; i.e., $h = (b')^{\min(L,L')}_{l=1}$. We say that $\theta$ is continuous at infinity (first defined by Fudenberg and Levine (1983)) iff for any $\varepsilon > 0$, there exists $L < \infty$ such that

$$\left| \theta_i(h) - \theta_i(\bar{h}) \right| < \varepsilon$$
for all $i \in N$ and all terminal histories $h, \bar{h} \in Z$. We say that a game $(\Gamma, \Theta, T, \kappa)$ is continuous at infinity if each $\theta \in \Theta$ is continuous at infinity.

We will confine ourselves throughout to games that are continuous at infinity. This includes all the standard cases of repeated games with discounting, bargaining games, etc. Our perturbations will also be continuous at infinity. Of course, our assumption that $B_i(h)$ is
finite restricts the games to finite stage games and finite set of possible offers in repeated games and bargaining, respectively.

3. Structure Theorem

In this section we will present our main result, which shows that in a game that is continuous at infinity, if an action \( a_i \) is rationalizable for a type \( t_i \), then there are perturbations of \( t_i \) for which following \( a_i \) for arbitrarily long future is the only rationalizable plan. As we will explain, we also prove a stronger version of the theorem for outcomes that occur in equilibrium.

In Weinstein and Yildiz (2007) we proved a version of this structure theorem for finite action games. We used a richness assumption on \( \Theta^* \) that is natural for static games but rules out fixing a dynamic extensive game form. Chen (2008) has proven this result for finite-horizon games under a weaker richness assumption that is satisfied in our formulation. The following result is implied by Chen’s theorem.

**Lemma 1** (Weinstein and Yildiz (2007) and Chen (2008)). For any finite-horizon game \( (\Gamma, \Theta, T, \kappa) \), for any type \( t_i \in T_i \) of any player \( i \in N \), any rationalizable action \( a_i \in S_i^\infty [t_i] \), and any neighborhood \( U_i \) of \( h_i(t_i) \) in the universal type space \( T^* \), there exists a hierarchy \( h_i(\hat{t}_i) \in U \) such that for each \( a'_i \in S_i^\infty [\hat{t}_i] \), \( a'_i \) is equivalent to \( a_i \), and \( \hat{t}_i \) is a type in some finite, common-prior model.

That is, if the game has finite horizon, then for any rationalizable action of a given type, we can make the given action uniquely rationalizable (in the reduced game) by perturbing the interim beliefs of the type. Moreover, we can do this by only considering perturbations that come from finite models with a common prior. In the constructions of Weinstein and Yildiz (2007) and Chen (2008), finiteness (or countability) of the action space \( A \) is used in a technical but crucial step that ensures that the constructed type has well-defined beliefs. The assumption ensures that a particular mapping is measurable, and there is no general condition that would ensure the measurability of the mapping when \( A \) is uncountable. Unfortunately, in infinite-horizon games, such as infinitely repeated games, there are uncountably
many histories and plans of action. Our main result in this section extends the above structure theorem to infinite-horizon games. Towards stating the result, we need to introduce one more definition.

**Definition 1.** An action $a_i$ is said to be $L$-equivalent to $a'_i$ iff $z(a_i, a_{-i})^L = z(a'_i, a_{-i})^L$ for all $a_{-i} \in A_{-i}$.

That is, two actions are $L$-equivalent if both actions prescribe the same moves in the first $L$ moves on the path against every action profile $a_{-i}$ by others. For the first $L$ moves, $a_i$ and $a'_i$ can differ only at the informations sets that they preclude. Of course, this is the same as the usual equivalence when the game has a finite horizon that is shorter than $L$. We are now ready to state our main result.

**Proposition 1.** For any game $(\Gamma, \Theta, T, \kappa)$ that is continuous at infinity, for any type $t_i \in T_i$ of any player $i \in N$, any rationalizable action $a_i \in S_i^\infty [t_i]$ of $t_i$, any neighborhood $U_i$ of $h_i(t_i)$ in the universal type space $T^*$, and any $L$, there exists a hierarchy $h_i(\hat{t}_i) \in U_i$, such that for each $a'_i \in S_i^\infty [\hat{t}_i]$, $a'_i$ is $L$-equivalent to $a_i$, and $\hat{t}_i$ is a type in some finite, common-prior model.

As in the introduction, imagine a researcher who wants to model a strategic situation with genuine incomplete information. He can somehow make some noisy observations about the players’ (first-order) beliefs about the payoffs, their (second-order) beliefs about the other players’ beliefs about the payoffs, etc., up to a finite order. The noise in his observation can be arbitrarily small, and he can observe arbitrarily many orders of beliefs. Suppose that given his information, he concludes that his information is consistent with a type profile $t$ that comes from a model that is continuous at infinity. Note that the set of hierarchies that is consistent with his information is an open subset $U = U_1 \times \cdots \times U_n$ of the universal type space, and $(h_1(t_1), \ldots, h_n(t_n)) \in U$. Hence, our proposition concludes that for every rationalizable action profile $a \in S^\infty [t]$ and any finite length $L$, the researcher cannot rule out the possibility that in the actual situation the first $L$ moves have to be as in the outcome of $a$ in *any* rationalizable outcome. That is, rationalizable outcomes can differ from $a$ only after $L$ moves. Since $L$ is arbitrary, for practical purposes he cannot practically rule out any rationalizable outcome as the unique solution.
Notice that Proposition 1 differs from Lemma 1 in two ways. First, instead of assuming that the game has a finite horizon, Proposition 1 assumes only that the game is continuous at infinity, allowing most games in economics. Second, it concludes that for the perturbed types all rationalizable actions are equivalent to \( a_i \) up to an arbitrarily long but finite horizon, instead of concluding that all rationalizable actions are equivalent to \( a_i \). These two statements are, of course, equivalent in finite-horizon games.

A main step in our proof is indeed Lemma 1. There are, however, many other steps that need to be spelled out carefully. Hence, we relegate the proof to the appendix. In order to illustrate the main idea, we now sketch out the proof for a simple but important case. Suppose that \( \Theta = \{ \bar{\theta} \} \) and \( T = \{ \bar{t} \} \), so that we have a complete information game, and \( a^* \) is a Nash equilibrium of this game. For each \( m \), perturb every history \( h \) at length \( m \) by assuming that thereafter the play will be according to \( a_i \), which prescribes a continuation at each history. Call the resulting history \( h^{m,a^*} \). This can also be described as a payoff perturbation: define the perturbed payoff function \( \hat{\theta}^m \) by setting \( \hat{\theta}^m (h) = \hat{\theta} (h^{m,a^*}) \) at every terminal history \( h \). Now consider the complete-information game with perturbed model \( \hat{\Theta}^m = \{ \hat{\theta}^m \} \) and \( T^m = \{ \hat{t}^m \} \), where according to \( \hat{t}^m \) it is common knowledge that the payoff function is \( \hat{\theta}^m \) (essentially, this means that players are forced to play according to \( a^* \) after the \( m \)th information set). We make three observations towards proving the proposition. We first observe that, since \( \hat{\theta} \) is continuous at infinity, by construction, \( \theta^m \to \hat{\theta} \), implying that \( h_i (\hat{t}_i^m) \to h_i (\bar{t}_i) \). Hence, there exists \( \bar{m} > L \) such that \( h_i (\hat{t}_i^m) \in U_i \). Second, there is a natural isomorphism between the payoff functions that do not depend on the moves after length \( \bar{m} \), such as \( \theta^m \), and the payoff functions for the finite-horizon extensive game form that is created by truncating the moves at length \( \bar{m} \). In particular, there is an isomorphism \( \varphi \) that maps the hierarchies in the universal type space \( T^{m*} \) for the truncated extensive game form to the types in universal type space \( T^* \) for the infinite-horizon game form that make the common-knowledge assumption that the moves after length \( \bar{m} \) are payoff-irrelevant. Moreover, the rationalizable moves for the first \( \bar{m} \) nodes do not change under the isomorphism, in that \( a_i \in S_i^\infty [\varphi (t_i)] \) if and only if the restriction \( a_i^{m*} \) of \( a_i \) to the truncated game is in \( S_i^\infty [t_i] \) for any \( t_i \in T^{m*} \). Third, since \( a^* \) is a Nash equilibrium, it remains a Nash equilibrium after the perturbation. This is because enforcing Nash equilibrium strategies after some histories does not give a new incentive to deviate. Therefore, \( a_i^* \) is a rationalizable strategy in the perturbed complete information game: \( a_i^* \in S_i^\infty [\hat{t}_i^m] \). Now, these three observations
together imply that the hierarchy $\varphi^{-1}(h_i((\tilde{t}_i^m)))$ for the finite-horizon game form is in an open neighborhood $\varphi^{-1}(U_i) \subset T_i^{m*}$ and the restriction $a_i^{*m}$ of $a_i^*$ to the truncated game form is rationalizable for $\varphi^{-1}(h_i((\tilde{t}_i^m)))$. Hence, by Lemma 1, there exists a type $\hat{t}_i$ such that (i) $h_i(\hat{t}_i) \in \varphi^{-1}(U_i)$ and (ii) all rationalizable actions of $\hat{t}_i$ are $\bar{m}$-equivalent to $a_i^{*m}$. Now consider a type $\hat{t}_i$ with hierarchy $h_i(\hat{t}_i) \equiv \varphi(h_i((\tilde{t}_i)))$, where $\hat{t}_i$ can be picked from a finite, common-prior model because the isomorphic type $\tilde{t}_i$ comes from such a type space. Type $\hat{t}_i$ has all the properties in the proposition. First, by (i), $h_i(\hat{t}_i) \in U_i$ because

$$h_i(\hat{t}_i) = \varphi(h_i((\tilde{t}_i))) \in \varphi(\varphi^{-1}(U_i)) \subset U_i.$$  

Second, by (ii) and the isomorphism in the second observation above, all rationalizable actions of $\hat{t}_i$ are $\bar{m}$-equivalent to $a_i^*$.

There are two limitations of Proposition 1. First, it is silent about the tails. Given a rationalizable action $a_i$, it does not ensure that there is a perturbation under which $a_i$ is the unique rationalizable plan—although it does ensure for an arbitrary $L$ that there is a perturbation under which following $a_i$ is the uniquely rationalizable plan up to $L$. The second limitation, which applies equally to Chen’s result, is as follows. Given any rationalizable path $z(a)$ and $L$, Proposition 1 establishes that there is a profile $t = (t_1, \ldots, t_n)$ of perturbed types for which $z^L(a)$ is the unique rationalizable path up to $L$. Nevertheless, as in the following example, this perturbation may rely crucially on the perturbed types’ all considering the path $z^L(a)$ unlikely at the start of play. We use a two-stage game for simplicity, since the relevant idea is the same as for infinite games.

**Cooperation in Twice-Repeated Prisoners’ Dilemma.** Consider a twice-repeated prisoners’ dilemma game with complete information and with no discounting. We shall need the standard condition $u(C, D) + u(D, C) > 2u(D, D)$, where $u$ is the payoff of player 1 in the stage game and $C$ and $D$ stand for the moves Cooperate and Defect, respectively. In the twice-repeated game, though of course there is a unique Nash equilibrium outcome, the following “tit-for-tat” strategy is rationalizable:

$a^{T4T}$: play Cooperate in the first round, and in the second round play what the other player played in the first round.

We show this rationalizability as follows. First, note that defection in every subgame, which we call $a^{DD}$, by both players is an equilibrium, so $a^{DD}$ is rationalizable. Next,
defection in the first period followed by tit-for-tat in the second period, which we call $a^{DT}$, is a best response to $a^{DD}$ and therefore rationalizable. Finally, under the inequality above, $a^{T4T}$ is a best response to $a^{DT}$ and so is rationalizable. This tells us that cooperation in both rounds is possible under rationalizable play.

This counterintuitive sort of conclusion is one reason standard rationalizability is not ordinarily used for extensive-form games; it is extremely permissive. This makes the results of Chen (2008) more surprising. By his theorem, there exists a perturbation $t^{T4T}$ of the common-knowledge type for which $a^{T4T}$ is the unique rationalizable action. If both players have type $t^{T4T}$, the unique rationalizable action profile $(a^{T4T}, a^{T4T})$ leads to cooperation in both rounds. However, we can deduce that the constructed type will necessarily have certain odd properties. Cooperation in the first round must make him update his beliefs about the payoffs in such a way that Cooperate becomes a better response than Defect in the second round. Since the definition of perturbation requires that, ex ante, he believes with high probability the payoffs are similar to the repeated prisoner dilemma, under which Defect is dominant in the second round, this drastic updating implies that $t^{T4T}$ finds it unlikely that the other player will play Cooperate in the first round. Hence, when both players have type $t^{T4T}$, the story must be as follows: they each cooperate in the first round even though they think they are playing Prisoners’ Dilemma, motivated by a belief that the other player has plan $a^{DT}$. Then, when they see the other player cooperate, they drastically update their payoffs (which they believe to be correlated with the other player’s type) and believe that it is optimal to cooperate in the second period.

This sort of perturbation, in which the induced behavior can only occur on a path the players themselves assign low probability, is to some extent unconvincing. As mentioned above, this motivates our Proposition 2 which shows that equilibrium outcomes can be induced by perturbations without this property. This reinforces the natural view that rationalizability is a weak solution concept in a dynamic context.

**Stronger Structure Theorem for Equilibrium Outcomes.** These limitations of Proposition 1 are the motivation for our next proposition, a stronger version of the structure theorem for which we need an outcome to be a Nash equilibrium rather than merely rationalizable. We fix a payoff function $\theta^*$, and consider the game in which $\theta^*$ is common knowledge. This game is represented by type profile $t^{CK}(\theta^*)$ in the universal type space. For
any Nash equilibrium $a^*$ of this game, we find a profile of perturbations under which $a^*$ is the unique rationalizable action and all players’ rationalizable beliefs assign high probability to the equilibrium outcome $z(a^*)$. In order to state the result, we need to introduce some new formalism. We call a probability distribution $\pi \in \Delta (\Theta \times T_{-i} \times A_{-i})$ a rationalizable belief of type $t_i$ if $\text{marg}_{\Theta \times T_{-i}} \pi = \kappa_{t_i}$ and $\pi (a_{-i} \in S^\infty_t [t_{-i}]) = 1$. We write $\Pr (\cdot | \pi, a_i)$ and $E \left[ \cdot | \pi, a_i \right]$ for the probability measure on terminal histories and expected payoff operator resulting from playing $a_i$ against belief $\pi$.

**Proposition 2.** Let $(\Gamma, \{\theta^*\}, \{t^{\text{CK}}(\theta^*)\}, \kappa)$ be a complete-information game that is continuous at infinity, and $a^*$ be a pure-strategy Nash equilibrium of this game. For any $i \in N$, for any neighborhood $U_i$ of $h_i(t^{\text{CK}}(\theta^*))$ in the universal type space $T^*$, and any $\varepsilon > 0$, there exists a hierarchy $h_i(\hat{t}_i) \in U_i$, such that for every rationalizable belief $\pi$ of $\hat{t}_i$,

1. $a_i \in S^\infty_t [\hat{t}_i]$ iff $a_i$ is equivalent to $a^*_i$;
2. $\Pr (z(a^*) | \pi, a^*_i) \geq 1 - \varepsilon$, and
3. $| E [u_j(\theta, a) | \pi, a^*_i] - u_j(\theta^*, a^*) | \leq \varepsilon$ for all $j \in N$.

The first conclusion states that the equilibrium action $a^*_i$ is the only rationalizable action for the perturbed type in reduced form. Hence, the first limitation of Proposition 1 does not apply. The second conclusion states that the perturbed type $\hat{t}_i$ finds it highly likely that the equilibrium outcome prevails in any rationalizable strategy profile. Hence, the second limitation of Proposition 1 does not apply, either. Finally, the last conclusion states that the perturbed type $\hat{t}_i$ expects that everybody enjoys nearly equilibrium payoffs under rationalizability. All in all, Proposition 2 establishes that no equilibrium outcome can be ruled out as the unique rationalizable outcome without knowledge of infinite hierarchy of beliefs, both in terms of actual realization and in terms of players’ ex-ante expectations.

### 4. Application: An Unusual Folk Theorem

In this section, we consider infinitely repeated games with complete information. Under the standard assumptions for the folk theorem, we prove an unusual folk theorem, which concludes that for every individually rational and feasible payoff vector $v$, there exists a perturbation of beliefs under which there is a unique rationalizable outcome and players expect to enjoy approximately the payoff vector $v$ under any rationalizable belief.
For simplicity, we consider a simultaneous-action stage game $G = (N, B, g)$ where $B = B_1 \times \cdots \times B_n$ is the set of profiles $b = (b_1, \ldots, b_n)$ of moves and $g : B \to [0, 1]^n$ is the vector of stage payoffs. We have perfect monitoring. Hence, a history is a sequence $h = (b^l)_{l \in \mathbb{N}}$ of profiles $b^l = (b^l_1, \ldots, b^l_n)$. In the complete-information game, the players maximize the average discounted stage payoffs. That is, the payoffs function is

$$
\theta^*_\delta(h) = (1 - \delta) \sum_{l=0}^{n} \delta^l g(b^l) \quad \forall h = (b^l)_{l \in \mathbb{N}}
$$

where $\delta \in (0, 1)$ is the discount factor, which we will let vary. Denote the repeated game by $G_\delta = (\Gamma, \{\theta^*_\delta\}, \{t_{CK} (\theta^*_\delta)\}, \kappa)$.

Let $V = \text{co} (g(B))$ be the set of feasible payoff vectors (from correlated mixed action profiles), where co takes the convex hull. Define also the pure-action minmax payoff as

$$
v_i = \min_{b_{-i} \in B_{-i}} \max_{b_i \in B_i} g(b)
$$

for each $i \in N$. We define the set of feasible and individually rational payoff vectors as

$$
V^* = \{ v \in V | v_i > v_i \text{ for each } i \in N \}.
$$

We denote the interior of $V^*$ by $\text{int}V^*$. The interior will be non-empty when a weak form of full-rank assumption holds. The following lemma states a typical folk theorem (see Proposition 9.3.1 in Mailath and Samuelson (2006) and also Fudenberg and Maskin (1991)).

**Lemma 2.** For every $v \in \text{int}V^*$, there exists $\tilde{\delta} < 1$ such that for all $\delta \in (\tilde{\delta}, 1)$, $G_\delta$ has a subgame-perfect equilibrium $a^*$ in pure strategies, such that $u(\theta^*_\delta, a^*) = v$.

The lemma states that every feasible and individually rational payoff vector in the interior can be supported as the subgame-perfect equilibrium payoff when the players are sufficiently patient. Given such a large multiplicity, both theoretical and applied researchers often focus on efficient equilibria (or extremal equilibria). By combining the lemma with Proposition 2, our next result establishes that the multiplicity is irreducible:

**Proposition 3.** For all $v \in \text{int}V^*$ and $\varepsilon > 0$, there exists $\tilde{\delta} < 1$ such that for all $\delta \in (\tilde{\delta}, 1)$, every open neighborhood $U$ of $t_{CK} (\theta^*_\delta)$ contains a type profile $\hat{t} \in U$ such that

1. each $\hat{t}_i$ has a unique rationalizable action $a^*_i$ in reduced form, and
(2) under every rationalizable belief $\pi$ of $\hat{t}_i$, the expected payoffs are all within $\varepsilon$ neighborhood of $v$:

$$\left| E \left[ u_j (\theta, a) \mid \pi, a^*_i \right] - u_j (\theta^*, a^*) \right| \leq \varepsilon \quad \forall j \in N.$$ 

Proof. Fix any $v \in \text{int} V^*$ and $\varepsilon > 0$. By Lemma 2, there exists $\delta < 1$ such that for all $\delta \in (\delta, 1)$, $G_\delta$ has a subgame-perfect equilibrium $a^*$ in pure strategies, such that $u (\theta^*_\delta, a^*) = v$. Then, by Proposition 2, for any $\delta \in (\delta, 1)$ and any open neighborhood $U$ of $t^{CK} (\theta^*_\delta)$, there exists a type profile $\hat{t} \in U$ such that each $\hat{t}_i$ has a unique rationalizable action $a^*_i$ in reduced form (Part 1 of Proposition 2), and under every rationalizable belief $\pi$ of $\hat{t}_i$, the expected payoffs are all within $\varepsilon$ neighborhood of $u (\theta^*_\delta, a^*) = v$ (Part 3 of Proposition 2).}

Proposition 3 states that every individually rational and feasible payoff $v$ in the interior is the unique rationalizable outcome for some perturbation. Moreover, in the actual situation described by the perturbation, all players play according to the subgame-perfect equilibrium that supports $v$ and all players anticipate that the payoffs are within an $\varepsilon$-neighborhood of $v$. That is, the complete-information game is surrounded by types with a unique solution, but the unique solution varies in such a way that it traces all individually rational and feasible payoffs. While the multiplicity in usual folk theorems may suggest a need for a refinement, the multiplicity in our unusual folk theorem emphasizes the impossibility of a robust refinement.

Chassang and Takahashi (2009) examine the question of robustness in repeated games from an ex ante perspective. That is, following Kajii and Morris (1997), they define an equilibrium as robust if approximately the same outcome is possible in a class of elaborations. (An elaboration is an incomplete-information game in which each player believes with high probability that the original game is being played.) They consider specifically elaborations with serially independent types, so that the moves of players do not reveal any information about their payoffs and behavior in the future. They obtain a useful one-shot robustness result—to paraphrase, an equilibrium of the repeated game is robust if the equilibrium at each stage game, augmented with continuation values, is risk-dominant. There are two major distinctions. First, their perturbations are defined from an ex ante perspective, by what players believe before receiving information. Ours are from an interim perspective, based on what players believe just before play begins. This could be subsequent to receiving
information, but our setup does not actually require reference to a particular information structure (type space with prior). For more on the distinction between these approaches, see our 2007 paper. Second, while they focus on serially independent types, whose moves do not reveal any information about the future payoffs, the moves of our perturbed types reveal information about both their own and the other players’ payoffs in the future stage games.

5. Application: Incomplete Information in Bargaining

In a model of bilateral bargaining with complete information, Rubinstein (1982) shows that there exists a unique subgame-perfect equilibrium. Subsequent research illustrates that the equilibrium result is sensitive to incomplete information. In this section, using Proposition 2, we show quite generally that the equilibrium must be highly sensitive: every bargaining outcome can be supported as the unique rationalizable outcome for a nearby model.

We consider Rubinstein’s alternating-offer model with finite set of divisions. There are two players, \( N = \{1, 2\} \), who want to divide a dollar. The set of possible shares is \( X = \{0, 1/m, 2/m, \ldots , 1\} \) for some \( m > 1 \). At date 0, Player 1 offers a division \((x, 1-x)\), where \( x \in X \) is the share of Player 1 and \( 1-x \) is the share of Player 2. Player 2 decides whether to accept or reject the offer. If he accepts, the game ends with division \((x, 1-x)\). Otherwise, we proceed to the next date. At date 1, Player 2 offers a division \((y, 1-y)\), and Player 1 accepts or rejects the offer. In this fashion, players make offers back and forth until an offer is accepted. We denote the bargaining outcome by \((x, l)\) if players reach an agreement on division \((x, 1-x)\) at date \( l \). In the complete-information game, the payoff function is

\[
\theta^* = \begin{cases} 
\delta^l (x, 1-x) & \text{if the outcome is } (x, l) \\
0 & \text{if players never agree}
\end{cases}
\]

for some \( \delta \in (0, 1) \).

When \( X = [0, 1] \), in the complete information game \( G^* = (\Gamma, \{\theta^*\}, \{t^{CK} (\theta^*)\}, \kappa) \), there is a unique subgame perfect equilibrium, and the bargaining outcome in the unique subgame-perfect equilibrium is

\[
(x^*, 0) = (1/(1+\delta), 0).
\]

That is, the players immediately agree on division \((x^*, 1-x^*)\). When \( X = \{0, 1/m, \ldots , 1\} \) as in here, there are more subgame-perfect equilibria due to multiple equilibrium behavior.
in the case of indifference. Nevertheless, the bargaining outcomes of these equilibria all converge to \((x^*, 0)\) as \(m \to \infty\).

In contrast with the unique subgame-perfect equilibrium, there is a large multiplicity of non-subgame-perfect Nash equilibria, but these equilibria are ignored as they rely on incredible threats or sequentially irrational moves off the path. Building on such non-subgame-perfect Nash equilibria and Proposition 2, the next result shows that each bargaining outcome is the outcome of the unique rationalizable plan under some perturbation.

**Proposition 4.** For any bargaining outcome \((x, l) \in X \times \mathbb{N}\) and any \(\varepsilon > 0\), every open neighborhood \(U\) of \(t^{CK}(\theta^*_x)\) contains a type profile \(\hat{t} \in U\) such that

1. each \(\hat{t}_i\) has a unique rationalizable action \(a^*_i\) in reduced form;
2. the bargaining outcome under \(a^*\) is \((x, l)\), and
3. every rationalizable belief of \(\hat{t}_i\) assigns at least probability \(1 - \varepsilon\) on \((x, l)\).

**Proof.** We will show that the complete-information game has a Nash equilibrium \(a^*\) with bargaining outcome \((x, l)\). Proposition 2 then establishes the existence of type profile \(\hat{t}\) as in the statement of the proposition. Consider the case of even \(l\), at which Player 1 makes an offer; the other case is identical. Define \(a^*\) in reduced-form as

- \((a^*_1)\) at any date \(l' \neq l\), offer only \((1, 0)\) and reject all offers; offer \((x, 1 - x)\) at date \(l\);
- \((a^*_2)\) at any date \(l' \neq l\), offer only \((0, 1)\) and reject all offers; accept only \((x, 1 - x)\) at \(l\).

It is clear that \(a^*\) is a Nash equilibrium, and the bargaining outcome under \(a^*\) is \((x, l)\). \(\square\)

That is, for every bargaining outcome \((x, l)\), one can introduce a small amount of incomplete information in such a way that the resulting type profile has a unique rationalizable action profile and it leads to the bargaining outcome \((x, l)\). Moreover, in the perturbed type profile, players are all nearly certain that \((x, l)\) will be realized. Unlike in the case of non-subgame-perfect equilibria, one cannot rule out these outcomes by refinement because there is a unique rationalizable outcome. In order to rule out these outcomes, one either needs to introduce irrational behavior or rule out the information structure that leads to the perturbed type profile by fiat (as he cannot rule out these structures by observation of finite-order beliefs without ruling out the original model). Therefore, despite the unique subgame-perfect outcome in the original model, and despite the fact that this outcome has
generated many important and intuitive insights, one cannot make any prediction on the outcome without introducing irrational behavior or making informational assumptions that cannot be verified by observing finite-order beliefs.

The existing literature illustrates already that the subgame-perfect equilibrium is sensitive to incomplete information. For example, for high $\delta$, the literature on the Coase conjecture establishes that if one party has private information about his own valuation, then he gets everything—in contrast to the nearly equal sharing in the complete information game. This further leads to delay due to reputation formation in bargaining with two-sided incomplete information on payoffs (Abreu and Gul (2000)) or on players’ second-order beliefs (Feinberg and Skrzypacz (2005)).

Proposition 4 differs from these results in many ways. The first difference is in the scope of sensitivity: while the existing results show that another outcome may occur under a perturbation, Proposition 4 shows that any outcome can be supported by a perturbation. The second difference is in the solution concept: while the existing result show sensitivity with respect to a sequential equilibrium or all sequential equilibria, there is a unique rationalizable outcome in Proposition 4, ruling out reinstating the original outcome by a refinement. Third, the existing results often consider the limit $\delta \to 0$, which is a point of discontinuity for the complete-information model already. In contrast, $\delta$ is fixed in Proposition 4. Finally, existing results consider simple perturbations, and these perturbations may correspond to the specification of economic parameters, such as valuation, or may be commitment types. In contrast, given the generality of the results, the types constructed in our paper are complicated, and it is not easy to interpret how they are related to the economic parameters. (In specific examples, the same results could be obtained using simple types that correspond to economic parameters, as in Izmalkov and Yildiz (2010)).

6. Concluding Remarks and Literature Review

The early literature on robustness identified two mechanisms through which a small amount of incomplete information can have a large effect: reputation formation (Kreps, Milgrom, Roberts, and Wilson (1982)) and contagion (Rubinstein (1989)). In reputation formation, one learns about the other players’ payoffs from their unexpected moves. As we
saw in the twice-repeated prisoners’ dilemma game, our perturbed types exhibit a more extreme version of this property: they learn not only about the other players’ payoffs but also about their own payoffs from the others’ moves. Moreover, our perturbations are explicitly constructed using a generalized contagion argument. Hence, the perturbations here and in Chen (2008) combine the two mechanisms in order to obtain a very strong conclusion: any rationalizable action can be made uniquely rationalizable under some perturbation.

As the above example illustrates, in our perturbations a player $i$ may learn about the payoffs of $j$ from the moves of another player $k$. This is reasonable in many contexts due to interdependence of preferences. Nevertheless, it may also be reasonable to keep it common knowledge that some parameters are only known by some players. For example, one may wish to assume common knowledge that a firm’s cost is its own private information, so that one does not update his beliefs about the firm’s cost by observing some other player’s move. Penta (2008) offers a framework for determining the set of possible outcomes under such assumptions. Assuming common knowledge that some parameters are known only by some players, he obtains an identical structure theorem for what he calls interim sequential rationalizability (ISR) instead of interim correlated rationalizability (ICR). Of course, ISR depends on what is kept common knowledge and is equal to ICR when nothing is kept common knowledge. We believe that one can extend Penta’s result to infinite-horizon games by modifying our construction, obtaining a more general result. We have not considered that extension for clarity, because ICR is a more transparent solution concept, and because we believe that the case of dropping all common knowledge assumptions is an important benchmark. Also, it is not clear how to extend ISR to infinite horizon.

Other papers have considered perturbations which are restricted to keep some structure common knowledge. For “nice” games (static games with unidimensional action spaces and strictly concave utility functions), Weinstein and Yildiz (2008) obtain a characterization for sensitivity of Bayesian Nash equilibria in terms of a local version of ICR, keeping arbitrary common knowledge restrictions on payoffs.\footnote{Within the important class of nice games, Weinstein and Yildiz (2008) are able to cope with uncountable action spaces, as we do here. The special structure of the games which allows this and the arguments involved are very different in the two cases.} In the same vein, Oury and Tercieux (2007) allow arbitrarily small perturbations on payoffs to obtain an equivalence between continuous
partial implementation in Bayesian Nash equilibria and full implementation in rationalizable strategies.

**Appendix A. Proof of Structure Theorem**

We begin with some additional notation:

**Notation 1.** For any belief $\pi \in \Delta (\Theta \times A_{-i})$ and action $a_i$ and for any history $h$, write $E \left[ \cdot | h, a_i, \pi \right]$ for the expectation operator induced by action $a_i$ and $\pi$ conditional on reaching history $h$. For any strategy profile $s : T \rightarrow A$ and any type $t_i$, we write $\pi (\cdot | t_i, s_{-i}) \in \Delta (\Theta \times T_{-i} \times A_{-i})$ for the belief induced by $t_i$ and $s_{-i}$. Given any functions $f : W \rightarrow X$ and $g : Y \rightarrow Z$, we write $(f, g)^{-1}$ for the preimage of the mapping $(w, y) \mapsto (f(w), g(y))$.

A.1. **Preliminaries.** We now define some basic concepts and present some preliminary results. By a *Bayesian game in normal form*, we mean a tuple $(N, A, u, \Theta, T, \kappa)$ where $N$ is the set of players, $A$ is the set of action profiles, $(\Theta, T, \kappa)$ is a model, and $u : \Theta \times A \rightarrow [0, 1]^n$ is the payoff function. For any $G = (N, A, u, \Theta, T, \kappa)$, we say that $a_i$ and $a'_i$ are $G$-equivalent if

$$u(\theta, a_i, a_{-i}) = u(\theta, a'_i, a_{-i}) \quad (\forall \theta \in \Theta, a_{-i} \in A_{-i}).$$

By a *reduced-form game*, we mean a game $G_R = (N, \tilde{A}, u, \Theta, T, \kappa)$ where for each $i$, $\tilde{A}_i$ contains at least one representative action from each $G$-equivalence class. Rationalizability depends only on the reduced form:

**Lemma 3.** Given any game $G$ and a reduced form $G_R$ for $G$, for any type $t_i$, the set $S^\infty_i [t_i]$ of rationalizable actions in $G$ is the set of all actions that are $G$-equivalent to some rationalizable action of $t_i$ in $G_R$.

The lemma follows from the fact that in the elimination process, all members of an equivalence class are eliminated at the same time; i.e., one eliminates, at each stage, a union of equivalence classes. It implies the following isomorphism for rationalizability.

**Lemma 4.** Let $G = (N, A, u, \Theta, T, \kappa)$ and $G' = (N, A', u', \Theta', T', \kappa)$ be Bayesian games in normal form, $\mu_i : A_i \rightarrow A'_i$, $i \in N$, be onto mappings, and $\varphi : \Theta \rightarrow \Theta'$ and $\tau_i : T_i \rightarrow T'_i$, $i \in N$, be bijections. Assume (i) $\kappa_{\tau_i(t_i)} = \kappa_{t_i} \circ (\varphi, \tau_{-i})^{-1}$ for all $t_i$ and (ii) $u' (\varphi(\theta), \mu(a)) = u(\theta, a)$ for all $(\theta, a)$. Then, for any $t_i$ and $a_i$,

$$(A.1) \quad a_i \in S^\infty_i [t_i] \iff \mu_i (a_i) \in S^\infty_i [\tau_i(t_i)].$$
Note that the bijections $\varphi$ and $\tau$ are a renaming, and (i) ensures that the beliefs do not change under the renaming. On the other hand, $\mu_i$ can map many actions to one action, but (ii) ensures that all those actions are $G$-equivalent. The lemma concludes that rationalizability is invariant to such a transformation.

**Proof.** First note that (ii) implies that for any $a_i, a'_i \in A_i$,

(A.2) \[ a_i \text{ is } G\text{-equivalent to } a'_i \iff \mu_i(a_i) \text{ is } G'\text{-equivalent to } \mu_i(a'_i). \]

In particular, if $\mu_i(a_i) = \mu_i(a'_i)$, then $a_i$ is $G$-equivalent to $a'_i$. Hence, there exists a reduced-form game $G_R = (N, \bar{A}, u, \Theta, T, \kappa)$ for $G$, such that $\mu$ is a bijection on $\bar{A}$, which is formed by picking a unique representative from each $\mu^{-1}(\mu(a))$. Then, by (A.2) again, $G'_R = (N, \mu(\bar{A}), u', \Theta', T', \kappa)$ is a reduced form for $G'$.$^5$ Note that $G_R$ and $G'_R$ are isomorphic up to the renaming of actions, parameters, and types by $\mu$, $\varphi$, and $\tau$, respectively. Therefore, for any $a'_i \in \bar{A}_i$ and $t_i$, $a'_i$ is rationalizable for $t_i$ in $G_R$ iff $\mu_i(a'_i)$ is rationalizable for $\tau_i(t_i)$ in $G'_R$. Then, Lemma 3 and (A.2) immediately yields (A.1). \hfill \Box

We will also apply a lemma from Mertens-Zamir (1985) stating that the mapping from types in any type space to their hierarchies is continuous, provided the belief mapping $\kappa$ defining the type space is continuous.

**Lemma 5** (Mertens and Zamir (1985)). Let $(\Theta, T, \kappa)$ be any model, endowed with any topology, such that $\Theta \times T$ is compact and $\kappa_t$ is a continuous function of $t_i$. Then, $h$ is continuous.

**A.2. Truncated Games.** We now formally introduce an equivalence between finitely-truncated games and payoff functions that implicitly assume such a truncation. For any positive integer $m$, define a truncated extensive game form $\Gamma^m = (N, H^m, (I_i)_{i \in N})$ by

$H^m = \{h^m|h \in H\}$.

The set of terminal histories in $H^m$ is $Z^m = \{z^m|z \in Z\}$.

We define $\bar{\Theta}^m = ([0,1]^{Z^m})^n$.

$^5$Proof: Since $\mu_i$ is onto, $A'_i = \mu_i(A_i)$. Moreover, for any $\mu_i(a_i) \in A'_i$, there exists $a'_i \in \bar{A}_i$ that is $G$-equivalent to $a_i$. By (A.2), $\mu_i(a_i)$ is $G'$-equivalent to $\mu_i(a'_i) \in \mu_i(\bar{A}_i)$. 

as the set of payoff functions for truncated game forms. Since $Z^m$ is not necessarily a subset of $Z$, $\tilde{\Theta}^m$ is not necessarily a subset of $\Theta^*$. We will now embed $\tilde{\Theta}^m$ into $\Theta^*$ through an isomorphism to a subset of $\Theta^*$. Define the subset

$$\Theta^m = \{ \theta \in \Theta^* | \theta (h) = \theta (\tilde{h}) \text{ for all } h \text{ and } \tilde{h} \text{ with } h^m = \tilde{h}^m \}.$$ 

This is the set of payoff functions for which moves after period $m$ are irrelevant. Games with such payoffs are nominally infinite but inherently finite, as we formalize via the isomorphism $\varphi_m : \tilde{\Theta}^m \rightarrow \Theta^m$ defined by setting

(A.3) $$\varphi_m (\theta) (h) = \theta (h^m)$$

for all $\theta \in \tilde{\Theta}^m$ and $h \in Z$, where $h^m \in H^m$ is the truncation of $h$ at length $m$. Clearly, under the product topologies, $\varphi_m$ is an isomorphism, in the sense that it is one-to-one, onto, and both $\varphi_m$ and $\varphi_m^{-1}$ are continuous. For each $a_i \in A_i$, let $a_i^m$ be the restriction of action $a_i$ to the histories with length less than or equal to $m$. The set of actions in the truncated game form is $A_i^m = \{ a_i^m | a_i \in A_i \}$.

**Lemma 6.** Let $G = (\Gamma, \Theta, T, \kappa)$ and $G^m = (\Gamma^m, \Theta^m, T^m, \kappa)$ be such that (i) $\Theta^m \subset \tilde{\Theta}^m$, (ii) $\Theta = \varphi_m (\Theta^m)$ and (iii) $T_i = \tau_i (T_i^m)$ for some bijection $\tau_i^m$ and such that $\kappa_{\tau_i^m} (t_i^m) = \kappa_{t_i} \circ (\varphi_m, t_i^m)^{-1}$ for each $t_i^m \in T_i^m$. Then, the set of rationalizable actions are $m$-equivalent in $G$ and $G^m$:

$$a_i \in S_i^\infty [\tau_i^m (t_i^m)] \iff a_i^m \in S_i^\infty [t_i^m] \quad (\forall i, t_i^m, a_i).$$

**Proof.** In Lemma 4, take $\varphi = \varphi_m^{-1}$, $\tau_i = (\tau_i^m)^{-1}$, and $\mu : a_i \mapsto a_i^m$. We only need to check that $u^m (\varphi_m^{-1} (\theta), a^m) = u (\theta, a)$ for all $(\theta, a)$ where $u^m$ denotes the utility function in the truncated game $G^m$. Indeed, writing $z^m (a^m)$ for the outcome of $a^m$ in $G^m$, we obtain

$$u^m (\varphi_m^{-1} (\theta), a^m) = \varphi_m^{-1} (\theta) (z^m (a^m)) = \varphi_m^{-1} (\theta) (z (a)^m) = \varphi_m (\varphi_m^{-1} (\theta)) (z (a)) = \theta (z (a)) = u (\theta, a).$$

Here, the first and the last equalities are by definition; the second equality is by definition of $a^m$, and the third equality is by definition (A.3) of $\varphi_m$. \(\square\)

Let $T^*^m$ be the $\tilde{\Theta}^m$-based universal type space, which is the universal type space generated by the truncated extensive game form. This space is distinct from the universal type space, $T^*$, for the original infinite-horizon extensive form. We will now define an embedding between the two type spaces, which will be continuous and one-to-one and preserve the rationalizable actions in the sense of Lemma 6.

**Lemma 7.** For any $m$, there exists a continuous, one-to-one mapping $\tau^m : T^*^m \rightarrow T^*$ with $\tau^m (t) = (\tau_1^m (t_1), \ldots, \tau_n^m (t_n))$ such that for all $i \in N$ and $t_i \in T_i^*^m$. 

(1) \( t_i \) is a hierarchy for a type from a finite model if and only if \( \tau_i^m (t_i) \) is a hierarchy for a type from a finite model;
(2) \( t_i \) is a hierarchy for a type from a common-prior model if and only if \( \tau_i^m (t_i) \) is a hierarchy for a type from a common-prior model, and
(3) for all \( a_i, a_i \in S_i^\infty [\tau_i^m (t_i)] \) if and only if \( a_i^m \in S_i^\infty [t_i] \).

Proof. Since \( T^* \) and \( T^m \) do not have any redundant type, by the analysis of Mertens and Zamir (1985), there exists a continuous and one-to-one mapping \( \tau^m \) such that

\[
(\text{A.4}) \quad \kappa_{\tau_i^m (t_i)} = \kappa_{t_i} \circ (\varphi_m, \tau_i^m)^{-1}
\]

for all \( i \) and \( t_i \in T_i^{*m} \). First two statements immediately follow from (A.4). Part 3 follows from (A.4) and Lemma 6. \( \square \)

A.3. Proof of Proposition 1. We will prove the proposition in several steps.

Step 1. Fix any positive integer \( m \). We will construct a perturbed incomplete information game with an enriched type space and truncated time horizon at \( m \) under which each rationalizable action of each original type remains rationalizable for some perturbed type. For each rationalizable action \( a_i \in S_i^\infty [t_i] \), let

\[
X [a_i, t_i] = \{ a_i' \in S_i^\infty [t_i] | a_i' \text{ is } m\text{-equivalent to } a_i \}
\]

and pick a representative action \( r_{t_i} (a_i) \) from each set \( X [a_i, t_i] \). We will consider the type space \( \tilde{T}^m = \tilde{T}_1^m \times \cdots \times \tilde{T}_n^m \) with

\[
\tilde{T}_i^m = \{ (t_i, r_{t_i} (a_i), m) | t_i \in T_i, a_i \in S_i^\infty [t_i] \}.
\]

Note that each type here has two dimensions, one corresponding to the original type the second corresponding to an action. Note also that \( \tilde{T}^m \) is finite because there are finitely many equivalence classes \( X [a_i, t_i] \), allowing only finitely many representative actions \( r_{t_i} (a_i) \). Towards defining the beliefs, recall that for each \( (t_i, r_{t_i} (a_i), m) \), since \( r_{t_i} (a_i) \in S_i^\infty [t_i] \), there exists a belief \( \pi_{t_i, r_{t_i} (a_i)} \in \Delta (\Theta \times T_{-i} \times A_{-i}) \) under which \( r_{t_i} (a_i) \) is a best reply for \( t_i \) and \( \pi_{t_i, r_{t_i} (a_i)} = \kappa_{t_i} \). Define a mapping \( \phi_{t_i, r_{t_i} (a_i), m} : \Theta^* \rightarrow \Theta^* \) between the payoff functions by setting

\[
(\text{A.5}) \quad \phi_{t_i, r_{t_i} (a_i), m} (\theta) (h) = E \left[ \theta (h) | h^m, r_{t_i} (a_i), \pi_{t_i, r_{t_i} (a_i)} \right]
\]

\[\text{If one writes } t_i = (t_i^1, t_i^2, \ldots) \text{ and } \tau_i^m (t_i) = (\tau_i^{m,1} (t_i^1), \tau_i^{m,2} (t_i^2), \ldots) \text{ as a hierarchy, we define } \tau_i^m \text{ inductively by setting } \tau_i^{m,1} (t_i^1) = t_i^1 \circ \varphi_m^1 \text{ and } \tau_i^{m,k} (t_i^k) = t_i^k \circ \left( \varphi_m, \tau_{-i}^{m,1}, \ldots, \tau_{-i}^{m,k-1} \right)^{-1} \text{ for } k > 1.\]
at each $\theta \in \Theta^*$ and $h \in Z$. Define a joint mapping
\begin{equation}
\Phi_{t_i, r_{t_i}(a_i), m}(\theta, t_{-i}, a_{-i}) \mapsto \left( \phi_{t_i, r_{t_i}(a_i), m}(\theta), (t_{-i}, r_{t_{-i}}(a_{-i}), m) \right) 
\end{equation}
on tuples for which $a_{-i} \in S_{i}^{S_{t_{-i}}[t_{-i}]}$. We define the belief of each type $(t_i, r_{t_i}(a_i), m)$ by
\begin{equation}
\kappa_{t_i, r_{t_i}(a_i), m} = \pi_{t_i, r_{t_i}(a_i)} \circ \Phi_{t_i, r_{t_i}(a_i), m}^{-1}.
\end{equation}

Note that $\kappa_{t_i, r_{t_i}(a_i), m}$ has a natural meaning. Imagine a type $t_i$ who wants to play $r_{t_i}(a_i)$ under a belief $\pi_{t_i, r_{t_i}(a_i)}$ about $(\theta, t_{-i}, a_{-i})$. Suppose he assumes that payoffs are fixed as if after $m$ the continuation will be according to him playing $r_{t_i}(a_i)$ and the others playing according to what is implied by his belief $\pi_{t_i, r_{t_i}(a_i)}$. Now he considers the outcome paths up to length $m$ in conjunction with $(\theta, t_{-i})$. His belief is then $\kappa_{t_i, r_{t_i}(a_i), m}$. Let $\hat{\Theta}^m = \cup_{t_i, r_{t_i}(a_i)} \Phi_{t_i, r_{t_i}(a_i), m}(\Theta)$. The perturbed model is $\left( \hat{\Theta}^m, \hat{T}^m, \kappa \right)$. We write $\hat{G}^m = \left( \Gamma, \hat{\Theta}^m, \hat{T}^m, \kappa \right)$ for the resulting Bayesian game, which we will sometimes refer to as a normal-form game.

**Step 2.** For each $t_i$ and $a_i \in S_{i}^{\infty} [t_i]$, the hierarchies $h_i(t_i, r_{t_i}(a_i), m)$ converge to $h_i(t_i)$.

**Proof:** Let $\hat{T}^\infty = \bigcup_{m=1}^{\infty} \hat{T}^m \cup T$ be a type space with beliefs as in each component of the union, and topology defined by the basic open sets being singletons $\{(t_i, r_{t_i}(a_i), m)\}$ together with sets $\{(t_i, r_{t_i}(a_i), m) : a_i \in S_{i}^{\infty} [t_i], m > k \} \cup \{t_i\}$ for each $t_i \in T$ and integer $k$. That is, the topology is almost discrete, except that there is non-trivial convergence of sequences $(t_i, r_{t_i}(a_i), m \to t_i)$. Note $\hat{T}^\infty$ is compact under this topology: from any basic open cover, we extract the sets containing the finitely many elements of $T$, and only finitely many elements will remain. Lemma 5 will now give the desired result, once we prove that the map $\kappa$ from types to beliefs is continuous. This continuity is the substance of the proof – if not for the need to prove this, our definition of the topology would have made the result true by fiat.

At types $(t_i, r_{t_i}(a_i), m)$ the topology is discrete and continuity is trivial, so it suffices to shows continuity at types $t_i$. Since $\Theta$ is finite, by continuity at infinity, for any $\epsilon$ we can pick an $m$ such that for all $\theta \in \Theta$, $\left| \theta_i(h) - \theta_i(\hat{h}) \right| < \epsilon$ whenever $h^m = \hat{h}^m$. Hence, by (A.5),
\begin{equation}
\left| \phi_{t_i, r_{t_i}(a_i), m}(\theta)(h) - \theta(h) \right| = \left| E \left[ \theta(\hat{h}) \left| h^m = h^m, r_{t_i}(a_i), \pi_{t_i, r_{t_i}(a_i)} \right. \right] - \theta(h) \right| 
\leq E \left[ \left| \theta(\hat{h}) - \theta(h) \right| \left| h^m = h^m, r_{t_i}(a_i), \pi_{t_i, r_{t_i}(a_i)} \right. \right] < \epsilon.
\end{equation}
Thus, $\phi_{t_i, r_{t_i}(a_i), m}(\theta)(h) \to \theta(h)$ for each $h$, showing that $\phi_{t_i, r_{t_i}(a_i), m}(\theta) \to \theta$. From the definition (A.6) we see that this implies $\phi_{t_i, r_{t_i}(a_i), m}(\theta, t_{-i}, a_{-i}) \to (\theta, t_{-i})$ as $m \to \infty$. (Recall that $(t_{-i}, r_{t_{-i}}(a_{-i}), m) \to t_{-i}$) Therefore, by (A.7), as $m \to \infty$,
\begin{equation}
\kappa_{t_i, r_{t_i}(a_i), m} \to \pi_{t_i, r_{t_i}(a_i)} \circ \text{proj}_{\Theta \times T_{-i}}^{-1} \text{marg}_{\Theta \times T_{-i}}(\pi_{t_i, r_{t_i}(a_i)}) = \kappa_{t_i}.
\end{equation}
which is the desired result.

Step 3. The strategy profile \( s^* : \tilde{T}^m \to A \) with \( s^*_i (t_i, r_{t_i} (a_i), m) = r_{t_i} (a_i) \) is a Bayesian Nash equilibrium in \( \tilde{\Theta}^m \).

**Proof:** Towards defining the belief of a type \( (t_i, r_{t_i} (a_i), m) \) under \( s^*_{-i} \), define mapping

\[
\gamma : (\theta, t_{-i}, r_{t_{-i}} (a_{-i}), m) \mapsto (\theta, t_{-i}, r_{t_{-i}} (a_{-i}), m, r_{t_{-i}} (a_{-i})) ;
\]

which describes \( s^*_{-i} \). Then, given \( s^*_{-i} \), his beliefs about \( \Theta \times \tilde{T}_{-i} \times A_{-i} \) is

\[
\pi \left( \cdot | t_i, r_{t_i} (a_i), m, s^*_{-i} \right) = \kappa_{t_i, r_{t_i} (a_i), m} \circ \gamma^{-1} = \pi_{t_i, r_{t_i} (a_i)} \circ \varphi_{t_i, r_{t_i} (a_i), m} \circ \gamma^{-1} ,
\]

where the second equality is by (A.7). His induced belief about \( \Theta \times A_{-i} \) is

\[
marg_{\Theta \times A_{-i}} \pi \left( \cdot | t_i, r_{t_i} (a_i), m, s^*_{-i} \right) = \pi_{t_i, r_{t_i} (a_i)} \circ \varphi_{t_i, r_{t_i} (a_i), m} \circ \gamma^{-1} \circ \text{proj}_{\Theta \times A_{-i}}^{-1} \quad (A.8)
\]

where \( r_{-i} : (t_{-i}, a_{-i}) \mapsto r_{t_{-i}} (a_{-i}) \). To see this, note that

\[
\text{proj}_{\Theta \times A_{-i}} \circ \gamma \circ \varphi_{t_i, r_{t_i} (a_i), m} : (\theta, t_{-i}, a_{-i}) \mapsto \left( \varphi_{t_i, r_{t_i} (a_i), m} (\theta), r_{t_{-i}} (a_{-i}) \right).
\]

Now consider any deviation \( a'_i \) such that \( a'_i (h) = r_{t_i} (a_i) (h) \) for every history longer than \( m \). It suffices to focus on such deviations because the moves after length \( m \) are payoff-irrelevant under \( \tilde{\Theta}^m \) by (A.5). The expected payoff vector from any such \( a'_i \) is

\[
E \left[ u (\theta, a'_i, s^*_{-i}) \right] = E \left[ u \left( \varphi_{t_i, r_{t_i} (a_i), m} (\theta), a'_i, r_{t_{-i}} (a_{-i}) \right) \right] = \pi_{t_i, r_{t_i} (a_i)} \left( \varphi_{t_i, r_{t_i} (a_i), m} (\theta), \pi_{t_i, r_{t_i} (a_i)} (a'_i, r_{t_{-i}} (a_{-i})) m, r_{t_i} (a_i) \right) \left[ \pi_{t_i, r_{t_i} (a_i)} \right]
\]

where the first equality is by (A.8); the second equality is by definition of \( u \); the third equality is by definition of \( \varphi_{t_i, r_{t_i} (a_i), m} \), which is (A.5); the fourth equality is by the fact that \( a'_i \) is equal to \( r_{t_i} (a_i) \) conditional on history \( z (a'_i, r_{t_{-i}} (a_{-i})) m \), and the fifth equality is by the law of iterated
expectations. Hence, for any such $a'_i$,

$$E\left[u_i(\theta, r_{t_i}(a_i), s_{-i}) | \kappa_{t_i, r_{t_i}(a_i), m}\right] = E \left[\theta_i(z(r_{t_i}(a_i), r_{t_{-i}}(a_{-i})) | \pi_{t_i, r_{t_i}(a_i)})\right]$$

$$\geq E \left[\theta_i(z(a'_i, r_{t_{-i}}(a_{-i})) | \pi_{t_i, r_{t_i}(a_i)})\right]$$

$$= E \left[u_i(\theta, a'_i, s^*_{-i}) | \kappa_{t_i, r_{t_i}(a_i), m}\right],$$

where the inequality is by the fact that $r_{t_i}(a_i)$ is a best reply to $\pi_{t_i, r_{t_i}(a_i)}$, by definition of $\pi_{t_i, r_{t_i}(a_i)}$. Therefore, $r_{t_i}(a_i)$ is a best reply for type $(t_i, r_{t_i}(a_i), m)$, and hence $s^*$ is a Bayesian Nash equilibrium.

**Step 4.** Referring back to the statement of the proposition, by Step 2, pick $m$, $t_i$, and $a_i$ such that $m > L$ and $h_i((t_i, r_{t_i}(a_i), m)) \in U_i$. By Step 3, $a_i$ is rationalizable for type $(t_i, r_{t_i}(a_i), m)$.

**Proof:** Since $h_i((t_i, r_{t_i}(a_i), m)) \to h_i(t_i)$ and $U_i$ is an open neighborhood of $t_i$, $h_i((t_i, r_{t_i}(a_i), m)) \in U_i$ for sufficiently large $m$. Hence, we can pick $m$ as in the statement. Moreover, by Step 3, $r_{t_i}(a_i)$ is rationalizable for type $(t_i, r_{t_i}(a_i), m)$ (because it is played in an equilibrium). This implies also that $a_i$ is rationalizable for type $(t_i, r_{t_i}(a_i), m)$, because $m$-equivalent actions are payoff-equivalent for type $(t_i, r_{t_i}(a_i), m)$.

The remaining steps will show that a further perturbation makes $a_i$ uniquely rationalizable.

**Step 5.** Define hierarchy $h_i(\tilde{t}_i) \in T^{m}_i$ for the finite-horizon game form $\Gamma^m$ by

$$h_i(\tilde{t}_i) = (\tau^m_i)^{-1}(h_i((t_i, r_{t_i}(a_i), m)))$$

where $\tau^m_i$ is as defined in Lemma 7 of Section A.2. Type $\tilde{t}_i$ comes from a finite game $G^m = (\Gamma^m, \Theta^m, T^m, \kappa)$ and $a^m_i \in S^\infty_i[\tilde{t}_i]$.

**Proof:** By Lemma 7, since type $(t_i, r_{t_i}(a_i), m)$ is from a finite model, so is $\tilde{t}_i$. Since $a_i$ is rationalizable for type $(t_i, r_{t_i}(a_i), m)$, by Lemma 7, $a^m_i$ is rationalizable for $h_i(\tilde{t}_i)$ and hence for type $\tilde{t}_i$ in $G^m$.

**Step 6.** By Step 5 and Lemma 1, there exists a hierarchy $h_i(\tilde{t}^m_i)$ in the open neighborhood $(\tau^m_i)^{-1} (U_i)$ of $h_i(\tilde{t}_i)$ such that each element of $S^\infty_i[\tilde{t}^m_i]$ is $m$-equivalent to $a^m_i$, and $\tilde{t}^m_i$ is a type in a finite, common-prior model.

**Proof:** By the definition of $h_i(\tilde{t}_i)$ in Step 5, $h_i(\tilde{t}_i) \in (\tau^m_i)^{-1}(U_i)$. Since $U_i$ is open and $\tau^m_i$ is continuous, $(\tau^m_i)^{-1}(U_i)$ is open. Moreover, $\tilde{t}_i$ comes from a finite game, and $a^m_i$ is rationalizable for $\tilde{t}_i$. Therefore, by Lemma 1, there exists a hierarchy $h_i(\tilde{t}^m_i)$ in $(\tau^m_i)^{-1}(U_i)$ as in the statement above.
Step 7. Define the hierarchy $h_i (\hat{t}_i)$ by $$h_i (\hat{t}_i) = \tau_i^m (h_i (\hat{p}^m_i)).$$ The conclusion of the proposition is satisfied by $\hat{t}_i$.

Proof: Since $h_i (\hat{p}^m_i) \in (\tau_i^m)^{-1} (U_i)$, $$h_i (\hat{t}_i) = \tau_i^m (h_i (\hat{p}^m_i)) \in \tau_i^m \left( (\tau_i^m)^{-1} (U_i) \right) \subseteq U_i.$$ Since $\hat{p}^m_i$ is a type from a finite, common-prior model, by Lemma 7, $\hat{t}_i$ can also be picked from a finite, common-prior model. Finally, take any $\hat{a}^m_i$ is $m$-equivalent to $a^m_i$. It then follows that $\hat{a}_i$ is and $m$-equivalent to $a_i$. Since $m > L$, $\hat{a}_i$ is also $L$-equivalent to $a_i$.

Appendix B. Proof of Proposition 2

Using Proposition 1, we first establish that every action is uniquely rationalizable for some type. This extends the lemma of Chen from equivalence at histories of bounded length to equivalence at histories of unbounded length.

Lemma 8. For all plans of action $a_i$, there is a type $t^{a_i}$ of player $i$ such that $a_i$ is the unique rationalizable action for $t^{a_i}$, up to reduced-form equivalence.

Proof. The set of non-terminal histories is countable, as each of them has finite length. Fixing any $i$ and $a_i$, index the set of histories where it is $i$’s move and the history thus far is consistent with $a_i$ as $\{h_k : k \in \mathbb{Z}^+\}$. By Proposition 1, for each $k$ there is a type $t^k_i$ whose rationalizable actions are always consistent with history $h_k$. We construct type $t^{a_i}$ as follows: his belief about $t_{-i}$ assigns probability $2^{-k}$ to type $t^k_i$. His belief about $\theta$ is a point-mass on the function $\theta_{a_i}$, defined as 1 if all of $i$’s actions were consistent with $a_i$ and $1 - 2^{-k}$ if his first inconsistent move was at history $h_k$. Now, if type $t^{a_i}$ plays action $a_i$ he receives a certain payoff of 1. If his plan $b_i$ is not reduced-form equivalent to $a_i$, let $h_k$ be the shortest history in the set $\{h_k : k \in \mathbb{Z}^+\}$ where $b_i(h_k) \neq a_i(h_k)$. By construction, there is probability at least $2^{-k}$ of reaching this history if he believes the other player’s action is rationalizable, so his expected payoff is at most $1 - 2^{-2k}$. This completes the proof.

Proof of Proposition 2. Construct a family of types $t_{j,m,\lambda}$, $j \in \mathbb{N}$, $m \in \mathbb{N}$, $\lambda \in [0,1]$, by

$$t_{j,0,\lambda} = t^{a_j},$$

$$\kappa_{t_{j,m,\lambda}} = \lambda \kappa_{t^{a_j}} + (1 - \lambda) \delta(\sigma_{t_{-i,m-1,\lambda}}) \quad \forall m > 0,$$
where $\delta(\theta^*, t_{-i,m-1,\lambda})$ is the Dirac measure at $(\theta^*, t_{-i,m-1,\lambda})$. For large $m$ and small $\lambda$, $t_{i,m,\lambda}$ satisfies all the desired properties of $\hat{t}_i$, as we establish below.

First note that for $\lambda = 0$, under $t_{i,m,0}$, it is $m$th-order mutual knowledge that $\theta = \theta^*$. Hence, as $m \to \infty$, $t_{i,m,0} \to t^C (\theta^*)$. Therefore, there exists $\tilde{m} > 0$ such that $h_i (t_{i,\tilde{m},0}) \in U_i$. Moreover, for $j \in N$, $m \leq \tilde{m}$, and $\lambda \in [0,1]$, the beliefs of $t_{j,m,\lambda}$ are continuous in $\lambda$. Hence, by Lemma 5, as $\lambda \to 0$, $h_i (t_{i,\tilde{m},\lambda}) \to h_i (t_{i,\tilde{m},0})$. Thus, there exists $\tilde{\lambda} > 0$ such that $h_i (t_{i,\tilde{m},\lambda}) \in U_i$ for all $\lambda < \tilde{\lambda}$.

Next, we use mathematical induction on $m$ to show that for all $\lambda > 0$ and for all $m$ and $j$, $a_j \in S_j^\infty [t_{j,m,\lambda}]$ if and only if $a_j$ is equivalent to $a^*_j$, establishing the first conclusion. This statement is true for $m = 0$ by definition of $t_{j,0,\lambda}$ and Lemma 8. Now assume that it is true up to some $m - 1$. Consider the type $t_{j,m,\lambda}$. Under any rationalizable belief, with probability $\lambda$ his belief is the same as that of $t^\theta_j$ to which $a^*_j$ is the unique best response in reduced form actions, and with probability $1 - \lambda$ the true state is $\theta^*$ and the other players play an action that is equivalent to $a^*_{-j}$, in which case $a^*_j$ is a best reply, as $a^*$ is a Nash equilibrium under $\theta^*$. Therefore, in reduced form $a^*_j$ is the unique best response to any of his rationalizable beliefs, proving the statement.

Now, for any $m > 0$ and any rationalizable belief $\pi$ of $t_{i,m,\lambda}$, observe that by the previous statement and the definition of $t_{i,m,\lambda}$, the type $t_{i,m,\lambda}$ assigns at least probability $1 - \lambda$ on $(\theta^*, a^*_{-i})$. Hence, $\Pr (z (a^*_j) | \pi, a^*_j) \geq 1 - \lambda$. Since the payoffs are all in $[0,1]$, this further implies that $E [u_i (\theta, a) | \pi, a^*_j] - u_i (\theta^*, a^*) \in [-\lambda, \lambda]$. Hence, $\hat{t}_i = t_{i,m,\lambda}$ for $\lambda \in (0, \min \{\tilde{\lambda}, \varepsilon\})$ satisfies all the desired properties.

References


\[7\] To ensure compactness, put all of the types in construction of types $t^{\theta_j}$ together and for $t_{j,m,\lambda}$ with $j \in N$, $m \in \{0,1,\ldots, \tilde{m}\}$, $\lambda \in [0,1]$, use the usual topology for $(j,m,\lambda)$. 


*Weinstein: Northwestern University; Yildiz: MIT*