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

Virtually all economic models make a common-knowledge assumption. That is, they assume that the payoff functions have a particular structure, every player knows this, every player knows that every player knows this, and so on, ad infinitum. Moreover, in the resulting games there are often a large number of Nash equilibria and even more rationalizable strategies. The sets of equilibrium strategies and rationalizable strategies are especially large in dynamic games with infinite horizon. For example, the literature on repeated games is filled with folk theorems, concluding that every individually rational payoff can be supported by a subgame-perfect equilibrium. For a less prominent example, consider Rubinstein's (1982) bargaining game. Although there is a unique subgame-perfect equilibrium, in which there is immediate agreement, many outcomes with long delays can be supported in Nash equilibrium. In this game, virtually any outcome, including perpetual disagreement, can be realized in ex ante rationalizable strategy profiles. Moreover, some of these equilibria may involve counterintuitive moves, such as sequentially irrational moves. For this reason, ex ante rationalizability or Nash equilibrium are considered too permissive, and economists usually focus on refinements of equilibrium. For example, in the analysis of the bargaining game, one focuses on the unique subgame-perfect equilibrium, and in the analysis of repeated games one makes a stricter refinement by focusing on efficient subgame-perfect equilibria. This approach is so common that we rarely think about rationalizable strategies in many extensively-analyzed dynamic games.

In this paper, building on existing theorems for finite games, we prove a structure theorem for rationalizability that questions the premises of the above approach. We consider an arbitrary dynamic game that is continuous at infinity, and has finitely many moves at each information set and a finite type space. Note that virtually all of the games analyzed in economics, such as repeated games with discounting and bargaining games, are continuous at infinity. For any type profile 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.[1] Fix some arbitrary 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 action. The 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 that 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.

[1]The usual notation in dynamic games and games of incomplete information clash; action a_i stands for a move in dynamic games but for an entire contingent plan in incomplete-information games; t stands for time in dynamic games but type profile in incomplete-information games; h_i stands for history in dynamic games but hierarchy in incomplete-information games, etc. Following Chen, we will use the notation customary in incomplete information games, so a_i is a complete contingent plan of action. We will sometimes use "move" to distinguish an action at a single node.

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. 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, an assumption that is made in almost all 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 infinitely-repeated prisoners dilemma, there are uncountably many strategies, even in reduced form.

We now briefly explain the implications of our structure theorem to the research agenda at the beginning.[2] 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 $10^{1,000,000}$ 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 specification of interim beliefs.

One can formally derive this from our result 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

[2]For a more detailed discussion of the ideas in this paragraph, we refer to Weinstein and Yildiz (2007). In particular, there, we have extensively discussed the meaning of perturbing interim beliefs from the perspective of economic modelling and compared alternative perturbations, such as the ex-ante perturbation of Kajii and Morris (1997).

about the payoff functions), the players' second-order beliefs (which are about the first-order beliefs), ... , 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 exist 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, in which case h is the only outcome for the first L moves *according to his refinement* as well.

We conclude our discussion by explaining the implications of our result for repeated games, assuming that (before our perturbations) there is complete information. Under appropriate conditions on such games we have folk theorems, which conclude that for every individually rational payoff profile v there exists a subgame-perfect equilibrium that leads to v as the average discounted payoff. The next step is often to find the structure and the implications of the most efficient subgame-perfect equilibria. This of course, in effect, employs a further refinement by ruling out all the other equilibria, which span most of the payoff space. This refinement is the basis for most of the theoretical predictions in applications of repeated games. Now, consider an uninteresting or unwanted pure-strategy equilibrium a^* that leads to a payoff v^* and is disregarded by the refinement which selects for efficiency. Fix an arbitrarily long L. Now, our result establishes that there is a perturbation of the original model, represented by some type profile $\hat{t}$, under which all players assign high probability to the event that the payoffs are close to the one as described in the original model, all players assign high probability to such beliefs, and so on, up to arbitrarily high orders, and yet *all* rationalizable plans play according to a^* for the first L periods. Hence, a fortiori, under any refinement of equilibrium, the actual behavior of the players will be almost exactly that of the equilibrium which the refinement tried to eliminate. Moreover, for large L, the expected payoff of these types will be approximately v^* in any rationalizable profile. Therefore, our result has two implications. First, it implies a stronger folk theorem: any individually rational payoff vector can be approximated as the unique rationalizable payoff under some perturbation. Second, although one may be disturbed by the large number of possible outcomes in the original common-knowledge case, one cannot refine the solution any further without making extremely specific assumptions on information structures.

2. Basic Definitions

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 $\Gamma = \left(N, H, (\mathcal{I}_i)_{i\in N}\right)$ with perfect recall where $N = \{1, 2, \ldots, n\}$ is a finite set of players, H is a set of histories, and $\mathcal{I}_i$ is the set of information sets at which player $i \in N$ moves. We designate $i \in N$ and $h \in H$ as the generic player and history, respectively. We write $I_i(h)$ for the information set that contains history h, at which player i moves. Here, $I_i(h)$ is 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* a_i of i is defined as any contingent plan 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 $a_{-i} \in A_{-i}$.

[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 $f_j : X_j \to Y_j$, we define $f_{-i} : X_{-i} \to X_{-i}$ 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 σ-algebra and the weak topology.

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 Θ^* for the set of all payoff functions $\theta : Z \to [0,1]^n$. 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 Θ^* with the product topology (i.e. the topology of pointwise convergence). Note that Θ^* is compact and u_i is continuous in θ. Note, however, that Θ^* 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 (Θ, T, κ) 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 Γ.

Given any type t_i in a type space T, we can compute the first-order belief $h_i^1(t_i) \in \Delta(\Theta^*)$ of t_i (about θ), second-order belief $h_i^2(t_i) \in \Delta(\Theta^* \times \Delta(\Theta^*)^n)$ *of* t_i (about θ and the first-order beliefs), etc., using the joint distribution of the types and θ. Using the mapping $h_i : t_i \mapsto (h_i^1(t_i), h_i^2(t_i), \ldots)$, 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 usual weak convergence. We say that a sequence of types $t_i(m)$ converges to a type t_i, denoted by $t_i(m) \to t_i$, if and only if $h_i^k(t_i(m)) \to h_i^k(t_i)$ for each k, where the latter convergence is in weak topology, i.e., "convergence in distribution."

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 π.

Interim Correlated Rationalizability. For each i and t_i, set $S_i^0[t_i] = A_i$, and define sets $S_i^k[t_i]$ for $k > 0$ iteratively, by letting $a_i \in S_i^k[t_i]$ if and only if $a_i \in BR_i\left(\text{marg}_{\Theta \times A_{-i}}\pi\right)$ 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 $\pi\left(a_{-i} \in S_{-i}^{k-1}[t_{-i}]\right) = 1$. That is, a_i is a best response to a belief of t_i that puts positive probability only to the actions that survive the elimination in round $k-1$. We write $S_{-i}^{k-1}[t_{-i}] = \prod_{j \neq i} S_j^{k-1}[t_j]$ and

$S^k[t] = S_1^k[t_1] \times \cdots \times S_n^k[t_n]$. The set of all rationalizable actions for player i with type t_i is

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

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 $h = (b^l)_{l=1}^L$ is formed by L moves for some finite L, then h is said to be *finite* and have length L. If h contains infinitely many moves, then 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 $h = (b^l)_{l=1}^L$ and any L', we write $h^{L'}$ for the subhistory of h that is truncated at length L'; i.e. $h = (b^l)_{l=1}^{\min\{L,L'\}}$. We say that θ is *continuous at infinity* iff for any $\varepsilon > 0$, there exists $L < \infty$, such that

$$\left|\theta_i(h) - \theta_i(\tilde{h})\right| < \varepsilon \text{ whenever } h^L = \tilde{h}^L \tag{2.1}$$

for all $i \in N$ and all terminal histories $h, \tilde{h} \in Z$. 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 to the games that are continuous at infinity throughout, including our perturbations. Note that most games analyzed in economics are continuous at infinity. This includes all finite-horizon games, repeated games with discounting, games of sequential bargaining, and so on. 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.

Weinstein and Yildiz (2007) have proven a version of this structure theorem for finite action games under a richness assumption on Θ^* that is natural for static games but rules out fixing a dynamic extensive game form. Chen (2008) has proven this result for finite 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 $a_i \in S_i^\infty[t_i]$ of 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 action space A is used in a technical but crucial step that ensures that the constructed type is indeed well-defined, having 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 actions. (Recall that an action here is a complete contingent plan of a type, not a move.) 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 $h_i\left(\hat{t}_i\right) \in U_i$, such that for each $a_i' \in S_i^\infty\left[\hat{t}_i\right]$, a_i' is L-equivalent to a_i, and $\hat{t}_i$ is a type in some finite, common-prior model.*

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, ..., upto 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, he cannot practically rule out any rationalizable outcome as the unique solution.

Notice that Proposition 1 differs from Lemma 1 only 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 involved 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 $\Theta = \left\{\bar{\theta}\right\}$ 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^*, which describes different continuations at different histories. Call the resulting history h^{m,a^*}. This can also be described as a payoff perturbation: define the perturbed payoff function θ^m by setting $\theta^m(h) = \bar{\theta}\left(h^{m,a^*}\right)$ at every terminal history h. Now consider the complete-information game with perturbed model $\tilde{\Theta}^m = \{\theta^m\}$ and $T^m = \{\bar{t}^m\}$, where according to $\bar{t}^m$ it is common knowledge that the payoff function is θ^m (essentially, players are forced to play according to a^* after the mth information set). We make three observations towards proving the proposition. We first observe that, since $\bar{\theta}$ is continuous at infinity, by construction, $\theta^m \to \bar{\theta}$, implying that $h_i(\bar{t}_i^m) \to h_i(\bar{t}_i)$. Hence, there exists $\bar{m} > L$ such that $h_i(\bar{t}_i^{\bar{m}}) \in U_i$. Second, there is a natural isomorphism between the payoff functions that do not depend on the moves after length $\bar{m}$, such as $\theta^{\bar{m}}$, 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 φ that maps the hierarchies in the universal type space $T^{\bar{m}*}$ 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 $t_i \in T^{\bar{m}*}$. We third observe that, 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: $a_i^* \in S_i^\infty[\bar{t}_i^{\bar{m}}]$. Now, these three observations together imply that the hierarchy $\varphi^{-1}(h_i(\bar{t}_i^{\bar{m}}))$ for the finite-horizon game form is in an open neighborhood $\varphi^{-1}(U_i) \subset T_i^{\bar{m}*}$ and the restriction $a_i^{*\bar{m}}$ of a_i^* to the truncated game form is rationalizable for $\varphi^{-1}(h_i(\bar{t}_i^{\bar{m}}))$. Hence, by Lemma 1, there exists a type $\tilde{t}_i$ such that (i) $h_i\left(\tilde{t}_i\right) \in \varphi^{-1}(U_i)$ and (ii) all rationalizable actions of $\tilde{t}_i$ are $\bar{m}$-equivalent to $a_i^{*\bar{m}}$. Now consider a type $\hat{t}_i$ with hierarchy $h_i\left(\hat{t}_i\right) \equiv \varphi\left(h_i\left(\tilde{t}_i\right)\right)$, 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), $h_i\left(\hat{t}_i\right) \in U_i$ because

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

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 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 equally applies to Chen's (2008) 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, these perturbed types may all find the path $z^L(a)$ unlikely, as the following example illustrates.

Example 1. Consider a twice-repeated prisoners' dilemma game with complete information. Recall that, in the stage game, the move Defect dominates the move Cooperate, although (Cooperate, Cooperate) is better than (Defect, Defect). In the repeated game 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.*

By Chen (2008), there exists a perturbation t^{T4T} of the common-knowledge type for which a^{T4T} is the unique rationalizable action. The unique rationalizable action profile (a^{T4T}, a^{T4T}) of type profile (t^{T4T}, t^{T4T}) leads to (Cooperate, Cooperate) in both rounds. Note that since t^{T4T} has a unique best reply, he does assign positive probability to the event that the other player cooperates in the first round, leading him to cooperate in the last round. Hence, it must be that the other player's cooperation in the first round makes him update his beliefs about the payoffs in such a way that Cooperate becomes a better response than Defect. Nevertheless, since his ex ante belief assigns high probability to the event that the payoffs are similar to those in repeated prisoner dilemma payoffs, under which Defect is a better response, it must be that t^{T4T} finds it unlikely that the other player will in fact play Cooperate in the first round. Hence, he assigns nearly probability one on the paths in which he plays Cooperate in the first round and Defect in the second round.

Therefore, in the situation described by the perturbation that leads to a unique rationalizable outcome, the players may anticipate a quite different scenario. One may then seek for perturbations in which the perturbed types assign high probability to the unique rationalizable path.

Considering the Nash equilibria of complete information games, we next establish a version of the structure theorem that does not have the limitations above. We fix a payoff function θ^*, and consider the game in which θ^* 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 formally, 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 $\pi\left(a_{-i} \in S_{-i}^{\infty}[t_{-i}]\right) = 1$. We write $\Pr(\cdot|\pi, a_i)$ and $E[\cdot|\pi, a_i]$ for the resulting probability measure and expectation operator from playing a_i against belief π, respectively. With this formalism, our result is stated as follows.

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

(1) $a_i \in S_i^{\infty}\left[\hat{t}_i\right]$ *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 hierarch of beliefs, both in terms of actual realization and in terms of players' ex-ante conjectures.

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 $h = \left(b^l\right)_{l \in \mathbb{N}}$ of profiles $b^l = \left(b_1^l, \ldots, b_n^l\right)$. In the complete-information game, the players maximize the average discounted stage payoffs. That is, the payoff function is

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

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

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

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

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

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

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. *Assume that V^* has a non-empty interior $intV^*$. Then, for every $v \in intV^*$, there exists $\bar{\delta} < 1$ such that for all $\delta \in \left(\bar{\delta}, 1\right)$, G_δ has a subgame-perfect equilibrium a^* in pure strategies, such that $u\left(\theta_\delta^*, a^*\right) = v$.*

Under a weak full-rank assumption, excluding the boundary, the lemma concludes that every feasible and individually rational payoff vector 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). Combining such a folk theorem with Proposition 2, our next result establishes that the multiplicity is irreducible.

Proposition 3. *Assume that V^* has a non-empty interior $intV^*$. Then, for all $v \in intV^*$ and $\varepsilon > 0$, there exists $\bar{\delta} < 1$ such that for all $\delta \in (\bar{\delta}, 1)$, every open neighborhood U of $t^{CK}(\theta^*_\delta)$ contains a type $\hat{t} \in U$ such that*

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

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

Proof. Fix any $v \in intV^*$ and $\varepsilon > 0$. By Lemma 2, there exists $\bar{\delta} < 1$ such that for all $\delta \in (\bar{\delta}, 1)$, G_δ has a subgame-perfect equilibrium a^* in pure strategies, such that $u(\theta^*_\delta, a^*) = v$. Then, by Proposition 2, for any $\delta \in (\bar{\delta}, 1)$ 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 π of $\hat{t}_i$, the expected payoffs are all within ε neighborhood of $u(\theta^*_\delta, a^*) = v$ (Part 3 of Proposition 2). □

Under the usual full-rank assumption for standard folk theorems, Proposition 3 establishes an unusual folk theorem. It concludes that every individually rational and feasible payoff v in the interior can be supported by 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 ε 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.

Appendix A. Proof of Structure Theorem

We start with describing the notation we use in the appendix.

Notation 1. For any belief $\pi \in \Delta(\Theta \times A_{-i})$ and action a_i and for any history h, write $E[\cdot|h, a_i, \pi]$ for the expectation operator induced by action a_i and π conditional on reaching history h. For any strategy profile $s : T \to A$ 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 $f : W \to X$ and $g : Y \to Z$, 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, (Θ, T, κ) is a model, and $u : \Theta \times A \to [0, 1]^n$ is the payoff function. While this notation is consistent with our formulation, we will also define some auxiliary Bayesian games with different action spaces, payoff functions and parameter spaces. For any $G = (N, A, u, \Theta, T, \kappa)$, we say that a_i and a_i' are *G-equivalent* if

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

By a *reduced-form* game, we mean a game $G_R = (N, \bar{A}, u, \Theta, T, \kappa)$ where $\bar{A}_i$ contains at least one representative action from each G-equivalence class for each i. 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 $S_i^\infty[t_i]$ 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, $\mu_i : A_i \to A_i'$, $i \in N$, be onto mappings, and $\varphi : \Theta \to \Theta'$ and $\tau_i : T_i \to T_i'$, $i \in N$, 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 (θ, a). Then, for any t_i and a_i,*

$$a_i \in S_i^\infty[t_i] \iff \mu_i(a_i) \in S_i^\infty[\tau_i(t_i)]. \tag{A.1}$$

Note that the bijections φ and τ are a renaming, and (i) ensures that the beliefs do not change under the renaming. On the other hand, μ_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 $a_i, a_i' \in A_i$,

$$a_i \text{ is } G\text{-equivalent to } a_i' \iff \mu_i(a_i) \text{ is } G'\text{-equivalent to } \mu_i(a_i'). \tag{A.2}$$

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 μ 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'.[4] Note that G_R and G_R' are isomorphic up to the renaming of actions, parameters, and types by μ, φ, and τ, 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). □

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 κ defining the type space is continuous.

Lemma 5 (Mertens and Zamir (1985)). *Let (Θ, T, κ) be any model, endowed with any topology, such that $\Theta \times T$ is compact and κ_{t_i} 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 $\Gamma^m = (N, H^m, (\mathcal{I}_i)_{i \in N})$ 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

$$\bar{\Theta}^m = \left([0,1]^{Z^m}\right)^n$$

[4]Proof: Since μ_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, $\bar{\Theta}^m$ is not necessarily a subset of Θ^*. We will now embed $\bar{\Theta}^m$ into Θ^* through an isomorphism to a subset of Θ^*. Define the subset

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

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 $\varphi_m : \bar{\Theta}^m \to \hat{\Theta}^m$ defined by setting

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

for all $\theta \in \bar{\Theta}^m$ and $h \in Z$, where $h^m \in H^m$ is the truncation of h at length m. Clearly, under the product topologies, φ_m is an isomorphism, in the sense that it is one-to-one, onto, and both φ_m and φ_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) $\Theta^m \subset \bar{\Theta}^m$, (ii) $\Theta = \varphi_m(\Theta^m)$ and (iii) $T_i = \tau_i(T_i^m)$ for some bijection τ_i^m and such that $\kappa_{\tau_i^m(t_i^m)} = \kappa_{t_i^m} \circ \left(\varphi_m, \tau_{-i}^m\right)^{-1}$ for each $t_i^m \in T_i^m$. Then, the set of rationalizable actions are m-equivalent in G and G^m:*

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

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 $u^m\left(\varphi_m^{-1}(\theta), a^m\right) = u(\theta, a)$ for all (θ, 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

$$\begin{aligned} u^m\left(\varphi_m^{-1}(\theta), a^m\right) &= \varphi_m^{-1}(\theta)(z^m(a^m)) = \varphi_m^{-1}(\theta)(z(a)^m) \\ &= \varphi_m\left(\varphi_m^{-1}(\theta)\right)(z(a)) = \theta(z(a)) = u(\theta, a). \end{aligned}$$

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 φ_m. □

Let T^{*m} be the $\bar{\Theta}^m$-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 $\tau^m : T^{*m} \to T^*$ 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^{*m} and T^* do not have any redundant type, by the analysis of Mertens and Zamir (1985), there exists a continuous and one-to-one mapping τ^m such that

$$\kappa_{\tau_i^m(t_i)} = \kappa_{t_i} \circ \left(\varphi_m, \tau_{-i}^m\right)^{-1} \tag{A.4}$$

for all i and $t_i \in T_i^{*m}$.[5] First two statements immediately follow from (A.4). Part 3 follows from (A.4) and Lemma 6. □

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

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

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

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

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 $\text{marg}_{\Theta \times T_{-i}}(\pi^{t_i, r_{t_i}(a_i)}) = \kappa_{t_i}$. Define a mapping $\phi_{t_i, r_{t_i}(a_i), m} : \Theta^* \to \Theta^*$ between the payoff functions by setting

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

[5] If one writes $t_i = \left(t_i^1, t_i^2, \ldots\right)$ and $\tau_i^m(t_i) = \left(\tau_i^{m,1}\left(t_i^1\right), \tau_i^{m,2}\left(t_i^2\right), \ldots\right)$ as a hierarchies, we define τ_i^m inductively by setting $\tau_i^{m,1}\left(t_i^1\right) = t_i^1 \circ \varphi_m^{-1}$ and $\tau_i^{m,k}\left(t_i^k\right) = t_i^k \circ \left(\varphi_m, \tau_{-i}^{m,1}, \ldots, \tau_{-i}^{m,k-1}\right)^{-1}$ for $k > 1$.

at each $\theta \in \Theta^*$ and $h \in Z$. Define a joint mapping

$$\bar{\phi}_{t_i,r_{t_i}(a_i),m} : (\theta, t_{-i}, a_{-i}) \mapsto \left(\phi_{t_i,r_{t_i}(a_i),m}(\theta), (t_{-i}, r_{t_{-i}}(a_{-i}), m)\right) \tag{A.6}$$

on tuples for which $a_{-i} \in S_{-i}^{\infty}[t_{-i}]$. We define the belief of each type $(t_i, r_{t_i}(a_i), m)$ by

$$\kappa_{t_i,r_{t_i}(a_i),m} = \pi^{t_i,r_{t_i}(a_i)} \circ \bar{\phi}^{-1}_{t_i,r_{t_i}(a_i),m}. \tag{A.7}$$

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 (θ, 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 (θ, t_{-i}). His belief is then $\kappa_{t_i,r_{t_i}(a_i),m}$. Let $\tilde{\Theta}^m = \cup_{t_i,r_{t_i}(a_i)} \phi_{t_i,r_{t_i}(a_i),m}(\Theta)$. The perturbed model is $\left(\tilde{\Theta}^m, \tilde{T}^m, \kappa\right)$. We write $\tilde{G}^m = \left(\Gamma, \tilde{\Theta}^m, \tilde{T}^m, \kappa\right)$ 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 $\tilde{T}^{\infty} = \bigcup_{m=1}^{\infty} \tilde{T}^m \cup T$ 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 $(t_i, r_{t_i}(a_i), m) \to t_i$. Since $\tilde{T}^{\infty}$ is compact under this topology, Lemma 5 will now give the desired result, once we prove that the map κ 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 Θ is finite, by continuity at infinity, for any ε we can pick an m such that for all $\theta \in \Theta$, $\left|\theta_i(h) - \theta_i(\tilde{h})\right| < \varepsilon$ whenever $h^m = \tilde{h}^m$. Hence, by (A.5),

$$\begin{aligned} \left|\phi_{t_i,r_{t_i}(a_i),m}(\theta)(h) - \theta(h)\right| &= \left|E\left[\theta\left(\tilde{h}\right) | \tilde{h}^m = h^m, r_{t_i}(a_i), \pi^{t_i,r_{t_i}(a_i)}\right] - \theta(h)\right| \\ &\leq E\left[\left|\theta\left(\tilde{h}\right) - \theta(h)\right| | \tilde{h}^m = h^m, r_{t_i}(a_i), \pi^{t_i,r_{t_i}(a_i)}\right] < \varepsilon. \end{aligned}$$

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 $\bar{\phi}_{t_i,r_{t_i}(a_i),m}(\theta, t_{-i}, a_{-i}) \to (\theta, t_{-i})$ as $m \to \infty$. (Recall that $\left(t_{-i}, r_{t_{-i}}(a_{-i}), m\right) \to t_{-i}$.) Therefore, by (A.7), as $m \to \infty$,

$$\kappa_{t_i,r_{t_i}(a_i),m} \to \pi^{t_i,r_{t_i}(a_i)} \circ proj^{-1}_{\Theta \times T_{-i}} = \text{marg}_{\Theta \times T_{-i}}(\pi^{t_i,r_{t_i}(a_i)}) = \kappa_{t_i},$$

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 $\tilde{G}^m$.

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

$$\pi(\cdot | t_i, r_{t_i}(a_i), m, s_{-i}^*) = \kappa_{t_i, r_{t_i}(a_i), m} \circ \gamma^{-1} = \pi^{t_i, r_{t_i}(a_i)} \circ \bar{\phi}_{t_i, r_{t_i}(a_i), m}^{-1} \circ \gamma^{-1},$$

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

$$\begin{aligned} \text{marg}_{\Theta \times A_{-i}} \pi(\cdot | t_i, r_{t_i}(a_i), m, s_{-i}^*) &= \pi^{t_i, r_{t_i}(a_i)} \circ \bar{\phi}_{t_i, r_{t_i}(a_i), m}^{-1} \circ \gamma^{-1} \circ proj_{\Theta \times A_{-i}}^{-1} \\ &= \pi^{t_i, r_{t_i}(a_i)} \circ \left(\phi_{t_i, r_{t_i}(a_i), m}, r_{-i}\right)^{-1} \end{aligned} \tag{A.8}$$

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

$$proj_{\Theta \times A_{-i}} \circ \gamma \circ \bar{\phi}_{t_i, r_{t_i}(a_i), m} : (\theta, t_{-i}, a_{-i}) \mapsto \left(\phi_{t_i, r_{t_i}(a_i), m}(\theta), r_{t_{-i}}(a_{-i})\right).$$

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

$$\begin{aligned} E\left[u(\theta, a_i', s_{-i}^*) | \kappa_{t_i, r_{t_i}(a_i), m}\right] &= E\left[u\left(\phi_{t_i, r_{t_i}(a_i), m}(\theta), a_i', r_{t_{-i}}(a_{-i})\right) | \pi^{t_i, r_{t_i}(a_i)}\right] \\ &= E\left[\phi_{t_i, r_{t_i}(a_i), m}(\theta)\left(z\left(a_i', r_{t_{-i}}(a_{-i})\right)\right) | \pi^{t_i, r_{t_i}(a_i)}\right] \\ &= E\left[E\left[\theta\left(z\left(a_i', r_{t_{-i}}(a_{-i})\right)\right) | z\left(a_i', r_{t_{-i}}(a_{-i})\right)^m, r_{t_i}(a_i), \pi^{t_i, r_{t_i}(a_i)}\right] | \pi^{t_i, r_{t_i}(a_i)}\right] \\ &= E\left[E\left[\theta\left(z\left(a_i', r_{t_{-i}}(a_{-i})\right)\right) | z\left(a_i', r_{t_{-i}}(a_{-i})\right)^m, a_i', \pi^{t_i, r_{t_i}(a_i)}\right] | \pi^{t_i, r_{t_i}(a_i)}\right] \\ &= E\left[\theta\left(z\left(a_i', r_{t_{-i}}(a_{-i})\right)\right) | \pi^{t_i, r_{t_i}(a_i)}\right], \end{aligned}$$

where the first equality is by (A.8); the second equality is by definition of u; the third equality is by definition of $\phi_{t_i, r_{t_i}(a_i), m}$, 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 $z\left(a_i', r_{t_{-i}}(a_{-i})\right)^m$, and the fifth equality is by the law of iterated expectations. Hence, for any such a_i',

$$\begin{aligned} E\left[u_i\left(\theta, r_{t_i}\left(a_i\right), s_{-i}^*\right)|\kappa_{t_i, r_{t_i}(a_i), m}\right] &= E\left[\theta_i\left(z\left(r_{t_i}\left(a_i\right), r_{t_{-i}}\left(a_{-i}\right)\right)\right)|\pi^{t_i, r_{t_i}(a_i)}\right] \\ &\geq E\left[\theta_i\left(z\left(a_i', r_{t_{-i}}\left(a_{-i}\right)\right)\right)|\pi^{t_i, r_{t_i}(a_i)}\right] \\ &= E\left[u_i\left(\theta, a_i', s_{-i}^*\right)|\kappa_{t_i, r_{t_i}(a_i), m}\right], \end{aligned}$$

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 $h_i(\tilde{t}_i) \in T_i^{*m}$ for the finite-horizon game form Γ^m by

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

where τ_i^m 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 $a_i^m \in S_i^\infty[\tilde{t}_i]$.

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_i^m 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 $h_i(\bar{t}_i^m)$ in open neighborhood $(\tau_i^m)^{-1}(U_i)$ of $h_i(\tilde{t}_i)$ such that each element of $S_i^\infty[\bar{t}_i^m]$ is m-equivalent to a_i^m, and $\bar{t}_i^m$ is a type in a finite, common-prior model.

Proof: By the definition of $h_i(\tilde{t}_i)$ in Step 5, $h_i(\tilde{t}_i) \in (\tau_i^m)^{-1}(U_i)$. Since U_i is open and τ_i^m is continuous, $(\tau_i^m)^{-1}(U_i)$ is open. Moreover, $\tilde{t}_i$ comes from a finite game, and a_i^m is rationalizable for $\tilde{t}_i$. Therefore, by Lemma 1, there exists a hierarchy $h_i(\bar{t}_i^m)$ in $(\tau_i^m)^{-1}(U_i)$ as in the statement above.

Step 7. Define the hierarchy $h_i\left(\hat{t}_i\right)$ by

$$h_i\left(\hat{t}_i\right) = \tau_i^m\left(h_i\left(\bar{t}_i^m\right)\right).$$

The conclusion of the proposition is satisfied by $\hat{t}_i$.

Proof: Since $h_i\left(\bar{t}_i^m\right) \in \left(\tau_i^m\right)^{-1}\left(U_i\right)$,

$$h_i\left(\hat{t}_i\right) = \tau_i^m\left(h_i\left(\bar{t}_i^m\right)\right) \in \tau_i^m\left(\left(\tau_i^m\right)^{-1}\left(U_i\right)\right) \subseteq U_i.$$

Since $\bar{t}_i^m$ 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 $\hat{a}_i \in S_i^\infty\left[\hat{t}_i\right]$. By Lemma 7, $\hat{a}_i^m \in S_i^\infty\left[\hat{t}_i\right]$. Hence, by Step 6, $\hat{a}_i^m$ is m-equivalent to a_i^m. 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 can be made rationalizable for some type. This extend 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. Index the set of histories where it is i's move and the history thus far is consistent with a_i as $\{h_k : k \in Z^+\}$. By Proposition 1, for each k there is a type t_{-i}^k 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_{-i}^k. His belief about θ is a point-mass on the function θ_{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 $\{h_k : k \in Z^+\}$ 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 $t_{j,m,\lambda}$, $j \in N$, $m \in \mathbb{N}$, $\lambda \in [0,1]$, by

$$\begin{aligned} t_{j,0,\lambda} &= t^{a_j^*}, \\ \kappa_{t_{j,m,\lambda}} &= \lambda \kappa_{t^{a_j^*}} + (1-\lambda)\,\delta_{\left(\theta^*, t_{-i,m-1,\lambda}\right)} \qquad \forall m > 0, \end{aligned}$$

where $\delta_{(\theta^*, t_{-i,m-1,\lambda})}$ is the Dirac measure at $(\theta^*, t_{-i,m-1,\lambda})$. For large m and small λ, $t_{i,m,\lambda}$ satisfies all the properties of $\hat{t}_i$, as we establish below.

First note that for $\lambda = 0$, under $t_{i,m,0}$, it is mth-order mutual knowledge that $\theta = \theta^*$. Hence, as $m \to \infty$, $t_{i,m,0} \to t_i^{CK}(\theta^*)$. Therefore, there exists $\bar{m} > 0$ such that $h_i(t_{i,\bar{m},0}) \in U_i$. Moreover, for $j \in N$, $m \leq \bar{m}$, and $\lambda \in [0,1]$, beliefs of $t_{j,m,\lambda}$ are continuous in λ. Hence, by Lemma 5,[6] as $\lambda \to 0$, $h_i(t_{i,\bar{m},\lambda}) \to h_i(t_{i,\bar{m},0})$. Thus, there exists $\bar{\lambda} > 0$ such that $h_i(t_{i,\bar{m},\lambda}) \in U_i$ for all $\lambda < \bar{\lambda}$.

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 λ his belief is the same as that of $t^{a_j^*}$ to which a_j^* is the unique best response in reduced form actions, and with probability $1-\lambda$ the true state is θ^* 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 θ^*. Therefore, a_j^* is the unique best response to any of his rationalizable belief in reduced form, proving the statement.

Now, for any $m > 0$ and any rationalizable belief π 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 (θ^*, a_{-i}^*). Hence, $\Pr(z(a^*)|\pi, a_i^*) \geq 1-\lambda$. Since the payoffs are all in $[0,1]$, this further implies that $E[u_i(\theta, a)|\pi, a_i^*] - u_i(\theta^*, a^*) \in [-\lambda, \lambda]$. Hence, $\hat{t}_i = t_{i,\bar{m},\lambda}$ for $\lambda \in (0, \min\{\bar{\lambda}, \varepsilon\})$ satisfies all the desired properties. $\square$

References

[1] Battigalli, P. and M. Siniscalchi (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics* 3-1, Article 3.

[2] Brandenburger, A. and E. Dekel (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189-198.

[3] Carlsson, H. and E. van Damme (1993): "Global Games and Equilibrium Selection," *Econometrica*, 61, 989-1018.

[4] Chen, Y. (2008): "A Structure Theorem for Rationalizability in Dynamic Games", Northwestern University Working Paper.

[5] Dekel, E. D. Fudenberg, S. Morris (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2, 15-40.

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

[6] Fudenberg, D. and E. Maskin (1991): "On the Dispansibility of Public Randomization in Discounted Repeated Games," *Journal of Economic Theory* 53, 428-438.

[7] Harsanyi, J. (1967): "Games with Incomplete Information played by Bayesian Players. Part I: the Basic Model," *Management Science* 14, 159-182.

[8] Kajii, A. and S. Morris (1997): "The Robustness of Equilibria to Incomplete Information," *Econometrica*, 65, 1283-1309.

[9] Mailath, G. and L. Samuelson (2006): *Repeated Games and Reputations*, Oxford University Press.

[10] Mertens and Zamir (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 10, 619-632.

[11] Rubinstein, A. (1989): "The Electronic Mail Game: Strategic Behavior Under 'Almost Common Knowledge'," *The American Economic Review*, Vol. 79, No. 3, 385-391.

[12] Weinstein, J. and M. Yildiz (2004): "Finite-order Implications of Any Equilibrium," MIT Working Paper.

[13] Weinstein, J. and M. Yildiz (2007): "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica*, 75, 365-400.

Weinstein: Northwestern University and Institute for Advanced Study; Yildiz: MIT and Institute for Advanced Study