

Research Article

Learning in General Games with Nature's Moves

Patrick L. Leoni

Kedge Business School, Domaine de Luminy, BP 921, 13 288 Marseille Cedex 9, France

Correspondence should be addressed to Patrick L. Leoni; patrick.leoni@kedgebs.com

Received 10 October 2013; Accepted 21 December 2013; Published 19 January 2014

Academic Editor: Takashi Matsuhisa

Copyright © 2014 Patrick L. Leoni. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigates simultaneous learning about both nature and others' actions in repeated games and identifies a set of sufficient conditions for which Harsanyi's doctrine holds. Players have a utility function over infinite histories that are continuous for the sup-norm topology. Nature's drawing after any history may depend on any past actions. Provided that (1) every player maximizes her expected payoff against her own beliefs, (2) every player updates her beliefs in a Bayesian manner, (3) prior beliefs about both nature and other players' strategies have a grain of truth, and (4) beliefs about nature are independent of actions chosen during the game, we construct a Nash equilibrium, that is, realization-equivalent to the actual plays, where Harsanyi's doctrine holds. Those assumptions are shown to be tight.

1. Introduction

Consider a finite number of agents interacting simultaneously. Every agent possibly plays an infinite number of times, and her payoff depends on the joint choice of actions as well as events beyond agents' control (called *choices of nature*). To analyze such interactions, it is always assumed that the players share a common prior about the probability distribution of nature. Such an approach is known as *Harsanyi's doctrine*, introduced in Harsanyi [1].

We provide a learning foundation for this doctrine. We consider a class of games where nature's choices may (or not) depend on any past actions by players, and payoff functions are continuous for the sup-norm topology over the set of infinite histories. Provided that Bayesian players have a grain of truth, we show that resulting outcomes converge for the sup-norm topology to a Nash equilibrium that we construct, where Harsanyi's doctrine holds.

Kalai and Lehrer [2, 3] consider a similar type of learning model, without nature's choice and with a weaker topology of convergence that significantly restricts their class of games. More precisely, those last references use a structure, that is, not a topology, in sharp contrast with the general sup-norm topology. The notion in Kalai and Lehrer states that, for any two probability measures μ and $\tilde{\mu}$ and for some small real $\epsilon > 0$, the measure μ is ϵ -close to $\tilde{\mu}$ if two conditions hold. There

must exist a measurable set Q that is assigned a measure of at least $1 - \epsilon$ both by μ and $\tilde{\mu}$, and for any measurable set $A \subset Q$ it must be true that

$$(1 - \epsilon) \tilde{\mu}(A) \leq \mu(A) \leq (1 + \epsilon) \tilde{\mu}(A). \quad (1)$$

A given strategy profile f with associated probability measure μ_f is then said to play ϵ -like another strategy profile g with associated probability measure μ_g if μ_f is ϵ -close to μ_g . The main problem with this concept is the lack of symmetry; that is, if μ_f is ϵ -close to μ_g , then μ_g is not ϵ -close to μ_f in general. This issue excludes any form of topology of convergence, and the restriction of the set of infinite play paths for which this property must hold severely restricts the class of games under consideration.

To derive a convergence result for the sup-norm topology with choices of nature, a significantly different approach than that in Kalai and Lehrer [2, 3] is needed. We explicitly construct the Nash equilibrium toward which convergence occurs, where Harsanyi's doctrine holds. The (almost) Nash equilibrium is as follows:

- (1) along every equilibrium play path, every player chooses the (possibly randomized) action that maximizes her subjective expected payoff against her beliefs on others' strategies;
- (2) in case of unilateral deviation, every player plays the strategy described in the beliefs of the deviator;

- (3) in case of multilateral deviation, strategies are defined arbitrarily.

Continuity for the sup-norm topology allows for a form of control over future payoffs. We use this property to approximate resulting plays by a Nash equilibrium for a game with a finite number of histories derived from the original game (a somewhat equivalent notion of truncated games for our setting). This approximation is a direct consequence of continuity for the sup-norm topology, and global approximation by a Nash equilibrium for the whole game is also a consequence of the control over subsequent payoffs derived from continuity. The resulting equilibrium is also proven to satisfy Harsanyi's doctrine.

The proof requires the introduction of the concept of *stochastic subjective equilibrium*, which generalizes the concept of *subjective equilibrium* (see Battigalli et al. [4] for a history and a discussion of this concept), as well as the concept of *self-confirming equilibrium* (see [5]). We show that, in finite time and for almost every infinite history, players' behaviors become identical to behaviors described in a stochastic subjective equilibrium. Given the properties of (stochastic) subjective equilibria discussed in the above references, this last result provides a second decision-theoretic foundation for the concept of the Nash equilibrium. We then show that any stochastic subjective equilibrium is realization-equivalent to the strategy profile above, which is an almost-Nash equilibrium for accurate enough beliefs, where Harsanyi's doctrine holds. The proof of this last statement uses the continuity of the utility functions to ensure convergence of payoffs as beliefs become correct. Finally, eventual correctness of beliefs follows from the grain of truth assumption, together with Blackwell-Dubins' theorem (see [6]).

The literature typically deals with more restrictive settings than ours; for instance, most follow Kalai and Lehrer by assuming that the same game is repeated over time and that the payoff function is a discounted sum of one-shot payoffs. Continuity for the sup-norm topology, as considered here, goes far beyond this setting. Absolute continuity of beliefs, an issue of paramount importance in our work, is not a necessary condition for convergence in general. Sandroni [7] shows that this assumption does not rule out plays that are asymptotically consistent with the Nash equilibrium play. Convergence to the Nash equilibrium is difficult to obtain in broader settings, even without choice of nature. Noguchi [8] considers the case of bounded rationality, where agents smoothly approximate optimal behavior, and he gives a set of prior beliefs leading to convergence to an approximate Nash equilibrium. Foster and Young [9] show that rational agents may not always learn others' strategies when payoffs are random; Nachbar [10] also shows that under a weak learning condition, if players are potentially able to learn to predict opponents' strategies for rich enough strategy sets, and if their learning process is symmetric, then one player cannot learn to predict others' optimal strategies given their own beliefs.

The paper is organized as follows. In Section 2 we formally describe the class of games and the equilibrium concepts; in Section 3 we give the main result; Section 4 contains some intermediary results; Section 5 presents the

counterexamples and concluding remarks; and finally all the technical proofs are in the Appendices.

2. The Model

The model and some assumptions needed to obtain the main result of the paper are now described.

Time is discrete and continues forever. A period is denoted by the letter t ($t = 1, 2, 3, \dots$). There are n players ($n \geq 1$), who play forever. In every period t ($t \in \mathbb{N}$), every player i has a finite set of actions Σ_i . Let Σ be defined as $\prod_{i=1}^n \Sigma_i$, the set of *action combinations*. There is a nonempty and finite (or possibly countable) set \mathcal{S} of *states of nature* in every period.

For every $s_t \in (\Sigma \times \mathcal{S})^t$ ($t \in \mathbb{N}$), a cylinder with base s_t is defined to be the set $C(s_t) = \{s \in (\Sigma \times \mathcal{S})^\infty \mid s = (s_t, \dots)\}$ of all infinite histories whose t initial elements coincide with s_t . We define the set Γ_t ($t \in \mathbb{N}$) to be the σ -algebra that consists of all finite unions of cylinders with base on $(\Sigma \times \mathcal{S})^t$, and Γ_0 is defined to be the trivial σ -algebra. The sequence $(\Gamma_t)_{t \in \mathbb{N}}$ generates a filtration, and we define Γ to be the σ -algebra generated by $\cup_{t \in \mathbb{N}} \Gamma_t$.

Let H^t be the set of all histories of length t ($t \in \mathbb{N}$); that is, $H^t = (\Sigma \times \mathcal{S})^t$ (The set H^0 is defined to be the singleton consisting of the null history.) and let \bar{H} be the set of all finite histories; that is, the set $\bar{H} = \cup_{t \in \mathbb{N}} H^t$.

A (*behavioral*) *strategy* for every player assigns to every possible finite history a (possibly randomized) action. We represent a (behavioral) strategy for player i ($i = 1, \dots, n$) as a function

$$f_i : \bar{H} \longrightarrow \Delta(\Sigma_i), \quad (2)$$

where $\Delta(\Sigma_i)$ is the set of all probability distributions on Σ_i .

Nature draws a state in every period, after every possible history. We thus represent nature's choices by a behavioral strategy

$$v : \bar{H} \longrightarrow \Delta(\mathcal{S}), \quad (3)$$

where $\Delta(\mathcal{S})$ is the set of all probability distributions on \mathcal{S} . The state of nature is not revealed to the players before they have simultaneously made their choices of actions.

The game is played with *perfect monitoring*; that is, the players know all realized past action combinations actually played, as well as all the states of nature drawn prior to the current period. The state of nature drawn in the current period is revealed to the players at the end of this period (without necessarily revealing the payoff of the other players).

2.1. Realized Play Paths. The concept of *infinite play path* is now defined. (What follows is described in Kalai and Lehrer [3].) This notion of infinite (or realized) play path represents the actual actions chosen by the players over time, in the following sense.

Consider a $(n + 1)$ -vector of behavioral strategies $f = (v, (f_i)_{i=1, \dots, n})$. The null history h_0 leads to the realized action combination z^1 in the support of $f(h_0)$. Defined recursively, in period $t + 1$, the players will choose the randomizations

$f(z(t))$, which will result in the action combination z^{t+1} . The vector (z^1, z^2, z^3, \dots) is called the *realized play path*. A realized play path, finite or infinite, will be denoted by the letter z .

Denote by A_f the support of f ; that is, the set of infinite play paths was assigned strictly positive probability by f .

2.2. Beliefs. The beliefs of the players about others' strategies and the realizations of the states of nature are now formally described.

Every player is assumed to have subjective prior beliefs about both other players' strategies and nature. Those prior beliefs are formed before the first period of the game, and they will be updated in every subsequent period in a Bayesian manner, according to available information (see Kalai and Lehrer [2, 3] for a formal definition, as we use the same approach here).

Formally, the beliefs of player i ($i = 1, \dots, n$) regarding the strategies of the other players are represented by a n -vector of strategies $f^i = (f_j^i)_{j=1, \dots, n}$. The belief f_j^i represents the belief of player i about player j 's strategy. Moreover, player i knows her own strategy (i.e., $f_i^i = f_i$ for every i).

The belief of player i about nature is represented by a behavioral strategy $\nu^i : \bar{H} \rightarrow \Delta(\mathcal{S})$, for every i ($i = 1, \dots, n$).

We consider the following probabilistic representation of beliefs. We associate a n -vector of strategies f and a choice of nature ν to a unique probability measure $\mu_{f,\nu}$ on the set on infinite play paths, as follows.

First, the measure $\mu_{f,\nu}$ is defined inductively on the set of finite play paths and then uniquely extended to the set of infinite play paths.

Define $\mu_{f,\nu}$ to be 1 for the null history. Consider now a finite history $h \in \bar{H}$ whose corresponding realized play path is given by the vector z , a vector of actions $a \in \Sigma$, and a state of nature $s \in \mathcal{S}$. The value $\mu_{f,\nu}(z, a, s)$ is inductively defined to be

$$\mu_{f,\nu}(z) \cdot \nu(h)(s) \cdot \prod_i f^i(h)(a_i). \quad (4)$$

So defined, we now uniquely extend this measure to $(\Sigma \times \mathcal{S})^\infty$. Any finite history $h \in \bar{H}$ is now interpreted as the cylinder $C(h)$, which is by definition the set of infinite paths whose initial segment is h .

Define the probability measure $\tilde{\mu}_{f,\nu}$ as $\tilde{\mu}_{f,\nu}(C(h)) \equiv \mu_{f,\nu}(h)$ for every such cylinder, and consider its unique extension to $((\Sigma \times \mathcal{S})^\infty, \Gamma)$ given by Caratheodory's theorem. The extension of the probability measure $\tilde{\mu}_{f,\nu}$ to $((\Sigma \times \mathcal{S})^\infty, \Gamma)$ is the unique extension of $\mu_{f,\nu}$ on $((\Sigma \times \mathcal{S})^\infty, \Gamma)$. See for instance Kalai and Lehrer [2] for more details.

The above representation implicitly requires that the belief of every player about nature is independent (in a probabilistic sense) of the actions chosen by other players. Moreover, it is also that every player believes that other players choose their actions independently of each other.

In Section 5.1, an example is given showing that none of the current results hold without those last two assumptions. The reader is also referred to Section 5.2 for a discussion of

those assumptions on beliefs and their behavioral implications.

For sake of notational convenience, we shall denote by the same symbol (f^i, ν^i) the prior belief about the nature and others' strategies of player i and her updated beliefs obtained by iterated applications of Bayes' formula.

2.3. Payoffs. The intertemporal payoff functions of the players are now described. Every player has the utility function over the set of infinite histories

$$u^i : (\Sigma \times \mathcal{S})^\infty \rightarrow \mathbb{R}. \quad (5)$$

We assume that, for every player i ,

- (1) the function u^i is continuous with respect to the sup-norm topology on $(\Sigma \times \mathcal{S})^\infty$, and
- (2) u^i is uniformly bounded above and below.

For any (n) -vector of strategies f , player i ($i = 1, \dots, n$), whose belief about the states of nature is ν^i , receives the expected payoff

$$U^i(f, \nu^i) = E^{f,\nu^i} [u^i(z)], \quad (6)$$

where E^{f,ν^i} is the expected value with respect to the probability measure μ_{f,ν^i} induced by the strategies f and nature' choices ν^i .

Moreover, every player is assumed to maximize the above expression, namely, her (subjective) expected payoff given her subjective belief about nature' drawings and against her subjective belief about other players' strategies.

The above specification of payoffs encompasses the case treated in Kalai and Lehrer [2, 3] and Sandroni [7], where the payoff over infinite histories takes the form of expected discounted sum of one-period payoff.

2.4. Equilibrium Concepts. This section is devoted to defining the solution concepts that will be used throughout.

First, the concept of best-response against others' strategies, given a belief about nature, is defined. Pick any player i ($i = 1, \dots, n$) and consider her belief about nature ν^i and a $(n-1)$ -vector strategies f_{-i} (this last vector can represent either beliefs about others' strategies or actual strategies). For every $\alpha \geq 0$, a strategy f_i is a α -best response to (f_{-i}, ν^i) if

$$U^i(g, f_{-i}, \nu^i) - U^i(f^i, f_{-i}, \nu^i) \leq \alpha, \quad (7)$$

for every other strategy g available to player i .

Fix now $\alpha > 0$. A Nash α -equilibrium is a n -vector of strategies such that f_i is an α -best response to (f_{-i}, ν) for every i ($i = 1, \dots, n$), where ν is the *true* probability of the states of nature. In particular, in any (almost) Nash equilibrium, beliefs about nature and others' strategies are exact.

The next notion allows us to specify a concept of closeness, in a probabilistic sense, between two vectors of strategies and for two particular choices of nature. Define first, for ν

and ψ' probability measures on the same measurable space (Ω, \mathcal{F}) , the *sup-norm* to be

$$\|\psi - \psi'\|_{\infty} = \sup_{A \in \mathcal{F}} |\psi(A) - \psi'(A)|. \quad (8)$$

Definition 1. Fix $\varepsilon > 0$. The strategy profile (f, ν_1) plays ε -like the strategy profile (g, ν_2) if

$$\|\mu_{f, \nu_1} - \mu_{g, \nu_2}\|_{\infty} \leq \varepsilon. \quad (9)$$

The concept of “playing ε -like” for two given strategies measures how distant those strategies are from each other in a probabilistic sense. It is preferable to approach this issue from a probabilistic standpoint, as explained in detail in Kalai and Lehrer [2, 3], even though the authors use a different concept.

With the above definitions, it is now possible to introduce the concept of *stochastic subjective equilibrium* (up to some constants), which generalizes the concept of subjective equilibrium introduced in Kalai and Lehrer [2, 3] and the concept of self-confirming equilibrium introduced in Fudenberg and Levine [5, 11]. This notion will play an important role in the proof of the main result of this paper.

Definition 2. Fix $\varepsilon > 0$. A stochastic ε -subjective equilibrium is a matrix of beliefs $(f^j, \nu^j)_{j=1, \dots, n}$, satisfying for every i ($i = 1, \dots, n$) that

- (i) the strategy f_i^i is a best-response to (f_{-i}^i, ν^i) , and
- (ii) the strategy profile $(f_j^j, \nu)_{j=1, \dots, n}$ plays ε -like (f^i, ν^i) .

In other words, in any stochastic subjective equilibrium, the following requirements hold: (i) every player maximizes her intertemporal utility function against her beliefs about others' strategies and given her beliefs about nature, and (ii) the beliefs about others' strategies and nature are realization-equivalent (up to ε) to actual plays. No particular condition on beliefs about nature is required in such equilibrium; for instance there is no need at this point for requirement ensuring that beliefs about nature are consistent with available information and learning processes.

The above definition extends the notion of subjective equilibrium, as introduced in Kalai and Lehrer [2, 3], in that uncertainty about nature is added. The point of the paper is to show that, even when facing this type of uncertainty, Bayesian players will choose (in finite time) a behavior described by a Nash equilibrium.

3. The Main Result

In this section, the main result of the paper is stated and discussed. That is, the set of sufficient conditions leading to convergence toward the Nash equilibria in finite time is given.

We first introduce a definition, which captures the concept of *induced strategy* resulting from a given strategy after a particular finite history.

Definition 3. Consider a n -vector of strategies f , a period $t \in \mathbb{N}$, and a finite history $h \in H^t$. The induced strategy f_h is defined as

$$f_h(h') = f(hh') \quad \text{for any } h' \in H^r \ (r \in \mathbb{N}), \quad (10)$$

where hh' is the history of length $t + r$ resulting from the concatenation of the history h (first) and (followed by) the history h' .

For any p -vector of strategies $\tilde{f} = (f_1, \dots, f_p)$ (with $1 \leq p \leq n + 1$), the induced p -vector of strategies \tilde{f}_h are defined as

$$\tilde{f}_h = ((f_1)_h, \dots, (f_p)_h). \quad (11)$$

Before stating the main result of this paper, a notion in Measure Theory is first defined. Consider two measures λ and $\tilde{\lambda}$ on the same measurable space (Ω, \mathcal{F}) . The measure λ is said to be *absolutely continuous* with respect to $\tilde{\lambda}$, denoted by $\lambda \ll \tilde{\lambda}$, if for every $A \in \mathcal{F}$ such that $\tilde{\lambda}(A) > 0$ it is true that $\lambda(A) > 0$.

Finally, for any realized play path z and time $t > 0$, denote by $z(t)$ the truncation of z to its t first elements (thus $z(t) \in H^t$).

Theorem 4. Consider a n -vector of strategies f representing actual plays and for every player j the belief (f^j, ν^j) with $f_j^j = f^j$ such that, for every i ,

- (i) the strategy f^i is a best-response to (f_{-i}^i, ν^i) ,
- (ii) the beliefs are such that $\mu_{f, \nu} \ll \mu_{f^i, \nu^i}$,
- (iii) player i updates her beliefs in a Bayesian manner, and
- (iv) the belief of player i about nature is not correlated with observations of actions taken by other players.

Fix now any arbitrary $\alpha > 0$. For $\mu_{f, \nu}$ -almost every path z , there exists a time T such that, for every $t \geq T$, there exists a strategy profile \bar{f} such that

- (1) \bar{f} is a Nash α -equilibrium, and
- (2) \bar{f} plays 0-like f .

The above theorem says that, if (1) players maximize their intertemporal utility functions against their own beliefs, and if (2) beliefs are updated in a Bayesian manner, as long as the independence requirement is satisfied and the grain of truth holds, actual plays are realization-equivalent to an almost Nash equilibrium in finite time.

One of the keys to proving the above result is that, when Assumptions (i)–(iv) are satisfied, along the realized play path actions satisfy the properties of a stochastic (almost-) subjective equilibrium in finite time. Since also any (almost) Nash equilibrium is an almost stochastic subjective equilibrium, and since also any (almost) Nash equilibrium trivially satisfies Assumptions (i)–(iv) above, Theorem 4 implicitly establishes some form of equivalence between those three different concepts.

The assumptions used in Theorem 4 are discussed next.

Assumptions (ii) and (iv) above are tight; for instance the reader is referred to Kalai and Lehrer [2] for counterexamples violating any of those assumptions, in the case where beliefs about nature are correct.

When all assumptions in Theorem 4 but (iv) above are satisfied, Bayesian players may not be able to learn anything about others' strategies and/or nature. In this case, resulting plays can become chaotic and convergence is not to be expected. In Section 5.1, such an example is presented. The example is taken from an early study by Jordan [12], and it is very similar to the framework under study. The only difference is that players have now private and partial information about the realizations of the states of nature. Using this example, Kalai and Lehrer [2] show that, when $n \geq 3$, convergence toward a Nash equilibrium is not obtained even when players have exact beliefs about nature. A counterexample, matching exactly the current framework, can be easily derived from Jordan's example.

In terms of possible extensions to Theorem 4, it is conjectured that Bayesian learning is not the only (non-trivial) learning process for which the above result holds. The characterization of all learning processes for which convergence toward a Nash equilibrium obtains is an open problem.

The proof of Theorem 4 is given in the next section.

4. Proof and Intermediary Results

This section is devoted to giving the main line of the proof of Theorem 4. Since the proof is technical, only the main intermediary results are presented here.

The strategy of the proof of Theorem 4 goes as follows. First, it is shown that beliefs and strategies satisfying Assumptions (i)–(iv) become, in finite time, identical to an almost subjective stochastic equilibrium. Second, almost subjective stochastic equilibria are shown to be realization-equivalent to an almost Nash equilibrium when beliefs are accurate enough. Finally, it is shown that beliefs become accurate enough, in finite time. Overall, this leads to the approximation of initial strategies and beliefs by an almost Nash equilibrium, as in Theorem 4.

The first proposition makes the link between strategies and beliefs satisfying Assumptions (i)–(iv) in Theorem 4 and the concept of (almost-) stochastic subjective equilibrium. Its proof is given in the Appendix.

Proposition 5. Consider a n -vector of strategies f representing actual plays and for every player j the belief (f^j, ν^j) with $f_j^j = f^j$ such that, for every i ,

- (i) the strategy f_i is a best-response to (f_{-i}^i, ν^j) ,
- (ii) the beliefs are such that $\mu_{f, \nu} \ll \mu_{f^i, \nu^j}$,
- (iii) player i updates her beliefs in a Bayesian manner, and
- (iv) the belief of player i about nature is not correlated with observations of actions taken by other players.

For every $\varepsilon > 0$ and for $\mu_{f, \nu}$ -almost every play path z , there exists a time T such that, for every $t \geq T$, the strategy profile $(f_{z(t)}^i, \nu_{z(t)}^j)_{i=1, \dots, n}$ is stochastic ε -subjective equilibrium for the repeated game starting after $z(t)$.

The above result implies that, as long as Assumptions (i)–(iv) hold, actual plays and beliefs about others' strategies along almost every path will become, in finite time, an (almost-) stochastic subjective equilibrium.

The next proposition makes the link between (almost-) stochastic subjective equilibrium and (almost-) Nash equilibrium. Its proof is given in the Appendix.

Proposition 6. Fix any vector of beliefs $(f^j, \nu^j)_{j=1, \dots, n}$. For every $\alpha > 0$, there exists a constant $\bar{\varepsilon} > 0$ such that for every $\varepsilon < \bar{\varepsilon}$, if $(f^j, \nu^j)_{i=1, \dots, n}$ is a stochastic ε -subjective equilibrium, then there exists a strategy profile \bar{f} satisfying

- (1) \bar{f} is a Nash α -equilibrium, and
- (2) \bar{f} plays 0-like f .

The above result mainly states that, provided that beliefs are accurate enough, every stochastic subjective equilibrium is an (almost-) Nash equilibrium. Given the conclusion of Proposition 6, and in order to prove the main result, it is enough to ensure that beliefs become arbitrarily correct.

Arbitrary accuracy of beliefs follows from the next proposition, which is the well-known and important result proved by Blackwell and Dubins [6]. It is a convergence result for conditional probabilities, stating that as information increases conditional probabilities of two different measures will converge, as long as a requirement of absolute continuity is satisfied by those two measures.

Before stating the result, let (Ω, \mathcal{F}) be a measurable space and let $(\mathcal{P}_t)_{t \in \mathbb{N}}$ be an increasing sequence of countable partitions of Ω , also called filter. This sequence of partitions represents the information available to an agent in any given period, and the set Ω represents the set of all the choices of nature. For any $w \in \Omega$ and any period t , let $P_t(w)$ be the unique set in \mathcal{P}_t such that $w \in P_t(w)$. Its proof can be found in Blackwell and Dubins [6].

Theorem 7 (Blackwell and Dubins [6]). Consider two σ -additive measures μ and $\bar{\mu}$ on (Ω, \mathcal{F}) such that $\mu \ll \bar{\mu}$. For μ -almost every $w \in \Omega$ and for every $\varepsilon > 0$, there exists a time T such that

$$|\mu(A | P_t(w)) - \bar{\mu}(A | P_t(w))| < \varepsilon \tag{12}$$

for every $A \in \mathcal{F}$ and for every $t \geq T$.

With all the above intermediary results, we next move to the proof of Theorem 4.

Proof of Theorem 4. Fix the strategies f and beliefs $(f^i, \nu^j)_{i=1, \dots, n}$ such that $f^i = f_i^i$ for every i , satisfying Assumptions (i)–(iv) of Theorem 4. Fix any $\alpha > 0$ and consider also $\bar{\varepsilon}$ associated with α and $(f^i, \nu^j)_{i=1, \dots, n}$ by Proposition 6 and any $\varepsilon < \bar{\varepsilon}$.

By Proposition 5, for $\mu_{f,\nu}$ -every path there exists a time t_0 after which $(f^j, \nu^j)_{j=1,\dots,n}$ is a stochastic ε -subjective equilibrium. By Proposition 6, there exists also a Nash α -equilibrium for the repeated game starting after $z(t_0)$ that plays 0-like f .

Thus, we have found a period t_0 such that the original strategies play 0-like a Nash α -equilibrium in the repeated game starting after $z(t_0)$.

The proof is now complete. \square

5. Concluding Remarks

This section provides some extended discussions of the assumptions in Theorem 4.

5.1. Independence of Beliefs. In this section, an example is given showing that Assumption (iv) in Theorem 4 cannot be relaxed. The game below is taken from Jordan [12], and it is discussed in detail in Kalai and Lehrer [2].

Consider two players engaged in an infinitely repeated game. The repeated game is similar to the one studied so far, with the difference that a randomly generated (fixed-size) pair of payoff matrices $(A_i, B_j)_{(i,j) \in I \times J}$ is drawn by nature before the first period (the sets I and J are assumed to be finite). A pair $(\bar{i}, \bar{j}) \in I \times J$ is drawn according to the probability distribution Π . The probability distribution Π is common knowledge among the players; that is, their beliefs about nature are exact.

Player 1 is told privately the realized value \bar{i} , and player 2 is told privately the realized value \bar{j} . Private information is not to be revealed to the other player at any point in the game. In this game, uncertainty about nature faced by any player is on the private information owned by the other player. Thus, the belief about nature of player 1 can be represented by the conditional probability $\Pi(\cdot | \bar{i})$ and similarly for player 2. All other assumptions but Assumption (iv) in Theorem 4 also hold. Assume also that the players maximize the expected sum of discounted one-period payoffs of the same game repeated over time (see Jordan, [12], for more details).

Kalai and Lehrer [2] show that, when the number of players is two, convergence toward a Nash equilibrium obtains when players update their beliefs about nature according to realized actions. However, when the number of players is strictly greater than two, convergence fails.

The intuition of such a result is that, when $n \geq 3$, players' information sets are no longer filters as more information becomes available, due mostly to strategic attempt to hide private information and mislead others players through their own actions. Such behaviors immediately lead to an impossibility of learning, and convergence as in Theorem 4 does not usually obtain.

5.2. Equivalent Representation of Beliefs and Kuhn's Theorem. Instead of representing the belief of player i about player j by the behavioral strategy f_j^i , we could have represented instead player i belief by a probability distribution over the set of behavioral strategies available to player j . Kuhn's theorem

(see Aumann, [13], and Kuhn, [14]) applies to our setting and ensures that the two approaches above are equivalent.

Of importance is the assumption that every player believes that others' actions are uncorrelated with each other. Informally, this last assumption ensures that any player's beliefs are represented by a measure product over beliefs about others' strategies. When this assumption is not present, it is easy to find examples where convergence toward a Nash equilibrium does not obtain (see for instance Kalai and Lehrer, [2], for a discussion of this issue).

Appendices

The Appendices are devoted to proving technical results left aside earlier in the paper. In what follows, we consider nature as an additional player maximizing a constant utility function. This does not yield any loss of generality.

A. Proof of Proposition 5

Proposition 5 is proved as follows.

Fix $\varepsilon > 0$. Consider now Γ_t to be the set of all cylinders up to time t ($t \in \mathbb{N}$) and their extensions to the infinitely repeated-game; that is, each element of Γ_t is the set of infinite play paths with the same basis of actions up to time t . Clearly, each Γ_t is a partition of the set of infinite play paths, and the family $(\Gamma_t)_{t \in \mathbb{N}}$ is a filtration of this set.

Therefore, Theorem 7 applies to the probability measures $\mu_{f,\nu}$ and μ_{f^i,ν^i} for every i and to the filtration $(\Gamma_t)_{t \in \mathbb{N}}$. It follows that for $\mu_{f,\nu}$ -almost every infinite play path z , there exists a time t such that for every $s \geq t$ it is true that

$$\left| \mu_{f,\nu}(\cdot | \Gamma_s(z)) - \mu_{f^i,\nu^i}(\cdot | \Gamma_s(z)) \right| < \varepsilon \quad (\text{A.1})$$

for every i . It follows that $(f, \nu)_{z(t)}$ plays ε -like $(f^i, \nu^i)_{z(t)}$, for every i .

Consider any realized play paths z described above and the corresponding time t . By the law of iterated expectations, the actions chosen by players i after the history $z(t)$, with the belief $\mu_{f^i,\nu^i}(\cdot | \Gamma_t(z))$ and taking it as given behaviors outside of $z(t)$, are identical to the actions chosen after the history $z(t)$ in the first period with belief μ_{f^i,ν^i} . Since f_{-i}^i is a best-response to (f^i, ν^i) for every i , this implies that $(f_i)_{z(t)}$ is best-response to $(f_{-i}^i, \nu^i)_{z(t)}$ for every i .

The proof of Proposition 5 is now complete.

B. Proof of Proposition 6

We first start with a technical lemma, stating that when two measures become eventually similar for the sup-norm, the expectations of any continuous functions according to those measures also become eventually similar. Consider a complete metric space $(\Omega, d(\cdot))$, denote by Φ the σ -algebra generated by the open ball of the metric, and denote by Υ the topology generated by the same open balls.

Lemma B.1. *Consider two positive and finite measures λ and $\tilde{\lambda}$, defined on the measurable space (Ω, Φ) . Let $u : \Omega \rightarrow \mathbb{R}$ be*

a continuous function for the topology Y , uniformly bounded above and below. It is true that

$$\left| E^\lambda(u) - E^{\tilde{\lambda}}(u) \right| \leq \sup_{x \in \Omega} |u(x)| \cdot \|\lambda - \tilde{\lambda}\|_\infty. \quad (\text{B.1})$$

Proof. Consider any such function u and any such measures λ and $\tilde{\lambda}$. We have that

$$\begin{aligned} \left| E^\lambda(u) - E^{\tilde{\lambda}}(u) \right| &= \left| \int_\Omega u \cdot d\lambda - \int_\Omega u \cdot d\tilde{\lambda} \right| \\ &= \left| \int_\Omega u \cdot d(\lambda - \tilde{\lambda}) \right|. \end{aligned} \quad (\text{B.2})$$

Since u is bounded and since the difference $\lambda - \tilde{\lambda}$ is a finite measure, it follows from the previous equation that

$$\left| E^\lambda(u) - E^{\tilde{\lambda}}(u) \right| \leq \sup_{x \in \Omega} |u(x)| \cdot \|\lambda - \tilde{\lambda}\|_\infty. \quad (\text{B.3})$$

The proof is complete. \square

We next state another technical lemma, related to the notion of stochastic subjective equilibrium.

For every i ($i = 1, \dots, n$), define first $\bar{u}^i = \sup_{s \in (\Sigma \times \mathcal{D})^\infty} |u^i(s)|$ and then $\bar{u} = \max_{i=1, \dots, n} \bar{u}^i$.

Lemma B.2. Fix any vector of beliefs $(f^j, \nu^j)_{j=1, \dots, n}$. For every $\alpha > 0$, there exists $\bar{\varepsilon} > 0$ such that, if $(\nu^j, f^j)_{j=1, \dots, n}$ is a stochastic $\bar{\varepsilon}$ -subjective equilibrium, then for every i and for every behavioral strategy g such that $A_{g, f_{-i}, \nu} \subseteq A_{f, \nu}$, the following holds:

$$\left| U^i(g, f_{-i}, \nu) - U^i(g, f_{-i}^i, \nu^i) \right| \leq \alpha. \quad (\text{B.4})$$

Proof. To prove the result, we first truncate the infinite repeated game to a finitely repeated game, show that the result holds within this truncated game, and then extends the result to the original framework.

Fix $\alpha > 0$. Consider any strategy vector $(f^j, \nu^j)_{j=1, \dots, n}$, any i , and any behavioral strategy g such that $A_{g, f_{-i}, \nu} \subseteq A_{f, \nu}$.

First, we have that, for every i , the function u^i is continuous for the product topology. This implies that there exists a period t_0 such that the contribution of any strategy profile to the overall payoff of every player after period t_0 is no greater than $\alpha/4$.

We restrict our attention to the truncated game of length t_0 in the following way: we consider the original strategy profile up to period t_0 , and leave payoff constant thereafter by extending the original strategy profile to a constant arbitrary strategy profile after t_0 . By our previous remark, the difference in payoff between the original strategy profile and the newly formed one is no greater than $\alpha/4$ for every player.

Formally, for any behavioral strategy q , we denote by q^{t_0} the restriction of q to H^{t_0} and by q^{-t_0} the restriction of q to the set of infinite histories starting at $t_0 + 1$. To truncate strategies, fix also any (dummy) strategy profile d such that $d_j(h_t) = d_j(h_{t'})$ for every $h_t, h_{t'}$ and j . For any strategy profile

$p = (p_1, \dots, p_{n+1})$, define now for every j the truncated strategy $\bar{p}_j = (p_j^{t_0}, d_j^{-t_0})$ and let $\bar{p} = (\bar{p}_1, \dots, \bar{p}_{n+1})$.

Consider also the function

$$\tilde{U}^i(p) = U^i(\bar{p}). \quad (\text{B.5})$$

With this last function, only changes of individual strategy within the truncated game of length t_0 can affect the value of \tilde{U}^i .

In a first step, we show that $|\tilde{U}^i(g, f_{-i}, \nu) - \tilde{U}^i(g, f_{-i}^i, \nu^i)| \leq \alpha/2$ for any unilateral deviation g from player i in the support of his initial strategy.

By applying Lemma B.1 applied to \tilde{U}^i , we get, for every i and g such that $A_{g, f_{-i}, \nu} \subseteq A_{f, \nu}$, that

$$\left| \tilde{U}^i(g, f_{-i}, \nu) - \tilde{U}^i(g, f_{-i}^i, \nu^i) \right| \leq \bar{u} \cdot \left\| \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}} - \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}^i} \right\|_{H^{t_0}}, \quad (\text{B.6})$$

where $\|\cdot\|_{H^{t_0}}$ is the sup-norm restricted to the σ -algebra generated by H^{t_0} .

We next analyze the right-hand side of (B.6). To simplify notations, we define for every $s \in H^{t_0}$ the function

$$\begin{aligned} \Phi^i(s) &= \left| \prod_{0 \leq t \leq t_0-1} \prod_{j \neq i} f_j(s^1, \dots, s^t) (s_j^{t+1}) \right. \\ &\quad \left. - \prod_{0 \leq t \leq t_0-1} \prod_{j \neq i} f_j^i(s^1, \dots, s^t) (s_j^{t+1}) \right|. \end{aligned} \quad (\text{B.7})$$

For every history $s \in H^{t_0}$, we have by construction of the beliefs that

$$\begin{aligned} &\left| \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}}(s) - \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}^i}(s) \right| \\ &= \left| \Phi^i(s) \cdot \prod_{0 \leq t \leq t_0-1} g(s^1, \dots, s^t) (s_i^{t+1}) \right|. \end{aligned} \quad (\text{B.8})$$

Further, for any given $\varepsilon > 0$, if $(\nu^j, f^j)_{j=1, \dots, n}$ is a stochastic ε -subjective equilibrium, we have for every history $s \in H^{t_0}$ that

$$\left| \Phi^i(s) \cdot \prod_{0 \leq t \leq t_0-1} f_i(s^1, \dots, s^t) (s_i^{t+1}) \right| \leq \varepsilon. \quad (\text{B.9})$$

Consider now the set of such histories assigned strictly positive probability by f_i and denote it by F . The last inequality implies for every $s \in F$ that

$$\Phi^i(s) \leq \frac{\varepsilon}{\prod_{0 \leq t \leq t_0-1} f_i(s^1, \dots, s^t) (s_i^{t+1})}. \quad (\text{B.10})$$

Moreover, since $A_{g, f_{-i}, \nu} \subseteq A_{f, \nu}$, for every history $s \in F$, it must be true that $g(s)$ is assigned strictly positive probability.

We next use the above remark to find a uniform upper-bound to the right-hand side of (B.6). Define

$$\rho = \min_{s \in F} \left| \prod_{0 \leq t \leq t_0-1} f_i(s^1, \dots, s^t) (s_i^{t+1}) \right|, \quad (\text{B.11})$$

which is strictly positive, and let c denote the (finite) cardinal of H^{t_0} .

For every set B of finite histories of length t_0 , from the above, we have that

$$\begin{aligned} & \left| \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}}(B) - \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}^i}(B) \right| \\ &= \sum_{s \in B \cap F} \left| \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}}(s) - \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}^i}(s) \right| \\ &\leq \sum_{s \in B \cap F} \left(\Phi^i(s) \cdot \prod_{0 \leq t \leq t_0 - 1} g(s^1, \dots, s^t)(s_i^{t+1}) \right) \quad (\text{B.12}) \\ &\leq \sum_{s \in B \cap F} \Phi^i(s) \\ &\leq c \frac{\varepsilon}{\rho}. \end{aligned}$$

Taking the maximum over such sets, we have that

$$\left\| \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}} - \mu_{\bar{g}, \bar{f}_{-i}, \bar{\nu}^i} \right\|_{H^{t_0}} \leq c \frac{\varepsilon}{\rho}. \quad (\text{B.13})$$

Setting $\bar{\varepsilon} = \alpha\rho/(2c\bar{u})$, and together with (B.6), the previous analysis implies that if $(\nu^j, f^j)_{j=1, \dots, n}$ is a stochastic $\bar{\varepsilon}$ -subjective equilibrium then

$$\left| \bar{U}^i(g, f_{-i}, \nu) - \bar{U}^i(g, f_{-i}, \nu^i) \right| \leq \frac{\alpha}{2}. \quad (\text{B.14})$$

Moreover, since the contribution of any strategy profile after period t_0 is no greater than $\alpha/4$, if $(f^j, \nu^j)_{j=1, \dots, n}$ is a stochastic $\bar{\varepsilon}$ -subjective equilibrium then

$$\begin{aligned} & \left| U^i(g, f_{-i}, \nu) - U^i(g, f_{-i}, \nu^i) \right| \\ &\leq \left| \bar{U}^i(g, f_{-i}, \nu) - \bar{U}^i(g, f_{-i}, \nu^i) \right| + \frac{\alpha}{2} \quad (\text{B.15}) \\ &\leq \alpha. \end{aligned}$$

We have thus derived the desired inequality, and the proof is now complete. \square

With the two previous lemmas, we can now prove Proposition 6.

The proof goes as follows. Fix $\alpha > 0$ and consider any vector of beliefs $(f^j, \nu^j)_{j=1, \dots, n}$.

We associate to $(f^j, \nu^j)_{j=1, \dots, n}$ the following strategy profile \bar{f} :

- (1) for every $h \in A_f$, define $\bar{f}_i(h) = f_i(h)$ for every i ,
- (2) for every $h \notin A_f$, consider the shortest prefix of h , say \bar{h} , such that $\bar{h} \notin A_f$ and consider two cases:

- (i) if \bar{h} corresponds to an unilateral deviation by player j from the support of her strategy, define $\bar{f}_i(h) = f_i^j(h)$ for every $j \neq i$,

- (ii) if \bar{h} does not correspond to an unilateral deviation, define $\bar{f}_i(h)$ arbitrarily.

To prove Proposition 6, it is enough to show that there exists $\bar{\varepsilon} > 0$ such that, if $(\nu^j, f^j)_{j=1, \dots, n}$ is a stochastic $\bar{\varepsilon}$ -subjective equilibrium, then its associated strategy profile \bar{f} is a Nash α -equilibrium.

We first claim that there exists $\varepsilon^1 > 0$ such that, if $(\nu^j, f^j)_{j=1, \dots, n}$ is a stochastic ε^1 -subjective equilibrium, then for every i ,

$$\left| U^i(f^i, \nu^i) - U^i(f, \nu) \right| \leq \frac{\alpha}{2}. \quad (\text{B.16})$$

Indeed, by Lemma B.1, we have for every i that

$$\left| U^i(f^i, \nu^i) - U^i(f, \nu) \right| \leq \bar{u} \cdot \left\| \mu_{f^i, \nu^i} - \mu_{f, \nu} \right\|_{\infty}. \quad (\text{B.17})$$

Define $\varepsilon^1 = (1/\bar{u})(\alpha/2)$. Then for every stochastic ε^1 -subjective equilibrium $(\nu^j, f^j)_{j=1, \dots, n}$ and for every i , the inequality (B.16) holds.

We next use the previous claim to get our result. In a first step we first prove the property for every individual deviation in the support of (f, ν) , and then we extend this result to any arbitrary individual deviation.

By Lemma B.2, there exists $\varepsilon^2 > 0$ such that, if $(\nu^j, f^j)_{j=1, \dots, n}$ is a stochastic ε^2 -subjective equilibrium, then for every i and for every behavioral strategy g such that $A_{g, f_{-i}, \nu} \subseteq A_{f, \nu}$, the following holds:

$$\left| U^i(g, f_{-i}, \nu) - U^i(g, f_{-i}, \nu^i) \right| \leq \frac{\alpha}{2}. \quad (\text{B.18})$$

Define $\bar{\varepsilon} = \min(\varepsilon_1, \varepsilon_2)$. Clearly, for every stochastic $\bar{\varepsilon}$ -subjective equilibrium $(\nu^j, f^j)_{j=1, \dots, n}$, for every i , and every g such that $A_{g, f_{-i}, \nu} \subseteq A_{f, \nu}$, the following holds:

$$\begin{aligned} U^i(g, \bar{f}_i, \nu) - U^i(\bar{f}, \nu) &= U^i(g, f_i, \nu) - U^i(g, f_{-i}^i, \nu^i) \\ &\quad + U^i(f^i, \nu^i) - U^i(f, \nu) \\ &\quad + U^i(g, f_{-i}^i, \nu^i) - U^i(f^i, \nu^i). \end{aligned} \quad (\text{B.19})$$

Combining (B.16) and (B.18) into this last relation gives

$$U^i(g, f_{-i}, \nu) - U^i(f, \nu) \leq \alpha + U^i(g, f^i, \nu^i) - U^i(f^i, \nu^i). \quad (\text{B.20})$$

Moreover, since f_i is best-response to (f_{-i}^i, ν^i) for player i , the above implies that

$$U^i(g, f_{-i}, \nu) - U^i(f, \nu) \leq \alpha. \quad (\text{B.21})$$

Equivalently, in terms of strategy profile \bar{f} , for any i and g such that $A_{g, f_{-i}, \nu} \subseteq A_{f, \nu}$ we have just shown that

$$U^i(g, \bar{f}_{-i}, \nu) - U^i(\bar{f}, \nu) \leq \alpha. \quad (\text{B.22})$$

We now extend this result to any arbitrary behavioral strategy g . Fix any player i and any strategy g . Assume that $A_{g, f_{-i}, v}$ differs from $A_{f, v}$. This implies that there exists a history h such that $g(h)$ is not in the support of $f_i(h)$. By construction of \bar{f} , in the subgames starting at any such corresponding unilateral deviation by player i , all the other players play according to f_{-i}^i . Since f_i is best-response to f_{-i}^i in those subgames, player i can improve upon g by playing in any such subgame according to f_i , and leave behaviors on $A_{f, v}$ unchanged. We are therefore in the previous case, and the result follows.

All together, we have shown that \bar{f} is a Nash α -equilibrium. Moreover, since as shown above there is no incentive to deviate from the original play paths, the strategy profile \bar{f} plays 0-like f . The proof is now complete.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

- [1] J. C. Harsanyi, "Games with incomplete information played by "Bayesian" players. I. The basic model," *Management Science*, vol. 14, pp. 159–182, 1967.
- [2] E. Kalai and E. Lehrer, "Rational learning leads to Nash equilibrium," *Econometrica*, vol. 61, no. 5, pp. 1019–1045, 1993.
- [3] E. Kalai and E. Lehrer, "Subjective equilibrium in repeated games," *Econometrica*, vol. 61, no. 5, pp. 1231–1240, 1993.
- [4] P. Battigalli, M. Gilli, and M. C. Molinari, "Learning and convergence to equilibrium in repeated strategic interactions: an introductory survey," *Ricerche Economiche*, vol. 46, pp. 335–377, 1992.
- [5] D. Fudenberg and D. K. Levine, "Self-confirming equilibrium," *Econometrica*, vol. 61, no. 3, pp. 523–545, 1993.
- [6] D. Blackwell and L. Dubins, "Merging of opinions with increasing information," *Annals of Mathematical Statistics*, vol. 33, pp. 882–886, 1962.
- [7] A. Sandroni, "Does rational learning lead to Nash equilibrium in finitely repeated games?" *Journal of Economic Theory*, vol. 78, no. 1, pp. 195–218, 1998.
- [8] Y. Noguchi, *Bayesian Learning with Bounded Rationality: Convergence to ϵ -Nash Equilibrium*, Kanto Gakuin University, 2007.
- [9] D. P. Foster and H. P. Young, "On the impossibility of predicting the behavior of rational agents," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 22, pp. 12848–12853, 2001.
- [10] J. H. Nachbar, "Beliefs in repeated games," *Econometrica*, vol. 73, no. 2, pp. 459–480, 2005.
- [11] D. Fudenberg and D. K. Levine, "Steady-state learning and self-confirming equilibrium," *Econometrica*, vol. 61, pp. 547–573, 1993.
- [12] J. S. Jordan, "Bayesian learning in normal form games," *Games and Economic Behavior*, vol. 3, no. 1, pp. 60–81, 1991.
- [13] R. J. Aumann, "Mixed and behavior strategies in infinite extensive games," in *Advances in Game Theory*, M. Dresher, L. Shapley, and A. Tucker, Eds., vol. 52 of *Annals of Mathematics Studies*, pp. 627–650, Princeton University Press, Princeton, NJ, USA, 1964.
- [14] H. W. Kuhn, "Extensive games and the problem of information," in *Contributions to the Theory of Games*, H. Kuhn and A. Tucker, Eds., vol. 28 of *Annals of Mathematics Studies*, pp. 193–216, Princeton University Press, Princeton, NJ, USA, 1953.