An Epistemic Approach to Stochastic Games*

Andrés Perea[†] and Arkadi Predtetchinski[‡]
Maastricht University

This version: October 2016

Abstract

In this paper we focus on stochastic games with finitely many states and actions. For this setting we study the epistemic concept of *common belief in future rationality,* which is based on the condition that players always believe that their opponents will choose rationally *in the future.* We distinguish two different versions of the concept – one for the *discounted case* with a fixed discount factor δ, and one for the case of *uniform optimality,* where optimality is required for "all discount factors close enough to 1".

We show that both versions of common belief in future rationality are always possible in every stochastic game. That is, for both versions we can always find belief hierarchies that express common belief in future rationality. We also provide an epistemic characterization of subgame perfect equilibrium for 2-player stochastic games, showing that it is equivalent to common belief in future rationality together with mutual belief in Bayesian updating and some "correct beliefs assumption".

JEL Classification: C72

Key words: Epistemic game theory, stochastic games, common belief in future rationality.

*We would like to thank János Flesch, some anonymous referees, and the audience at the Workshop on Correlated Information Change in Amsterdam, for very valuable feedback on this paper.

[†] *Address: EpiCenter* & Dept. of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. *E-mail:* a.perea@maastrichtuniversity.nl *Web:* http://www.epicenter.name/Perea/

[‡] *Address:* Dept. of Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. *E-mail:* a.predtetchinski@maastrichtuniversity.nl *Web:* http://researchers-sbe.unimaas.nl/arkadipredtetchinski/

1 Introduction

The literature on stochastic games is massive, and has concentrated mostly on the question whether Nash equilibria, subgame perfect equilibria, or other types of equilibria exist in such games. To the best of our knowledge, this paper is the first to analyze stochastic games from an epistemic point of view.

A distinctive feature of an equilibrium approach to games is the assumption that every player believes that the opponents are correct about his beliefs (see Brandenburger and Dekel (1987, 1989), Tan and Werlang (1988), Aumann and Brandenburger (1995), Asheim (2006) and Perea (2007)). The main idea of this paper is to analyze stochastic games without imposing the correct beliefs assumption, while at the same time preserving the spirit of subgame perfection. This leads to a concept called *common belief in future rationality* – an extension of the corresponding concept by Perea (2014) which has been defined for dynamic games of *finite* duration. Very similar concepts have been introduced in Baltag, Smets and Zvesper (2009) and Penta (2015).

Common belief in future rationality states that, after every history, the players continue to believe that their opponents will choose rationally in the future, that they believe that their opponents believe that their opponents will choose rationally in the future, and so on, *ad infinitum.* The crucial feature that common belief in future rationality has in common with subgame perfect equilibria is that the players uphold the belief that the opponents will be rational in the future, even if this belief has been violated in the past. What distinguishes common belief in future rationality from subgame perfect equilibrium is that the former allows the players to have erroneous beliefs about their opponents, while the latter incorporates the condition of correct beliefs in the sense that we make precise.

We introduce our solution concept using the language of epistemic models with types, following Harsanyi (1967–1968). An epistemic model specifies, for each player, the set of possible types, and for each type and each history of the game, a probability distribution over the opponents' strategy–type combinations. An epistemic model succinctly describes the entire belief hierarchy after each history of the game. This model is essentially the same as the epistemic models used by Ben-Porath (1997), Battigalli and Siniscalchi (1999, 2002) and Perea (2012, 2014) to encode conditional belief hierarchies in finite dynamic games.

For a given discount factor δ, we say that a type in the epistemic model believes in the opponents' future δ–rationality if, at every history, it assigns probability 1 to the set of opponents' strategy–type combinations where the strategy is optimal for the type's beliefs, given the discount factor δ, at every *future* history. We say that the type believes in the opponents' future *uniform* rationality if it assigns probability 1 to the set of opponents' strategy–type combinations where the strategy is uniformly optimal – that is, optimal for all δ close to 1 – for the type's beliefs at every future history. *Common* belief in future δ-rationality requires that the type not only believes in the opponents' future δ-rationality, but also believes, throughout the game, that his opponents always believe in *their* opponents' future δ-rationality, and so on, *ad infinitum.* Similarly, we can define common belief in future *uniform* rationality.

In this paper we show that common belief in future rationality is always possible in a stochastic game with finitely many states. More precisely, we prove in Theorem 5.1 that for every discount factor $\delta < 1$, we can always construct an epistemic model in which all types express common belief in future δ–rationality. A similar result holds for the uniform optimality case – see Theorem 5.2.

A second objective of this paper is to relate common belief in future rationality in stochastic games to the well-known concept of *subgame perfect equilibrium* (Selten (1965)). In Theorems 6.1 and 6.2 we provide an epistemic characterization of subgame perfect equilibrium for 2–player stochastic games. We show that a behavioral strategy profile (σ_1, σ_2) is a subgame perfect equilibrium, if and only if, there is an epistemic model and a type t_1 for player 1 such that (a) t_1 expresses common belief in future rationality, (b) t_1 believes at every history that player 2 will play σ_2 in the future, and believes that player 2 believes, after every history, that player 1 will play σ_1 in the future, (c) t_1 satisfies Bayesian updating, and believes that player 2 satisfies Bayesian updating, and (d) t_1 believes that player 2 is correct about 1's beliefs, and believes that 2 believes that 1 is correct about 2's beliefs.

Item (d) expresses the correct beliefs assumption mentioned at the beginning of this introduction, stating that player 1 believes that player 2 is correct about 1's beliefs, and that player 1 believes that player 2 believes that player 1 is correct about 2's beliefs. This is the main condition that separates subgame perfect equilibrium from common belief in future rationality, at least for the case of two players. Our characterization result is analogous to the epistemic characterizations of Nash equilibrium as presented in Brandenburger and Dekel (1987, 1989), Tan and Werlang (1988), Aumann and Brandenburger (1995), Asheim (2006) and Perea (2007).

The equilibrium counterpart of common belief in future uniform rationality is the concept we term uniform subgame perfect equilibrium. A uniform subgame perfect equilibrium is a strategy profile that is a subgame perfect perfect equilibrium under a discounted evaluation for all sufficiently high values of the discount factor. Uniform subgame perfect equilibria may fail to exist in some of the stochastic games considered in this paper. Indeed, every uniform subgame perfect equilibrium is also a subgame perfect equilibrium under the limiting average reward. It is well–known that subgame perfect equilibria, and in fact even Nash equilibria, may fail to exist in stochastic games under the limiting average reward criterion. We examine in some detail two examples that admit no uniform subgame perfect equilibria: the Big Match (Gillette, 1957) and the quitting game in Solan and Vieille (2003). Our existence result in Theorems 5.1 and 5.2, which guarantee that common belief in future rationality is always possible in a stochastic game – even for the uniform optimality case – do not rely on any form of equilibrium existence. Instead, we explicitly construct an epistemic model where each type exhibits common belief in future (δ- or uniform) rationality.

Epistemic game theory has been developed largely within the realm of finite games, i.e. games with finitely many stages. One notable exception is Battigalli (2003), who considers games with infinite duration and focuses on the concepts of weak and strong Δ–rationalizability. Some important differences between Battigalli's approach and ours are that (a) Battigalli considers games with *incomplete* information, whereas we stick to the case of complete information, (b) Battigalli considers exogenous restrictions on the players' first-order beliefs, whereas we do not, and (c) Battigalli's concepts of weak and strong Δ-rationalizability are both different from common belief in future rationality.

More precisely, weak Δ-rationalizability states that players choose rationally after every history, given their conditional beliefs, and that this event is commonly believed at the *beginning* of the game (but not necessarily when the game is under way). It may be viewed as an extension of Ben-Porath's (1997) concept of *common certainty of rationality at the beginning of the game* – which has been defined for finite dynamic games with perfect and complete information – to Battigalli's framework of infinite dynamic games with incomplete information and exogenous restrictions on first-order beliefs. Strong Δ-rationalizability is a *forward induction* concept which requires a player to believe, whenever possible, that all opponents are choosing optimal strategies. It is a generalization of Battigalli and Siniscalchi's (2002) notion of *common strong belief in rationality* – which has been defined for finite dynamic games with complete information – to Battigalli's (2003) setting. In contrast, the notion of common belief in future rationality we use is a *backward induction* concept, as it requires players to only reason about the opponents' *future* moves, not about their past moves as in strong Δ-rationalizability. If we apply weak and strong Δ-rationalizability to our setting of stochastic games with complete information and no exogenous restrictions on the first-order beliefs, then both strong Δ-rationalizability and common belief in future rationality are refinements of weak Δ-rationalizability, whereas there is no logical relationship – in terms of induced strategy choices – between the concepts of strong Δ-rationalizability and common belief in future rationality. Indeed, even in *finite* dynamic games the concepts of common strong belief in rationality (which in such games is equivalent to strong Δ-rationalizability) and common belief in future rationality may induce different sets of strategy choices for a player (see, for instance, Perea (2010, 2014)).

The paper is structured as follows. In Section 2 we provide a preliminary discussion of the concept of common belief in future rationality, and its relation to subgame perfect equilibrium, by means of two examples: the Big Match (Gillette, 1957) and a quitting game (Solan and Vieille, 2003). In Section 3 we introduce Markov decision problems and stochastic games. In Section 4 we introduce epistemic models and define the concept of common belief in future rationality. In Section 5 we prove that common belief in future δ- (and uniform) rationality is always possible in a stochastic game, whereas in Section 6 we present our epistemic characterization of subgame perfect equilibrium. All proofs are collected in Section 7.

2 Two Examples

Before presenting our formal model and definitions, we will illustrate the concept of *common belief in future rationality,* and its relation to subgame perfect equilibrium, by means of two well-known examples in the literature on stochastic games: the Big Match by Gillette (1957)

	L	R
C	$(0,0)$	$(1,-1)$
S	$(1,-1)^*$	$(0,0)^*$

Figure 1: The Big Match

and a quitting game by Solan and Vieille (2003). Both games have been originally considered under the limiting average reward criterion. The Big Match has no Nash equilibrium, and hence no subgame perfect equilibrium, under this criterion. The quitting game of Solan and Vieille (2003), on the other hand, does have a Nash equilibrium but not a subgame perfect equilibrium under this criterion.

In dynamic games of finite duration, subgame perfect equilibrium can be viewed as the equilibrium analogue to common belief in future rationality. Similarly, within stocahstic games, uniform subgame perfect equilibrium is the equilibrium counterpart to common belief in future uniform rationality. Uniform subgame perfect equilibrium is defined as a strategy profile that is a subgame perfect equilibrium for all sufficiently high values of the discount factor. As uniform optimality implies optimality under the limiting average reward criterion, each uniform subgame perfect equilibrium is also a subgame perfect equilibrium under the limiting average reward criterion. Thus neither the Big Match by Gillette (1957) nor the quitting game by Solan and Vieille (2003) admit a uniform subgame perfect equilibrium. Nevertheless, we will show that for both games we can construct belief hierarchies that express common belief in future rationality with respect to the uniform optimality criterion.

Example 1. The Big Match.

The Big Match, introduced by Gillette (1957), has become a real classic in the literature on stochastic games. It is a two-player zero-sum game with three states, two of which are absorbing. Here, by "absorbing" we mean that if the game reaches this state, it will never leave this state thereafter. In state 1 each player has only one action, and the instantaneous utilities are $(1,-1)$. From state 1 the transition to state 1 occurs with probability 1, so state 1 is absorbing. In state 2 each player has only one action, and the instantaneous utilities are $(0,0)$. From state 2 the transition to state 2 occurs with probability 1, so also state 2 is absorbing. In state 0 player 1 can play C (continue) or S (stop), while player 2 can play L (left) or R (right), the instantaneous utilities being given by the table in Figure 1. After actions (C,L) or (C,R), the transition to state 0 occurs, after (S,L) transition to state 1 occurs, while after (S,R) transition to state 2 occurs. So, the $*$ in the table above represents a situation where the game enters an absorbing state.

It is well-known that for the limiting average reward case – and hence also for the uniform optimality case – there is no subgame perfect equilibrium, nor a Nash equilibrium, in this game.

Blackwell and Ferguson (1968) have shown, however, how to construct an ε-(subgame perfect) equilibrium for the limiting average reward case for every $\varepsilon > 0$.

Consider now the belief hierarchy for player 1 in which

(a) player 1 always believes that player 2 will always choose L at state 0 in the future,

(b) player 1 always believes that player 2 always believes that player 1 will always choose C at state 0 in the future,

(c) player 1 always believes that player 2 always believes that player 1 always believes that player 2 will always choose R at state 0 in the future,

(d) player 1 always believes that player 2 always believes that player 1 always believes that player 2 always believes that player 1 will choose S at state 0 in the future,

(e) player 1 always believes that player 2 always believes that player 1 always believes that player 2 always believes that player 1 always believes that player 2 will always choose L at state 0 in the future,

and so on.

Then, it can be verified that player 1 always believes that player 2 will choose rationally in the future, that player 1 always believes that player 2 always believes that player 1 will always choose rationally in the future, and so on. Here, rationality is taken with respect to the uniform optimality criterion. That is, the belief hierarchy above expresses *common belief in future rationality* with respect to the uniform optimality criterion. In a similar way, we can construct a belief hierarchy for player 2 that expresses common belief in future rationality with respect to the uniform optimality criterion.

Note, however, that in player 1's belief hierarchy above, player 1 believes that player 2 is *wrong* about his actual beliefs: on the one hand, player 1 believes that player 2 will always choose L in the future, but at the same time player 1 believes that player 2 believes that player 1 believes that player 2 will always choose R in the future. This is something that can never happen in a subgame perfect equilibrium: there, players are always assumed to believe that the opponent is *correct* about the actual beliefs they hold. We will see in Section 6 of this paper that this correct beliefs assumption is essentially what separates the concept of common belief in future rationality from subgame perfect equilibrium.

Example 2. A quitting game.

We next consider a quitting game that has been introduced by Solan and Vieille (2003). It is a two-player stochastic game with four states, denoted 1, 2, 1^*, and 2^*. The states 1^* and 2^* are absorbing, and the players have only one action in both of these states. In state $x \in \{1, 2\}$ player x can choose actions S or C (stop or continue), and the other player has only one action. If player x plays S, the game moves to state x^*, while if he plays C, the game moves to the state $3 - x$. The instantaneous utilities in state 1^* are $(-1, 2)$, in state 2^* they are $(-2, 1)$, and in states 1 and 2 they are $(0, 0)$. See Figure 2. Also for this game, subgame perfect equilibria fail to exist if we use the uniform optimality criterion.

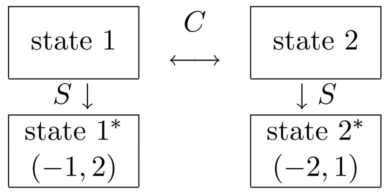

Figure 2: Quitting game

However, we can construct belief hierarchies for both players that express common belief in future rationality with respect to the uniform optimality criterion. Consider, for instance, the belief hierarchy for player 1 where

(a) player 1 always believes that player 2 will always choose C in the future,

(b) player 1 always believes that player 2 always believes that player 1 will choose S in the future,

(c) player 1 always believes that player 2 always believes that player 1 always believes that player 2 will choose S in the future,

(d) player 1 always believes that player 2 always believes that player 1 always believes that player 2 always believes that player 1 will always choose C in the future,

(e) player 1 always believes that player 2 always believes that player 1 always believes that player 2 always believes that player 1 always believes that player 2 will always choose C in the future,

and so on.

Then, similarly as in the previous example, it may be verified that this belief hierarchy expresses common belief in future rationality with respect to the uniform optimality criterion. In the same fashion, we can also construct a belief hierarchy for player 2 that expresses common belief in future rationality.

Notice that, similarly to the previous example, the belief hierarchy for player 1 is such that player 1 believes that player 2 is *wrong* about player 1's actual beliefs, and hence this belief hierarchy violates the correct beliefs assumption that underlies subgame perfect equilibrium.

3 Model

In this section we first introduce (one person) Markov decision problems, and subsequently show how stochastic games can be defined as a multi-person generalization of it.

3.1 Markov Decision Problems

A *finite Markov decision problem* consists of (1) a finite, non-empty set of states X, (2) a finite, non-empty set of actions $A(x)$ for every state $x \in X$, (3) an instantaneous utility $u(x, a)$ for every state $x \in X$ and action $a \in A(x)$, and (4) a transition probability $p(y|x,a) \in [0,1]$ for every two states $x, y \in X$ and every action $a \in A(x)$. Here, the transition probabilities should be such that

$$\sum_{y \in X} p(y|x,a) = 1$$

for every $x \in X$ and every $a \in A(x)$.

Suppose we start at some fixed state $x^1 \in X$. Then, the decision maker chooses at period 1 some action $a^1 \in A(x^1)$, which moves the system to some new state $x^2 \in X$ at period 2, according to the transition probabilities $p(y|x^1,a^1)$. If the system is at state x^2 in period 2, then the decision maker chooses some action $a^2 \in A(x^2)$ at period 2, which moves the system to some new state $x^3 \in X$ at period 3, according to the transition probabilities $p(y|x^2,a^2)$, and so on.

A *history* of length k is a sequence $h = ((x^1,a^1),...,(x^{k-1},a^{k-1}),x^k)$, where (1) $x^m \in X$ for all $m \in \{1,...,k\}$, (2) $a^m \in A(x^m)$ for all $m \in \{1,...,k-1\}$, and where (3) for every period $m \in \{2,...,k\}$ the state x^m can be reached with positive probability given that at period $m-1$ state x^{m-1} and action $a^{m-1} \in A(x^{m-1})$ have been realized. By $x(h) := x^k$ we denote the last state that occurs in history h. Let H^k denote the set of all possible histories of length k. Let $H := \cup_{k \in \mathbb{N}} H^k$ be the set of all (finite) histories.

A *strategy* s is a function that assigns to every history $h \in H$ some action $s(h) \in A(x(h))$. The strategy s is called *stationary* if $s(h) = s(h')$ for every two histories $h, h' \in H$ with $x(h) = x(h')$. Hence, the prescribed action only depends on the state reached, not on the specific period or history. In that case, we may write $s = (s(x))_{x \in X}$.

Consider a strategy s, a history $h \in H^k$ and a history $h' \in H^m$ with $m \geq k$. Then we denote by $p(h'|h,s)$ the probability that history h' will be realized, conditional on the event that h has been realized and that the decision maker chooses according to s. By

$$U^m(h,s) := \sum_{h' \in H^m} p(h'|h,s)\ u(x(h'),s(h'))$$

we denote the expected utility achieved at period m by the decision maker, conditional on the event that history h has been realized and that the decision maker uses strategy s.

For a given discount factor $\delta \in (0,1)$, we denote by

$$U^\delta(h,s) := \sum_{m \geq k} \delta^m U^m(h,s)$$

the *discounted expected utility* for the decision maker. We say that a strategy s is δ-*optimal* if

$$U^\delta(h,s) \geq U^\delta(h,s')$$

for all histories $h \in H$ and all strategies s'.

The strategy s is said to be *uniformly optimal* if there is some $\bar{\delta} \in (0,1)$ such that s is δ-optimal for all $\delta \in [\bar{\delta},1)$. Every strategy which is uniformly optimal is also optimal under the *limiting average reward* criterion, which is also often used in Markov decision problems. This result can be found, for instance, in Filar and Vrieze (1997), Theorem 2.8.3.

The following classical results state that for every finite Markov decision problem, we can always find a *stationary* strategy that is optimal – both for the δ-discounted and the uniform optimality case.

Theorem 3.1 (Optimal strategies in Markov decision problems) *Consider a finite Markov decision problem.*

(a) For every $\delta \in (0,1)$, there is a δ-optimal strategy which is stationary.

(b) There is a uniformly optimal strategy which is stationary.

Part (a) follows from Shapley (1953) and has later been shown in Howard (1960), but Blackwell (1962) provides a simpler proof. The proof for part (b) can be found in Blackwell (1962).

3.2 Stochastic Games

A *finite stochastic game* Γ consists of the following ingredients: (1) a finite set of players I, (2) a finite, non-empty set of states X, (3) for every state x and player $i \in I$, there is a finite, non-empty set of actions $A_i(x)$, (4) for every state x and every profile of actions a in $\times_{i\in I} A_i(x)$, there is an instantaneous utility $u_i(x,a)$ for every player i, and (5) a transition probability $p(y|x,a) \in [0,1]$ for every two states $x, y \in X$ and every action profile a in $\times_{i\in I} A_i(x)$. Here, the transition probabilities should be such that

$$\sum_{y\in X} p(y|x,a) = 1$$

for every $x \in X$ and every action profile a in $\times_{i\in I} A_i(x)$.

At every state x, we write $A(x) := \times_{i\in I} A_i(x)$. A *history* of length k is a sequence $h = ((x^1, a^1), ..., (x^{k-1}, a^{k-1}), x^k)$, where (1) $x^m \in X$ for all $m \in \{1, ..., k\}$, (2) $a^m \in A(x^m)$ for all $m \in \{1, ..., k-1\}$, and where (3) for every period $m \in \{2, ..., k\}$ the state x^m can be reached with positive probability given that at period $m-1$ state x^{m-1} and action profile $a^{m-1} \in A(x^{m-1})$ have been realized. By $x(h) := x^k$ we denote the last state that occurs in history h. Let H^k denote the set of all possible histories of length k. Let $H := \cup_{k\in\mathbb{N}} H^k$ be the set of all (finite) histories.

A *strategy* for player i is a function s_i that assigns to every history $h \in H$ some action $s_i(h) \in A_i(x(h))$. By S_i we denote the set of all strategies for player i. Note that the set S_i of strategies is typically uncountably infinite. We say that the strategy s_i is *stationary* if $s_i(h) = s_i(h')$ for all $h, h' \in H$ with $x(h) = x(h')$. So, the prescribed action only depends on the state, and not on the specific history. A stationary strategy can thus be summarized as $s_i = (s_i(x))_{x\in X}$.

During the game, players always observe what their opponents have done in the past, but face uncertainty about what the opponents will do now and in the future, and also about what these opponents would have done at histories that are no longer possible. That is, after every history h all players know that their opponents have chosen a combination of strategies that could have resulted in this particular history h. To model this precisely, consider a history $h^k = ((x^1, a^1), ..., (x^{k-1}, a^{k-1}), x^k)$ of length k. For every $m \in \{1, ..., k-1\}$ let $h^m := ((x^1, a^1), ..., (x^{m-1}, a^{m-1}), x^m))$ be the induced history of length m. For every player i, we denote by $S_i(h)$ the set of strategies $s_i \in S_i$ such that $s_i(h^m) = a_i^m$ for every $m \in \{1, ..., k-1\}$. Here, a_i^m is the action of player i in the action profile $a^m \in A(x^m)$. Hence, $S_i(h)$ contains precisely those strategies for player i that are compatible with the history h.

So, after every history h, every player i knows that each of his opponents j is implementing a strategy from $S_j(h)$, without knowing precisely which one. This uncertainty can be modelled by conditional belief vectors. Formally, a *conditional belief vector* b_i for player i specifies for every history $h \in H$ some probability distribution $b_i(h) \in \Delta(S_{-i}(h))$. Here, $S_{-i}(h) := \times_{j \neq i} S_j(h)$ denotes the set of opponents' strategy combinations that are compatible with the history h, and $\Delta(S_{-i}(h))$ is the set of probability distributions on $S_{-i}(h)$.

To define the space $\Delta(S_{-i}(h))$ formally we must first specify a σ-algebra $\Sigma_{-i}(h)$ on $S_{-i}(h)$, since $S_{-i}(h)$ is typically an uncountably infinite set. Let $h \in H^k$ be a history of length k. For a given player j, strategy $s_j \in S_j(h)$, and $m \geq k$, let $[s_j]_m$ be the set of strategies that coincide with s_j at all histories of length at most m. As $m \geq k$, every strategy in $[s_j]_m$ must in particular coincide with s_j at all histories that precede h, and hence every strategy in $[s_j]_m$ will be in $S_j(h)$ as well. Let $\Sigma_j(h)$ be the σ-algebra on $S_j(h)$ generated by the sets $[s_j]_m$, with $s_j \in S_j(h)$ and $m \geq k$.[1] By $\Sigma_{-i}(h)$ we denote the product σ-algebra generated by the σ-algebras $\Sigma_j(h)$ with $j \neq i$. Hence, $\Sigma_{-i}(h)$ is a σ-algebra on $S_{-i}(h)$, and this is precisely the σ-algebra we will use. So, when we say $\Delta(S_{-i}(h))$ we mean the set of probability distributions on $S_{-i}(h)$ with respect to this specific σ-algebra $\Sigma_{-i}(h)$.

Suppose that the game has reached history $h \in H^k$. Consider for every player i some strategy $s_i \in S_i(h)$ which is compatible with the history h. Let $s = (s_i)_{i \in I}$. Then, for every $m \geq k$, and every history $h' \in H^m$, we denote by $p(h'|h, s)$ the probability that history $h' \in H^m$ will be realized, conditional on the event that the game has reached history $h \in H^k$ and the players choose according to s. The corresponding expected utility for player i at period $m \geq k$ would be given by

$$U_i^m(h, s) := \sum_{h' \in H^m} p(h'|h, s)\ u_i(x(h'), s(h')),$$

where $s(h') \in A(x(h'))$ is the combination of actions chosen by the players at state $x(h')$ after history h', if they choose according to the strategy profile s. The *expected discounted utility* for

[1] This is arguably the most natural σ–algebra on the set of strategies.

player i would be

$$U_i^\delta(h,s) := \sum_{m \geq k} \delta^m U_i^m(h,s).$$

Suppose now that player i, after history h, holds the conditional belief $b_i(h) \in \Delta(S_{-i}(h))$. Then, the *expected discounted utility* of choosing strategy $s_i \in S_i(h)$ after history h, under the belief $b_i(h)$, is given by

$$U_i^\delta(h, s_i, b_i(h)) := \int_{S_{-i}(h)} U_i^\delta(h,(s_i,s_{-i}))\ db_i(h).$$

The strategy s_i is δ-*optimal* under the conditional belief vector b_i if

$$U_i^\delta(h, s_i, b_i(h)) \geq U_i^\delta(h, s_i', b_i(h))$$

for every history $h \in H$ and every strategy $s_i' \in S_i(h)$.

The strategy s_i is said to be *uniformly optimal* under b_i if there is some $\bar{\delta} \in (0,1)$ such that s_i is δ-optimal under b_i for every $\delta \in [\bar{\delta}, 1)$. Note that every strategy s_i which is uniformly optimal under the conditional belief vector b_i, will also be optimal under b_i with respect to the *limiting average reward* criterion – an optimality criterion which is widely used in the literature on stochastic games. This result follows from Theorem 2.8.3 in Filar and Vrieze (1997).

4 Common Belief in Future Rationality

In this section we define the central notion in this paper – *common belief in future rationality*. In words, the concept states that a player always believes, after every history, that his opponents will choose rationally in the future, that his opponents always believe that their opponents will choose rationally in the future, and so on. Before we define this concept formally, we first introduce epistemic models with types *à la* Harsanyi (1967–1968) as a possible way to encode belief hierarchies.

4.1 Epistemic Model

We do not only wish to model the beliefs of players about the opponents' strategy choices, but also the beliefs about the opponents' beliefs about the other players' strategy choices, and so on. One way to do so is by means of an epistemic model with types *à la* Harsanyi (1967–1968).

Definition 4.1 (Epistemic model) *Consider a finite stochastic game Γ. A* ***finite epistemic model*** *for Γ is a tuple $M = (T_i, \beta_i)_{i \in I}$ where*

(a) T_i is a finite set of types for player i, and

(b) β_i is a mapping that assigns to every type $t_i \in T_i$, and every history $h \in H$, some conditional belief $\beta_i(t_i, h) \in \Delta(S_{-i}(h) \times T_{-i})$.

Here, the σ-algebra on $S_{-i}(h) \times T_{-i}$ that we use is the product σ-algebra generated by the σ-algebra $\Sigma_{-i}(h)$ on $S_{-i}(h)$, and the discrete σ-algebra on the finite set T_{-i}, containing all subsets. The probability distribution $\beta_i(t_i, h)$ encodes the belief that type t_i holds, after history h, about the opponents' strategies and the opponents' conditional beliefs. In particular, by taking the marginal of $\beta_i(t_i, h)$ on $S_{-i}(h)$, we obtain the *first-order* belief $b_i(t_i, h) \in \Delta(S_{-i}(h))$ of type t_i about the opponents' strategies. As $\beta_i(t_i, h)$ also specifies a belief about the opponents' types, and every opponent's type holds conditional beliefs about his opponents' strategies, we can also derive, for every type t_i and history h, the *second-order* belief that type t_i holds, after history h, about the opponents' conditional first-order beliefs.

By continuing in this fashion, we can derive for every type t_i in the epistemic model his first-order beliefs, second-order beliefs, third-order beliefs, and so on. That is, we can derive for every type t_i a complete *belief hierarchy.* The epistemic model just represents a very easy and compact way to *encode* such belief hierarchies. The epistemic model above is very similar to models used in Ben-Porath (1997), Battigalli and Siniscalchi (1999, 2002) and Perea (2012, 2014) for finite dynamic games.

The reader may wonder why we restrict to *finitely many types* in the epistemic model. The reason is purely pragmatic: it is easier to work with finitely many types, since we do not need additional topological or measure-theoretic machinery. At the same time, our analysis and results in this paper would not change if we would allow for infinitely many types. For instance, in order to prove the existence of common belief in future rationality in both the discounted and the uniform case, it is sufficient to build *one* epistemic model in which all types express common belief in future rationality, and we show that we can always build an epistemic model with *finitely many types* that has this property.

A property that we will be especially interested in throughout this paper is *Bayesian updating.* Here is a formal definition.

Definition 4.2 (Bayesian updating) *A type t_i satisfies* ***Bayesian updating*** *if for every history h, and every history h' following h with $\beta_i(t_i, h)(S_{-i}(h') \times T_{-i}) > 0$, we have that*

$$\beta_i(t_i, h')(E_{-i} \times \{t_{-i}\}) = \frac{\beta_i(t_i, h)(E_{-i} \times \{t_{-i}\})}{\beta_i(t_i, h)(S_{-i}(h') \times T_{-i})}$$

for every set $E_{-i} \in \Sigma_{-i}(h')$ and every $t_{-i} \in T_{-i}$.

Remember that $\Sigma_{-i}(h')$ is the σ-algebra on $S_{-i}(h')$ we have introduced in Section 3.

4.2 Belief in Future Rationality

Consider a type t_i, and let $b_i(t_i)$ be the induced first-order belief vector. That is, $b_i(t_i)$ specifies for every history h the first-order belief $b_i(t_i, h) \in \Delta(S_{-i}(h))$ that t_i holds about the opponents' strategies. Note that $b_i(t_i)$ is a conditional belief vector as defined in the previous section. We say that strategy s_i is δ*-optimal* for type t_i at history h if s_i is δ-optimal at h for the conditional belief $b_i(t_i, h)$. More precisely, s_i is δ-optimal for type t_i at history h if

$$U_i^\delta(h, s_i, b_i(t_i, h)) \geq U_i^\delta(h, s_i', b_i(t_i, h))$$

for every $s_i' \in S_i(h)$.

We say that type t_i believes in his opponents' future δ-rationality if at every stage of the game, type t_i assigns probability 1 to the set of those opponents' strategy-type pairs where the opponent's strategy is δ-optimal for the opponent's type at all *future stages.* To formally define this, let

$$(S_i \times T_i)^{h,\delta-opt} := \{(s_i, t_i) \in S_i \times T_i \mid s_i \text{ is } \delta\text{-optimal for } t_i \text{ at every } h' \text{ that weakly follows } h\}.$$

Here, we say that h' weakly follows h if h' follows h, or $h' = h$. Moreover, let $(S_{-i} \times T_{-i})^{h,\delta-opt} := \times_{j \neq i}(S_j \times T_j)^{h,\delta-opt}$ be the set of opponents' strategy-type combinations where the strategies are δ-optimal for the types at all stages weakly following h.

Similar definitions can be given for the case of uniform optimality. We define

$$\begin{aligned} (S_i \times T_i)^{h,u-opt} \quad : \quad & = \{(s_i, t_i) \in S_i \times T_i \mid \text{there is some } \bar{\delta} \in (0,1) \text{ such that for all } \delta \in [\bar{\delta}, 1), \\ & \quad s_i \text{ is } \delta\text{-optimal for } t_i \text{ at every } h' \text{ that weakly follows } h\}, \end{aligned}$$

and let $(S_{-i} \times T_{-i})^{h,u-opt} := \times_{j \neq i}(S_j \times T_j)^{h,u-opt}$.

Definition 4.3 (Belief in future rationality) *Consider a finite epistemic model $M = (T_i, \beta_i)_{i \in I}$, and a type $t_i \in T_i$.*

*(a) Type t_i **believes in the opponents' future δ-rationality** if for every history h we have that $\beta_i(t_i, h)(S_{-i} \times T_{-i})^{h,\delta-opt} = 1$.*

*(b) Type t_i **believes in the opponents' future uniform rationality** if for every history h we have that $\beta_i(t_i, h)(S_{-i} \times T_{-i})^{h,u-opt} = 1$.*

With this definition at hand, we can now define "common belief in future δ-rationality", which means that players do not only believe in their opponents' future δ-rationality, but also always believe that the other players believe in their opponents' future δ-rationality, and so on. We do so by recursively defining, for every player i, smaller and smaller sets of types $T_i^1, T_i^2, T_i^3, ...$

Definition 4.4 (Common belief in future rationality) *Consider a finite epistemic model $M = (T_i, \beta_i)_{i \in I}$, and some $\delta \in (0, 1)$. Let*

$$T_i^1 := \{t_i \in T_i \mid t_i \textit{ believes in the opponents' future } \delta\textit{-rationality}\}$$

for every player i. For every $m \geq 2$, recursively define

$$T_i^m := \{t_i \in T_i^{m-1} \mid \beta_i(t_i, h)(S_{-i} \times T_{-i}^{m-1}) = 1 \text{ for all } h \in H\}.$$

A type t_i expresses ***common belief in future δ-rationality*** *if $t_i \in T_i^m$ for all m.*

That is, T_i^2 contains those types that believe in the opponents' future δ-rationality, and which only deem possible opponents' types that believe in their opponents' future δ-rationality. Similarly for T_i^3, T_i^4, and so on. This definition is based on the notion of "common belief in future rationality" as presented in Perea (2014), which has been designed for dynamic games of finite duration. Baltag, Smets and Zvesper (2009) and Penta (2015) present concepts that are very similar to "common belief in future rationality". In the same way, we can define "common belief in future *uniform* rationality" for stochastic games.

5 Existence Result

In this section we will show that "common belief in future δ-rationality" and "common belief in future uniform rationality" are possible in every finite stochastic game. The proof will be constructive, as we will explicitly construct an epistemic model in which all types express common belief in future δ- (or uniform) rationality.

5.1 Common Belief in Future Rationality is Always Possible

We first show the following important result, for which we need some new notation. For a given strategy s_i and history h, let $S_i[s_i, h]$ be the set of strategies in $S_i(h)$ that coincide with s_i on histories that weakly follow h. Similarly, for a given combination of strategies $s_{-i} \in S_{-i}$ and history h, we denote by $S_{-i}[s_{-i}, h] := \times_{j \neq i} S_j[s_j, h]$ the set of opponents' strategy combinations in $S_{-i}(h)$ that coincide with s_{-i} on histories that weakly follow h.

Lemma 5.1 (Stationary strategies are optimal under stationary beliefs) *Consider a finite stochastic game Γ. Let s_{-i} be a profile of stationary strategies for i's opponents. Let b_i be a conditional belief vector that assigns, at every history h, probability 1 to $S_{-i}[s_{-i}, h]$. Then,*

(a) for every $\delta \in (0, 1)$ there is a stationary strategy for player i that is δ-optimal under b_i, and

(b) there is a stationary strategy for player i that is uniformly optimal under b_i.

That is, if we always assign full probability to the same stationary continuation strategy for each of our opponents, then there will be a stationary strategy for us that is optimal after every history. We are now in a position to prove that common belief in future δ-rationality is always possible in every finite stochastic game.

Theorem 5.1 (Common belief in future δ-rationality is always possible) *Consider a finite stochastic game* Γ, *and some* $\delta \in (0, 1)$. *Then, there is a finite epistemic model* $M = (T_i, \beta_i)_{i \in I}$ *for* Γ *such that*

(a) every type in M *expresses common belief in future* δ*-rationality, and*

(b) every type in M *satisfies Bayesian updating.*

Smilarly, we can prove that common belief in future *uniform* rationality is always possible as well.

Theorem 5.2 (Common belief in future uniform rationality is always possible) *Consider a finite stochastic game* Γ. *Then, there is a finite epistemic model* $M = (T_i, \beta_i)_{i \in I}$ *for* Γ *such that*

(a) every type in M *expresses common belief in future uniform rationality, and*

(b) every type in M *satisfies Bayesian updating.*

The proof for this theorem is almost identical to the proof of Theorem 5.1. The only difference is that we must use part (b), instead of part (a), in Lemma 5.1. For that reason, this proof is omitted.

Suppose that, instead of restricting to finitely many types, we would start from a *terminal* epistemic model (Friedenberg (2010)) in which *all* possible belief hierarchies are present. Then, Theorems 5.1 and 5.2 would imply that within this terminal epistemic model we can always find belief-closed submodels with *finitely many types* in which every type expresses common belief in future rationality. Hence, the message of these two theorems would not change if we would consider such terminal epistemic models with infinitely many types.

5.2 Examples Revisited

We will now illustrate the existence result by means of the two examples we discussed in Section 2. For both examples, it has been shown that subgame perfect equilibria fail to exist if we use the uniform optimality criterion. Nevertheless, our Theorem 5.2 guarantees that common belief in future uniform rationality is possible for both games. In fact, for both games we will explicitly construct epistemic models where all types express common belief in future uniform rationality, and all types satisfy Bayesian updating.

Example 1 continued. The Big Match.

Recall the game from Figure 1. We will now construct an epistemic model in which all types express common belief in future uniform rationality. With a slight abuse of notation we write C to denote player 1's stationary strategy in which he always plays action C in state 0, and similarly for S, L, and R. Now consider the chain of stationary strategy pairs:

$$(S, R) \to (C, R) \to (C, L) \to (S, L) \to (S, R).$$

In this chain, each stationary strategy is δ-optimal, for every $\delta \in (0,1)$, under the belief that the opponent will play the preceding strategy in the future. For instance, "$(S,R) \to (C,R)$" indicates that for player 1 it is optimal to play C if he believes that player 2 will play R in the future, and for player 2 it is optimal to play R if he believes that player 1 will play S in the future. Similarly for the other arrows in the chain. In particular, each of these strategies is uniformly optimal as well for these beliefs. This chain leads to the following epistemic model with types

$$T_1 = \{t_1^C, t_1^S\},\ T_2 = \{t_2^L, t_2^R\}$$

and beliefs

$$\begin{array}{rcl} b_1(t_1^S, h) & = & (L, t_2^L) \\ b_1(t_1^C, h) & = & (R, t_2^R) \\ b_2(t_2^L, h) & = & (C, t_1^C) \\ b_2(t_2^R, h) & = & (S, t_1^S) \end{array}.$$

Here, $b_1(t_1^S, h) = (L, t_2^L)$ means that type t_1^S, after every possible history h, assigns probability 1 to player 2 choosing the stationary strategy L in the remainder of the game, and to player 2 having type t_2^L. Similarly for the other types. The full beliefs of these types can be constructed such that they all satisfy Bayesian updating.

Note that type t_2^R always believes that player 1 will choose S in the current stage, even though it is evident that player 1 has always chosen C in the past. This degree of stubbornness is typical for backward induction concepts such as common belief in future rationality or subgame perfect equilibrium. Think, for instance, of Rosenthal's (1981) centipede game, where in a subgame perfect equilibrium a player always believes that his opponent will opt out in the next round, whereas it is evident that the opponent has not opted out at any point in the past.

It may be verified that every type in the epistemic model above satisfies Bayesian updating, and that every type believes in the opponent's future δ- (and uniform) rationality. As a consequence, every type expresses *common* belief in future δ- (and uniform) rationality.

Note that the type t_1^S for player 1 induces exactly the belief hierarchy we have described in Section 2.

Example 2 continued. A quitting game.

Recall the game from Figure 2. We write C_i to denote player i's stationary strategy in which he always plays action C in state i, and similarly for S_i. Now consider the chain of stationary strategy pairs

$$(S_1, S_2) \to (S_1, C_2) \to (C_1, C_2) \to (C_1, S_2) \to (S_1, S_2).$$

In this chain, each stationary strategy is δ-optimal, for every $\delta \in [\frac{1}{2}, 1)$, under the belief that the opponent plays the preceding stationary strategy in the future. In particular, each of these strategies is uniformly optimal under these beliefs. This leads to the following epistemic model with types

$$T_1 = \{t_1^{C_1}, t_1^{S_1}\},\ T_2 = \{t_2^{C_2}, t_2^{S_2}\}$$

and beliefs

$$\begin{aligned} b_1(t_1^{C_1}, h) &= (C_2, t_2^{C_2}) \\ b_1(t_1^{S_1}, h) &= (S_2, t_2^{S_2}) \\ b_2(t_2^{C_2}, h) &= (S_1, t_1^{S_1}) \\ b_2(t_2^{S_2}, h) &= (C_1, t_1^{C_1}). \end{aligned}$$

Again, the full beliefs of these types can be constructed in such a way that all types satisfy Bayesian updating.

Note that types $t_1^{S_1}$ and $t_2^{C_2}$ always believe that the opponent will choose S during the next stage, although it is evident that the same opponent has always chosen C in the past. Again, as we have already explained above, this type of stubbornness is not uncommon for backward induction concepts like common belief in future rationality and subgame perfect equilibrium.

It may be verified that all types believe in the opponents' future δ-rationality for all $\delta \in [\frac{1}{2}, 1)$. Consequently, all types express *common* belief in future δ-rationality as well, for all $\delta \in [\frac{1}{2}, 1)$. Similarly, it can be shown that all types express common belief in future *uniform* rationality.

Note that the type $t_1^{C_1}$ for player 1 induces exactly the belief hierarchy we have described in Section 2.

6 Relation to Subgame Perfect Equilibrium

In the literature on stochastic games, the concepts which are most commonly used are Nash equilibrium (Nash (1950, 1951)) and subgame perfect equilibrium (Selten (1965)). In this section we will explore the precise relation between common belief in future rationality on the one hand, and subgame perfect equilibrium on the other hand. We will show that in two-person stochastic games, subgame perfect equilibrium can be characterized by the conditions in common belief in future rationality, together with Bayesian updating and some "correct beliefs conditions". In particular, it follows that subgame perfect equilibrium can be viewed as a refinement of common belief in future rationality, as it implicitly assumes each of its conditions.

In Section 5 we have seen that common belief in future rationality (in combination with Bayesian updating) is always possible in every finite stochastic game, even if we use the uniform optimality criterion. Hence, the reason that subgame perfect equilibrium fails to exist in some of these games is that the conditions in common belief in future rationality and Bayesian updating are logically inconsistent with the "correct beliefs conditions" in those games.

6.1 From Types to Behavioral Strategies

The concepts of common belief in future rationality and subgame perfect equilibrium are defined within two different languages: The first concept is defined within an epistemic model with types, whereas the latter is defined by the use of behavioral strategies. How can we then formally relate these two concepts? We will see that, under certain conditions, a type within an epistemic model will naturally *induce* a profile of behavioral strategies.

From now on, we assume that there are only two players in the game. Formally, a *behavioral strategy* for player i is a function σ_i that assigns to every history h some probability distribution $\sigma_i(h) \in \Delta(A_i(x(h)))$ on the set of actions available at state $x(h)$. Now, consider an epistemic model $M = (T_i, \beta_i)_{i \in I}$, and a type t_i within that epistemic model. For every history h and every action $a_j \in A_j(x(h))$ for opponent j at h, let $S_j(h, a_j)$ denote the set of strategies $s_j \in S_j(h)$ with $s_j(h) = a_j$. We define the behavioral strategy $\sigma_j^{t_i}$ induced by type t_i for opponent j by

$$\sigma_j^{t_i}(h)(a_j) := \beta_i(t_i, h)(S_j(h, a_j) \times T_j)$$

for every history h and every action $a_j \in A_j(x(h))$. Hence, $\sigma_j^{t_i}(h)(a_j)$ is the probability that type t_i assigns, after history h, to the event that player j will choose action a_j after h. In this way, every type t_i naturally induces a behavioral strategy $\sigma_j^{t_i}$ for his opponent j. So, $\sigma_j^{t_i}$ represents t_i's conditional beliefs about j's *future* behavior.

But what does it mean that a type t_i for player i induces a behavioral strategy σ_i for player i himself? This is more subtle, as t_i holds no belief about his own actions in the game, only about the actions of his opponent. But t_i *does* hold a belief about j's beliefs about i's actions, and this second-order belief will constitute the link to σ_i. More precisely, we will say that type t_i induces a behavioral strategy σ_i for himself if, after any history, he only assigns positive probability to opponent's types t_j where $\sigma_i^{t_j} = \sigma_i$. This naturally leads to the following definition.

Definition 6.1 (From types to behavioral strategies) *A type t_i* ***induces a behavioral strategy pair*** *(σ_i, σ_j) if*

(1) $\sigma_j^{t_i} = \sigma_j$, and

(2) after every history h, the conditional belief $\beta_i(t_i, h) \in \Delta(S_j(h) \times T_j)$ only assigns positive probability to types t_j for which $\sigma_i^{t_j} = \sigma_i$.

Condition (2) thus states that, after every history h, type t_i believes – with probability 1 – that player j believes that i's future behavior is given by σ_i.

With this definition at hand it is now clear what it means that a type induces a subgame perfect equilibrium, since a subgame perfect equilibrium is just a behavioral strategy pair satisfying some special conditions. In order to define a subgame perfect equilibrium formally, we need some additional notation first. Take some behavioral strategy pair (σ_i, σ_j), and some history h. We denote by $U_i^\delta(h, \sigma_i, \sigma_j)$ the δ-discounted expected utility for player i, if the game would start after history h, and if the players choose according to (σ_i, σ_j) in the subgame that starts after history h.

Definition 6.2 (Subgame perfect equilibrium) *(a) A behavioral strategy pair (σ_1, σ_2) is a δ-**subgame perfect equilibrium** if after every history h, and for both players i, we have that $U_i^\delta(h, \sigma_i, \sigma_j) \geq U_i^\delta(h, \sigma_i', \sigma_j)$ for every behavioral strategy σ_i'.*

*(b) A behavioral strategy pair (σ_1, σ_2) is a **uniform subgame perfect equilibrium** if there is some $\bar{\delta} \in (0,1)$ such that for every $\delta \in [\bar{\delta}, 1)$, for every history h, and for both players i, we have that $U_i^\delta(h, \sigma_i, \sigma_j) \geq U_i^\delta(h, \sigma_i', \sigma_j)$ for every behavioral strategy σ_i'.*

Hence, a δ-subgame perfect equilibrium constitutes a δ-Nash equilibrium in each of the subgames. A behavioral strategy pair is thus a uniform subgame perfect equilibrium if it is a subgame perfect equilibrium under a discounted evaluation for all sufficiently high values of the discount factor. The concept of uniform ϵ–equilibrium (e.g. Jaśkiewicz and Nowak (2016)) features prominently in the literature on stochastic games. While uniform subgame perfect equilibrium is not logically related to the uniform ϵ–equilibrium, it is somewhat similar in spirit. Both concepts entail a requirement of robustness of the solution within a small range of the parameters of the game.

6.2 Epistemic Characterization of Subgame Perfect Equilibrium

We will now characterize those types t_i within an epistemic model that induce a δ-subgame perfect equilibrium. We will see that these are precisely the types that satisfy all the conditions in common belief in future rationality, together with Bayesian updating and some "correct beliefs conditions". Before we state this result formally, we must first define what we mean by these "correct beliefs conditions".

We say that a type t_i believes that opponent j is correct about i's beliefs, if t_i only assigns positive probability to opponent's types who are correct about his full belief hierachy. Similarly, we say that t_i believes that j believes that i is correct about j's beliefs, if t_i only assigns positive probability to opponent's types t_j that believe that i is correct about j's beliefs. Since t_i's belief hierarchy is encoded by his type, this leads to the following two definitions.

Definition 6.3 (Correct beliefs assumption) *Consider a finite epistemic model $M = (T_i, \beta_i)_{i \in I}$.*

*(1) Type t_i **believes that** j **is correct about** i**'s beliefs**, if after every history h, the conditional belief $\beta_i(t_i, h) \in \Delta(S_j(h) \times T_j)$ only assigns positive probability to types t_j that, after every history h', assign probability 1 to type t_i.*

*(2) Type t_i **believes that** j **believes that** i **is correct about** j**'s beliefs**, if after every history h, type t_i only assigns positive probability to types t_j that believe that i is correct about j's beliefs.*

Remember the definition of Bayesian updating that we gave in Section 4. We say that type t_i believes that j satisfies Bayesian updating if, after every history h, the conditional belief $\beta_i(t_i, h)$ only assigns positive probability to types t_j that satisfy Bayesian updating.

We are now ready to state our epistemic characterization of δ-subgame perfect equilibrium in two-player stochastic games.

Theorem 6.1 (Characterization of δ-subgame perfect equilibrium) *Consider a finite two-player stochastic game Γ, and a behavioral strategy pair (σ_1, σ_2) in Γ. Then, (σ_1, σ_2) is a δ-subgame perfect equilibrium, if and only if, there is a finite epistemic model $M = (T_i, \beta_i)_{i \in I}$ and for both players i a type $t_i \in T_i$, such that*

(1) t_i induces (σ_1, σ_2),

(2) t_i expresses common belief in future δ-rationality,

(3) t_i believes that j is correct about i's beliefs, and believes that j believes that i is correct about j's beliefs,

(4) t_i satisfies Bayesian updating, and believes that j satisfies Bayesian updating.

In a similar way we can prove the following characterization of *uniform* subgame perfect equilibrium.

Theorem 6.2 (Characterization of uniform subgame perfect equilibrium) *Consider a finite two-player stochastic game Γ, and a behavioral strategy pair (σ_1, σ_2) in Γ. Then, (σ_1, σ_2) is a uniform subgame perfect equilibrium, if and only, there is a finite epistemic model $M = (T_i, \beta_i)_{i \in I}$ and for both players i a type $t_i \in T_i$, such that*

(1) t_i induces (σ_1, σ_2),

(2) t_i expresses common belief in future uniform rationality,

(3) t_i believes that j is correct about i's beliefs, and believes that j believes that i is correct about j's beliefs,

(4) t_i satisfies Bayesian updating, and believes that j satisfies Bayesian updating.

The proof is almost identical to the proof of Theorem 6.1, and is therefore omitted.

Note that the two theorems above would not change if we would allow for epistemic models with *infinitely* many types. For instance, if we would start from a *terminal* epistemic model in which all belief hierarchies are present, then the two theorems above state that (σ_1, σ_2) is a subgame perfect equilibrium exactly when we can find, for both players, a type within that model which satisfies conditions (1)–(4).

7 Proofs

Proof of Lemma 5.1. We construct the following Markov decision problem MDP for player i. The set of states X in MDP is simply the set of states in the stochastic game Γ, and for every state x the set of actions $A(x)$ in MDP is simply the set of actions $A_i(x)$ for player i in Γ. For every state x and action $a \in A(x)$, let the utility $u(x,a)$ in MDP be the utility that player i would obtain in Γ if the game reaches x, player i chooses a at x, and the opponents choose according to s_{-i} at x. Note that s_{-i} is a profile of stationary strategies, and hence the behavior induced by s_{-i} at x is independent of the history. So, $u(x,a)$ is well-defined. Finally, we define the transition probabilities $q(y|x,a)$ in MDP. For every two states x,y and every action $a \in A(x)$, let $q(y|x,a)$ be the probability that state y will be reached in Γ next period if the game is at x, player i chooses a at x, and i's opponents choose according to s_{-i} at x. Again, $q(y|x,a)$ is well-defined since, by stationarity of s_{-i}, the behavior of s_{-i} at x is independent of the history. This completes the construction of MDP.

We will now prove part (a) of the theorem. Take some $\delta \in (0,1)$. By part (a) in Theorem 3.1, we know that player i has a δ-optimal strategy $\hat{s}_i$ in MDP which is stationary. So, we can write $\hat{s}_i = (\hat{s}_i(x))_{x \in X}$. Now, let s_i be the stationary strategy for player i in the game Γ which prescribes, after every history h, the action $\hat{s}_i(x(h))$. Then, it may easily be verified that the stationary strategy s_i is δ-optimal for player i in Γ, given the conditional belief vector b_i.

Part (b) of the theorem can be shown in a similar way, by relying on part (b) in Theorem 3.1. ■

Proof of Theorem 5.1. We start by recursively defining profiles of stationary strategies, as follows. Let $s^1 = (s_i^1)_{i \in I}$ be an arbitrary profile of stationary strategies for the players. Let $b_i[s_{-i}^1]$ be a conditional belief vector for player i that assigns, after every history h, probability 1 to some strategy combination $s_{-i}^*[h]$ in $S_{-i}[s_{-i}^1, h]$. Moreover, these strategy combinations $s_{-i}^*[h]$ can be chosen in such a way that $s_{-i}^*[h] = s_{-i}^*[h']$ whenever h follows h' and $s_{-i}^*[h'] \in S_{-i}(h)$. In that way, we guarantee that $b_i[s_{-i}^1]$ satisfies Bayesian updating.

We know from Lemma 5.1 that for every player i there is a stationary strategy s_i^2 which is δ-optimal, given the conditional belief vector $b_i[s_{-i}^1]$. Let $s^2 := (s_i^2)_{i \in I}$ be the new profile of stationary strategies thus obtained. By recursively applying this step, we obtain an infinite sequence $s^1, s^2, s^3, ..$ of profiles of stationary strategies.

As there are only finitely many states in Γ, and finitely many actions at every state, there are also only finitely many stationary strategies for the players in the game. Hence, there are also only finitely many profiles of stationary strategies. Therefore, the infinite sequence $s^1, s^2, s^3, ...$ must go through a cycle

$$s^m \to s^{m+1} \to s^{m+2} \to ... \to s^{m+R} \to s^{m+R+1}$$

where $s^{m+R+1} = s^m$. We will now transform this cycle into an epistemic model where (a) all types express common belief in future δ-rationality, and (b) all types satisfy Bayesian updating.

For every player i, we define the set of types

$$T_i = \{t_i^m, t_i^{m+1}, ..., t_i^{m+R}\},$$

where t_i^{m+r} is a type that, after every history h, holds belief $b_i[s_{-i}^{m+r-1}](h)$ about the opponents' strategies, and assigns probability 1 to the event that every opponent j is of type t_j^{m+r-1}. If $r = 0$, then type t_i^m, after every history h, holds belief $b_i[s_{-i}^{m+R}](h)$ about the opponents' strategies, and assigns probability 1 to the event that every opponent j is of type t_j^{m+R}. This completes the construction of the epistemic model M.

Then, by construction, every type in the epistemic model satisfies Bayesian updating, since the conditional belief vectors $b_i[s_{-i}^k]$ are chosen such that they satisfy Bayesian updating. Moreover, every type t_i^{m+r} holds the conditional belief vector $b_i[s_{-i}^{m+r-1}]$ about the opponents' strategies. By construction, the stationary strategy s_i^{m+r} is δ-optimal under the conditional belief vector $b_i[s_{-i}^{m+r-1}]$, and hence s_i^{m+r} is δ-optimal for the type t_i^{m+r}, for every type t_i^{m+r} in the model.

By construction, every type t_i^{m+r} assigns, after every history h, and for every opponent j, probability 1 to the set of opponents' strategy-type pairs $S_j[s_j^{m+r-1}, h] \times \{t_j^{m+r-1}\}$. As every strategy $s_j' \in S_j[s_j^{m+r-1}, h]$ coincides with s_j^{m+r-1} at all histories weakly following h, and strategy s_j^{m+r-1} is δ-optimal for type t_j^{m+r-1} at all histories weakly following h, it follows that every strategy $s_j' \in S_j[s_j^{m+r-1}, h]$ is δ-optimal for type t_j^{m+r-1} at all histories weakly following h. That is,

$$S_j[s_j^{m+r-1}, h] \times \{t_j^{m+r-1}\} \subseteq (S_j \times T_j)^{h,\delta-opt} \text{ for all histories } h.$$

Since $\beta_i(t_i^{m+r}, h)(S_{-i}[s_{-i}^{m+r-1}, h] \times \{t_{-i}^{m+r-1}\}) = 1$ for all histories h, it follows that $\beta_i(t_i^{m+r}, h)(S_{-i} \times T_{-i})^{h,\delta-opt} = 1$ for all histories h. This means, however, that t_i^{m+r} believes in the opponents' future δ-rationality.

As this holds for every type t_i^{m+r} in the model M, we conclude that all types in M believe in the opponents' future δ-rationality. Hence, as a consequence, all types in M express *common* belief in future δ-rationality. This completes the proof. ■

Proof of Theorem 6.1. **(a)** Take first a δ-subgame perfect equilibrium (σ_1, σ_2). We will construct an epistemic model $M = (T_i, \beta_i)_{i \in I}$, and choose for both players i a type $t_i \in T_i$ within it, that satisfies the conditions (1) – (4) in the statement of the theorem.

Let $T_1 = \{t_1\}$ and $T_2 = \{t_2\}$, so we only consider one type for each player. Fix a player i. We transform σ_j into a conditional belief vector $b_i^{\sigma_j}$ for player i about j's strategy choice, as follows. Consider a history $h = ((x^1, a^1), ..., (x^{k-1}, a^{k-1}), x^k)$ of length k, and for every $m \leq k-1$ let $h^m = ((x^1, a^1), ..., (x^{m-1}, a^{m-1}), x^m)$ be the induced history of length m. Let σ_j^h be a modified behavioral strategy such that

(i) $\sigma_j^h(h^m)(a_j^m) = 1$ for every $m \leq k-1$, and

(ii) $\sigma_j^h(h') = \sigma_j(h')$ for all other histories h'.

Hence, σ_j^h assigns probability 1 to all the player j actions leading to h, and coincides with σ_j otherwise.

Remember that, for every strategy $s_j \in S_j(h)$ and every $m \geq k$, we denote by $[s_j]_m$ the set of strategies in $S_j(h)$ that coincide with s_j on histories up to length m. The σ-algebra $\Sigma_j(h)$ we use is generated by these sets $[s_j]_m$, with $s_j \in S_j(h)$ and $m \geq k$. Let $H^{\leq m}$ be the finite set of histories of length at most m. Then, let $b_i^{\sigma_j}(h) \in \Delta(S_j(h))$ be the unique probability distribution on $S_j(h)$ such that

$$b_i^{\sigma_j}(h)([s_j]_m) := \prod_{h' \in H^{\leq m}} \sigma_j^h(h')(s_j(h')) \tag{1}$$

for every strategy $s_j \in S_j(h)$ and every $m \geq k$. Note that $b_i^{\sigma_j}(h)$ is indeed a probability distribution on $S_j(h)$ as, by construction, σ_j^h assigns probability 1 to all player j actions leading to h. In this way, the behavioral strategy σ_j induces a conditional belief vector $b_i^{\sigma_j} = (b_i^{\sigma_j}(h))_{h \in H}$ for player i about j's strategy choices. Moreover, the conditional belief $b_i^{\sigma_j}(h) \in \Delta(S_j(h))$ has the property that the induced belief about j's *future* behavior is given by σ_j.

For both players i, we define the conditional beliefs $\beta_i(t_i, h) \in \Delta(S_j(h) \times T_j)$ about the opponent's strategy-type pairs as follows. At every history h of length k, let $\beta_i(t_i, h) \in \Delta(S_j(h) \times T_j)$ be the unique probability distribution such that

$$\beta_i(t_i, h)([s_j]_m \times \{t_j\}) := b_i^{\sigma_j}(h)([s_j]_m) \tag{2}$$

for every strategy $s_j \in S_j(h)$ and all $m \geq k$. So, type t_i believes, after every history h, that player j is of type t_j, and that player j will choose according to σ_j in the game that lies ahead. This completes the construction of the epistemic model $M = (T_i, \beta_i)_{i \in I}$.

Choose an arbitrary player i. We show that type t_i satisfies the conditions (1) – (4) above.

(1) We first show that $\sigma_j^{t_i} = \sigma_j$. Take some history $h = ((x^1, a^1), ..., (x^{k-1}, a^{k-1}), x^k)$ of length k, and some action $a_j \in A_j(x^k)$. Let

$$[S_j(h, a_j)]_k := \{[s_j]_k \mid s_j \in S_j(h, a_j)\}$$

be the finite collection of equivalence classes that partitions $S_j(h, a_j)$. Then,

$$\begin{aligned}
\sigma_j^{t_i}(h)(a_j) &= \beta_i(t_i, h)(S_j(h, a_j) \times T_j) \\
&= b_i^{\sigma_j}(h)(S_j(h, a_j)) \\
&= \sum_{[s_j]_k \in [S_j(h,a_j)]_k} b_i^{\sigma_j}(h)([s_j]_k) \\
&= \sum_{[s_j]_k \in [S_j(h,a_j)]_k} \prod_{h' \in H^{\leq k}} \sigma_j^h(h')(s_j(h')) \\
&= \sigma_j^h(h)(a_j) \\
&= \sigma_j(h)(a_j),
\end{aligned}$$

which implies that $\sigma_j^{t_i} = \sigma_j$. Here, the first equality follows from the definition of $\sigma_j^{t_i}$. The second equality follows from (2). The third equality follows from the observation that $[S_j(h, a_j)]_k$ constitutes a finite partition of the set $S_j(h, a)$, and that each member of $[S_j(h, a_j)]_k$ is in the σ-algebra $\Sigma_j(h)$. The fourth equality follows from (1). The fifth equality follows from two observations: First, that $s_j \in S_j(h, a_j)$, if and only if, $s_j(h^m) = a_j^m$ for all $m \leq k-1$ and $s_j(h) = a_j$, where $h^m = ((x^1, a^1), ..., (x^{m-1}, a^{m-1}), x^m)$ for all $m \leq k-1$. The second observation is that $\sigma_j^h(h^m)(a_j^m) = 1$ for all $m \leq k-1$. The sixth equality follows from the fact that σ_j^h coincides with σ_j on histories that weakly follow h. In particular, this implies that $\sigma_j^h(h) = \sigma_j(h)$.

In a similar way, we can show that $\sigma_i^{t_j} = \sigma_i$. Since t_i only assigns positive probability to type t_j, it follows that type t_i induces the behavioral strategy pair (σ_1, σ_2).

(2) We start by showing that type t_i believes in j's future δ-rationality. Consider an arbitrary history h. We show that $\beta_i(t_i, h)(S_j \times T_j)^{h,\delta-opt} = 1$.

Since (σ_i, σ_j) is a subgame perfect equilibrium, we have at every history h' weakly following h that

$$U_j^\delta(h', \sigma_j, \sigma_i) \geq U_j^\delta(h', \sigma_j', \sigma_i)$$

for every behavioral strategy σ_j'. This implies that

$$U_j^\delta(h', \sigma_j, \sigma_i) \geq U_j^\delta(h', s_j', \sigma_i)$$

for all $s_j' \in S_j(h')$. By (1), this is equivalent to stating that

$$U_j^\delta(h', b_i^{\sigma_j}(h'), b_j^{\sigma_i}(h')) \geq U_j^\delta(h', s_j', b_j^{\sigma_i}(h')) \tag{3}$$

for every history h' weakly following h, and every $s_j' \in S_j(h')$. Let

$$S_j^{opt}(h') := \{s_j \in S_j \mid U_j^\delta(h', s_j, b_j^{\sigma_i}(h')) \geq U_j^\delta(h', s_j', b_j^{\sigma_i}(h')) \text{ for all } s_j' \in S_j(h')\},$$

and let

$$S_j^{h,opt} := \{s_j \in S_j(h) \mid s_j \in S_j^{opt}(h') \text{ for every history } h' \text{ weakly following } h\}.$$

Then, by (3) it follows that $b_i^{\sigma_j}(h)(S_j^{h,opt}) = 1$.

Since the conditional belief of type t_j at h' about i's strategy is given by $b_j^{\sigma_i}(h')$, it follows that $S_j^{h,opt}$ contains exactly those strategies $s_j \in S_j(h)$ that are δ-optimal for type t_j at all histories weakly following h. Moreover, the conditional belief that type t_i has at h about j's strategy is given by $b_i^{\sigma_j}(h)$, for which we have seen that $b_i^{\sigma_j}(h)(S_j^{h,opt}) = 1$. By combining these two insights, we obtain that

$$\beta_i(t_i, h)(S_j \times T_j)^{h,\delta-opt} = \beta_i(t_i, h)(S_j^{h,opt} \times \{t_j\}) = b_i^{\sigma_j}(h)(S_j^{h,opt}) = 1.$$

As this holds for every history h, we conclude that t_i believes in j's future δ-rationality.

Since this holds for both players i, and since $T_1 = \{t_1\}$ and $T_2 = \{t_2\}$, it follows that both types t_1 and t_2 express common belief in future δ-rationality, which was to show.

(3) By the construction of our epistemic model M, type t_i always assigns probability 1 to type t_j which, in turn, always assigns probability 1 to type t_i. Hence, type t_i believes that j is correct about i's beliefs, and believes that j believes that i is correct about j's beliefs.

(4) Take some history $h^k = ((x^1, a^1), ..., (x^{k-1}, a^{k-1}), x^k)$ in H^k, and some history $h^{k+1} = ((x^1, a^1), ..., (x^{k-1}, a^{k-1}), (x^k, a^k), x^{k+1})$ in H^{k+1} that immediately follows h^k, and for which $\beta_i(t_i, h^k)(S_j(h^{k+1}) \times \{t_j\}) > 0$. Consider some $m \geq k+1$, and some $s_j \in S_j(h^{k+1})$. Then,

$$\begin{aligned}
\beta_i(t_i, h^k)([s_j]_m \times \{t_j\}) &= b_i^{\sigma_j}(h^k)([s_j]_m) \\
&= \prod_{h \in H^{\leq m}} \sigma_j^{h^k}(h)(s_j(h)) \\
&= \sigma_j^{h^k}(h^k)(s_j(h^k)) \prod_{h \in H^{\leq m} \backslash \{h^k\}} \sigma_j^{h^{k+1}}(h)(s_j(h)) \\
&= \sigma_j^{h^k}(h^k)(a_j^k) \prod_{h \in H^{\leq m} \backslash \{h^k\}} \sigma_j^{h^{k+1}}(h)(s_j(h)). \qquad (4)
\end{aligned}$$

Here, the first equality follows from equation (2). The second equality follows from equation (1). The third equality follows from the observation that $\sigma_j^{h^k}$ and $\sigma_j^{h^{k+1}}$ coincide on all histories except h^k. The fourth equality follows from the fact that $s_j(h^k) = a_j^k$, since $s_j \in S_j(h^{k+1})$.

On the other hand,

$$\begin{aligned}
\beta_i(t_i, h^k)(S_j(h^{k+1}) \times \{t_j\}) &= \beta_i(t_i, h^k)(S_j(h^k, a_j^k) \times \{t_j\}) \\
&= \sigma_j^{t_i}(h^k)(a_j^k) \\
&= \sigma_j(h^k)(a_j^k) \\
&= \sigma_j^{h^k}(h^k)(a_j^k). \qquad (5)
\end{aligned}$$

The first equality follows from the observation that $S_j(h^{k+1}) = S_j(h^k, a_j^k)$. The second equality follows from the definition of $\sigma_j^{t_i}$. The third equality follows from the fact that $\sigma_j^{t_i} = \sigma_j$, as we have shown above. The fourth equality follows from the observation that $\sigma_j^{h^k}(h^k) = \sigma_j(h^k)$.

By equations (4) and (5) it follows, for every $s_j \in S_j(h^{k+1})$,

$$\begin{aligned}\frac{\beta_i(t_i,h^k)([s_j]_m \times \{t_j\})}{\beta_i(t_i,h^k)(S_j(h^{k+1}) \times \{t_j\})} &= \prod_{h \in H^{\leq m} \setminus \{h^k\}} \sigma_j^{h^{k+1}}(h)(s_j(h)) \\ &= \prod_{h \in H^{\leq m}} \sigma_j^{h^{k+1}}(h)(s_j(h)). \\ &= b_i^{\sigma_j}(h^{k+1})([s_j]_m) \\ &= \beta_i(t_i,h^{k+1})([s_j]_m \times \{t_j\}).\end{aligned}$$

Here, the second equality follows from the fact that $\sigma_j^{h^{k+1}}(h^k)(s_j(h^k)) = \sigma_j^{h^{k+1}}(h^k)(a_j^k) = 1$, by construction of $\sigma_j^{h^{k+1}}$. The third and fourth equality follow from equations (1) and (2), respectively.

Hence, we have shown that

$$\beta_i(t_i,h^{k+1})([s_j]_m \times \{t_j\}) = \frac{\beta_i(t_i,h^k)([s_j]_m \times \{t_j\})}{\beta_i(t_i,h^k)(S_j(h^{k+1}) \times \{t_j\})}$$

for every $s_j \in S_j(h^{k+1})$ and every $m \geq k+1$. As the σ-algebra $\Sigma_j(h^{k+1})$ is generated by these sets $[s_j]_m$, it follows that

$$\beta_i(t_i,h^{k+1})(E_j \times \{t_j\}) = \frac{\beta_i(t_i,h^k)(E_j \times \{t_j\})}{\beta_i(t_i,h^k)(S_j(h^{k+1}) \times \{t_j\})}$$

for every history $h \in H^k$, every history $h^{k+1} \in H^{k+1}$ following h^k with $\beta_i(t_i,h^k)(S_j(h^{k+1}) \times \{t_j\}) > 0$, and every set $E_j \in \Sigma_j(h^{k+1})$. But then, it follows that this equality also holds for every history h, every history h' following h with $\beta_i(t_i,h)(S_j(h') \times \{t_j\}) > 0$, and every $E_j \in \Sigma_j(h')$. So, type t_i indeed satisfies Bayesian updating, as was to show. In the same way, it can be shown that also t_j satisfies Bayesian updating, so t_i believes that j satisfies Bayesian updating too.

Summarizing, we have shown that both types t_i and t_j satisfy the conditions (1) – (4).

(b) Assume next that there is a finite epistemic model $M = (T_i, \beta_i)_{i \in I}$, and for both players i a type $t_i \in T_i$, such that t_i induces (σ_i, σ_j), and satisfies conditions (2) – (4). We show that (σ_i, σ_j) must be a δ-subgame perfect equilibrium. We proceed in two steps.

Step 1: There is a type $t_j \in T_j$ such that t_i always assigns probability 1 to t_j.

Proof of Step 1. Suppose that there are two different types, t_j and t_j', and two histories h, h', such that $\beta_i(t_i,h)$ assigns positive probability to t_j, and $\beta_i(t_i,h')$ assigns positive probability to t_j'. Since, by the first part in condition (3), t_i believes that j is correct about i's beliefs, it must be the case that t_j always assigns probability 1 to type t_i. But then, t_j always believes with probability 1 that player i, at history h', assigns positive probability to type $t_j' \neq t_j$. This means that type t_j does *not* believe that i is correct about j's beliefs. As a consequence, t_i does not believe that j believes that i is correct about j's beliefs, which would contradict the second part of condition (3). Hence, there must be single type t_j such that t_i always assigns probability 1 to t_j. This completes the proof of step 1.

By Step 1, and the assumption that t_i believes that j is correct about i's beliefs, it follows that there is a single type t_j such that (i) t_i always assigns probability 1 to t_j, and (ii) t_j always assigns probability 1 to t_i. Since t_i induces (σ_i, σ_j), it must then be that $\sigma_j^{t_i} = \sigma_j$ and $\sigma_i^{t_j} = \sigma_i$. Moreover, as t_i satisfies Bayesian updating, and believes that j satisfies Bayesian updating, both t_i and t_j must satisfy Bayesian updating.

Step 2: The behavioral strategy pair (σ_i, σ_j) is a δ-subgame perfect equilibrium.

Proof of Step 2. Take a player i and a history h. We must show that

$$U_i^\delta(h, \sigma_i, \sigma_j) \geq U_i^\delta(h, \sigma_i', \sigma_j) \tag{6}$$

for every behavioral strategy σ_i'. By (1) this is equivalent to showing that

$$U_i^\delta(h, b_j^{\sigma_i}(h), b_i^{\sigma_j}(h)) \geq U_i^\delta(h, s_i', b_i^{\sigma_j}(h)) \tag{7}$$

for all $s_i' \in S_i(h)$. Let

$$S_i^{opt}(h) := \{s_i \in S_i(h) \mid U_i^\delta(h, s_i, b_i^{\sigma_j}(h)) \geq U_i^\delta(h, s_i', b_i^{\sigma_j}(h)) \text{ for all } s_i' \in S_i(h)\}.$$

Then, (7) is equivalent to showing that

$$b_j^{\sigma_i}(h)(S_i^{opt}(h)) = 1. \tag{8}$$

As $\sigma_j^{t_i} = \sigma_j$ and t_i satisfies Bayesian updating, it follows that the conditional belief of type t_i at h about j's continuation strategy is given by $b_i^{\sigma_j}(h)$. But then,

$$S_i^{opt}(h) = \{s_i \in S_i(h) \mid s_i \text{ is } \delta\text{-optimal for } t_i \text{ at history } h\}.$$

As t_i, by assumption, believes that j believes in i's future δ-rationality, it must be that t_j believes in i's future δ-rationality. In particular,

$$\beta_j(t_j, h)(S_i \times T_i)^{h,\delta-opt} = 1.$$

As t_j assigns probability 1 to t_i, and every strategy s_i which is δ-optimal for t_i at all histories weakly following h must be in $S_i^{opt}(h)$, it follows that

$$\beta_j(t_j, h)(S_i^{opt}(h) \times \{t_i\}) = 1. \tag{9}$$

Since $\sigma_i^{t_j} = \sigma_i$ and t_j satisfies Bayesian updating, it follows that the conditional belief of type t_j at h about i's continuation strategy is given by $b_j^{\sigma_i}(h)$. So, (9) implies that

$$b_j^{\sigma_i}(h)(S_i^{opt}(h)) = 1,$$

which establishes (8). This, as we have seen, implies (6), stating that

$$U_i^\delta(h, \sigma_i, \sigma_j) \geq U_i^\delta(h, \sigma_i', \sigma_j)$$

for every behavioral strategy σ_i'.

Since this holds for both players i and every history h, it follows that (σ_i, σ_j) is a δ-subgame perfect equilibrium. This completes the proof of Step 2, and therefore completes the proof of this theorem. ■

References

[1] Asheim, G.B. (2006), *The consistent preferences approach to deductive reasoning in games,* Theory and Decision Library, Springer, Dordrecht, The Netherlands.

[2] Aumann, R. and A. Brandenburger (1995), Epistemic conditions for Nash equilibrium, *Econometrica* **63,** 1161-1180.

[3] Baltag, A., Smets, S. and J.A. Zvesper (2009), Keep 'hoping' for rationality: a solution to the backward induction paradox, *Synthese* **169,** 301–333 (*Knowledge, Rationality and Action* 705–737).

[4] Battigalli, P. (2003), Rationalizability in infinite, dynamic games with incomplete information, *Research in Economics* **57,** 1–38.

[5] Battigalli, P. and M. Siniscalchi (1999), Hierarchies of conditional beliefs and interactive epistemology in dynamic games, *Journal of Economic Theory* **88**, 188–230.

[6] Battigalli, P. and M. Siniscalchi (2002), Strong belief and forward induction reasoning, *Journal of Economic Theory* **106,** 356–391.

[7] Ben-Porath, E. (1997), Rationality, Nash equilibrium and backwards induction in perfect-information games, *Review of Economic Studies* **64**, 23–46.

[8] Blackwell, D. (1962), Discrete dynamic programming, *The Annals of Mathematical Statistics* **33,** 719–726.

[9] Blackwell, D., and T.S. Ferguson (1968), The Big Match, *The Annals of Mathematical Statistics* **39**, 159–163.

[10] Brandenburger, A. and E. Dekel (1987), Rationalizability and correlated equilibria, *Econometrica* **55,** 1391–1402.

[11] Brandenburger, A. and E. Dekel (1989), The role of common knowledge assumptions in game theory, in *The Economics of Missing Markets, Information and Games,* ed. by Frank Hahn. Oxford: Oxford University Press, pp. 46–61.

[12] Filar, J. and K. Vrieze (1997), *Competitive Markov Decision Processes,* Springer-Verlag.

[13] Friedenberg, A. (2010), When do type structures contain all hierarchies of beliefs?, *Games and Economic Behavior* **68,** 108–129.

[14] Gillette, D. (1957), Stochastic games with zero stop probabilities, in A.W. Tucker, M. Dresher, and P. Wolfe (eds.), *Contributions to the Theory of Games,* Princeton University Press.

[15] Harsanyi, J.C. (1967–1968), Games with incomplete information played by "bayesian" players, I–III', *Management Science* **14**, 159–182, 320–334, 486–502.

[16] Howard, R.A. (1960), *Dynamic Programming and Markov Processes,* Technology Press and Wiley, New York.

[17] Jaśkiewicz, A., and A.S. Nowak (2017), Zero–Sum stochastic games, *Handbook of Dynamic Games,* Vol. I (Theory), Springer.

[18] Nash, J.F. (1950), Equilibrium points in N-person games, *Proceedings of the National Academy of Sciences of the United States of America* **36**, 48–49.

[19] Nash, J.F. (1951), Non-cooperative games, *Annals of Mathematics* **54**, 286–295.

[20] Penta, A. (2015), Robust dynamic implementation, *Journal of Economic Theory* **160**, 280–316.

[21] Perea, A. (2007), A one-person doxastic characterization of Nash strategies, *Synthese* **158**, 251–271 (*Knowledge, Rationality and Action* 341–361).

[22] Perea, A. (2010), Backward induction versus forward induction reasoning, *Games* **1**, 168–188.

[23] Perea, A. (2012), *Epistemic Game Theory: Reasoning and Choice,* Cambridge University Press.

[24] Perea, A. (2014), Belief in the opponents' future rationality, *Games and Economic Behavior* **83,** 231–254.

[25] Rosenthal, R.W. (1981), Games of perfect information, predatory pricing and the chain-store paradox, *Journal of Economic Theory* **25,** 92–100.

[26] Selten, R. (1965), Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragezeit, *Zeitschrift für die Gesammte Staatswissenschaft* **121,** 301–324, 667–689.

[27] Shapley, L.S. (1953), Stochastic games, *Proceedings of the National Academy of Science USA* **39,** 1095–1100.

[28] Solan, E., and N. Vieille (2003), Deterministic multi–player Dynkin games, *Journal of Mathematical Economics* **39**, 911–929.

[29] Tan, T. and S.R.C. Werlang (1988), The bayesian foundations of solution concepts of games, *Journal of Economic Theory* **45,** 370–391.