Subgame-perfection in stochastic games with perfect information and recursive payoffs

Please check the document version of this publication:

János Flesch, Jeroen Kuipers,
Gijs Schoenmakers, Koos Vrieze

Subgame-Perfection in Stochastic Games with
Perfect Information and Recursive Payoffs

METE⊚R

Subgame-Perfection in Stochastic Games with Perfect Information and Recursive Payoffs

J. Flesch,[*] J. Kuipers,[†] G. Schoenmakers,[†] K. Vrieze[†]

November 21, 2008

Abstract

We consider a class of n-player stochastic games with the following properties: (1) in every state, the transitions are controlled by one player, (2) the payoffs are equal to zero in every non-absorbing state, (3) the payoffs are non-negative in every absorbing state. With respect to the expected average reward, we provide a constructive proof that a subgame-perfect ε-equilibrium exists in pure strategies, for every $\varepsilon > 0$. Moreover, if all transitions of a game in our class are deterministic, then the game has a subgame-perfect 0-equilibrium in pure strategies.

Key words: stochastic games; perfect information games; recursive games; subgame-perfect equilibria

MSC2000 subject classification: 91A15, 91A06

OR/MS subject classification: games/group decisions, stochastic

1 Introduction

In the theory of average-reward stochastic games (with finite state and action spaces), a major open problem is whether or not ε-equilibria exist for all $\varepsilon > 0$. The famous game called the Big Match, which was introduced by Gillette [5] and solved by Blackwell and Ferguson [1], and the game in Sorin [16] demonstrated that 0-equilibria do not necessarily exist, and moreover, that history-dependent strategies are indispensable for establishing ε-equilibria. For two-player zero-sum games, Mertens and Neyman [8] showed the existence of ε-equilibria, in terms of ε-optimal strategies. Later, Vieille [18], [19] provided a proof that ε-equilibria exist in all two-player stochastic games. For more than two players, however, only partial results are available, under rather restrictive conditions. For results on the existence of ε-equilibria in special classes, we refer to Thuijsman and Raghavan [17], Solan [12], Solan and Vieille [14], Simon [11], Flesch et al. [2] and Flesch et al. [3], [4].

About subgame-perfect ε-equilibria even less is known. Some recent results fell into the class $\mathcal{G}$ of n-player perfect information games with recursive payoffs. Here, perfect information refers to the property that, in any state of the game, at most one player has

---

[*]Dept. of Quantitative Economics, Maastricht University, P.O. Box 616, 6200MD, Maastricht, The Netherlands.

[†]Department of Mathematics, Maastricht University, P.O. Box 616, 6200MD, Maastricht, The Netherlands.

more than one action (this player controls the state). We assume that the payoffs in absorbing states are non-negative, so that the players have an incentive to eventually reach an absorbing state. It follows from Thuijsman and Raghavan [17] that equilibria exist in the class $\mathcal{G}$, but these equilibria are not subgame-perfect, since they involve punishments where $n-1$ players minimize a deviating player, without regard to their own payoffs.

Solan [13] showed for a subclass of games in $\mathcal{G}$ that subgame-perfect ε-equilibria exist, for all $\varepsilon > 0$ (without assuming the non-negativity of the payoffs). His proof requires two restrictions for games in $\mathcal{G}$: (1) each player controls exactly one non-absorbing state, (2) each player in his own state has the choice between two actions: one "absorbing" action leading immediately to an absorbing state, and a "non-absorbing" one leading to the n non-absorbing states according to a probability distribution. It is essential that this probability distribution is the same for all n non-absorbing actions. His proof is based on the analysis of a specific type of differential inclusions. A generalization of this result to periodic probability distributions can be found in Mashiah-Yaakovi [7].

For another subclass of games in $\mathcal{G}$, called free transition games, Kuipers et al. [6] proved the existence of subgame-perfect 0-equilibria in pure strategies. A game $G \in \mathcal{G}$, with T denoting the set of non-absorbing states of G, is called a free transition game if it satisfies the following condition: in any state $s \in T$, the action space of the controlling player i_s is exactly $\{0\} \cup T$, where action 0 is "absorbing" and leading immediately to an absorbing state, and where action $a \in T$ is "non-absorbing" and leads to non-absorbing state a with probability 1.

In this paper, we provide a constructive proof that subgame-perfect ε-equilibria exist in pure strategies, for all $\varepsilon > 0$, within the whole class $\mathcal{G}$. For the subclass of games with deterministic transitions the proof simplifies, and the existence of subgame-perfect 0-equilibria in pure strategies is guaranteed. Our construction uses history-dependent strategies, which is necessary, as stationary strategies are in general insufficient for ε-equilibria (cf., for example, Solan and Vieille [15] or Kuipers et al. [6]). Although our approach relies on the properties of the games in the class $\mathcal{G}$, it is very natural and may initiate further results outside this class. In this respect, we wish to mention two related models, in which there is an ongoing investigation regarding the existence of equilibria and subgame-perfect equilibria. First, we mention a class of Dynkin games, where in each state (of a possibly infinite state space), the controlling player has one non-absorbing action, called "continue", and one absorbing action, called "quit" (cf. for example, Solan and Vieille [15] and Solan [13]). Another related model is the class of so-called stopping games, where in contrast with Dynkin games, several players may choose simultaneously between actions "continue" and "quit" (cf. for example, Shmaya et al. [10] and Shmaya and Solan [9]).

2 The model and the main results

The class $\mathcal{G}$ of stochastic games An n-player stochastic game in class $\mathcal{G}$ is given by (1) a nonempty set of players $N = \{1, \ldots, n\}$, (2) a nonempty and finite set of states S, (3) for each state $t \in S$, an associated 'controlling' player $i_t \in N$, (4) for each state $t \in S$, a nonempty and finite set of actions A_t, (5) for each state $t \in S$ and each action $a \in A_t$,

a transition probability distribution $p_t(a) = (p_t(a, u))_{u \in S}$, (6) for each state $t \in S$ and each action $a \in A_t$, a payoff $r_t^i(a) \in \mathbb{R}$ to each player i, such that the payoffs are equal to 0 in all non-absorbing states and the payoffs are non-negative in all absorbing states. Here, a state t is called absorbing if $p_t(a, t) = 1$ for all actions $a \in A_t$; otherwise t is called non-absorbing.

The game is to be played at stages in $\mathbb{N}$ in the following way. At any stage m, in the present state $s_m \in S$, the controlling player i_{s_m} has to choose an action a_m from the action set A_{s_m}. The chosen action a_m induces a payoff $r_{s_m}^j(a_m)$ to each player j, and a transition to a new state according to the transition probability distribution $p_{s_m}(a_m)$, where play will continue at stage $m + 1$. We assume complete information (i.e. the players know all the data of the stochastic game), full monitoring (i.e. the players observe the present state and the action chosen by the controlling player), and perfect recall (i.e. the players remember all previous states and actions). The game starts in an initial state $s \in S$.

Strategies A mixed action in state $t \in S$ for player i_t is a probability distribution on A_t. The set of these mixed actions is denoted by X_t. For $a \in A_t$, let $S_t(a) = \{u \in S \mid p_t(a, u) > 0\}$ denote the set of states to which transition occurs with a positive probability when action a is taken at state t. Let $H_{s,t}$ denote the set of all possible sequences $(s = s_1, a_1, \ldots, s_m, a_m, s_{m+1} = t)$ of arbitrary but finite length, where for every $k = 1, \ldots, m$ we have that (a) s_k is a state and a_k is an action of the controlling player player i_{s_k} in state s_k, (b) $s_{k+1} \in S_{s_k}(a_k)$. Thus, $H_{s,t}$ is the set of all possible histories starting in initial state s and ending in state t.

A strategy π^i for player i and initial state s is a decision rule that, for any history $h \in H_{s,t}$ with $i = i_t$, prescribes a mixed action $\pi^i(h) \in X_t$. We use the notation Π_s^i for the set of strategies for player i and initial state s.[1] A strategy $\pi^i \in \Pi_s^i$ is called pure if every prescription $\pi^i(h)$ places probability 1 on one action. If the mixed actions prescribed by a strategy only depend on the final state of h, then the strategy is called stationary. We use the notation Π_s for the set of joint strategies $\pi = (\pi^i)_{i \in N}$ with $\pi^i \in \Pi_s^i$ for $i \in N$. A joint strategy $\pi = (\pi^i)_{i \in N}$ is pure if π^i is pure for all $i \in N$, and it is stationary if π^i is stationary for all $i \in N$.

Consider a strategy $\pi^i \in \Pi_s^i$ and a history $h \in H_{s,t}$. The <u>continuation strategy</u> $\pi^i[h] \in \Pi_t^i$ for player i and initial state t prescribes mixed actions for histories $h' \in H_{t,u}$ with $i_u = i$ according to π^i, but as if h had happened before h'. More formally, the continuation strategy $\pi^i[h]$ prescribes the mixed action $\pi^i[h](h') = \pi^i(h \oplus h')$ in state u. Here, $h \oplus h'$ is the history obtained by concatenation of h and h' (where t, the final state of h and the initial state of h', merge to one state t in $h \oplus h'$). We use the notation $\pi[h]$ to denote the joint continuation strategy, associated with $\pi = (\pi^i)_{i \in N}$ and $h \in H_{s,t}$.

Rewards For initial state $s \in S$ and a joint strategy $\pi \in \Pi_s$, the sequences of payoffs are evaluated by the expected average reward, which is given for player i by

$$\phi^i(\pi) := \liminf_{M \to \infty} \mathbb{E}_\pi \left(\frac{1}{M} \sum_{m=1}^M R_m^i \right) = \liminf_{M \to \infty} \frac{1}{M} \sum_{m=1}^M \mathbb{E}_\pi \left(R_m^i \right),$$

---

[1]In this article, a strategy provides prescriptions for histories with a given initial state, and it provides no prescriptions for other initial states. This is to conform with the concept of a continuation strategy, to be defined later on.

where R_m^i is the random variable for the payoff for player i at stage m, and where $\mathbb{E}_\pi$ stands for expectation with respect to play according to the joint strategy π.

Equilibria A joint strategy $\pi = (\pi^i)_{i \in N} \in \Pi_s$ is called a (Nash) ε-equilibrium for initial state s, for some $\varepsilon \geq 0$, if

$$\phi^i \left(\sigma^i, (\pi^j)_{j \in N - \{i\}} \right) \leq \phi^i (\pi) + \varepsilon \qquad \forall \sigma^i \in \Pi_s^i, \ \forall i \in N,$$

which means that no player can gain more than ε by a unilateral deviation. A strategy profile $\pi = (\pi_s)_{s \in S}$, with $\pi_s \in \Pi_s$ for all $s \in S$, is called a (Nash) ε-equilibrium, for some $\varepsilon \geq 0$, if π_s is a Nash-equilibrium for initial state s for all $s \in S$. As we mentioned in the introduction, a 0-equilibrium exists in every game in the class $\mathcal{G}$, but not in stationary strategies in general.

Subgame-perfect equilibria A joint strategy $\pi \in \Pi_s$ is called a subgame-perfect ε-equilibrium for initial state s, for some $\varepsilon \geq 0$, if for any $t \in S$ and history $h \in H_{s,t}$, the joint continuation strategy $\pi[h]$ is an ε-equilibrium for initial state t. A strategy profile $\pi = (\pi_s)_{s \in S}$, with $\pi_s \in \Pi_s$ for all $s \in S$, is called a subgame-perfect ε-equilibrium, for some $\varepsilon \geq 0$, if π_s is a subgame-perfect ε-equilibrium for initial state s for all $s \in S$. Subgame-perfection is a refinement of the equilibrium concept.

Our main results concern the existence of subgame-perfect ε-equilibria in the class $\mathcal{G}$ of stochastic games. We will pay special attention to the case when all transitions are deterministic, i.e. when for every state $s \in S$ and action $a \in A_s$, there exists a state $t \in S$ such that $p_s(a, t) = 1$.

Main Theorem *Every stochastic game G in class $\mathcal{G}$ has a subgame-perfect ε-equilibrium in pure strategies, for every $\varepsilon > 0$. Moreover, if all transitions in G are deterministic, then G has a subgame-perfect 0-equilibrium in pure strategies.*

The article is structured as follows. In section 3, we prove a useful lemma claiming that, for the main theorem, it suffices to consider games in a subclass $\mathcal{G}'$ of $\mathcal{G}$. In section 4, we introduce notions that will play an important role in our analysis. Finally, in section 5, we present the proof of the main theorem for the subclass $\mathcal{G}'$.

3 A reduction

In this section, we will show that, for our main results, it is sufficient to guarantee the existence of pure subgame-perfect ε-equilibria in a subclass of $\mathcal{G}$. Let $\mathcal{G}'$ denote the class of games in $\mathcal{G}$ which satisfy:

(1) From any state, the players have a joint strategy such that absorption eventually occurs with probability 1. (One can show that it is sufficient to consider pure stationary strategies.)

(2) In any absorbing state s, player i_s has precisely one action.

(3) In any non-absorbing state s, any action of player i_s either leads to non-absorbing states only, or it leads to one absorbing state with probability 1. In this sense, we will speak of non-absorbing and absorbing actions.

(4) In any non-absorbing state s, player i_s has precisely one absorbing action.

Now we will provide a natural transformation which reduces an arbitrary game $G \in \mathcal{G}$ into a game $G' \in \mathcal{G}'$. So, take a game $G \in \mathcal{G}$. We obtain G' in four steps as follows:

Step 1 for property (1): If G does not satisfy this property, then there must exist a state s such that, irrespective of the strategies of the players, the probability that play ever absorbs is zero. This means the payoffs are zeros regardless what happens. For this reason, we replace all such states by absorbing states in which the controlling player has one action and all payoffs are equal to zero.

Step 2 for property (2): In any absorbing state s, we delete all actions of player i_s except one that offers player i_s the highest payoff in state s.

Step 3 for property (3): Suppose that, in some non-absorbing state s, some action a of player i_s does not satisfy this property. Then, action a leads to the set of absorbing states with a positive probability. Given absorption takes place through action a, let w^j denote the conditional expected payoff for player j in the set of absorbing states. More formally,

$$
w^j = \frac{\displaystyle\sum_{u \in S, \ u \text{ is absorbing}} p_s(a, u) \cdot r_u^j}{\displaystyle\sum_{u \in S, \ u \text{ is absorbing}} p_s(a, u)},
$$

where r_u^j denotes the unique payoff (cf. property (2)) for player j in absorbing state u. Then, (i) we add two new states: (ia) we add a new absorbing state s^* in which the controlling player has one action and the payoff for every player j is equal to w^j, (ib) we add a new non-absorbing state s' with one action for the controlling player from which transition occurs to state s^* with probability 1, and (ii) we replace each transition through action a to an absorbing state by a transition to state s'. We apply the same transformation to all actions that violate property (3).

Step 4 for property (4): If in some non-absorbing state s, player i_s has more than one absorbing action, then we delete all absorbing actions of player i_s except one that offers player i_s the highest payoff at absorption. We apply the same transformation to all such states. Suppose now on the other hand that there is a non-absorbing state s in which player i_s has no absorbing action. Let S^0 denote the set of these states. Then, (i) we raise all payoffs in the absorbing states by 1, (ii) we add a new absorbing state t^* in which the controlling player has one action and all payoffs are equal to 0, (iii) in each state in S^0, we add a new absorbing action leading to state t^*.

It is clear that the new game G' satisfies all four properties, thus $G' \in \mathcal{G}'$.

Lemma 1 *Take an arbitrary game $G \in \mathcal{G}$, and transform G, according to the rules above, into a game $G' \in \mathcal{G}'$. If a subgame-perfect ε-equilibrium exists in G' in pure strategies for all $\varepsilon > 0$, then one also exists in G. Moreover, if a subgame-perfect 0-equilibrium exists in G' in pure strategies, then one also exists in G.*

Proof. We only show the second part, as the proof for the first part on subgame-perfect ε-equilibria is almost identical. Now let π denote a pure subgame-perfect 0-equilibrium in the game G'. We will show that π induces a pure subgame-perfect 0-equilibrium for the original game G.

We will only argue for the transformation in step 4, as it is obvious for the other three steps. It is enough to show that π never prescribes to choose the new absorbing actions in states in S^0. Suppose $S^0 \neq \emptyset$, otherwise the statement is obvious. Let T denote the set of states after the transformation in step 3 (T can differ from the original state space S due to step 3). Due to the transformation in step 1, $T - S^0 \neq \emptyset$. Consider first a state $s \in T - S^0$. In this state, player i_s can obtain a payoff of at least 1 by playing his absorbing action. Hence, his continuation reward from state s with regard to π is always at least 1. Notice that the payoffs in any absorbing state are either all at least 1 or all equal to 0. Hence, each player's (expected) continuation reward from state s with regard to π is always at least some $\rho_s > 0$. Due to the transformation in step 1, there exists a state $t \in S^0$ with an action a such that action a leads to $S - S^0$ with a positive probability, say q_a. Then, if player i_t plays action a in state t, his continuation reward with respect to π is at least $q_a \cdot \min_{s \in S-S^0} \rho_s > 0$. Hence, the continuation reward of player i_t from state t with regard to π, and consequently for all other players as well, is at least some $\rho_t > 0$. This means in particular that player i_t never plays his new absorbing action in state t. If $S^0 - \{t\} = \emptyset$, then we are done. On the other hand, if $S^0 - \{t\} \neq \emptyset$, then we can continue as before, since there exists again a state $u \in S^0 - \{t\}$ with an action b such that action b leads to $(T - S^0) \cup \{t\}$ with a positive probability. Since the number of states is finite, we obtain in finitely many steps a $\rho > 0$ such that the continuation rewards of the players with regard to π are always at least ρ. Hence, the new absorbing actions in states in S^0 are never chosen, as desired. $\qquad\square$

Due to the above lemma, we may restrict our investigation to games in the class $\mathcal{G}'$.

An equivalent interpretation of games in $\mathcal{G}'$. Note that, in every state s of a game belonging to $\mathcal{G}'$, if player i_s plays his absorbing action, then play moves with probability 1 to the absorbing state associated with s, say $t(s)$. Play is over from a strategic point of view once $t(s)$ is reached, and each player j receives an (expected) reward equal to $r^j_{t(s)}$. From now on, we will use the following equivalent interpretation of games in $\mathcal{G}'$: in every non-absorbing state $s \in S$, if player i_s plays his absorbing action then play terminates, with payoff $r^j_s := r^j_{t(s)}$ to every player j. If the players never use their absorbing actions, play continues forever with payoff 0 to every player. In this interpretation, the absorbing states play no role. Playing the absorbing action in a non-absorbing state will be called quitting. Moreover, the absorbing actions will be called quitting actions, and all other actions will be called non-quitting actions.

4 Game-plans and viability

Take a stochastic game $G \in \mathcal{G}'$ and an initial state $s \in S$. For a joint pure strategy $\pi \in \Pi_s$, let H^π denote the set of those histories which have a positive probability with respect to π. A function g from histories to actions is called a complete game-plan for initial state s if g equals the restriction $\pi_{|H^\pi}$ of some joint pure strategy $\pi \in \Pi_s$ to the set H^π. Let

$H^g = H^\pi$ denote the domain of g. Thus, $g(h) = \pi(h)$ for all $h \in H^g$. Clearly, different joint pure strategies can induce the same complete game-plan. The idea of a complete game-plan is that it provides a prescription for the choice of an action during the whole play, if no player ever deviates from the plan. We say that a game-plan g is stationary if $g(h)$ only depend on the final state of h.

It is also possible to construct a complete game-plan g for initial state s inductively, as follows.

(1) At stage 1, an action $g(s) \in A_s$ is chosen. Define $H^{g,1} = \{(s)\}$.

(2) At stage $m+1$ ($m \geq 1$), an action $g(h) \in A_u$ is chosen for all histories h of the form $h = h^m \oplus (t, g(h^m), u)$, where $h^m \in H^{g,m}$, where t denotes the final state of h^m, and where $u \in S_t(g(h^m))$. Let $H^{g,m+1}$ denote the set of such histories h.

The domain of such a constructed game-plan g is given by $H^g = \cup_{m \in \mathbb{N}} H^{g,m}$.

A function g from histories to actions is called a truncated game-plan for initial state s if g equals the restriction $\pi_{|W}$ of a joint pure strategy π to a set $W \subsetneq H^\pi$ of histories such that W satisfies: if $h \in W$ is an arbitrary history, say up to stage m, then, for any stage $l < m$, the part of h up to stage l also belongs to W. An equivalent formulation of this property of the set W is that if a history does not belong to W, then neither does any extension of this history up to larger stages. Note that if the players follow the prescriptions of a truncated game-plan g, then it may happen that a history $h \in H^\pi - W$ occurs. In this case, g provides no further prescriptions, and we will say that g expires. Also for the truncated game-plan g, let $H^g = W$ denote the domain of g.

Example 1 Consider a game G with three players and three states s, t and u. In state s, player 1 can either play a non-quitting action a_s leading to state s with probability $\frac{1}{2}$ and to state t with probability $\frac{1}{2}$, or quit. In state t, player 2 can either play a non-quitting action a_t leading to state u with probability 1, or play the quitting action. In state u, player 3 can only quit. Consider state s as the initial state. An example of a complete game-plan is $\hat{g}$, which prescribes action a_s in state s until play arrives at state t and then quits in state t. More formally, $H^{\hat{g}}$ consists of all histories that are either of the form $(s, a_s, s, a_s, \ldots, s, a_s, s)$ or of the form $(s, a_s, s, a_s, \ldots, s, a_s, t)$, and $\hat{g}$ assigns action a_s to the former ones and quitting to the latter ones. An example of a truncated game-plan is g' which prescribes action a_s in state s until play arrives at state t. More formally, $H^{g'}$ consists of all histories of the form $(s, a_s, s, a_s, \ldots, s, a_s, s)$, and g' assigns action a_s to all these histories. ∎

Consider a (complete or truncated) game-plan g for initial state s. For any $h \in H^g$, we can define the continuation game-plan $g[h]$ of g with respect to h by $g[h](h') = g(h \oplus h')$ for all h' with $h \oplus h' \in H^g$, just as in the case of strategies. Notice that $g[h]$ is a game-plan for the final state of h. In the following we will sometimes use the notation H^g_t to denote the histories in H^g with final state t. Thus, if $h \in H^g_t$, then $g[h]$ is a game-plan for state t. Note that $g[h]$ is complete if g is complete.

A complete game-plan g is called quitting, if playing according to g eventually leads to quitting, with probability 1. For a quitting game-plan g, the expected payoff to a player

i is denoted by $\phi^i(g)$. Note that any continuation game-plan $g[h]$ of a quitting game-plan g is also quitting.

Take a real vector $\alpha = (\alpha_t)_{t \in S}$. A quitting game-plan g is called <u>viable</u> with respect to α, if $\phi^{i_t}(g[h]) \geq \alpha_t$ holds for all $t \in S$ and all $h \in H_t^g$. This means that, whenever play is in some state t, and play is according to g, the controlling player i_t can expect a payoff of at least α_t. In particular, viability of g with respect to α implies that termination can only take place at states t with the property $r_t^{i_t} \geq \alpha_t$. A state t with this property is called a <u>quitting state</u> with respect to α.

Let $V_s(\alpha)$ denote the set of game-plans for initial state s that are viable with respect to α. A specific viable game-plan will be denoted by v. Note that, for a game-plan $v \in V_s(\alpha)$ and a history $h \in H^v$ with final state t, the continuation game-plan $v[h]$ is a viable game-plan in $V_t(\alpha)$. Notice that viable game-plans are, by definition, always quitting and complete.

Example 2 Consider the game G from example 1 and game-plan $\hat{g}$ for initial state s as defined in that example. Let the payoff vectors be given by $(2,2,2)$, $(3,1,2)$, and $(1,2,2)$ for s, t and u respectively. Let $\alpha_s = \alpha_u = 2$ and $\alpha_t = 1$. According to $\hat{g}$, whenever play is in state s, player 1's expected payoff is equal to $3 > 2 = \alpha_s$, and whenever play is in state t, player 2 receives $1 = \alpha_t$. Hence, $\hat{g}$ is viable with respect to α. ∎

Consider two states $t, u \in S$, and a real vector $\alpha = (\alpha_t)_{t \in S}$. We define

$$\beta_t(u, \alpha) = \begin{cases} \inf\limits_{v \in V_u(\alpha)} \phi^{i_t}(v) & \text{when } V_u(\alpha) \neq \emptyset \\ \infty & \text{otherwise,} \end{cases}$$

which is the highest lower-bound for the payoff to player i_t with regard to viable game-plans starting in state u. Note that we cannot guarantee that the infimum is attained (cf. the concluding remarks).

Now, we can also define such a bound for an action $a \in A_t$ of player i_t in state t as follows:

$$\gamma_t(a, \alpha) = \begin{cases} \sum\limits_{u \in S} p_t(a, u)\beta_t(u, \alpha) & \text{if action } a \text{ is non-quitting} \\ r_t^{i_t} & \text{if action } a \text{ is quitting.} \end{cases}$$

Let further

$$\delta_t(\alpha) = \max_{a \in A_t} \gamma_t(a, \alpha).$$

The main motivation for these definitions is the following. Let $\varepsilon \geq 0$. It is easily proved (cf. Lemma 3) that $\delta_t(\alpha) \geq \alpha_t$ for all $t \in S$. Suppose that also $\alpha_t \geq r_t^{i_t}$ and $\delta_t(\alpha) \leq \alpha_t + \frac{1}{2}\varepsilon$ for some state t and some vector $\alpha \in \mathbb{R}^{|S|}$. Consider the situation that play is in state t, according to some game-plan v which is viable with respect to α. If the players follow the prescriptions of v, then due to viability, player i_t will receive a payoff of at least α_t in expectation. Suppose, however, that player i_t deviates by playing a non-prescribed action a. If a is quitting then player i_t receives payoff $r_t^{i_t} \leq \alpha_t$, which is no improvement. Otherwise, if a is non-quitting, then notice the following. As $\delta_t(\alpha) \leq \alpha_t + \frac{1}{2}\varepsilon$, it is possible to switch to a new game-plan $v_u \in V_u(\alpha)$ from the next state u such that

$$\sum_{u \in S} p_t(a, u)\phi^{i_t}(v_u) \leq \sum_{u \in S} p_t(a, u)\beta_t(u, \alpha) + \frac{1}{2}\varepsilon = \gamma_t(a, \alpha) + \frac{1}{2}\varepsilon \leq \delta_t(\alpha) + \frac{1}{2}\varepsilon \leq \alpha_t + \varepsilon.$$

This means that, after any deviation, player i_t's expected payoff, with regard to the new game-plan, increases by at most ε. This idea will later be used when deriving the existence of subgame-perfect ε-equilibria.

We finally define $B_t(\alpha)$ as the set of those actions a for player i_t in state t for which

$$\gamma_t(a, \alpha) = \delta_t(\alpha).$$

By the definition of $\delta_t(\alpha)$, the set $B_t(\alpha)$ is always nonempty. Observe that if $\delta_t(\alpha) = \infty$, then $B_t(\alpha)$ consists of those non-quitting actions a in state t for which there is a state $u \in S_t(a)$ with $V_u(\alpha) = \emptyset$.

The following lemma provides useful properties of these functions. Its proof follows straightforwardly from the definitions.

Lemma 2 *If $\alpha \geq \overline{\alpha}$, then we have for all states $t \in S$ that*

1. $V_t(\alpha) \subseteq V_t(\overline{\alpha})$;

2. $\beta_t(u, \alpha) \geq \beta_t(u, \overline{\alpha})$ *for all* $u \in S$;

3. $\gamma_t(a, \alpha) \geq \gamma_t(a, \overline{\alpha})$ *for all* $a \in A_t$;

4. $\delta_t(\alpha) \geq \delta_t(\overline{\alpha})$.

5 Subgame-perfect ε-equilibria in pure strategies in $\mathcal{G}'$

In this section, we prove the existence of subgame-perfect ε-equilibria, in pure strategies, for all games in the class $\mathcal{G}'$. So, consider an arbitrary game $G \in \mathcal{G}'$. We start with an iterative scheme, which will appear to converge to a finite limit. We define a sequence $\alpha^0, \alpha^1, \ldots$ of vectors, where $\alpha^k = (\alpha_t^k)_{t \in S} \in \mathbb{R}^{|S|}$, as follows. For $k = 0$, we set $\alpha_t^0 = r_t^{i_t}$ for all states $t \in S$. Given α^k, we define $\alpha_t^{k+1} = \delta_t(\alpha^k)$ for all states $t \in S$.

Lemma 3 *For all $k \geq 0$ it holds that $\alpha^{k+1} \geq \alpha^k$.*

Proof. The proof is by complete induction. Let a_t denote the quitting action in state t. For $k = 0$ we have for any state $t \in S$ that

$$\alpha_t^1 = \delta_t(\alpha^0) \geq \gamma_t(a_t, \alpha^0) = r_t^{i_t} = \alpha_t^0.$$

Assume now that $\alpha^k \geq \alpha^{k-1}$ holds for some $k \geq 1$. Then, by lemma 2-4, we obtain for any state $t \in S$ that

$$\alpha_t^{k+1} = \delta_t(\alpha^k) \geq \delta_t(\alpha^{k-1}) = \alpha_t^k,$$

which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Observe the following. For any state $t \in S$, at any iteration level k, either $\alpha_t^k = \infty$ or α_t^k is bounded from above by the maximal payoff of the game. Since the sequence α_t^k

9

$(k = 0, 1, \ldots)$ is nondecreasing, α_t^k will either converge to a finite limit, or, after a finite number of iterations, $\alpha_t^k = \infty$. We will call this limit α_t^*, and set $\alpha^* = (\alpha_t^*)_{t \in S}$.

Example 3 Consider the following game G. There are 4 players, so $N = \{1, 2, 3, 4\}$. The state space is $S = \{s_1, s_2, s_3, s_4\}$, with player i controlling state s_i. In state s_i, player i's action space is $A_{s_i} = \{a_1^i, a_2^i, a_3^i, a_4^i\}$, where action a_j^i, with $j \neq i$, leads to state s_j with probability 1 whereas action a_i^i is quitting. The payoffs for quitting are:

$$r_{s_1} = (1, 1, 3, 1), \ r_{s_2} = (1, 1, 1, 1), \ r_{s_3} = (1, 2, 2, 1), \ r_{s_4} = (2, 1, 1, 2).$$

Obviously, $G \in \mathcal{G}'$, and all transitions in G are deterministic. We are going to perform the iteration scheme defined above, and we will see that it allows us to construct a pure subgame-perfect 0-equilibrium. For simplicity, a quitting game-plan for initial state $t_1 \in S$ will be denoted by $(t_1, t_2, \ldots, t_m, *)$, where $t_l \neq t_{l+1}$ for all $l = 1, \ldots, m-1$, with the interpretation that in state t_1 action $a_{t_2}^{t_1}$ should be played, leading to state t_2, then in state t_2 action $a_{t_3}^{t_2}$ should be played, leading to state t_3, and so on until quitting should take place in state t_m.

Step 0: Initially, $\alpha_{s_1}^0 = \alpha_{s_2}^0 = 1$ and $\alpha_{s_3}^0 = \alpha_{s_4}^0 = 2$. Notice that, since player 4 receives less than $\alpha_{s_4}^0 = 2$ in all states except his own state s_4, all game-plans which are viable with regard to α^0 and which start in state s_4, will eventually quit in state s_4.

Regarding player 1: Due to the above observation, we obtain $\beta_{s_1}(s_4, \alpha^0) = 2$ and $a_4^1 \in B_{s_1}(\alpha^0)$. Hence, $\alpha_{s_1}^1 = 2$.

Regarding player 2: We have $\beta_{s_2}(s_1, \alpha^0) = 1$ as the game-plan $(s_1, s_4, *)$ (or simply quitting in state s_1 immediately) is viable with regard to α^0. Also, $\beta_{s_2}(s_3, \alpha^0) = 1$, as $(s_3, s_1, *)$ is viable with regard to α^0. Clearly, $\beta_{s_2}(s_4, \alpha^0) = 1$. Thus, $\beta_{s_2}(t, \alpha^0) = 1$ for all states $t \in S$, yielding $\alpha_{s_2}^1 = 1$.

Regarding player 3: Since $(s_1, s_2, *)$ is viable with respect to α^0, it follows that $\alpha_{s_3}^1 = 2$.

Regarding player 4: Of course, $\alpha_{s_4}^1 = 2$.

Step 1: We obtained $\alpha_{s_1}^1 = \alpha_{s_3}^1 = \alpha_{s_4}^1 = 2$ and $\alpha_{s_2}^1 = 1$. The main difference is that quitting in state s_1 is no longer viable for player 1. This means that $(s_3, s_1, *)$ is not viable with regard to α^1. Hence, $\beta_{s_2}(s_3, \alpha^1) = 2$ yielding $\alpha_{s_2}^2 = 2$. Thus, $\alpha_t^2 = 2$ for all states $t \in S$.

Step 2: We obtained $\alpha_t^2 = 2$ for all states $t \in S$. It is easy to see that $\alpha^k = \alpha^2$ for all $k > 2$, implying $\alpha_t^* = 2$ for all states $t \in S$.

Consider the following game-plans, which are all viable with respect to α^*:

$$g_{s_1} = (s_1, s_4, *), \ g_{s_2} = (s_2, s_3, *), \ g_{s_3} = (s_3, *), \ g_{s_4} = (s_4, *).$$

Now we obtain a pure subgame-perfect 0-equilibrium as follows. From any initial state $s \in S$, game-plan g_s should be played. If any player along the way deviates from g_s to a non-quitting action, by which play moves to state t, then game-plan g_t should be played from state t. And similarly, if a deviation occurs from g_t to a non-quitting action leading to state u, then game-plan g_u should be played from state u, and so on. It is easy to check that this prescription provides a subgame-perfect 0-equilibrium. ∎

The following lemma considers concatenations of an action with viable game-plans.

Lemma 4 *Suppose $V_t(\alpha^k) \neq \emptyset$ and let $v_t \in V_t(\alpha^k)$ for all $t \in S$. Choose an initial state $s \in S$, and an action $a \in A_s$. Construct a game-plan v for initial state s as follows.*

(1) At stage 1, player i_s plays action a;

(2) If action a is non-quitting and play reaches state $t \in S_s(a)$, then from stage 2 onwards, game-plan v_t will be played (with forgetting the history before stage 2 and considering state t as the initial state).

If either $a \in B_s(\alpha^{k-1})$ or $a \in B_s(\alpha^k)$, then $v \in V_s(\alpha^k)$.

Proof. First assume that a is the quitting action. If $a \in B_s(\alpha^{k-1})$, then $\delta_s(\alpha^{k-1}) = \gamma_s(a, \alpha^{k-1})$, hence we obtain

$$\alpha_s^k = \delta_s(\alpha^{k-1}) = \gamma_s(a, \alpha^{k-1}) = r_s^{i_s}.$$

If $a \in B_s(\alpha^k)$, then $\delta_s(\alpha^k) = \gamma_s(a, \alpha^k)$, hence

$$\alpha_s^k \leq \alpha_s^{k+1} = \delta_s(\alpha^k) = \gamma_s(a, \alpha^k) = r_s^{i_s},$$

where the inequality follows from lemma 3. In either case, quitting is viable with respect to α^k.

Now assume that a is non-quitting. Since $v_t \in V_t(\alpha^k)$ for every state $t \in S_s(a)$, we conclude that the game-plan v is quitting and that $\phi^{i_u}(v[h]) \geq \alpha_u^k$ holds for any $u \in S$ and any history $h \in H_u^v$ that reaches stage 2. It remains to show that $\phi^{i_s}(v) \geq \alpha_s^k$. As

$$\phi^{i_s}(v) = \sum_{t \in S_s(a)} p_s(a, t) \phi^{i_s}(v_t) \geq \sum_{t \in S_s(a)} p_s(a, t) \beta_s(t, \alpha^k) = \gamma_s(a, \alpha^k),$$

it suffices to verify $\gamma_s(a, \alpha^k) \geq \alpha_s^k$. We distinguish two cases: $a \in B_s(\alpha^{k-1})$ or $a \in B_s(\alpha^k)$. If $a \in B_s(\alpha^{k-1})$, we obtain

$$\alpha_s^k = \delta_s(\alpha^{k-1}) = \gamma_s(a, \alpha^{k-1}) \leq \gamma_s(a, \alpha^k),$$

where the inequality follows from lemmas 2-3 and 3. On the other hand, if $a \in B_s(\alpha^k)$, we have

$$\alpha_s^k \leq \alpha_s^{k+1} = \delta_s(\alpha^k) = \gamma_s(a, \alpha^k),$$

where the inequality follows from lemma 3. Hence, $\phi^{i_s}(v) \geq \alpha_s^k$, as desired. $\qquad \square$

The following lemma considers concatenations of a truncated game-plan with viable game-plans.

Lemma 5 *Suppose $V_t(\alpha^k) \neq \emptyset$ and let $v_t \in V_t(\alpha^k)$ for all $t \in S$. Choose an initial state $s \in S$, and let g be a truncated game-plan for s that only uses actions from the sets $B_u(\alpha^{k-1})$ and $B_u(\alpha^k)$ for every state $u \in S$, and that expires with probability 1. Let v denote the complete game-plan for initial state s, according to which*

(1) from stage 1 onwards, game-plan g is executed, and

11

(2) when g expires, say in state t, game-plan v_t will be played (with forgetting the history induced by g and considering state t as the initial state).

Then $v \in V_s(\alpha^k)$.

Proof. It is clear that v is a quitting game-plan. We need to prove that v is viable with respect to α^k. For any $m \in \mathbb{N}$, consider the following game-plan v_m, by adapting v: if, at stage m, the game-plan g has not expired yet, and play is in some state t, then start game-plan v_t. Note that v_m is viable with respect to α^k for all m, which follows by repeated application of lemma 4. Let $\varepsilon > 0$. Since g expires with probability 1, we can choose m so large that the probability that g expires before stage m is so close to 1 that

$$\phi^{i_s}(v_m) \le \phi^{i_s}(v) + \varepsilon.$$

Since $v_m \in V_s(\alpha^k)$, we have

$$\phi^{i_s}(v_m) \ge \alpha_s^k.$$

Therefore,

$$\phi^{i_s}(v) \ge \phi^{i_s}(v_m) - \varepsilon \ge \alpha_s^k - \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we proved $\phi^{i_s}(v) \ge \alpha_s^k$.

Observe that, for any $t \in S$ and any history $h \in H_t^v$, the above reasoning can also be given for $v[h]$ to show that $\phi^{i_t}([h]) \ge \alpha_t^k$. Hence, $v \in V_s(\alpha^k)$. $\square$

Let $k \in \mathbb{N}$. For two states s and t, we write $s \preceq^k t$ if state t can eventually be visited with a positive probability when starting in s, by only using actions in the sets $B_u(\alpha^k)$, $u \in S$. This relation $\preceq^k$ is obviously transitive. With respect to $\preceq^k$, a nonempty set $Q \subseteq S$ is called closed, if for every $s \in Q$ there is no $t \in S - Q$ such that $s \preceq^k t$. A closed set $Q \subseteq S$ is called minimal closed if Q contains no proper subset which is closed. We will call every minimal closed set a <u>persistent set</u> and its elements persistent states with respect to α^k. It is clear that there always exists a persistent set. Thus, we have the following properties for persistent states: (i) from any non-persistent state, we can eventually reach the set of persistent states with probability 1, by only using actions in the sets $B_u(\alpha^k)$, $u \in S$, (ii) a persistent set P^k cannot be left through actions in the sets $B_u(\alpha^k)$, $u \in P^k$, and (iii) if s and $t \ne s$ belong to the same persistent set P^k, then t can eventually be visited when starting in s with probability 1, by only using actions in the sets $B_u(\alpha^k)$, $u \in P^k$.

The next lemma derives an essential property of persistent states.

Lemma 6 *Suppose that $V_u(\alpha^k) \ne \emptyset$ for all $u \in S$. Take a persistent state $s \in S$ with respect to α^k. Then*

  1. *If $a \in B_s(\alpha^{k-1})$ (which implies $k > 0$), then $a \in B_s(\alpha^k)$;*

  2. *If a is quitting and $k = 0$, then $a \in B_s(\alpha^0)$.*

Proof. Let $a \in B_s(\alpha^{k-1})$ if $k > 0$ or let a be quitting if $k = 0$. We will show that $a \in B_s(\alpha^k)$. Let P^k denote the persistent set with respect to α^k that state s belongs to. For every state $t \in P^k$, we define a game-plan v_t, starting in state t, according to which:

(1) From stage 1 onwards, starting in state t, a truncated game-plan will be used to eventually visit state s, by only using actions in the sets $B_u(\alpha^k)$, $u \in P^k$. Such a game-plan exists by property (iii) of persistent states. (If $t = s$, then this truncated game-plan is empty.) Notice that the game-plan expires in state s with probability 1.

(2) When s is reached, say at stage $m \geq 1$, action a will be played.

(3) If a is non-quitting, choose viable game-plans $w_u \in V_u(\alpha^k)$ for all $u \in S_s(a)$. From stage $m+1$ onwards, if $u' \in S_s(a)$ denotes the state at stage $m+1$, the viable game-plan $w_{u'}$ will be played (with forgetting the history before stage $m+1$ and considering state u' as the initial state).

We have $v_t \in V_t(\alpha^k)$ for all $t \in P^k$. If $k > 0$, then $a \in B_s(\alpha^{k-1})$, and viability follows by lemma 5. If $k = 0$, then a is quitting, and as quitting is always viable with respect to α^0, viability of v_t follows by lemma 5.

We claim that

$$\beta_s(t, \alpha^k) \leq \gamma_s(a, \alpha^k)$$

for all $t \in P^k$. As $v_t \in V_t(\alpha^k)$, for all $t \in P^k$, we have $\beta_s(t, \alpha^k) \leq \phi^{i_s}(v_t)$. If a is quitting, then each game-plan v_t terminates at s with payoff $r_s^{i_s}$ for i_s, in which case the claim follows from

$$\beta_s(t, \alpha^k) \leq \phi^{i_s}(v_t) = r_s^{i_s} = \gamma_s(a, \alpha^k).$$

If a is non-quitting, then each game-plan v_t induces the payoff

$$\phi^{i_s}(v_t) = \sum_{u \in S_s(a)} p_s(a, u) \phi^{i_s}(w_u).$$

The claim now follows from

$$\beta_s(t, \alpha^k) \leq \sum_{u \in S_s(a)} p_s(a, u) \left[\inf_{w_u \in V_u(\alpha^k)} \phi^{i_s}(w_u) \right] = \sum_{u \in S_s(a)} p_s(a, u) \beta_s(u, \alpha^k) = \gamma_s(a, \alpha^k),$$

since the choice of w_u was arbitrary in $V_u(\alpha^k)$ for all $u \in S_s(a)$.

Now take $b \in B_s(\alpha^k)$. We claim that

$$\gamma_s(b, \alpha^k) \leq \gamma_s(a, \alpha^k).$$

If b is quitting, this claim follows from $\gamma_s(b, \alpha^k) = r_s^{i_s} \leq \gamma_s(a, \alpha^k)$. If b is non-quitting, then, due to $s \in P^k$ and $b \in B_s(\alpha^k)$, we have $S_s(b) \subseteq P^k$. Therefore, the claim follows from

$$\gamma_s(b, \alpha^k) = \sum_{t \in S_s(b)} p_s(b, t) \beta_s(t, \alpha^k) \leq \sum_{t \in S_s(b)} p_s(b, t) \gamma_s(a, \alpha^k) = \gamma_s(a, \alpha^k).$$

Since $b \in B_s(\alpha^k)$, we may derive

$$\delta_s(\alpha^k) = \gamma_s(b, \alpha^k) \leq \gamma_s(a, \alpha^k),$$

which implies $a \in B_s(\alpha^k)$, as desired. $\qquad\square$

Lemma 7 *If t is a quitting state with respect to α^k and if t is persistent with respect to α^k, then t is a quitting state with respect to α^{k+1}.*

Proof. Let t be a persistent quitting state with respect to α^k and let a denote the quitting action. If $k = 0$, then $a \in B_t(\alpha^k)$ by lemma 6. If $k > 0$, then $a \in B_t(\alpha^{k-1})$, since $\delta_t(\alpha^{k-1}) = \alpha_t^k = r_t^{it} = \gamma_t(a, \alpha^{k-1})$. So if $k > 0$, $a \in B_t(\alpha^k)$ also follows by lemma 6. Consequently, $\alpha_t^{k+1} = \delta_t(\alpha^k) = \gamma_t(a, \alpha^k) = r_t^{it}$. $\qquad\square$

We are now ready to prove that the iterative scheme converges to a finite limit.

Lemma 8

1. $V_s(\alpha^k) \neq \emptyset$ for all $k \geq 0$ and all $s \in S$,

2. Every persistent set with respect to α^k contains quitting states with respect to α^k.

3. The limit α^* of the iterative process α^k is finite.

Proof. We prove 1 and 2 by induction on k. Notice that both 1 and 2 are trivially true for $k = 0$. Now assume 1 and 2 are true for some $k \geq 0$.

To prove claim 1 for $k + 1$, notice that a truncated game-plan for $s \in S$ exists that only uses actions in the sets $B_u(\alpha^k)$ ($u \in S$), and that expires with probability 1 in a persistent set with respect to α^k. This follows by property (i) of persistent states. Moreover, the truncated game-plan can be extended to a truncated game-plan that expires with probability 1 at a quitting state with respect to α^k, only by actions in the sets $B_u(\alpha^k)$, by property (iii) of persistent sets and by the assumption that 2 holds for k. We then complete the game-plan, which we denote by w_s, by choosing the quitting action when a quitting state with respect to α^k is reached. By lemma 7, the state is also quitting with respect to α^{k+1}, and we may apply lemma 5, to derive that $w_s \in V_s(\alpha^{k+1})$.

To prove 2 for $k + 1$, let P^{k+1} be a persistent set with respect to α^{k+1}, choose $s \in P^{k+1}$ arbitrarily, and construct the game-plan $w_s \in V_s(\alpha^{k+1})$ as above. We claim that the game-plan w_s only visits states in P^{k+1}. This is trivial for state s visited at stage 1. Assume it is true for a state t visited at stage m. By construction of w_s, the action at state t is chosen from $B_t(\alpha^k)$. By lemma 6, this action is also in $B_t(\alpha^{k+1})$. Therefore, the state visited at stage $m + 1$ is again in P^{k+1}, by property (ii) of persistent sets. This demonstrates that w_s terminates with probability 1 in P^{k+1}. Since $w_s \in V_s(\alpha^{k+1})$, it follows that P^{k+1} contains quitting states with respect to α^{k+1}.

To prove 3, notice that due to 1, each α_s^k is finite for all $s \in S$ and $k \in \mathbb{N}$. Moreover, as each α_s^k is a convex combination of payoffs in the game, it is bounded from above by the maximal payoff. Since the sequence $(\alpha^k)_{k \in \mathbb{N}}$ is non-decreasing, it converges to a finite limit. $\qquad\square$

Now we are ready to show Main Theorem 1 restricted to games in $\mathcal{G}'$.

Theorem 9 *In every stochastic game G in class $\mathcal{G}'$, there exists a subgame-perfect ε-equilibrium in pure strategies, for every $\varepsilon > 0$. Moreover, if all transitions in G are deterministic, then G has a subgame-perfect 0-equilibrium in pure strategies.*

Proof. <u>General transitions:</u> We start by showing the case of general transitions. Take a stochastic game G in $\mathcal{G}'$. We assume that all payoffs at quitting are at least 1; otherwise we can raise all payoffs at quitting by 1 (for any $\varepsilon \geq 0$, any subgame-perfect ε-equilibrium in this modified game is also a subgame-perfect ε-equilibrium in the original game). Take an initial state $s \in S$. Let $\varepsilon > 0$ and let $k \in \mathbb{N}$ be so large that

$$\left\| \alpha^* - \alpha^k \right\| \leq \frac{\varepsilon}{4|S|},$$

where the norm is the maximum-norm, and where $|S|$ equals the number of states.

We will now define a joint pure strategy π^ε, and show that π^ε is a subgame-perfect ε-equilibrium for initial state s.

<u>Step 1:</u> Definition of π^ε. We will define the joint pure strategy π^ε inductively. Let $s^1 = s$ and take an arbitrary game-plan $v^1 \in V_{s^1}(\alpha^k)$. When starting in state s^1, the joint strategy π^ε prescribes to play according to the game-plan v^1, as long as all players follow the prescriptions of v^1. If, on the other hand, in some state, the controlling player ignores the prescription by v^1 and deviates to a non-quitting action, by which play moves to some state s^2, then π^ε prescribes to switch to a certain new game-plan $v^2 \in V_{s^2}(\alpha^{k+1})$. This game-plan v^2, to be specified later, will be used as long as all players follow the prescriptions by v^2. And similarly, if deviation occurs at some point to a non-quitting action, then π^ε prescribes another new game-plan $v^3 \in V_{s^3}(\alpha^{k+2})$ from the state s^3 right after the deviation, and so on. Thus, with respect to π^ε, a game-plan is active at any point during play.

We will now describe the choice of these game-plans after a deviation takes place. Suppose the players are expected to use game-plan v^m, but in state t, player i_t deviates to non-quitting action a. Let s^{m+1} denote the state to which transition occurs through action a. Then, the new game-plan v^{m+1} is chosen such that $v^{m+1} \in V_{s^{m+1}}(\alpha^{k+m})$ and the expected reward satisfies

$$\phi^{i_t}(v^{m+1}) \leq \beta_t(s^{m+1}, \alpha^{k+m}) + \frac{\varepsilon}{2^{m+1}}.$$

Such a game-plan exists by lemma 8 and by the definition of the function β.

<u>Step 2:</u> π^ε is a subgame-perfect ε-equilibrium for initial state s. In order to prove this, we will show that π^ε is an ε-equilibrium for initial state s. Since the structure of any continuation strategy $\pi^\varepsilon[h]$ is almost identical to that of π^ε (the only difference is that $\pi^\varepsilon[h]$ starts with a continuation game-plan of v^m, for some m), a similar proof can be given that $\pi^\varepsilon[h]$ is an ε-equilibrium in the subgame after an arbitrary history h.

Take a player i, and a pure strategy $\widetilde{\sigma}^i$ for player i. We will show that player i cannot improve his expected payoff by more than ε if he deviates from the strategy $\pi^{\varepsilon,i}$ to $\widetilde{\sigma}^i$, i.e.

$$\phi_s^i(\widetilde{\sigma}^i, \pi^{\varepsilon,-i}) \leq \phi_s^i(\pi^{\varepsilon,i}, \pi^{\varepsilon,-i}) + \varepsilon. \tag{1}$$

Note that we only consider pure deviations for player i, but one can show the same along similar lines for arbitrary deviations as well.

Let σ^i be the strategy for player i which follows the prescriptions of $\widetilde{\sigma}^i$ until, during play, a history h occurs, with a final state u controlled by player i, such that either

(1) the probability that $(\widetilde{\sigma}^i[h], \pi^{\varepsilon,-i}[h])$ ever prescribes quitting is less than $1/\bar{r}$, where $\bar{r}$ is the maximal payoff in the game, or

(2) $\widetilde{\sigma}^i[h]$ prescribes to quit at u.

In both cases, $\sigma^i[h]$ tells player i to play according to $\pi^{i,\varepsilon}[h]$. Notice that the expected payoff for player i with regard to $(\widetilde{\sigma}^i[h], \pi^{\varepsilon,-i}[h])$ is at most his expected payoff with regard to $(\sigma^i[h], \pi^{\varepsilon,-i}[h])$. In case $\widetilde{\sigma}^i[h]$ prescribes quitting, this follows from the fact that $\pi^\varepsilon[h]$ is viable with respect to α^*, and since $\alpha_u^* \geq r_u^i$. In case the probability on quitting is at most $1/\bar{r}$, then the expected payoff for player i is at most 1, while his payoff is at least 1 if he follows $\sigma^i[h]$. (Recall our assumption that all quitting payoffs are at least 1.) Notice that whenever $(\sigma^i, \pi^{\varepsilon,-i})$ deviates from π^ε, it is by a non-quitting action. Moreover, $(\sigma^i, \pi^{\varepsilon,-i})$ leads to quitting eventually with probability 1.

Since

$$\phi^i(\sigma^i, \pi^{\varepsilon,-i}) \geq \phi^i(\widetilde{\sigma}^i, \pi^{\varepsilon,-i}),$$

it suffices to show

$$\phi^i(\sigma^i, \pi^{\varepsilon,-i}) \leq \phi^i(\pi^{\varepsilon,i}, \pi^{\varepsilon,-i}) + \varepsilon. \tag{2}$$

to prove (1). For any $m \in \mathbb{N} \cup \{0\}$, let σ_m^i be the modification of σ^i which does not deviate from $\pi^{\varepsilon,i}$ any more if game-plan v^{m+1} becomes active. This means that σ_m^i deviates at most m times. Note that $\sigma_0^i = \pi^{\varepsilon,i}$. Let d_m denote the expected payoff for player i with respect to $(\sigma_m^i, \pi^{\varepsilon,-i})$ and initial state s, i.e.

$$d_m = \phi^i(\sigma_m^i, \pi^{\varepsilon,-i}).$$

Since $(\sigma^i, \pi^{\varepsilon,-i})$ from initial state s leads to quitting eventually, with probability 1, we must have

$$\phi^i(\sigma^i, \pi^{\varepsilon,-i}) = \lim_{m \to \infty} \phi^i(\sigma_m^i, \pi^{\varepsilon,-i}) = \lim_{m \to \infty} d_m.$$

Let $H(1) \subseteq H^{v^1}$ denote the set of histories h, such that (i) player i controls the final state of h, say u, (ii) $\sigma^i(h)$ prescribes to deviate by playing some action $a \in A_u$. For $h \in H(1)$, let $\tau(1, h)$ denote the event that h occurs. The construction of σ^i guarantees that a is non-quitting, hence some game-plan v^2 will be chosen after the deviation. With $\mathbb{E}$ denoting the expectation with respect to $(\sigma^i, \pi^{\varepsilon,-i})$, we have by the choice of v^2

$$
\begin{aligned}
\mathbb{E}\left(\phi^i(v^2)|\tau(1,h)\right) &= \sum_{s^2 \in S_u(a)} p_u(a, s^2)\beta_u(s^2, \alpha^{k+1}) + \tfrac{\varepsilon}{4} \\
&= \gamma_u(a, \alpha^{k+1}) + \tfrac{\varepsilon}{4} \\
&\leq \delta_u(\alpha^{k+1}) + \tfrac{\varepsilon}{4} \\
&= \alpha_u^{k+2} + \tfrac{\varepsilon}{4} \\
&\leq \alpha_u^k + (\alpha_u^{k+2} - \alpha_u^k) + \tfrac{\varepsilon}{4} \\
&\leq \phi^i(v^1[h]) + \left\|\alpha^{k+2} - \alpha^k\right\| + \tfrac{\varepsilon}{4},
\end{aligned}
$$

where the last inequality follows from the viability of v^1 with respect to α^k. Thus, with $\mathbb{P}$ denoting the probability of an event with respect to $(\sigma^i, \pi^{\varepsilon,-i})$, we obtain

$$d_1 - d_0 = \sum_{h \in H(1)} \mathbb{P}(\tau(1,h)) \cdot \left[\mathbb{E}\left(\phi^i(v^2)|\tau(1,h)\right) - \phi^i(v^1[h])\right] \leq \left\|\alpha^{k+2} - \alpha^k\right\| + \frac{\varepsilon}{4}.$$

16

In a similar fashion,

$$d_2 - d_1 \le \left\| \alpha^{k+3} - \alpha^{k+1} \right\| + \frac{\varepsilon}{8},$$

and in general

$$d_{m+1} - d_m \le \left\| \alpha^{k+m+2} - \alpha^{k+m} \right\| + \frac{\varepsilon}{2^{m+2}}.$$

Hence

$$d_m - d_0 \le \sum_{l=k}^{k+m-1} \left\| \alpha^{l+2} - \alpha^l \right\| + \sum_{l=2}^{m+1} \frac{\varepsilon}{2^l} \le \sum_{l=k}^{\infty} \left\| \alpha^{l+2} - \alpha^l \right\| + \frac{\varepsilon}{2}$$

The choice of k implies

$$
\begin{aligned}
\sum_{l=k}^{\infty} \left\| \alpha^{l+2} - \alpha^l \right\| &\le \sum_{l=k}^{\infty} \sum_{u \in S} (\alpha_u^{l+2} - \alpha_u^l) \\
&= \sum_{u \in S} \sum_{l=k}^{\infty} (\alpha_u^{l+2} - \alpha_u^l) \\
&= \sum_{u \in S} \left[(\alpha_u^* - \alpha_u^k) + (\alpha_u^* - \alpha_u^{k+1}) \right] \le \frac{\varepsilon}{2}.
\end{aligned}
$$

Then

$$\phi_s^i(\sigma^i, \pi^{\varepsilon,-i}) = \lim_{m \to \infty} d_m \le d_0 + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \phi_s^i(\pi^\varepsilon) + \varepsilon,$$

which completes the proof of (2).

<u>Deterministic transitions:</u> Now suppose, additionally, that all transitions in the game G are deterministic. Then every quitting game-plan induces one specific history, which ends when a specific player quits. Thus, quitting game-plans can only induce finitely many different payoffs. As a consequence, in the definition of $\beta_s(t, \alpha^*)$, the infimum is attained for all $s, t \in S$. Moreover, $\delta_t(\alpha^*) = \alpha_t^*$ for all $t \in S$, since $\alpha_t^* = \alpha_t^{k+1} = \delta(\alpha^k) = \delta(\alpha^*)$ for k sufficiently large. For these reasons, the proof for the general case can be applied with $\varepsilon = 0$. $\qquad\square$

Concluding Remarks. In all examples we have analyzed, we found a pure subgame-perfect 0-equilibrium. Whether or not this holds in general, is unclear. Nevertheless, in every game for which infimum is attained in the definition of the function β, and for which the sequence α^k converges in a finite nuber of steps, the existence of a pure subgame-perfect 0-equilibrium follows, just as in the case of deterministic transitions.

One can prove that $V_t(\alpha^*) \ne \emptyset$ for all $t \in S$. The reason is that the constructed game-plans in the proof of lemma 8 are all stationary. Thus, for a given state t, one can choose stationary game-plans $v_t^k \in V_t(\alpha^k)$ for all $k \ge 0$. Since there are only finitely many stationary game-plans, it follows that $(v_t^k)_{k \in \mathbb{N}}$ contains a constant subsequence, and it is straightforward to prove that this constant is a stationary game-plan in $V_t(\alpha^*)$. We did not prove in general that $\delta_t(\alpha^*) = \alpha_t^*$ for all $t \in S$, so it remains open whether the vector equation $\delta(\alpha) = \alpha$ has a solution (where $\delta(\alpha) = (\delta_t(\alpha))_{t \in S}$). Although the existence of such a fixed point would simplify the proof of theorem 9, it would not resolve the issue of subgame-perfect 0-equilibria.

In the case of deterministic transitions, a polynomial time algorithm exists to determine the vector α^*. Also, the game-plans in a subgame-perfect equilibrium can be determined during play when needed, in polynomial time. To see this, note that for every $t \in S$, the number α_t^k can only have $|S|$ different values. Since every such number is non-increasing, it follows that the vector α^k can change at most $(|S|-1)|S|$ times before $\alpha^{k+1} = \alpha^k$. Hence,

the calculation of α^* requires at most $(|S|-1)|S|$ iterations. As one iteration requires the calculation of $|S|^2$ numbers $\beta_s(t,\alpha^k)$, it suffices to show that $\beta_s(t,\alpha)$ can be calculated in polynomial time. Since transitions are deterministic, we have

$$\beta_s(t,\alpha) = \min\{r_u^{i_s} \mid v \in V_t(\alpha) \text{ exists that terminates in } u\}.$$

Hence, the calculation of $\beta_s(t,\alpha)$ can be done by a check for every $u \in S$ whether a viable game-plan for t with respect to α exists that terminates in u. To do the check, construct the digraph with vertex-set

$$V = \{u' \in S \mid \alpha_{u'} \le r_u^{i_{u'}}\}$$

and arc-set

$$A = \{(u',u'') \mid \text{an action } a \in A_{u'} \text{ exists such that } p_{u'}(a,u'') = 1\}.$$

Now observe that a viable game-plan for t with respect to α terminating in u exists if and only if the digraph (V,A) has a directed path from t to u. Since the construction of each digraph and the detection of a directed path can be done in polynomial time, it follows that the calculation of $\beta_s(t,\alpha)$ requires polynomial time.

References

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

[2] J. Flesch, F. Thuijsman, and O. J Vrieze, *Stochastic games with additive transitions*, European Journal of Operational Research, **179** (2007), 483–497.

[3] J. Flesch, G. Schoenmakers, and O. J. Vrieze, *Stochastic games on a product state space*, Mathematics of Operations Research, **33** (2008), 403–420.

[4] J. Flesch, G. Schoenmakers, and O. J. Vrieze, *Stochastic games on a product state space: The periodic case*, Research memorandum RM/08/016, University of Maastricht, the Netherlands, 2008.

[5] D. Gillette, *Stochastic games with zero stop probabilities*, Contributions to the theory of games III (M. Dresher, A. W. Tucker, and P. Wolfe, eds.), Princeton University Press, 1957, 179-187.

[6] J. Kuipers, J. Flesch J, G. Schoenmakers, and O. J. Vrieze, *Subgame-perfect equilibria in free transition games*, Research memorandum RM/08/027, University of Maastricht, the Netherlands, 2008.

[7] A. Mashiah-Yaakovi, *Periodic stopping games*, Preprint, forthcoming in International Journal of Game Theory.

[8] J. F. Mertens and A. Neyman, *Stochastic games*, International Journal of Game Theory, **10** (1981), 53–66.

[9] E. Shmaya and E. Solan, *Two Player Non Zerosum Stopping Games in Discrete Time*, The Annals of Probability, **32** (2004), 2733–2764.

[10] E. Shmaya, E. Solan, and N. Vieille, *An Application of Ramsey Theorem to Stopping Games*, Games and Economic Behavior, **42** (2003), 300–306.

[11] R. S Simon, *The Structure of Non-zero-sum Stochastic Games*, Advances in Applied Mathematics, **38** (2003), 1–26.

[12] E. Solan, *Three-Player Absorbing Games*, Mathematics of Operations Research, **24** (1999), 669–698.

[13] E. Solan, *Subgame-perfection in quitting games with perfect information*, Mathematics of Operations Research **30** (2005), 51–72.

[14] E. Solan and N. Vieille, *Quitting games*, Mathematics of Operations Research, **26** (2001), 265–285.

[15] E. Solan and N. Vieille, *Deterministic Multi-Player Dynkin Games*, Journal of Mathematical Economics, **39** (2003), 911–929.

[16] S. Sorin, *Asymptotic properties of a non-zerosum game*, International Journal of Game Theory, **15** (1986), 101–107.

[17] F. Thuijsman and T. E. S Raghavan, *Perfect information stochastic games and related classes*, International Journal of Game Theory **26** (1997), 403–408.

[18] N. Vieille, *Equilibrium in 2-person stochastic games I: A reduction*, Israel Journal of Mathematics, **119** (2000), 55–91.

[19] N. Vieille, *Equilibrium in 2-person stochastic games II: The case of recursive games*, Israel Journal of Mathematics, **119** (2000), 93–126.