Noname manuscript No.
(will be inserted by the editor)

Evolutionary Stochastic Games

János Flesch · Thiruvenkatachari Parthasarathy · Frank Thuijsman · Philippe Uyttendaele

Received: date / Accepted: date

Abstract We extend the notion of Evolutionarily Stable Strategies introduced by Maynard Smith & Price [6] in 1973 for models ruled by a single fitness matrix A, to the framework of stochastic games developed by Lloyd Shapley [13] in 1953 where, at discrete stages in time, players play one of finitely many matrix games, while the transitions from one matrix game to the next follow a jointly controlled Markov chain. We show that this extension from a single-state model to a multi-state model can be done on the assumption of having an irreducible transition law. In a similar way we extend the notion of Replicator Dynamics introduced by Taylor & Jonker [16] in 1978 to the multi-state model. These extensions facilitate the analysis of evolutionary interactions that are richer than the ones that can be handled by the original, single-state, evolutionary game model. Several examples are provided.

Keywords: evolutionary games, stochastic games, evolutionarily stable strategy, replicator dynamics.

1 Introduction

In his early 1928 work on Game Theory, John von Neumann [9] showed that all matrix games have a value and both players have optimal strategies. A quarter century later Lloyd Shapley [13] wrote his ancestral paper on the stochastic game model, in which at each of a possibly infinite number of stages two players play one of finitely many different matrix games, where at each stage the transition probabilities to go from one matrix game to the next are determined by the specific matrix game played and the specific actions chosen

Address for TP: Indian Statistical Institute, Chennai Centre, 37/110 Nelson Manikham Road, Chennai 600029, India.
Address for JF, FT and PU: Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
For correspondence on this paper: f.thuijsman@maastrichtuniversity.nl

at that stage. In the same paper Lloyd Shapley shows that, assuming there is a strictly positive probability of stopping for each pair of actions, whenever they are played, these stochastic games have a value and the players have stationary optimal strategies, i.e. strategies for which a player's action choices at each stage only depend on the specific matrix game being played while neither the stage number nor the history of play leading to that state, needs to be taken into account.

Evolutionary game theory, as started by the seminal paper of Maynard Smith & Price [6] in 1973, studies the dynamic development of populations. Here a population consists of randomly interacting individuals of finitely many different types. Interactions between individuals lead to 'fitness payoffs' for these individuals depending on their types (e.g. number of offspring), where these fitness payoffs are given by a single fitness matrix A. Every entry of A gives the fitness payoff to the row player. When taking the Darwinian viewpoint that the fractions of types that are doing better than average increase, while those doing worse than average decrease, we see that the population distribution is changing over time. The most widely studied dynamics is the so called Replicator Dynamics introduced by Taylor & Jonker [16] in 1978, that builds on the assumption that the rate of change of a population fraction of a specific type is proportional to the size of that fraction as well as to the difference between the fitness for individuals of that type and the current population average fitness.

In this paper, we extend the evolutionary game model to achieve an evolutionary stochastic game model. In this evolutionary stochastic game model we consider a population of individuals from different types, where at every stage these individuals are interacting with each other in one of finitely many environments (or circumstances). The transition probabilities between the environments determine the impact of each of these environments on the fitness of the individuals from specific types. Then, like in the single-state model, the fractions of those types that have a higher fitness than the population average fitness will increase, while the fractions of types that are doing less good decrease. In section 2 we give precise definitions and in the subsequent sections we analyze this evolutionary stochastic game model.

2 The Model

In this section we first describe the models of stochastic games and evolutionary games according to their original definitions and next we introduce a model that combines the features of each of these. We remark that some earlier work was done on introducing stochastic dynamics to evolutionary games and we refer to chapters 10 to 12 in Sandholm [12] for a recent survey. In addition we mention Altman et al. [1], who introduce a model where each individual is facing a Markov decision problem, and Pruett-Jones & Heifetz [10], who examine a model where each strategic interaction between two individuals is a stochastic game. However, our model is essentially different because for us

population members are characterized by their behavior in a finite collection of different circumstances and as such our population types correspond to pure stationary strategies in a stochastic game.

2.1 Stochastic Games

A two-person stochastic game Γ, introduced in a slightly different way by Lloyd Shapley [13] in 1953, can be described by a state space $S := \{1, \ldots, z\}$, and a corresponding collection $\{A_1, \ldots, A_z\}$ of matrices, where matrix A_s has size $m_s^1 \times m_s^2$ and, for $i \in I_s := \{1, \ldots, m_s^1\}$ and $j \in J_s := \{1, \ldots, m_s^2\}$, entry (i, j) of A_s consists of payoffs $r^1(s, i, j), r^2(s, i, j) \in \mathbb{R}$ and a probability vector $p(s, i, j) = (p(s'|s, i, j))_{s' \in S}$. The elements of S are called states and for each state $s \in S$ the elements of I_s and J_s are called (pure) actions of player 1 and player 2 respectively in state s. The game is to be played at stages in $\mathbb{N} = \{1, 2, 3, \ldots\}$ in the following way. Play starts at stage 1 in an initial state, say in state $s^1 \in S$, where, simultaneously and independently, both players are to choose an action: player 1 chooses an $i^1 \in I_{s^1}$, while player 2 chooses a $j^1 \in J_{s^1}$. These choices induce immediate payoffs $r^1(s^1, i^1, j^1), r^2(s^1, i^1, j^1)$ to players 1 and 2 respectively. Next, the play moves to a new state according to the probability vector $p(s^1, i^1, j^1)$, say to state s^2. At stage 2 new actions $i^2 \in I_{s^2}$ and $j^2 \in J_{s^2}$ are to be chosen by the players in state s^2. Then the players receive payoffs $r^1(s^2, i^2, j^2), r^2(s^2, i^2, j^2)$ respectively and play moves to some state s^3 according to the probability vector $p(s^2, i^2, j^2)$, and so on. The players are assumed to have complete information and perfect recall. The latter means that at every stage n they know the history of play up to that stage: $h^n = (s^1, i^1, j^1; \ldots; s^{n-1}, i^{n-1}, j^{n-1}, s^n)$.

A mixed action for a player in state s is a probability distribution on the set of his actions in state s. Mixed actions in state s will be denoted by x_s for player 1 and by y_s for player 2, and the sets of mixed actions in state s by X_s and Y_s respectively. A strategy is a decision rule that prescribes a mixed action for any past history of the play. Such general strategies, so-called behavior strategies, will be denoted by π for player 1 and by σ for player 2. We use the notations Π and Σ for the respective behavior strategy spaces of the players. A strategy is called pure if it specifies one pure action for each possible history. We denote the respective pure strategy spaces by Π^p and Σ^p. If for all past histories, the mixed actions prescribed by a strategy only depend on the current state then the strategy is called stationary. Thus the stationary strategy spaces are $X := \times_{s \in S} X_s$ for player 1 and $Y := \times_{s \in S} Y_s$ for player 2 and we write x and y for stationary strategies for players 1 and 2 respectively. For the spaces of pure stationary strategies we will use X^p and Y^p.

The stochastic game is called irreducible if for all pairs of stationary strategies the associated Markov chain on the state space is irreducible, i.e. all states will be visited infinitely often with probability 1.

For an infinite history $h = (s^n, i^n, j^n)_{n \in \mathbb{N}}$, player k will evaluate the sequence

of payoffs by the limiting average reward, defined by

$$\gamma^k(h) := \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} r^k(s^n, i^n, j^n).$$

Another commonly used evaluation is the β-discounted reward, where $\beta \in (0,1)$ is the discount factor, given by

$$\gamma^k_\beta(h) := (1-\beta) \cdot \sum_{n=1}^{\infty} \beta^{n-1} r^k(s^n, i^n, j^n).$$

However, in this paper we will only focus on the limiting average rewards. A pair of strategies (π, σ) together with an initial state $s \in S$, by Kolmogorov's existence theorem (cf. Kolmogorov [5]), determines a probability measure on the set of infinite histories with initial state s. By using this probability measure, for (π, σ) and initial state s, the sequences of payoffs are evaluated by the expected limiting average reward denoted by $\gamma^k(s, \pi, \sigma)$.

A stochastic game in which $r^2(s,i,j) = -r^1(s,i,j)$ for all triples (s,i,j), is called a zero-sum stochastic game. In such a game it is assumed that player 1 wants to maximize his limiting average reward, while player 2 tries to minimize player 1's limiting average reward. A zero-sum stochastic game has a limiting average value $v = (v_s)_{s\in S}$ if

$$\sup_{\pi\in\Pi} \inf_{\sigma\in\Sigma} \gamma^1(s,\pi,\sigma) = \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi} \gamma^1(s,\pi,\sigma) =: v_s \ \ \forall s \in S. \tag{1}$$

Although the seminal paper by Shapley [13] already implied the existence of the β-discounted value, and stationary β-discounted optimal strategies, the general existence of a limiting average value was only established in 1981 by Mertens & Neyman [7]. However, for the limiting average case the players need not have optimal strategies and behavior strategies may be indispensable for achieving ε-optimality. Here a strategy π of player 1 is called ε-optimal, where $\varepsilon \geq 0$, if for all initial states $s \in S$ we have

$$\gamma^1(s,\pi,\sigma) \geq v_s - \varepsilon \ \ \forall \sigma \in \Sigma,$$

and 0-optimal strategies are simply called optimal for player 1. Similar definitions apply for player 2.

For games that are not zero-sum, we use the notion of ε-equilibria, $\varepsilon \geq 0$, which are pairs of strategies that are ε-best replies against each other. For simplicity, if $\varepsilon = 0$ we speak of an equilibrium rather than of a 0-equilibrium. Here a stategy π^ε by player 1 is an ε-best reply against a strategy σ by player 2, when

$$\forall s \in S \ \forall \pi \in \Pi: \ \gamma^1(s,\pi^\varepsilon,\sigma) \geq \gamma^1(s,\pi,\sigma) - \varepsilon. \tag{2}$$

Against a fixed stationary strategy of player 2, there always exist pure stationary best replies for player 1, i.e.

$$\forall y \in Y \ \exists x \in X^p \ \forall s \in S \ \forall \pi \in \Pi: \ \gamma^1(s,x,y) \geq \gamma^1(s,\pi,y). \tag{3}$$

Obviously, similar statements hold for the best replies of player 2.

Rogers [11] and Sobel [14] have shown that for irreducible stochastic games there always exist stationary equilibria. Vieille [18], [19] has shown the existence of ε-equilibria for two-person stochastic games. The existence of ε-equilibria for stochastic games with more than 2 players is still an open problem.

Remark 1
The assumption of an infinite horizon is only used to approximate games with a sufficiently long, but possibly unknown, horizon. More precisely, in an irreducible game for every $\delta > 0$ there is a time horizon T_δ such that for any pair of stationary strategies (x, y) and for all $T > T_\delta$ we have $|\gamma(x, y) - \gamma_T(x, y)| < \delta$ where $\gamma_T(x, y)$ denotes the T-stage expected average payoff. Moreover, stationary strategies that are optimal for the infinite horizon game are ε-optimal in all T-stage games for T sufficiently large.

2.2 Evolutionary Games

An evolutionary game is determined by a fitness matrix, based on which a population distribution over different types will change. Here the population distribution at time t can be described by the vector $d(t) = (d_1(t), d_2(t), \ldots, d_m(t))$, where $d_\ell(t) > 0$ for all ℓ (all types are present) and $\sum_{\ell=1}^{m} d_\ell(t) = 1$. The fitness matrix is an $m \times m$ matrix A, that is to be interpreted as follows: The entry $a_{\ell k}$ is the fitness (or payoff or offspring) for an individual of type ℓ when interacting with an individual of type k. So, given the population distribution d, the average fitness of an individual of type ℓ is equal to $e_\ell A d^\top$ and the average fitness of an individual in the population is $dAd^\top$. Here the vector e_ℓ is a unit-vector with 1 in position ℓ and 0 elsewhere. The emphasis of the research in these games is on stability. Loosely speaking, a population distribution $\overline{d}$ is stable if for the process $\{d(t) : t \geq 0\}$ we have that, if it ever gets close to $\overline{d}$, then it will always stay close to $\overline{d}$, or even converge to it. The most commonly used stability concept in evolutionary games is the so-called Evolutionarily Stable Strategy (or ESS) (cf. Maynard Smith and Price [6]). A population distribution (or a strategy) $\overline{d}$ is an ESS if for all strategies $d \neq \overline{d}$ we have:

E1. $d A \overline{d}^\top \leq \overline{d} A \overline{d}^\top$;
E2. $d A \overline{d}^\top = \overline{d} A \overline{d}^\top \Rightarrow \overline{d} A d^\top > d A d^\top$.

Evolutionary stability is a refinement of the well-known Nash-equilibrium (cf. Nash [8]) for symmetric games, i.e. games (A, B) in which the payoffs for the players are symmetric in the sense that $B^\top = A$. Condition E1 says that $\overline{d}$ should be a best reply against itself, while condition E2 addresses the stability of $\overline{d}$. Namely, if d is also a best reply against $\overline{d}$ then, in order for the population distribution not to drift away in the direction of d, we need that $\overline{d}$ performs

better against d than d against itself.
The dynamics that are used most frequently, are the replicator dynamics, introduced by Taylor & Jonker [16]. According to the replicator dynamics, the proportion of population members of type ℓ, changes in time according to the following system of differential equations:

$$\dot{d_\ell} = d_\ell(e_\ell A d^\top - dAd^\top) \quad \text{for } \ell = 1, 2, \ldots, m.$$

So the replicator dynamics dictates, in a Darwinian way, that the population fraction of those types that perform better than average (or have more than average amount of offspring) will grow, while the fraction of types that perform below average, will fall. It can be shown that any ESS $\overline{d}$ is an asymptotically stable point for the corresponding replicator dynamics; i.e. if $d(0)$ is close enough to $\overline{d}$, then the population distribution converges to $\overline{d}$. However, we really need to be careful here, as the opposite is not always true, which can be seen from an example in Hofbauer & Sigmund [3] (page 71).
A different way of characterizing an ESS is by means of the concept of invasion. Suppose that we are dealing with a population distribution d and a fitness matrix A. If we replace a fraction $\varepsilon > 0$ of the population by mutants of types distributed as $\widetilde{d} \neq d$, then the new population would be

$$d_{\varepsilon\widetilde{d}} = (1-\varepsilon)d + \varepsilon\widetilde{d}.$$

We say that the mutants $\widetilde{d}$ cannot invade the population if for all $\varepsilon > 0$ sufficiently small we have

$$\widetilde{d}Ad^\top_{\varepsilon\widetilde{d}} < dAd^\top_{\varepsilon\widetilde{d}},$$

which can be interpreted as the mutants, the *new* members of the population, have a strict lower fitness than the *old* members of the population. It turns out that $\overline{d}$ is an ESS if and only if $\overline{d}$ cannot be invaded by any mutant $\widetilde{d} \neq \overline{d}$.

As a final word for this section we would like to stress that an ESS does not exist for every game, as can be seen from the game Rock-Paper-Scissors given by

$$\begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix},$$

which has only one symmetric equilibrium $\overline{d} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ to fit condition E1, but condition E2 fails to hold for $d = (1, 0, 0)$.

Remark 2
While we have presented the games in this section by giving just the payoff matrix A for the row player, we would like to stress that the games examined can be viewed as symmetric bimatrix games (A, B) where $B = A^\top$, because in the population there is no distinction between row and column players. In the next sections we will specify both payoffs, to avoid confusion with zero-sum games.

2.3 Evolutionary Stochastic Games

We identify an evolutionary stochastic game by an irreducible two-person stochastic game (as defined in Section 2.1) with the additional property that all matrices A_s are symmetric in payoffs as well as in transitions, i.e. it is an irreducible stochastic game for which:

ES1. $m_s^1 = m_s^2$ for each state s;
ES2. $r^2(s,i,j) = r^1(s,j,i)$ for each triple (s,i,j);
ES3. $p(s,i,j) = p(s,j,i)$ for each triple (s,i,j).

Notice that the symmetry assumptions are in line with Remark 2 and these are needed because we explicitly write the payoffs for row player and for column player and these players should be facing exactly the same strategic situation.

Types
The types in a symmetric irreducible stochastic game are identified by the pure stationary strategies in

$$X^p = \{e_1, e_2, \ldots, e_{|X^p|}\}.$$

As such the type of an individual specifies a certain action (behavior) for each of the states (environments) that an individual may encounter during its lifetime. Individuals of different types will choose different actions in at least one state, but they may choose the same action in some other states. Now a population distribution $d = (d_1, d_2, d_3, \ldots, d_{|X^p|})$ is a distribution over the set of pure stationary strategies, and it uniquely defines a stationary strategy x_d. More precisely,

$$x_d(s,i) = \sum_{k=1}^{|X^p|} d_k \cdot e_k(s,i),$$

where $x_d(s,i)$ and $e_k(s,i)$ denote the probability on action i in state s for the stationary strategies x_d and e_k respectively. So, given a population distribution d we will have a fraction $x_d(s,i)$ of individuals that choose action i in state s.
Please note that for population distributions d and d' with $d \neq d'$ we may well find $x_d = x_{d'}$, as can be seen from the following example. Consider a game with 2 states and with 2 pure actions in each state. For such a game there are 4 different pure stationary strategies: $e_1 = (1,1)$, $e_2 = (1,2)$, $e_3 = (2,1)$ and $e_4 = (2,2)$. Obviously the population distributions $d = (0.5, 0, 0, 0.5)$ and $d' = (0, 0.5, 0.5, 0)$ yield the same stationary strategy, i.e. $x_d = x_{d'} = ((0.5, 0.5), (0.5, 0.5))$. Thus, two population distributions may well consist of completely different types, but in terms of the actions observed in each of the states, one would not notice the difference.

Fitness
Consider a population distribution d and an individual of type k playing pure

stationary strategy e_k. We assume that during his lifetime the individual will visit all states of the stochastic game sufficiently often, while the population distribution is changing at a different, much larger time scale. This individual faces the combined behavior of all members of the whole population. This means that the individual plays against stationary strategy x_d and accordingly there are state payoffs and a transition function that takes the individual from state to state. The fitness of the individual will be the average of payoffs accumulated during his lifetime. As an example one may think of the individuals as members of a political party. The succes of the political party (the type) will depend on how good each of its members is doing in discussions on all kind of different issues, where media attention may move stochastically from one issue to the next based on the positions taken on the present issue. The succes of the political party will depend on how good its members are doing on each of the issues that have to be dealt with. As argued before, the limiting average is a good approximation for the finite horizon average for sufficiently long plays. At the same time the limiting average reward has the advantage that it does not depend on the initial state.

ESS

Based on the observation that different population distributions may yield the same stationary strategy, it seems more natural to define an Evolutionarily Stable Strategy in the space of stationary strategies rather than in the space of population distributions over the 'pure' types. We therefore define a stationary strategy $\overline{x}$ to be an ESS if for all stationary strategies $y \neq \overline{x}$:

E1*. $\gamma(y, \overline{x}) \leq \gamma(\overline{x}, \overline{x})$;
E2*. $\gamma(y, \overline{x}) = \gamma(\overline{x}, \overline{x}) \Rightarrow \gamma(\overline{x}, y) > \gamma(y, y)$.

Observe that for the single-state model this definition coincides with the original definition of an ESS for the classical evolutionary game model. This means that the game Rock-Paper-Scissors still applies as an example of a game without any ESS in the evolutionary stochastic game model.

Notice that if a population distribution d induces an ESS x_d, then there may well be other population distributions d' that also induce x_d, which could be interpreted as population distribution d being vulnerable to invasion by d'. However, as d' and d have to induce the same stationary strategy x_d, the density of suitable population distributions d' in population space is 0. As such, even for an ESS the population distribution d may change in time, but these changes have to remain within the class of distributions that induce x_d.

Replicator Dynamics

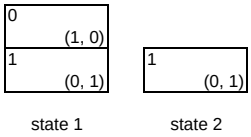
We can extend the approach using replicator dynamics to the evolutionary stochastic game model by taking the following system of differential equations:

$$\dot{d_\ell} = d_\ell(\gamma(e_\ell, x_d) - \gamma(x_d, x_d)) \quad \text{for } \ell = 1, 2, \ldots, |X^p|,$$

where, like mentioned before, d is the distribution over the pure stationary strategies e_ℓ and x_d is the stationary strategy induced by d. Again, for the single-state stochastic game model this definition of replicator dynamics coincides with its original definition.

Please note that for any stable point $\overline{d}$ of this replicator dynamics, we have that $\overline{d}_\ell > 0$ implies that $\gamma(e_\ell, x_{\overline{d}}) = \gamma(x_{\overline{d}}, x_{\overline{d}})$, which means that each of the prevailing types is playing a best reply to the induced stationary population strategy, because otherwise that type would not have survived the evolutionary competition. This observation suggests that for a Markov decison problem (MDP), which is what a player is facing when playing against a fixed stationary strategy, we have that if x^* is a stationary optimal strategy and x is a stationary strategy for which $\mathcal{C}(x_s) \subseteq \mathcal{C}(x_s^*)$ for each $s \in S$ (where $\mathcal{C}(x_s) = \{i : x_s(i) > 0\}$ is the carrier of x_s), then x is optimal as well. Such is not true in general, as can easily be seen by the following simple example.

Example 1



In this example of an MDP the vectors in the bottom right corners of each entry denote the transition probability vectors, while the upper left corners show the payoffs to the player. Any mixed stationary strategy is optimal for initial state 1, because it gives an average reward of 1. However, for the pure strategy $((1,0),1)$ the average reward is only 0.

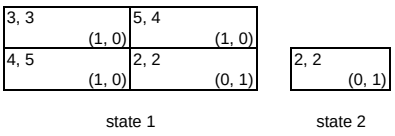
For irreducible MDP's we have the following theorem.

Theorem 1 *Consider an irreducible MDP. Suppose that x^* is a stationary optimal strategy, and x is a stationary strategy such that $\mathcal{C}(x_s) \subseteq \mathcal{C}(x_s^*)$ for every state $s \in S$. Then, x is optimal as well.*

A formal proof for this theorem is provided in the appendix.

We impose the condition of irreducibility on the stochastic game, because without this condition symmetric equilibria in stationary strategies may fail to exist as is shown by the following example:

Example 2



Again the transitions are given in the bottom right corners, while the upper left corners show the payoffs to the row and column players respectively. We

show that there can be no symmetric ε-equilibrium (x, x): If $x_1(2) > 0$, then, for any $\varepsilon > 0$ sufficiently small, the unique stationary ε-best reply is $((1, 0), 1)$; which rules out any ε-equilibrium (x, x) with $x_1(2) > 0$. However, against $((1, 0), 1)$, for any $\varepsilon > 0$ sufficiently small, the unique ε-best reply is $((0, 1), 1)$.

On the other hand, if we do impose the condition of irreducibility on a symmetric stochastic game, then we are guaranteed to have at least one symmetric stationary equilibrium, by the following theorem. This is important to know because the existence of symmetric equilibria is a necessary condition for the existence of evolutionarily stable strategies, just like in one state evolutionary games.

Theorem 2 *Every symmetric irreducible stochastic game admits a symmetric stationary equilibrium* (x^*, x^*).

Proof Take an arbitrary symmetric irreducible stochastic game. So, $X = Y$. For a discount factor $\beta \in (0, 1)$ and a stationary strategy $x \in X$ of player 1, let $B^2_\beta(x)$ denote the set of stationary strategies $y \in X$ of player 2 that are β-discounted best responses to x.

We have the following well-known properties (cf. e.g. Fink [2] or Takahashi [15]): (1) the set X is nonempty, compact and convex, (2) $B^2_\beta(x)$ is nonempty and convex for every $x \in X$, (3) the set-valued map $B^2_\beta : x \mapsto B^2_\beta(x)$ is upper semi-continuous, i.e. if sequences x_n and y_n in X converge to $x \in X$ and $y \in X$ respectively, and $y_n \in B^2_\beta(x_n)$ for every $n \in \mathbb{N}$, then $y \in B^2_\beta(x)$ must hold. Hence, by Kakutani's fixed point theorem [4], the map B^2_β has a fixed point $x^*_\beta \in X$, i.e. $x^*_\beta \in B^2_\beta(x^*_\beta)$. Due to symmetry, it follows that the stationary strategy pair (x^*_β, x^*_β) is a symmetric β-discounted equilibrium.

Since X is compact, there exists a sequence of discount factors β_n such that $\beta_n \to 1$ and $x^*_{\beta_n}$ converges to some $x^* \in X$. We now prove that (x^*, x^*) is an equilibrium. Consider an arbitrary stationary strategy $x \in X$ for player 1. Then, for every $n \in \mathbb{N}$, because $(x^*_{\beta_n}, x^*_{\beta_n})$ is a β_n-discounted equilibrium, we have

$$\gamma^1_{\beta_n}(x, x^*_{\beta_n}) \leq \gamma^1_{\beta_n}(x^*_{\beta_n}, x^*_{\beta_n}).$$

Since the game is irreducible and $x^*_{\beta_n}$ converges to x^*, we have (cf. Lemma 2.2.6 in [17])

$$\lim_{n\to\infty} \gamma^1_{\beta_n}(x, x^*_{\beta_n}) = \gamma^1(x, x^*)$$

$$\lim_{n\to\infty} \gamma^1_{\beta_n}(x^*_{\beta_n}, x^*_{\beta_n}) = \gamma^1(x^*, x^*).$$

Therefore, $\gamma^1(x, x^*) \leq \gamma^1(x^*, x^*)$. Since $x \in X$ was arbitrary, it follows that x^* is a best response for player 1 to x^*. Due to symmetry, (x^*, x^*) is an equilibrium, as claimed.

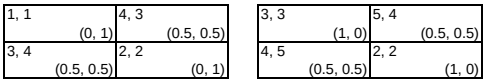
Moreover, as a consequence of Theorem 1 we also have the following result, which extends a well-known result for one-state evolutionary games.

Corollary 1 *If x^* is an ESS in an evolutionary stochastic game and $x \neq x^*$ is a stationary strategy with $C(x_s) \subseteq C(x^*_s)$ for each s, then x is no ESS.*

Proof Let x^* be an ESS in an evolutionary stochastic game and let x be a stationary strategy with $C(x_s) \subseteq C(x_s^*)$ for all s. Because x^* is an ESS, x^* is a best reply against itself. Therefore x^* is an optimal stationary strategy in the MDP that arises for player 1 if player 2 is playing x^*. This implies by Theorem 1 that $\gamma(x, x^*) = \gamma(x^*, x^*)$. Hence $\gamma(x^*, x) > \gamma(x, x)$ by the 2nd ESS condition. This means that x is no ESS.

For illustration we now look at a two-state evolutionary stochastic game:

Example 3



For this example the unique ESS is $((\frac{1}{2}, \frac{1}{2}), (\frac{5}{7}, \frac{2}{7}))$. The uniqueness follows from the fact that the stochastic game has only one symmetric Nash equilibrium in stationary strategies. Because the population space consists of the convex combinations of four pure stationary strategies, we cannot visualize population development in a two dimensional figure. We have therefore chosen to visualize population development by menas of two movies. These exhibit how the population develops under the replicator equation for two different intial population distributions. These movies can be viewed at
http:/www.youtube.com/watch?v=CuM3GtoyMM0

3 Concluding remarks

In this paper we have introduced a model that fuses the classical model of evolutionary games with that of stochastic games. This implies that there are still many issues open for future study. In terms of applications it is challenging to find a specific real life phenomenon of population development that would perfectly fit this model. At the theoretical level, there are also some fine challenges, like to characterize the class of symmetric two-person stochastic games for which symmetric equilibria exist in stationary strategies, or to address the more general question whether or not any symmetric two-person stochastic game always has a symmetric ε-equilibrium, when we also allow for non-stationary ones. It would also be very interesting to explore for evolutiuonary stochastic games other dynamics than the replicator dynamics. Some first studies on the fictitious play dynamics for the example in section 2.3 seem to indicate that it converges to the ESS. Again, an illustrative movie is available at
http://www.youtube.com/watch?v=P5QBTVCXXC0

4 Appendix

The following results may be known in MDP literature, but we have not been able to find a precise reference. For sake of completeness we provide the proofs.

Consider an irreducible MDP. Take an arbitrary stationary strategy x and a state $s \in S$. For the limiting average reward for x we simply write $\gamma(x)$ because the reward does not depend on the initial state s. For a mixed action α_s in state s, let $x[s, \alpha_s]$ be the stationary strategy which uses the mixed action α_s in state s and uses the mixed actions x_z in all other states $z \in S \setminus \{s\}$, i.e. $x[s, \alpha_s]_s = \alpha_s$ and $x[s, \alpha_s]_z = x_z$ for all $z \in S \setminus \{s\}$.

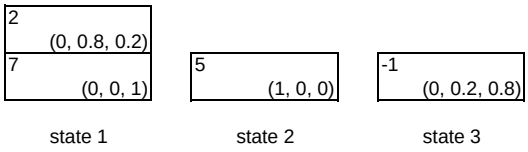
Let W denote the set of all finite histories $h = (s^1, i^1, \ldots, s^{\ell-1}, i^{\ell-1}, s^\ell)$, where $\ell \in \mathbb{N}$ and $s^1 = s^\ell = s$ and $s^k \neq s$ for every $k = 2, \ldots, \ell - 1$, such that the stationary strategy $x[s, \alpha_s]$ generates h with a positive probability when starting in state s. Note that W is countable. Let $q(h)$ denote the corresponding probability for every $h \in W$. Due to irreducibility, we have $\sum_{h \in W} q(h) = 1$. Let $t(h) = \ell - 1$, which is just the time it takes along h to visit state s again, and let $R(h)$ denote the sum of the payoffs along h during periods $1, \ldots, \ell - 1$.

Let $t^{x,s}_{\alpha_s} = \sum_{h \in W} q(h)\, t(h)$, which is the expected number of periods it takes to visit state s again when we use $x[s, \alpha_s]$ and start in state s. Clearly, $t^{x,s}_{\alpha_s} \geq 1$ and $t^{x,s}_{\alpha_s}$ is finite due to irreducibility. Let $R^{x,s}_{\alpha_s} = \sum_{h \in W} q(h)\, R(h)$ denote the corresponding expected sum of payoffs before visiting state s again. We define

$$r^{x,s}_{\alpha_s} = \frac{R^{x,s}_{\alpha_s}}{t^{x,s}_{\alpha_s}}.$$

These definitions are illustrated by the following example.

Example 4



Let $x = ((\frac{1}{3}, \frac{2}{3}), 1, 1)$, let $s = 1$ and let $\alpha_1 = (\frac{1}{2}, \frac{1}{2})$. When also applying the notations $e_1 = (1, 0)$ and $e_2 = (0, 1)$, then we have

$$r^{x,1}_{e_1} = \frac{R^{x,1}_{e_1}}{t^{x,1}_{e_1}} = \frac{6}{3} = 2$$

$$r^{x,1}_{e_2} = \frac{R^{x,1}_{e_2}}{t^{x,1}_{e_2}} = \frac{7}{7} = 1$$

$$r^{x,1}_{\alpha_1} = \frac{R^{x,1}_{\alpha_1}}{t^{x,1}_{\alpha_1}} = \frac{6.5}{5} = 1.3$$

Notice that, although α_1 uses e_1 and e_2 with probability $\frac{1}{2}$ each, we have $r^{x,1}_{\alpha_1} = 1.3 \neq 1.5 = \frac{1}{2} r^{x,1}_{e_1} + \frac{1}{2} r^{x,1}_{e_2}$. In view of the following lemma, this example shows that

$$\gamma((\frac{1}{2}, \frac{1}{2}), 1, 1) \neq \frac{1}{2} \cdot \gamma((1, 0), 1, 1) + \frac{1}{2} \cdot \gamma((0, 1), 1, 1).$$

Lemma 1 *Consider an irreducible MDP. Take an arbitrary stationary strategy x and a mixed action α_s in some state $s \in S$. Then, $\gamma(x[s, \alpha_s]) = r^{x,s}_{\alpha_s}$.*

Proof Suppose that we use $x[s, \alpha_s]$ and start in state s. Then, with probability 1, an infinite play $h^\infty = (s^m, i^m)_{m\in\mathbb{N}}$, with $s^1 = s$, is generated such that: (1) state s is visited infinitely many times, (2) between each two consecutive visits, a history in W is generated, (3) the relative frequency of every history $h \in W$ is exactly $q(h)$. Consequently, every history $h \in W$ is associated to a proportion

$$\frac{q(h) \cdot t(h)}{\sum_{h'\in W} q(h') \cdot t(h')}$$

of the set $\mathbb{N}$ of all periods, which yields

$$\gamma(x[s, \alpha_s]) = \sum_{h\in W} \frac{q(h) \cdot t(h)}{\sum_{h'\in W} q(h') \cdot t(h')} \cdot \frac{R(h)}{t(h)} = \frac{\sum_{h\in W} q(h) \cdot R(h)}{\sum_{h\in W} q(h) \cdot t(h)} = \frac{R^{x,s}_{\alpha_s}}{t^{x,s}_{\alpha_s}} = r^{x,s}_{\alpha_s},$$

which completes the proof.

The following lemma presents a useful expression for the reward $\gamma(x)$ induced by a stationary strategy x based on the quantities $t^{x,s}_i$ and $r^{x,s}_i$, with $i \in I_s$.

Lemma 2 *Consider an irreducible MDP. Take an arbitrary stationary strategy x and a state $s \in S$. Then,*

$$\gamma(x) = \frac{\sum_{i\in I_s} x(s,i) \cdot R^{x,s}_i}{\sum_{i\in I_s} x(s,i) \cdot t^{x,s}_i} = \frac{\sum_{i\in I_s} x(s,i) \cdot t^{x,s}_i \cdot r^{x,s}_i}{\sum_{i\in I_s} x(s,i) \cdot t^{x,s}_i}, \tag{4}$$

where $x(s,i)$ is the probability that the stationary strategy x places on action i in state s.

Proof Note that by definition we have $x = x[s, x_s]$. Hence, Lemma 1 tells us that

$$\gamma(x) = r^{x,s}_{x_s} = \frac{R^{x,s}_{x_s}}{t^{x,s}_{x_s}}.$$

The observation that

$$R^{x,s}_{x_s} = \sum_{i\in I_s} x(s,i) \cdot R^{x,s}_i \quad \text{and} \quad t^{x,s}_{x_s} = \sum_{i\in I_s} x(s,i) \cdot t^{x,s}_i,$$

completes the proof.

Notice that, in view of equation (4), the reward $\gamma(x)$ is a convex combination of the quantities $r^{x,s}_i$, with $i \in I_s$.

Lemma 3 *Consider an irreducible MDP. Take a stationary optimal strategy x^* and a state $s \in S$. Then $r^{x^*,s}_i = \gamma(x^*)$ for every $i \in \mathcal{C}(x^*_s)$ and $r^{x^*,s}_i \leq \gamma(x^*)$ for every $i \in I_s \setminus \mathcal{C}(x^*_s)$.*

Proof Take an arbitrary action $i \in I_s$, and consider the stationary strategy $x^*[s,i]$. By Lemma 1, we have $r_i^{x^*,s} = \gamma(x^*[s,i])$. Due to the optimality of x^*, we obtain $r_i^{x^*,s} \leq \gamma(x^*)$. Because this inequality holds for all $i \in I_s$, and because $\gamma(x^*)$ is a convex combination of $r_i^{x^*,s}$, $i \in I_s$, due to (4), we obtain $r_i^{x^*,s} = \gamma(x^*)$ for every $i \in \mathcal{C}(x_s^*)$.

Proof of Theorem 1

Proof We may assume that $\mathcal{C}(x_s) \subseteq \mathcal{C}(x_s^*)$ for some state $s \in S$ and $x_z = x_z^*$ for all other states $z \in S \setminus \{s\}$, because then the theorem follows if we iteratively apply it state by state. Note that, by lemma 3, it holds that $r_i^{x^*,s} = \gamma(x^*)$ for all $i \in \mathcal{C}(x_s^*)$. Because due to our assumption $r_i^{x,s} = r_i^{x^*,s}$ for all $i \in I_s$ and $\mathcal{C}(x_s) \subseteq \mathcal{C}(x_s^*)$, we obtain $r_i^{x,s} = \gamma(x^*)$ for all $i \in \mathcal{C}(x_s)$. Hence, by Lemma 2 we find $\gamma(x) = \gamma(x^*)$.

Acknowledgement

We would like to thank an anonymous referee for his valuable comments and suggestions that helped us to substantially improve the presentation of our results.

References

1. E. Altman, Y. Hayel, H. Tembine, R. El-Azouzi (2008): "Markov decision evolutionary games with expected average fitness", working paper, INRIA, Sophia-Antipolis.
2. A.M. Fink (1964): "Equilibrium in a stochastic n-person game", *Journal of Science of Hiroshima University* A-I 28, 89-93.
3. J. Hofbauer, K. Sigmund (1998): *Evolutionary Games and Population Dynamics*, Cambridge University Press.
4. S. Kakutani (1941): "A generalization of Brouwers fixed point theorem", *Duke Mathematical Journal* 8, 457-459.
5. A. Kolmogorov (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik* 2, No. 3, Springer-Verlag, Berlin.
6. J. Maynard Smith, G.R. Price (1973): "The logic of animal conflict", *Nature* 246, 15-18.
7. J.-F. Mertens, A. Neyman (1981): "Stochastic games", *International Journal of Game Theory* 10, 53-66.
8. J.F. Nash (1951): "Non-cooperative games", *Annals of Mathematics* 54, 286-295.
9. J. von Neumann (1928): "Zur Theorie der Gesellschaftsspiele", *Mathematische Annalen* 100, 295-320.
10. S. Pruett-Jones, A. Heifetz (2012): "Optimal marauding in bowerbirds", *Behavioral Ecology*, doi:10.1093/beheco/ars004.
11. P.D. Rogers (1969): "Non-zerosum stochastic games", PhD Thesis, report ORC 69-8, Operations Research Center, University of California, Berkeley.
12. W.H. Sandholm (2010): *Population Games and Evolutionary Dynamics*, MIT Press.
13. L.S. Shapley (1953): "Stochastic games", *Proceedings of the National Academy of Sciences U.S.A.* 39, 1095-1100.

14. M.J. Sobel (1971): "Noncooperative stochastic games", *Annals of Mathematical Statistics* 42, 1930-1935.
15. M. Takahashi (1964): "Equilibrium points of stochastic noncooperative n-person games", *Journal of Science of Hiroshima University* A-I 28, 95-99.
16. P.D. Taylor, L. Jonker (1978): "Evolutionarily stable strategies and game dynamics", *Mathematical Biosciences* 40, 145-156.
17. F. Thuijsman (1992): *Optimality and Equilibria in Stochastic Games*, CWI-Tract 82, CWI, Amsterdam.
18. N. Vieille (2000): "2-person stochastic games I: A reduction", *Israel Journal of Mathematics* 119, 55-91.
19. N. Vieille (2000): "2-person stochastic games II: The case of recursive games", *Israel Journal of Mathematics* 119, 93-126.