Stochastic Games with Hyperbolic Discounting

Please check the document version of this publication:

Stochastic Games with Hyperbolic Discounting

Ronald Peeters[*]

February 2004

Abstract

In this paper the hyperbolic discounted reward criterion is introduced for stochastic games. For different levels of sophistication an individual may have existence of a particular type of optimal strategy in a Markov decision problems is shown. Next, a particular type of equilibrium for hyperbolic discounted stochastic games with only one type of individual present is shown to exist. From this, the existence of a particular (delayed-stationary) type of equilibrium for general hyperbolic discounted stochastic games is obtained.

JEL Classification Codes: C61; C72; C73; D90.

Keywords: Markov decision problem; Stochastic game; Hyperbolic discounting.

1 Introduction

In the economic theory of intertemporal choice, as it is originally formulated by Irving Fisher and generalized by Paul Samuelson, all individuals behave similarly as they borrow or lend until the marginal rate of substitution between consumption today and consumption tomorrow is equal to the interest rate. Strotz (1955-1956) had a critical view on individual's behavior in the decision making process, by speculating that individuals act as if their discount rates vary with the length of time to be waited: individuals tend to "overvalue" the more proximate satisfaction relative to the more distant ones. When individuals indeed behave according to this discount function, dynamic inconsistent behavior is likely to occur.

The inconsistency above can be illustrated by the example coming from O'Donoghue and Rabin (1999) where an individual is presented a choice between doing seven hours of an unpleasant activity on April 1 versus eight hours on April 15. If asked on February 1 virtually everyone would prefer the seven hours on April 1. But come April 1, given the same choice, most of us are apt to put off the work until April 15. Individuals with sophisticated behavior, however, foresee that they will have self-control problems in the future and will for instance plan already on February 1 to do the unpleasant job on April 15.

Discounted stochastic games have been introduced by Shapley (1953). He considered two-person zero-sum finite stochastic games. Shapley proved that such games have a value and that both players possess optimal stationary strategies with respect to the discounted payoff criterion. Fink (1964), Takahashi (1964), and Sobel (1971) extended Shapley's model to general n-person stochastic games, and proved independently of one another the existence of a stationary equilibrium. For this general model Haller and Lagunoff (2000) proved generic finiteness of the number of stationary equilibria. Herings and Peeters (2004) extended this to generic oddness.

For the limiting average reward criterion that is originally proposed by Gillette (1957), Mertens and Neyman (1981) proved that every zero-sum stochastic game has a value. For the nonzero-sum case equilibria are known not always to exist. Thuijsman and Vrieze (1991) proved that for this class there always exists an initial state such that an ε-equilibrium exists. Later, various existence results have been obtained, but only under specific assumptions on the transition and payoff structure. Vieille (2001) proved that every two-player stochastic game has an ε-equilibrium. For more than two players, it is still an open question whether all stochastic games possess an ε-equilibrium or not.

Both types of evaluating streams of utilities mentioned above do not capture the notion of present-biasedness that is described in the first two paragraphs of this introduction. One type of payoff-evaluation that does capture this behavior of overvaluing the near future in comparison to the distant future is by hyperbolic discounting. Although this evaluation criterion has been introduced for individual decision making quite extensively (see Frederick, Loewenstein, and O'Donoghue (2002) and the many references to applications therein) and for (infinitely) repeated games quite recently and implicitly (see Arribas and Urbano (2003) and Vieille and Weibull (2002)), it has never been introduced and therefore also not been studied for the class of stochastic games in general.[1]

---

[1]In fact, implicitly this type of discounting is captured by Shapley's definition of a stochastic game. Namely, hyperbolic discounting is equivalent to having the probability that the game ends increasing over time—which implies that the end of the game gets closer over time (in a probabilistic sense).

This paper introduces hyperbolic discounting in stochastic games. First hyperbolic discounted stochastic games with one individual, Markov decision problems, are considered. With an example it is shown that an individual can suffer from the problem of self-control. An individual that is resolute pre-commits to play a certain strategy and therefore never suffers, but pre-commitments lead to irrational behavior in later stages. A naive individual suffers as he will revise its strategy any instance in time and therefore displays dynamic inconsistent behavior: in future stages he will not play the action as he planned to play at forehand. An individual who is sophisticated will realize eventual future self-control problems and therefore decides to play a dynamic consistent strategy. It is shown that resolute individuals have a delayed-stationary strategy that is optimal; that naive individuals have a delayed-stationary strategy that is optimal, but will after all play stationary; and that sophisticated individuals have an optimal dynamic consistent strategy that is stationary.

Finally the many-individual game is studied. First, it is shown that when there are only resolute individuals present, there exists a delayed-stationary equilibrium. Then, for the situation with only naive individuals present it is shown that a stationary equilibrium play-path exists. When only sophisticate individuals are present a stationary dynamic consistent equilibrium is shown to exist. The paper ends with a small discussion on what happens when different types of individuals are playing in one hyperbolic discounted stochastic game. This leads to the conclusion that for any hyperbolic discounted stochastic game an equilibrium exists that has a delayed-stationary equilibrium path.

Section 2 considers different types of utility evaluations and formulates them explicitly. In Section 3 the stochastic game with hyperbolic discounting is introduced. Section 4 concentrates on the Markov decision problem—which is a stochastic game with one individual—and contains an example that shows that dynamic inconsistency can occur. In Section 5 the multi-individual case is studied for three alternative situations: one in which all individuals are resolute, one in which all individuals are naive and one in which all individuals are sophisticated. Moreover, situations with a combination of different types of individuals present are shortly discussed on.

2   Utility evaluations

In period k, an individual does not only care about his present instantaneous utility u^k, but also about his future instantaneous utilities $u^{k+1}, u^{k+2}, \ldots$. Let $U^k(u^k, u^{k+1}, \ldots)$ represent the individual's *intertemporal preferences* from the perspective of period k, where U^k is continuous and increasing in all components. The individual's preferences are said to be *dynamic consistent* if the relative preference for well-being at an earlier date over a later

date is the same no matter when this person is asked.

One standard way to evaluate the stream of instantaneous utilities is by *exponential discounting*: for all k, $U^k(u^k, u^{k+1}, \ldots) = \sum_{\kappa=k}^{\infty} \delta^\kappa u^\kappa$, where $\delta \in (0,1)$ is the discount factor. Evaluating a stream of utilities by exponential discounting captures the notion of dynamic consistency. Another way to evaluate the stream of utilities is by *limit averaging*: for all k, $U^k(u^k, u^{k+1}, \ldots) = \liminf_{K \to \infty} \frac{1}{K-k+1} \sum_{\kappa=k}^{K} u^\kappa$.

Individual's preferences are often found to be *present-biased* (see Thaler (1981)), i.e. when considering trade-offs between two future moments stronger relative weight is given to the earlier moment as it gets closer. This behavior can be captured with a simple two-parameter model that modifies exponential discounting via $U^k(u^k, u^{k+1}, \ldots) = \delta^k u^k + \beta \sum_{\kappa=k+1}^{\infty} \delta^\kappa u^\kappa$ for all k, where $\delta, \beta \in (0,1)$ (as it is originally proposed by Phelps and Pollak (1968); see also O'Donoghue and Rabin (1999) or Frederick, Loewenstein, and O'Donoghue (2002)). The parameter δ represents long-run, dynamic consistent discounting; the parameter β represents a bias for the present—how you favor now versus later. For $\beta = 1$ this type of discounting is equivalent to exponential discounting. But $\beta < 1$ implies present-biased preferences: the person gives more relative weight to period κ in period κ than she did in any period prior to period κ. This way of discounting is also referred to as *hyperbolic discounting*.[2]

The present-biasedness exhibits a specific type of dynamic inconsistency depending on the level of sophistication in behavior (see the example in Section 4). Individuals who are *sophisticated* foresee that they will have self-control problems in the future; individuals who are *naive* do not foresee these self-control problems but do not make pre-commitments to their behavior; and finally individuals that are *resolute* do also not foresee future self-control problems but do make pre-commitments (see O'Donogue and Rabin (1999a and 1999b), and Harris and Laibson (2001)).

3 Stochastic game with hyperbolic discounting

A *finite stochastic game with hyperbolic discounting* is given by the tuple

$$\Gamma = \langle N, \Omega, \{S_\omega^i\}_{(i,\omega) \in N \times \Omega}, \{u^i\}_{i \in N}, \pi, \delta, \beta \rangle,$$

in which N denotes the finite set of players, Ω the finite set of states, and S_ω^i is the finite set of actions that player i has at its disposal in state $\omega \in \Omega$. The instantaneous payoff to player i in state ω when the players play $s_\omega = (s_\omega^i)_{i \in N}$ is given by $u^i(\omega, s_\omega)$. The probability

---

[2]Or discrete-time quasi-hyperbolic discounting as it is called in Harris and Laibson (2001).

of going from state ω to state $\bar{\omega}$ when the players play s_ω is given by $\pi(\bar{\omega} \,|\, \omega, s_\omega)$. The players are allowed to randomize their actions. Both the instantaneous payoff-function and the transition mapping are extended multi-linearly. Players maximize the total stream of expected payoffs, where the future payoffs are discounted by discount factor $\delta \in (0, 1)$ and where the distant future—as opposed to the near future—is further discounted by discount factor $\beta \in (0, 1)$.

The game proceeds as follows. Each player i selects at the initial state ω^0 an action $\sigma^i_{\omega^0} \in \Sigma^i_{\omega^0} = \Delta(S^i_{\omega^0})$. Then two things happen, both depending on the current state ω^0 and the action choices $\sigma_{\omega^0} = (\sigma^i_{\omega^0})_{i \in N}$:

(1) player i earns $u^i(\omega^0, \sigma_{\omega^0})$, and

(2) the system jumps to the next state ω^1 according to the outcome of the chance experiment given by $\pi(\cdot \,|\, \omega^0, \sigma_{\omega^0})$.

Subsequently, in the next period, all players are informed about the previous actions chosen by the players, and of the new state ω^1. In this next period, the above procedure is repeated, starting from the state ω^1.

The most elementary type of strategy one can formulate is a strategy which is a series of actions that depend on time, state, and history (of states revisited and the actions realized). Such a strategy is called a *behavior strategy*. Given initial state ω and a combination of behavior strategies σ, the stream of expected payoffs is evaluated by

$$U^i(\omega, \sigma) := U^{i0}(\omega, \sigma) + \beta \sum_{k=1}^{\infty} \delta^k U^{ik}(\omega^k, \sigma),$$

where $U^{ik}(\omega^k, \sigma)$ denotes the expected instantaneous utility at stage k. Here, $U^i(\omega, \sigma)$ equals the total hyperbolic discounted expected payoff of player i with discount factors δ and β, the starting state is ω and the strategy-tuple σ is played. Since the state and action spaces are assumed to be finite, $U^i(\omega, \sigma)$ exists. A combination of behavior strategies σ is a *Nash equilibrium* if and only if σ^i is a best response to σ^{-i} for all $i \in N$, where the superscript '$-i$' is used to indicate player i's opponents behavior strategies. So, formally, σ is a Nash equilibrium if for given initial state ω^0

$$U^i(\omega^0, \sigma) \geq U^i(\omega^0, \sigma^{-i}, \bar{\sigma}^i),$$

for any behavior strategy $\bar{\sigma}^i$ of player i and for all players $i \in N$. A combination of behavior strategies σ is a *subgame perfect Nash equilibrium* if it constitutes a Nash equilibrium for every subgame.

A *Markov strategy* is a strategy that does not depend on history, but on time and state only. A *stationary strategy* refines the Markov strategy in the sense that a stationary strategy does not depend on time, thus on the state only. The two strategic refinements induce refinements on the equilibrium concept. A *Markov (perfect) equilibrium* is a Nash equilibrium in Markov strategies. Similarly, a *stationary (perfect) strategy* is a Nash equilibrium (and also a Markov (perfect) equilibrium) in stationary strategies. The word 'perfect' within the brackets is active if the equilibrium is an equilibrium in every proper subgame. For a stationary equilibrium this perfectness condition is equivalently to being a Nash equilibrium for any possible initial state. To that end, imagine that some states might not be accessible from certain initial states. In the remaining of this paper, although not mentioned explicitly, all equilibria referred to are perfect.

4 One player: Markov decision problem

A stochastic game with only one player is a Markov decision problem and can be denoted by

$$\Gamma = \langle \Omega, \{S_\omega\}_{\omega \in \Omega}, u, \pi, \delta, \beta \rangle.$$

In Figure 1 a Markov decision problem is displayed where in the first state the individual can decide between the options T (top) and B (bottom). If T is chosen, the individual receives an immediate reward of 1 and is sure to stay in the first state. If B is chosen, the individual receives an immediate reward of 0 but is sure to be in the second state in the next period. Whence arrived in the second state, the individual does not have to make a choice; the player is simply rewarded by 3 and surely returns to the first state in the subsequent period. Without any loss of generality it is assumed that the first state is the initial state.
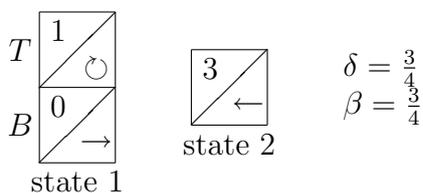

Figure 1: Markov decision problem.

Below six possible strategies are listed (only considering these six strategies will be sufficient for the analysis) with the corresponding payoffs by playing these strategies. In

brackets the corresponding payoffs in case of standard discounting (when β would be equal to 1) are given.

$$
\begin{array}{llllll}
TT\cdot\cdot\cdot\cdot & \longrightarrow & \frac{13}{4} & \approx & 3.2500 & (\ 4 & \approx & 4.0000\ ) \\
BB\cdot\cdot\cdot\cdot & \longrightarrow & \frac{27}{7} & \approx & 3.8571 & (\ \frac{36}{7} & \approx & 5.1429\ ) \\
TBB\cdot\cdot\cdot & \longrightarrow & \frac{109}{28} & \approx & 3.8929 & (\ \frac{34}{7} & \approx & 4.8571\ ) \\
BTT\cdot\cdot\cdot & \longrightarrow & \frac{27}{8} & \approx & 3.3750 & (\ \frac{9}{2} & \approx & 4.5000\ ) \\
TBTB\cdot\cdot & \longrightarrow & \frac{109}{28} & \approx & 3.8929 & (\ \frac{34}{7} & \approx & 4.8571\ ) \\
BTBT\cdot\cdot & \longrightarrow & \frac{27}{7} & \approx & 3.8571 & (\ \frac{36}{7} & \approx & 5.1429\ )
\end{array}
$$

The payoffs reveal that $TBB\cdot\cdot\cdot$ and $TBTB\cdot\cdot$ are optimal strategies. In fact, any strategy in which action T is played in the initial state and where B is played when the first state is the active state is optimal (note to this end that in the second optimal strategy in later periods T would have been played once the first state is not active). It is easily shown that when the Markov decision problem is slightly perturbed such that the transition dynamics is no longer determinate, $TBB\cdot\cdot\cdot$ will be the unique optimal strategy. In each moment in time the strategy $TBB\cdot\cdot\cdot$ is optimal and therefore the player plans to play $TBB\cdot\cdot\cdot$ every moment in time. This leads to the following decision matrix where in any moment in time (the rows) committing to play the strategy $TBB\cdot\cdot\cdot$ is optimal.

$$
\begin{bmatrix}
T & B & B & B & \cdots \\
* & T & B & B & \cdots \\
* & * & T & B & \\
* & * & * & \ddots & \ddots
\end{bmatrix}
$$

The resolute individual.   A resolute individual that can make strategic pre-commitments would commit playing the optimal strategy $TBB\cdot\cdot\cdot$ and will play action B in all periods except the initial one.

Theorem 4.1 *For all hyperbolic discounted Markov decision problems with a resolute individual an optimal strategy exists in which the individual plays stationary from the first period on.*

Proof.   From the first period on, continuously playing the action of the optimal stationary strategy of the underlying Markov decision problem with standard discounted reward criterion ($\beta = 1$) will be optimal. Given this future behavior, the optimal action in the initial period is easily determined and is not necessarily similar to the action played in all subsequent periods (see for instance the example above). $\square$

The naive individual. A naive individual would recognize in the second stage that succeeding with playing B is not optimal as it is optimal to play T at any instance in time given that in all future periods B will be played. Therefore, a naive individual will end up playing $TT \cdot \cdot \cdot \cdot$—the diagonal of the matrix above.

Theorem 4.2 *For all hyperbolic discounted Markov decision problems with a naive individual an optimal strategy such that the individual plays stationary exists.*

Proof. If the strategy of playing action σ_0 in the initial period and σ_r in all subsequent periods is the optimal strategy in the resolute case, then a naive individual will gradually revise its strategy such that its play path will be always playing σ_0. $\qquad\square$

The sophisticated individual. Sophisticated individuals will foresee the future self-control problems that naive individuals do not see. From the perspective of the naive individual, the sophisticated individual would realize that he would deviate in any instance and therefore would play T stationary. Knowing that T will be played in all future events, the optimal action in the present stage is to play B again (to this end consult the list of strategies and corresponding payoffs above). Apparently, this mode of reasoning leads to a cyclic process of thought.

One other way to find optimal strategies in infinite horizon models is to extrapolate from a finite horizon model. Suppose thereto that the game ends after stage K has been reached. Then in stage K, if the first state would be the active state, playing T is the optimal action. Regardless of the action that will be chosen in stage K, in stage $K - 1$, playing B is the optimal action. Knowing that in the final two stages action B and subsequently T will be chosen, it is optimal to play T in stage $K - 2$, as $1 + \frac{3}{4} \cdot \frac{3}{4} \cdot (0 + \frac{3}{4} \cdot 3) = \frac{145}{64} > \frac{135}{64} = 0 + \frac{3}{4} \cdot \frac{3}{4} \cdot (3 + \frac{3}{4} \cdot 1)$. Knowing that in the final three stages TBT will be played, action B is optimal in stage $K - 3$. The optimal dynamic consistent strategy in the finite horizon model is easily shown to be $BTB \cdots BT$ if K is odd (and the number of periods is even) and $TBT \cdots BT$ if K is even.

The finite horizon analysis above triggers the idea that both $TBTB \cdot \cdot$ and $BTBT \cdot \cdot$ are optimal dynamic consistent strategies for the infinite horizon game. And indeed, when the individual intends to play $BTB \cdot \cdot$ from tomorrow on, it is optimal to play action T today as $\frac{109}{28} > \frac{837}{224} = 0 + \frac{3}{4} \cdot \frac{3}{4} \cdot (3 + \frac{3}{4} \cdot \frac{34}{7})$; and when the individual intends to play $TBT \cdot \cdot$ from tomorrow on, it is optimal to play action B today as $\frac{27}{7} > \frac{209}{56} = 1 + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{34}{7}$.[3] Although both dynamic consistent optimal strategies are not stationary, they are Markovian as the strategies do not depend on history but on time only.

---

[3]For some numbers consult the list of strategies with corresponding payoffs.

One possibility left, to have a stationary optimal dynamic consistent strategy is one using mixed actions. Suppose the individual plays T stationary with probability p. Then the present value of the stream of payoffs from the second stage on, which depends on the state, is found by solving:

$$\nu(1,p) = p + \tfrac{3}{4} \cdot (p \cdot \nu(1,p) + (1-p) \cdot \nu(2,p)) \quad \text{and} \quad \nu(2,p) = 3 + \tfrac{3}{4} \cdot \nu(1,p).$$

This solves for

$$\nu(1,p) = \tfrac{36-20p}{7-3p} \quad \text{and} \quad \nu(2,p) = \tfrac{48-24p}{7-3p}.$$

So, the present value of the stream of payoffs from the first stage on when the individual chooses T respectively B in the first period equals

$$U_T = 1 + \tfrac{3}{4} \cdot \tfrac{3}{4} \cdot \nu(1,p) = \tfrac{109-57p}{28-12p} \quad \text{and} \quad U_B = 0 + \tfrac{3}{4} \cdot \tfrac{3}{4} \cdot \nu(2,p) = \tfrac{108-54p}{28-12p}.$$

The individual is indifferent between playing action T and B in the first stage—and therefore willing to play a mixed action—if these two utilities are equal, which is the case for $p = \tfrac{1}{3}$. So, the stationary strategy $(\tfrac{1}{3}, \tfrac{2}{3})$ is an optimal dynamic consistent strategy.

Theorem 4.3 *For hyperbolic discounted Markov decision problems with a sophisticated individual there exists a stationary optimal dynamic consistent strategy.*

Proof. Suppose that from the second stage on the sophisticated individual plays according to the stationary strategy $\bar{\rho} = (\bar{\rho}_\omega)_{\omega \in \Omega}$. Then the present values of the (standard) discounted stream of payoffs from the second stage on for each possible state that is met in the second stage, satisfies

$$\nu(\omega, \bar{\rho}) = u(\omega, \bar{\rho}_\omega) + \delta \sum_{\bar{\omega} \in \Omega} \pi(\bar{\omega} \mid \omega, \bar{\rho}_\omega)\, \nu(\bar{\omega}, \bar{\rho}) \qquad (\omega \in \Omega).$$

In vector matrix notation we can write the equations above as

$$\underline{\nu}(\bar{\rho}) = \underline{u}(\bar{\rho}) + \delta\, \underline{\pi}(\bar{\rho})\, \underline{\nu}(\bar{\rho}).$$

So,[4]

$$\underline{\nu}(\bar{\rho}) = [I - \delta\, \underline{\pi}(\bar{\rho})]^{-1} \underline{u}(\bar{\rho}).$$

The optimal response in the first stage against the strategy to play according to $\bar{\rho}$ in all future stages is found by maximizing the utility

$$U(\omega, \rho, \bar{\rho}) = u(\omega, \rho_\omega) + \beta\, \delta \sum_{\bar{\omega} \in \Omega} \pi(\bar{\omega} \mid \omega, \rho_\omega)\, \nu(\bar{\omega}, \bar{\rho}).$$

---

[4]The inverse is easily shown to exist as the matrix $\underline{\pi}(\bar{\rho})$ is stochastic and is multiplied by a factor smaller than 1.

for ρ. This leads to a best response correspondence $B : \Sigma \rightarrowtail \Sigma$ ($\bar{\rho} \mapsto \rho$). This correspondence is closed, upper hemi-continuous with nonempty convex compact values. According to Kakutani's fixed point theorem a fixed point exists. This shows the existence of a stationary optimal dynamic consistent strategy. $\square$

5 More players: stochastic game

In this section the previous section is extended to the multi-player case. This extension is not trivial as the current setting allows different types of individuals to interact. First three classes of stochastic games are considered to each of which a special subsection is devoted: (1) with only resolute individuals, (2) with only naive individuals, and (3) with only sophisticated individuals. For each case common knowledge in behavior is assumed, i.e. for the case with only resolute (naive/sophisticated) individuals, the resolution (naivety/sophistication) is common knowledge among the individuals. Finally, in the fourth subsection all types of individuals come together. Important here is the common knowledge of all individuals being either resolute, naive, or sophisticated and do not change type during the play of the game. It is however not necessary to now exactly who has which type.

5.1 When there are only resolute individuals

From Theorem 4.1 it follows that there might not exist a stationary best response against opponents playing stationary, as the decision problem of finding the optimal strategy against the opponents is Markovian due to the stationarity induced by the opponents behavior. But an optimal strategy that is stationary from the first period on exists. Let us call such a strategy *delayed-stationary*. Notice that this type of strategy is Markovian, since the strategy does not depend on the history, but only on state and time.

Theorem 5.1 *For hyperbolic discounted stochastic games with resolute individuals there is a delayed-stationary optimal best response against opponents playing delayed-stationary.*

Proof. Given that all opponents play delayed-stationary, the optimal response is to play stationary on the second period on, as from the second period on the game is equivalent to a stochastic game with (standard) discounted reward criterion. So, there is a delayed-stationary optimal response against opponents playing delayed-stationary. $\square$

In line with other refinements of the equilibrium concept, a delayed-stationary equilibrium

is defined as an equilibrium in delayed-stationary strategies. Quite naturally, a delayed-stationary equilibrium is also a Markov perfect equilibrium.

Theorem 5.2 *For every hyperbolic discounted stochastic game with resolute individuals a delayed-stationary equilibrium exists.*

Proof. From the second period on, stationary equilibria of the underlying stochastic game with (standard) discounted reward criterion are part of an equilibrium in the whole stochastic game with hyperbolic discounting. The decision in the initial period can be solved by backward induction and is equivalent to a normal form game. From the existence of both a stationary equilibrium for the underlying stochastic game with (standard) discounted reward criterion and a Nash equilibrium for the resulting normal form game, existence of a delayed-stationary equilibrium follows. $\square$

5.2 When there are only naive individuals

Theorem 5.3 *For hyperbolic discounted stochastic games with naive individuals there exists an optimal strategy against opponents playing stationary that gives rise to a stationary strategy in realization.*

Proof. From Theorem 4.2 it follows that, given that all opponents play stationary, there exists an optimal response that is delayed-stationary. As the individual is naive, he will gradually revise its strategy in every period. This leads to a stationary play-path. $\square$

Theorem 5.4 *For every hyperbolic discounted stochastic game with naive individuals there exists an equilibrium with a stationary equilibrium-path.*

Proof. Let $B^i : \Sigma^{-i} \twoheadrightarrow \Sigma^i$ be the correspondence that assigns to each combination of stationary strategies of the opponents the stationary play-path that is proved to exist in Theorem 5.3. This correspondence is closed, upper hemi-continuous with nonempty convex compact values. The compounded correspondence over all individuals $B : \Sigma \twoheadrightarrow \Sigma$ also has these properties. Since B is a correspondence from a set that is compact and convex valued into itself with the properties listed, a fixed point exists according to Kakutani's fixed point theorem. Fixed points of the correspondence B are stationary strategy combinations with the property that there exists an equilibrium having this stationary strategy combination as equilibrium-path. $\square$

5.3 When there are only sophisticated individuals

Theorem 5.5 *For hyperbolic discounted stochastic games with sophisticated individuals there exists a stationary optimal dynamic consistent strategy against opponents playing stationary.*

Proof. Given that all opponents play stationary, the decision to respond optimally is a Markov decision problem. From Theorem 4.3 it follows that a stationary optimal dynamic consistent response exists. □

Theorem 5.6 *For every hyperbolic discounted stochastic games with sophisticated individuals a stationary dynamic consistent equilibrium exists.*

Proof. Let $B^i : \Sigma^{-i} \twoheadrightarrow \Sigma^i$ be the best response correspondence of individual i over stationary (dynamic consistent) strategies only. As claimed in the proof to Theorem 4.3, this correspondence is closed, upper hemi-continuous with nonempty convex compact values. The compounded correspondence over all individuals $B : \Sigma \twoheadrightarrow \Sigma$ also has these properties. Again, applying Kakutani implies the existence of a fixed point and in turn the existence of a stationary dynamic consistent equilibrium. □

5.4 Mixtures of individuals

From the latter two subsection it follows that in a hyperbolic discounted stochastic game with only naive and sophisticated individuals present (and common knowledge of this information), an equilibrium inducing a stationary equilibrium-path exists. Namely, for a sophisticated individual, given stationary behavior of his opponents, there will be a stationary dynamic consistent best response. And, for a naive individual, given stationary behavior of his opponents, any optimal strategy will lead to a stationary strategy in realization.

When we add one resolute individual, this individual will have a delayed-stationary best response against all other individuals playing stationary. But for the naive and sophisticated individuals it will no longer be optimal to play stationary as the resolute individual does not play stationary (and therefore their private decision problem given the opponents strategies is no longer Markovian). But trivially, for a sophisticated individual, given delayed-stationary behavior of his opponents, there will be a delayed-stationary best response. And, also quite naturally, for a naive individual, given delayed-stationary behavior of his opponents, any optimal strategy will lead to a delayed-stationary strategy in realization. In return, for a resolute individual, given delayed-stationary behavior of his

opponents, there exists a delayed-stationary optimal response. This leads to the following theorem.

Theorem 5.7 *For hyperbolic discounted stochastic games an equilibrium exists that has a delayed-stationary equilibrium-path.*

6 References

1. ARRIBAS, I. AND A. URBANO (2003), "Repeated Games with Probabilistic Horizon," IVIE Working Paper: WP-AD 2000-29, Alicante.

2. FINK, A.M. (1964), "Equilibrium in a Stochastic n-Person Game," *Journal of Science of Hiroshima University Series A-I, 28,* 89-93.

3. FREDERICK, S., G. LOEWENSTEIN, AND T. O'DONOGHUE (2002), "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature, 40,* 351-401.

4. GILLETTE, D. (1957), "Stochastic Games with Zero-Stop Probabilities," in: M. Dresher, A.W. Tucker, and P. Wolfe, P. (eds.), *Contributions to the Theory of Games III, Annals of Mathematical Studies, 39,* Princeton University Press, Princeton, pp. 179-187.

5. HALLER, H. AND R. LAGUNOFF (2000), "Genericity and Markovian Behavior in Stochastic Games," *Econometrica, 68,* 1231-1248.

6. HARRIS, C. AND D. LAIBSON (2001), "Dynamic Choices of Hyperbolic Consumers," *Econometrica, 69,* 935-958.

7. HARRIS, C. AND D. LAIBSON (2001), "Hyperbolic Discounting and Consumption," *Forthcoming in Econometric Society Monographs.*

8. HERINGS, P.J.J. AND R.J.A.P. PEETERS (2004), "Stationary Equilibria in Stochastic Games: Structure, Selection, and Computation," *Forthcoming in Journal of Economic Theory.*

9. MERTENS, J.F. AND A. NEYMAN (1981), "Stochastic Games," *International Journal of Game Theory, 10,* 53-56.

10. O'DONOGHUE, T. AND M. RABIN (1999a), "Doing It Now or Later," *American Economic Review, 89,* 103-124.

11. O'DONOGHUE, T. AND M. RABIN (1999b), "Incentives and Procrastinators," *Quarterly Journal of Economics, 114,* 769-816.

12. PHELPS, E.S. AND R.A. POLLAK (1968), "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies, 35,* 185-199.

13. SHAPLEY, L.S. (1953), "Stochastic Games," *Proceedings of the National Academy of Sciences of the United States of America, 39,* 1095-1100.

14. SOBEL, M.J. (1971), "Non-cooperative Stochastic Games," *The Annals of Mathematical Statistics, 42,* 1930-1935.

15. STROTZ, R. (1955-1956), "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies, 23,* 165-180.

16. TAKAHASHI, M. (1964), "Equilibrium Points of Stochastic, Noncooperative n-Person Games," *Journal of Science of Hiroshima University Series A-I, 28,* 95-99.

17. THALER, R.H. (1981), "Some Empirical Evidence on Dynamic Inconsistency," *Economic Letters, 8,* 201-207.

18. THUIJSMAN, F. AND O.J. VRIEZE (1991), "Easy Initial States in Stochastic Games," in: T.E.S. Raghavan, T.S. Ferguson, T. Parthasarathy, and O.J. Vrieze (eds.), *Stochastic Games and Related Topics, in Honor of L.S. Shapley,* Kluwer Academic Publishers, Dordrecht, pp. 85-100.

19. VIEILLE, N. (2001), "The Existence of Equilibrium Payoffs in Two-Player Stochastic Games," in: E. Altman and O. Pourtallier (eds.), *Advances in Dynamic Games and Applications, (Annals of the International Society of Dynamic Games),* Birkhäuser, Boston, pp. 101-118.

20. VIEILLE, N. AND J.W. WEIBULL (2002), "Uniqueness in Infinitely Repeated Decision Problems," IUI Working Paper No. 577, Stockholm.