HARVARD | BUSINESS | SCHOOL

# The NTU-Value of Stochastic Games

**Elon Kohlberg**
**Abraham Neyman**

## Working Paper

**15-014**

**September 11, 2014**

# THE NTU-VALUE OF STOCHASTIC GAMES

ELON KOHLBERG* AND ABRAHAM NEYMAN+

## 1. Introduction

Since the seminal paper of Shapley [15], the theory of stochastic games has been developed in many different directions. However, there has been practically no work on the interplay between stochastic games and cooperative game theory.

Our purpose here is to make a first step in this direction. We show that the Harsanyi–Shapley–Nash cooperative solution to one-shot strategic games can be extended to stochastic games.

While this extension applies to general $n$-person stochastic games, it does not rely on Nash equilibrium analysis in such games. Rather, it only makes use of minmax analysis in two-person (zero-sum) stochastic games. This will become clear in the sequel.

## 2. The Shapley Value of Coalitional Games

A *coalitional game* is a pair $(N, v)$, where $N = \{1, \ldots, n\}$ is a finite set of players and $v\colon 2^N \to \mathbb{R}$ is a mapping such that $v(\emptyset) = 0$.

For any subset ("coalition") $S \subset N$, $v(S)$ may be interpreted as the total utility that the players in S can achieve on their own. Of course, such an interpretation rests on the assumption that utility is transferable among the players.

Shapley [15] introduced the notion of a "value", or an apriori assessment of what the play of the game is worth to each player. Thus a value is a mapping $\varphi\colon \mathbb{R}^{2^N} \to \mathbb{R}^N$ that assigns to each coalitional game $v$ a vector of individual utilities, $\varphi v$.

Shapley proposed four desirable properties, and proved that they imply a unique value mapping. This mapping – the Shapley Value – can be defined as follows:

$$\varphi_i v := \frac{1}{n!} \sum_{\mathcal{R}} (v(P_i^{\mathcal{R}} \cup i) - v(P_i^{\mathcal{R}})), \tag{1}$$

* Harvard Business School, Harvard University; *ekohlberg@hbs.edu*.

+ Institute of Mathematics, and Center for the Study of Rationality, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel; *aneyman@math.huji.ac.il*.

This research was supported in part by Israel Science Foundation grant 1596/10.

where the summation is over the $n!$ possible orderings of the set $N$ and where $P_i^{\mathcal{R}}$ denotes the subset of those $j \in N$ that precede $i$ in the ordering $\mathcal{R}$. From this formula, it is easy to see that the Shapley value has the following properties.

(2) **Efficiency** $$\sum_{i \in N} \varphi_i v = v(N).$$

(3) **Linearity** $$\varphi(\alpha v + \beta w) = \alpha \varphi v + \beta \varphi w \quad \forall \alpha, \beta \in \mathbb{R}.$$

**Note**: These are two of four properties that characterize the Shapley value. We spell them out because they will be used in the sequel.

Another property of the Shapley value that is used in the sequel is the following consequence of (1):

$$\varphi_i v \leq \max_{S \subset N} \left( v(S \cup i) - v(S) \right). \tag{4}$$

## 3. The NTU-Value of Strategic Games

A finite *strategic game* is a triple $G = (N, A, g)$, where

- $N = \{1, \ldots, n\}$ is a finite set of players,
- $A$ is the finite set of a player's pure strategies, and
- $g = (g^i)_{i \in N}$, where $g^i \colon A^N \to \mathbb{R}$, is player $i$'s payoff function.

**Remark**: In order to simplify the notation, we assume that the set of pure strategies is the same for all players. Since these sets are finite, there is no loss of generality.

We use the same notation, $g$, to denote the linear extension

- $g^i \colon \Delta(A^N) \to \mathbb{R}^N$,

where for any set K, $\Delta(K)$ denotes the probability distributions on $K$.

And we denote

- $A^i = A^S$ and $A^S = \prod_{i \in S} A^i$, and
- $X^S = \Delta(A^S)$ (correlated strategies of the players in $S$).

**Remark**: The notation $X^S = \Delta(A^S)$ is potentially confusing. Since $X = \Delta(A)$, it would seem that $X^S$ should stand for $(\Delta(A))^S$ (independent choices by the players in S) and not for $\Delta(A^S)$ (correlated choices). Still, we adopt this notation for its compactness.

**Remark**: Of course, $(\Delta(A))^S \subset \Delta(A^S)$.

In a strategic game, utility is not transferable between players; so there is no single number, $v(S)$, that captures what a coalition, $S$, can achieve on its own. In particular, then, there is no direct way to define the Shapley value of the game.

Nevertheless, inspired by the work of Harsanyi [7], Shapley [16], and Aumann and Kurz [1], we consider an indirect method for defining the value of a strategic game: Assume that utility becomes transferable after an appropriate multiplication by *scaling factors* $\lambda = (\lambda_1, \ldots, \lambda_n) \geq 0$, $\lambda \neq 0$. The total available to all players is then

$$v_\lambda(N) := \max_{x \in X^N} \sum_{i \in N} \lambda_i g^i(x). \tag{5}$$

**Note**: In a single-person maximization there is no advantage in using randomized strategies. So $v_\lambda(N) = \max_{a \in A^N} \sum_{i \in N} \lambda_i g^i(x)$ . We use the formulation in (5) merely in order to conform with (6).

In determining the amount that a coalition $S \neq N$ can achieve on its own, we apply the bargaining model of Nash [11]. In that model, the players in $S$ choose a "threat strategy", $x \in X^S$, which they commit to deploy if no agreement is reached; and similarly, the players in $N \setminus S$ choose a threat strategy $y \in X^{N \setminus S}$.

The model then prescribes that $S$ and $N \backslash S$ receive their "disagreement payoffs" $g^S(x, y) = \sum_{i \in S} \lambda_i g^i(x, y)$ and $g^{N \backslash S}(x, y) = \sum_{i \notin S} \lambda_i g^i(x, y)$, respectively, plus half the "surplus", $v_\lambda(N) - (g^S(x, y) + g^{N \backslash S}(x, y))$. In other words, $S$ receives $\frac{1}{2} v_\lambda(N) + \frac{1}{2}(g^S(x, y) - g^{N \backslash S}(x, y))$ while $N \backslash S$ receives $\frac{1}{2} v_\lambda(N) - \frac{1}{2}(g^S(x, y) - g^{N \backslash S}(x, y))$.

Since, in the context of the bargaining between $S$ and $N \backslash S$, the amount $v_\lambda(N)$ is fixed, $S$ will strive to maximize $g^S(x, y) - g^{N \backslash S}(x, y)$, while $N \backslash S$ will strive to minimize the same expression. Thus we define, for $S \subseteq N$:

$$v_\lambda(S) := \frac{1}{2} v_\lambda(N) + \frac{1}{2} \max_{x \in X^S} \min_{y \in X^{N \backslash S}} \left( \sum_{i \in S} \lambda^i g^i(x, y) - \sum_{i \notin S} \lambda^i g^i(x, y) \right). \tag{6}$$

**Note**: When $S = N$ this is the same formula as (5), considering that $N \setminus N = \emptyset$.

Having defined a game that captures in a single number the amount that each coalition can achieve on its own, we can use its Shapley value, $\varphi(v_\lambda)$, to define an "NTU (non-transferable-utility) value" of the strategic game.

There remains the question of choosing the scaling factors, $\lambda$. We do not specify a single $\lambda$ but rather accept any $\lambda$ for which the associated Shapley value can be implemented without actual transfers of utility. Thus we require that $\varphi(v_\lambda)$ be a rescaling of an allocation in the feasible set

$$F := \{g(x) : x \in X^N\} = \mathrm{conv}\{g(a) \colon a \in A^N\}.$$

In other words, using the notation

$$f * g := (f_i g_i)_{i \in N} \quad \forall f, g \in \mathbb{R}^N$$

we require that $\varphi v_\lambda = \lambda * \psi$, where $\psi \in F$.

Note that, by the linearity of the Shapley value, for every vector $\lambda$ of scaling factors, and for every $\alpha > 0$, if $\varphi(v_\lambda) = \lambda * \psi$ then $\varphi(v_{\alpha\lambda}) = \alpha\lambda * \psi$. Hence, we can normalize $\lambda$ to lie in the simplex

$$\Delta := \{\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n, \ \ \lambda \geq 0, \ \sum_{i \in N} \lambda_i = 1\}.$$

In summary:

**Definition 1.** $\psi \in F = \mathrm{conv}\{g(a) \colon a \in A^N\}$ is an NTU value of the strategic game $G$ if $\exists \lambda \in \Delta$ such that $\varphi(v_\lambda) = \lambda * \psi$, where $v_\lambda$ is the coalitional game defined by (6).

**Theorem 1.** *For every finite strategic game there exists an NTU value.*

This theorem is closely related to the results of Shapley [16] and Harsanyi [7]. A proof is provided in Appendix A. It is a special case of Neyman [12].

## 4. Stochastic Games

In a stochastic game, play proceeds in stages. At each stage, the game is in one of a finite number of states. Each one of $n$ players chooses an action from a finite set of possible actions. The players' actions and the state jointly determine a payoff to each player and transition probabilities to the succeeding state.

We assume that before making their choices, the players observe the current state and the previous actions.

**Definition 2.** A finite *stochastic game-form* is a tuple $\Gamma = (N, Z, A, g, p)$ , where

- $N = \{1, 2, \ldots, n\}$ is a finite set of players
- $Z$ is a finite set of states
- $A$ is the finite set of a player's stage actions
- $g = (g^i)_{i \in N}$, where $g^i \colon Z \times A^N \to \mathbb{R}$ is the stage payoff to player $i$, and

- $p\colon Z \times A \to \Delta(Z)$ are the transition probabilities.

**Remark**: We use the same notation $N$, $A$, $g$, as in a strategic game. The different meanings should be apparent from the context.

**Remark**: Again we make the simplifying assumption that the set of stage actions, A, is the same for all players; furthermore, we assume that the set of actions is independent of the state. In other words, if $A^i[z]$ denotes player $i$'s set of actions in state $z$, then $A^i[z] = A$ for all $i$ and $z$.

In order to define a specific *stochastic game* we must indicate the players' strategies and their payoffs. We denote the players' behavioral strategies in the infinite game by

- $\sigma^i_t\colon (Z \times A^N)^{t-1} \times Z \to \Delta(A)$ and
- $\sigma^i = (\sigma^i_t)^\infty_{t=1}$ , $\sigma = (\sigma^i)_{i\in N}$.

The strategies $\sigma$ along with the initial state $z$ determine a probability distribution $P^z_\sigma$ over the plays of the infinite game, and hence a probability distribution over the streams of payoffs. The expectation with respect to this distribution is dented by $E^z_\sigma$.

Of course, there are many possible valuations of the streams of payoffs. One standard valuation is obtained by fixing a number of stages, k. We denote:

- $\gamma^i_k(\sigma)[z] = E^z_\sigma \frac{1}{k}\sum_{t=1}^k g^i(z_t, a_t)$,
- $\gamma^i_k(\sigma) = (\gamma^i_k(\sigma)[z])_{z\in Z}$, and
- $\gamma_k(\sigma) = (\gamma^i_k(\sigma))_{i\in N}$.

We refer to the game with this valuation as the *k-stage game* and denote it by $\Gamma_k$.

Another standard valuation is obtained by applying a discount rate, $0 < r < 1$. We denote:

- $\gamma^i_r(\sigma)[z] = E^z_\sigma \Sigma^\infty_{t=1} r(1-r)^{t-1} g^i(z_t, a_t)$,
- $\gamma^i_r(\sigma) = (\gamma^i_r(\sigma)[z])_{z\in Z}$, and
- $\gamma_r(\sigma) = (\gamma^i_r(\sigma))_{i\in N}$.

We refer to the game with this valuation as the *r-discounted game* and denote it by $\Gamma_r$.

**Note**: In fact, $\Gamma_r$ is a family of games, $\Gamma^z_r$, parameterized by the initial state. Similarly for $\Gamma_k$.

We denote by $v_r$, respectively $v_k$, the minmax value of $\Gamma_r$, respectively $\Gamma_k$.

## 5. The $r$-discounted game: Two-person zero-sum

In a two-person zero-sum stochastic game, $N = \{1, 2\}$ and $g^2 = -g^1$. To simplify the notation, we denote $\sigma^1 = \sigma$, $\sigma^2 = \tau$ and $\gamma_r(\sigma, \tau) = \gamma^1_r(\sigma^1, \sigma^2)$ and similarly $\gamma_k(\sigma, \tau) = \gamma^1_k(\sigma^1, \sigma^2)$.

**Definition 3.** $v \in \mathbb{R}^Z$ is the minmax value of the $r$-discounted game (respectively, the $k$-stage game) if $\exists \sigma_0, \tau_0$ s.t. $\forall \sigma, \tau$

$$\gamma_r(\sigma_0, \tau) \geq v \geq \gamma_r(\sigma, \tau_0) \text{ (respectively, } \gamma_k(\sigma_0, \tau) \geq v \geq \gamma_k(\sigma, \tau_0)).$$

**Note**: The vector notation above says that, for all $z \in Z$, $v[z]$ is the minmax value of the game with initial state $z$.

We denote by $\mathrm{Val}(G)$ the minmax value of a two-person zero-sum strategic game $G$.

**Theorem 2.** *(Shapley 1953) Let $\Gamma_r$ be a two-person zero-sum r-discounted stochastic game.*

- *$\Gamma_r$ has a minmax value and stationary optimal strategies. Furthermore:*
- *$(v[z])_{z \in Z}$ is the minmax value of $\Gamma_r$ with initial state $z$ iff it is the (unique) solution of the equations*

$$v[z] = Val\, G_r[z, v] \ \ \forall z \in Z \tag{7}$$

  *where*

$$G_r[z, v](a) := rg(z, a) + (1 - r)\Sigma_{z'} p(z, a)[z']v[z'].$$

- *If $x_r[z]$ and $y_r[z]$ are optimal strategies for players 1 and 2, respectively, in the (one-shot) game $G_r[z, v]$, then the stationary strategies $\sigma_t = x_r$, $\tau_t = y_r$ $\forall t$ are optimal strategies in $\Gamma_r$, and*

We denote by $v_r$, respectively $v_k$, the minmax value of $\Gamma_r$, respectively $\Gamma_k$. (The existence of $v_k$ is obvious, as $\Gamma_k$ is a finite game.)

## 6. Markov Decision Processes

A single-person stochastic game is known as a Markov Decision Process (MDP). Since in a single-person one-shot game the player has a pure optimal strategy, Theorem 2 implies:

**Corollary 1.** In an $r$-discounted MDP there exists an optimal strategy that is stationary and pure.

**Note**: The same corollary applies to stochastic games with perfect information. In such games, at each state $z$ one player is restricted to a single action, i.e., $A^1[z]$ or $A^2[z]$ consists of a single point.

In fact, the corollary can be substantially strengthened:

**Theorem 3.** *(Blackwell 1962) In every MDP there exists a uniformly optimal pure stationary strategy. That is, there exists a pure stationary strategy $\sigma^*$ such that*

(i) $\sigma^*$ *is optimal in the $r$-discounted MDP for all $r < r_0$ for some $r_0 > 0$. Furthermore:*

(ii) $\forall \varepsilon > 0,\ \exists k_\varepsilon > 0$, *such that $\sigma^*$ is $\varepsilon$-optimal in the $k$-stage game for all $k > k_\varepsilon$, and*

(iii) $\bar{g}_k := \frac{1}{k}\sum_{t=1}^{k} g(a_t, z_t)$ *converges $P_{\sigma^*}$ a.e., and* $E_{\sigma^*} \lim_{k\to\infty} \bar{g}_k \geq E_\sigma \limsup_{k\to\infty} \bar{g}_k \ \ \forall \sigma$.

**Note**: For completeness, we provide a proof of Blackwell's theorem in Appendix C.

**Notes**:

- The limit of $\bar{g}_k$ exists $P_{\sigma^*}$ a.e. because a stationary strategy induces fixed transition probabilities on the states, resulting in a Markov chain.
- It follows that the $\lim_{k\to\infty} v_k = \lim_{k\to\infty} E_{\sigma^*}\bar{g}_k$ exists. This implies that $v_r$ converges to the same limit. (One way to see this is to apply Lemma 1 below.)
- Statement (ii) is, of course, equivalent to
  (ii') $\sigma^*$ is $\varepsilon$-optimal in the k-stage game for all but finitely many values of $k$.
- While the theorem guarantees the existence of a strategy that is optimal uniformly for all small $r$, it only guarantees the existence of a strategy that is $\varepsilon$-optimal uniformly for all large $k$. To see that the optimal strategy in the $k$-stage game might depend on $k$, consider the following example: In state 1, one action yields 0 and transition to state 2; the other action yields 1 and the state is unchanged. In state 2, there is a single action yielding 3 and with probability .9 the state is unchanged. The unique optimal strategy is to play the first action in the first $k-1$ stages and the second action in stage $k$.

Blackwell's Theorem establishes the existence of a stationary strategy that is optimal in a very strong sense. It is simultaneously optimal in all the $r$-discounted games with $r > 0$ sufficiently small, and (essentially) simultaneously optimal in all the $k$-stage games with $k$ sufficiently large; it is also optimal when infinite streams of payoffs are evaluated by their limiting average.

In other words, Blackwell's Theorem establishes the existence of a stationary strategy that is optimal in the MDP under any one of the three main interpretations of the infinite-stage model:

(i) Future payoffs are discounted at a very small positive but unspecified discount rate, or – equivalently – at every stage the game stops with some very small positive probability.

(ii) The "real" game is finite, with a large but unspecified number of stages.

(iii) There is an unspecified valuation of infinite streams of payoffs. This valuation lies between the lim inf and the lim sup of the average payoff in the first k stages.

Blackwell's Theorem also implies the existence of a *value*, i.e., a maximal payoff that can be (uniformly) guaranteed according to each of the three interpretations above.

Indeed, let $v := \lim_{r\to 0} v_r = lim_{k\to\infty} v_k$. Then $\forall \varepsilon > 0 \ \exists r'_\varepsilon,\ k'_\varepsilon$ s.t. $\forall\, \sigma$

(i') $\varepsilon + \gamma_r(\sigma^*) \geq v \geq \gamma_r(\sigma) - \varepsilon \ \ \forall \ \ 0 < r < r'_\varepsilon$,

(ii') $\varepsilon + \gamma_k(\sigma^*) \geq v \geq \gamma_k(\sigma) - \varepsilon \ \ \forall \ \ k > k'_\varepsilon$, and

(iii') $E_{\sigma^*} \liminf_{k\to\infty} \bar{g}_k \geq v \geq E_\sigma \limsup_{k\to\infty} \bar{g}_k$.

The left inequalities indicate that the payoff $v$ is guaranteed by the strategy $\sigma^*$; and the right inequalities indicate that no larger payoff can be guaranteed by any strategy.

## 7. The undiscounted game: two-person zero-sum

In an undiscounted two-person zero-sum stochastic game it is not obvious how to define the value and optimal strategies.

A natural first attempt is to proceed in analogy with Blackwell's Theorem for MDPs. First, define a pair of strategies $\sigma_0$, $\tau_0$ for player 1 and 2, respectively, to be optimal, if there exist $r_0 > 0$ and $k_0 > 0$ such that, for all $\sigma, \tau$,

(i) $\gamma_r(\sigma_0, \tau) \geq \gamma_r(\sigma, \tau_0) \ \ \forall\, 0 < r < r_0$.

(ii) $\gamma_k(\sigma_0, \tau) \geq \gamma_k(\sigma, \tau_0) \ \ \forall\, k > k_0$.

(iii) $E_{\sigma_0,\tau} \liminf_{k\to\infty} \bar{g}_k \geq E_{\sigma,\tau_0} \limsup_{k\to\infty} \bar{g}_k$.

(Note that (ii) holds in MDPs within $\varepsilon$.)
Next, prove the existence of stationary strategies satisfying these conditions.

However, it turns out that for some games there exist no stationary strategies that satisfy either (i), or (ii), or (iii), even within an $\epsilon$.

This is illustrated by the game known as the Big Match (Gilette [6]) where, moreover, there are even no Markov strategies that satisfy either (i), or (ii), or (iii) within an $\epsilon$ ([5]).

The main difficulty in the transition from MDPs to two-person games is this: In an r-discounted MDP, the same strategy that is optimal for some small $r$ is also optimal for other small $r$; but this is not so in two-person games. For example, the unique optimal strategy for Player 1 in the r-discounted Big Match, while guaranteeing the minmax value of that game, only guarantees (approximately) the maxmin in pure strategies in the $r^2$-discounted Big Match.

However, upon reflection, it appears that if we wish to define the notion of a player "guaranteeing" a certain payoff in the undiscounted game, then the essential requirement should be this: For any $\varepsilon > 0$ there is a strategy guaranteeing the payoff up to $\varepsilon$, simultaneously in all the r-discounted games with $r$ sufficiently small. It is *not* essential that this strategy be stationary or that it be independent of $\varepsilon$, as is the case in MDPs.

In other words, our requirement should be an analog of conditions (i') -(iii') above, where the strategy $\sigma^*$ may depend on $\varepsilon$ and it need not be stationary (or even Markov).

Thus we may define $v$ to be the minmax value of the game if Player 1 can guarantee $v$ and Player 2 can guarantee $-v$. Formally, we have:

Let $\sigma, \sigma_\varepsilon$ denote strategies of player 1 and $\tau, \tau_\varepsilon$ denote strategies of player 2.

**Definition 4.** $v \in \mathbb{R}^Z$ is the *(minmax) value* of a two-person zero-sum stochastic game if $\forall \varepsilon > 0,\ \exists \sigma_\varepsilon, \tau_\varepsilon,\ r_\varepsilon > 0,$ and $k_\varepsilon > 0$ s.t. $\forall \sigma, \tau$

(i) $\varepsilon + \gamma_r(\sigma_\varepsilon, \tau) \geq v \geq \gamma_r(\sigma, \tau_\varepsilon) - \varepsilon \quad \forall\, 0 < r < r_\varepsilon$.
(ii) $\varepsilon + \gamma_k(\sigma_\varepsilon, \tau) \geq v \geq \gamma_k(\sigma, \tau_\varepsilon) - \varepsilon \quad \forall\, k > k_\varepsilon$.
(iii) $\varepsilon + E_{\sigma_\varepsilon, \tau} \liminf_{k \to \infty} \bar{g}_k \geq v \geq E_{\sigma, \tau_\varepsilon} \limsup_{k \to \infty} \bar{g}_k - \varepsilon$,

**Notes:**

- Condition (i) can be dropped from the definition as it is a consequence of condition (ii). (See below.)
- $v \in \mathbb{R}$ is the *uniform*, respectively, the *limiting-average*, *value* of a two-person zero-sum stochastic game if $\forall \varepsilon > 0,\ \exists \sigma_\varepsilon, \tau_\varepsilon$ and $k_\varepsilon > 0$ s.t. $\forall \sigma, \tau$ (ii), respectively, (iii), holds.
- Obviously, if the value, respectively, the uniform value or the limiting-average exists, then it is unique.
- If a minmax value, $v$, exists then $v = \lim_{r \to 0} v_r = lim_{k \to \infty} v_k$.

We now show that (ii) implies (i). More generally, (ii) implies

(iv) $\forall \varepsilon > 0, \exists \sigma_\varepsilon, \tau_\varepsilon$ and $w_\varepsilon > 0$ s.t. $\forall \sigma, \tau$ and for any non-increasing sequence of non-negative numbers $(w_t)_{t=1}^\infty$ that sum to 1, if $w_1 < w_\varepsilon$, then

$$\varepsilon + \gamma_w(\sigma_\varepsilon, \tau) \geq v \geq \gamma_w(\sigma, \tau_\varepsilon) - \varepsilon \quad \forall (w_t)\ s.t.\ w_1 < w_\varepsilon,$$

where $\gamma_w(\sigma)[z] := E_\sigma^z \sum_{t=1}^\infty w_t g(z_t, a_t)$.

This follows from the Lemma below.

**Lemma 1.** Any non-increasing sequence of non-negative numbers $(w_t)$ that sum to 1 is an average of sequences of the form $e(k)_{t=1}^{\infty}$, where $e(k)_t = \frac{1}{k}$ for $t \leq k$ and $e(k)_t = 0$ for $t > k$.

*Proof.* It is easy to see that $(w_t) = \sum_{t=1}^{\infty} \alpha_k e_k$, where $\alpha_k = k(w_k - w_{k+1})$. Clearly, $\alpha_k \geq 0$ and $\sum_{k=1}^{\infty} \alpha_k = \sum_{k=1}^{\infty} w_k = 1$. □

**Theorem 4.** *(Mertens and Neyman 1981)*
*Every finite two-person zero-sum stochastic game has a minmax value.*

We denote the minmax value by VAL(Γ).

**Notes**:

- The first step towards a proof was taken by Blackwell and Ferguson [5]. They showed that in the Big Match, for any $\varepsilon > 0$, there exist non-Markov strategies that satisfy (iii) within an $\varepsilon$. This was extended by Kohlberg [8] to a special class of stochastic games – repeated games with absorbing states. The general definition and existence theorem were provided by Mertens and Neyman [9].
- A priori there is no reason to rule out the possibility that the uniform value exist while the limiting-average value does not, or vice versa, or that both exist but differ. However, the existence theorem for the value implies that (in a finite stochastic game) both the uniform and the limiting-average values exist and are equal.
- A consequence of the above is that our results apply to the undiscounted value, whether we consider the uniform or the limiting-average value.

**Corollary 2.** Let $v_r$ (respectively, $v_k$) denote the minmax value of the $r$-discounted game (respectively, the k-stage game). Then $v = \text{VAL}(\Gamma)$ iff $v = lim_{r\to 0} v_r$
(respectively, $v = lim_{k\to\infty} v_k$).

**Corollary 3.** If $\Gamma = (N, Z, A, g, p)$ and $\Gamma' = (N, Z, A, g', p)$ then

$$\|\text{VAL}(\Gamma) - \text{VAL}(\Gamma')\|_\infty := \max_{z\in Z} |\text{VAL}(\Gamma)[z] - \text{VAL}(\Gamma')[z]| \leq \|g - g'\|_\infty,$$

where $\|g\|_\infty := \max_{(z,a)\in Z\times A} |g(z, a)|$.

To prove the corollary, first note that the stage payoffs in the games Γ and Γ′ differ by at most $\|g - g'\|_\infty$. Therefore, an optimal strategy in $\Gamma_k$ guarantees $v_k - \|g - g'\|_\infty$ in $\Gamma'_k$, and vice versa; hence $\|v_k - v'_k\| \leq \|g - g'\|_\infty$. Next, let $k \to \infty$ and apply the previous corollary.

**Corollary 4.** Every MDP has a uniform value.

**Note**: Of course, this corollary also follows from Blackwell's Theorem.

## 8. NTU-value of the undiscounted game

We now proceed to define the NTU-value of a stochastic game in analogy with the definition for strategic games.

Let $\Gamma$ be a stochastic game. For every $S \subseteq N$, denote by

- $X^S = \Delta(A^S)$ the set of all correlated stage actions of the players in $S$.
- $\sigma_t^S\colon (Z \times A^N)^{t-1} \times Z \to X^S$ a correlated stage strategy of the players in $S$ at time $t$.
- $\sigma^S = (\sigma_t^S)_{t=1}^{\infty}$ a correlated behavior strategy of the players in $S$.
- $\Sigma^S = \{\sigma^S\}$ the set of all correlated behavior strategies of the players in $S$.

In addition, denote by

- $\Sigma^N_{s.p.}$ the finite set of stationary pure strategies in $\Sigma^N$.

We define the feasible set $F_0 \subset \mathbb{R}^N$ as follows:

$$F_0 := \{x\colon \exists\sigma \in \Sigma^N \text{ s.t. } x = \lim_{r\to 0} \gamma_r(\sigma)\}$$

**Lemma 2.**

$$\begin{aligned} F_0 = {} & \operatorname{conv}\{x\colon \exists\sigma \in \Sigma^N_{s.p.} \text{ s.t. } x = \lim_{r\to 0} \gamma_r(\sigma)\} \\ = {} & \{x\colon \exists\sigma \in \Sigma^N \text{ s.t. } x = \lim_{k\to \infty} \gamma_k(\sigma)\} \\ = {} & \operatorname{conv}\{x\colon \exists\sigma \in \Sigma^N_{s.p.} \text{ s.t. } x = \lim_{k\to \infty} \gamma_k(\sigma)\}. \qquad (8) \end{aligned}$$

**Note**: The lemma says that $F_0$ is a convex polytope spanned by the limiting expected payoffs of the finitely many pure stationary strategies, where the limits can be taken either as $\lim_{r\to 0} \gamma_r(\sigma)$ or as $\lim_{k\to\infty} \gamma_k(\sigma)$.

*Proof.* We first show that $F_0$ is convex. Let $x', x'' \in F_0$. Then $\exists \sigma', \sigma'' \in \Sigma^N$ s.t. $x' = \lim_{r\to 0} \gamma_r(\sigma')$ and $x'' = \lim_{r\to 0} \gamma_r(\sigma'')$. By Kuhn's Theorem $\exists \hat\sigma \in \Sigma^N$ inducing the same distribution on the plays of the game as the mixed strategy $\frac{1}{2}\sigma' + \frac{1}{2}\sigma''$. So $\gamma_r(\hat\sigma) = \gamma_r(\frac{1}{2}\sigma' + \frac{1}{2}\sigma'') = \frac{1}{2}\gamma_r(\sigma') + \frac{1}{2}\gamma_r(\sigma'')$ and therefore
$F_0 \ni \lim_{r\to 0} \gamma_r(\hat\sigma) = \frac{1}{2}\lim_{r\to 0} \gamma_r(\sigma') + \frac{1}{2}\lim_{r\to 0} \gamma_r(\sigma'') = \frac{1}{2}x' + \frac{1}{2}x''$.

Next we note that, since $F_0$ is convex, $F_0 \supseteq \operatorname{conv}\{x\colon \exists\sigma \in \Sigma^N_{s.p.} \text{ s.t. } x = \lim_{r\to 0} \gamma_r(\sigma)\}$. To prove the equality, assume $F_0 \ni x_0 \notin \operatorname{conv}\{x\colon \exists\sigma \in \Sigma^N_{s.p.} \text{ s.t. } x = \lim_{r\to 0} \gamma_r(\sigma)\}$.

Then there is a separating linear functional, $y \in \mathbb{R}^N$, such that

$$\langle y, x_0\rangle \;>\; \langle y, x\rangle \;\forall x = \lim_{r\to 0}\gamma_r(\sigma) \; s.t.\; \sigma \in \Sigma^N_{s.p.}.$$

But this contradicts Theorem 3 w.r.t. the MDF with stage payoff $\langle y, g\rangle$.

A similar argument shows that the second set of limits is also a convex polytope spanned by the limiting expected payoffs of the pure stationary strategies.

Finally, note that if $\sigma$ is a stationary strategy, then $\lim_{r\to 0}\gamma_r(\sigma) = \lim_{k\to\infty}\gamma_k(\sigma)$ (see Lemma 5). Thus the first and the third sets in (8) are equal, and therefore all three sets are identical to $F_0$.

□

For future reference, we note the following.

**Lemma 3.** Let $F_0(\lambda) := \{\lambda * x : x \in F_0\}$. Then

(i) If $y \in F_0(\lambda)$ then $y_i \leq \lambda_i\|g^i\| \;\; \forall i \in N$, and
(ii) The mapping $\lambda \to F_0(\lambda)$ is continuous.

We denote by $\Gamma^S_\lambda$ the two-person zero-sum stochastic game played between $S$ and $N\backslash S$, where the pure stage actions are $A^S$ and $A^{N\backslash S}$, respectively, and where the stage payoff to S is given by

$$\sum_{i\in S}\lambda_i g^i(z, a^S, a^{N\backslash S}) - \sum_{i\notin S}\lambda_i g^i(z, a^S, a^{N\backslash S}).$$

And we denote by VAL($\Gamma$) the uniform (minmax) value of the two-person zero-sum game, or MDP, $\Gamma$.

We can now define the NTU (Shapley) value for stochastic games analogously to Definition 1 for strategic games.

**Definition 5.** $\psi \in F_0$ is an NTU-value of the stochastic game $\Gamma$ if $\exists\lambda \in \Delta$ such that $\varphi(v_\lambda) = \lambda * \psi$, where $v_\lambda$ is the coalitional game defined by

$$v_\lambda(S) := \frac{1}{2}\mathrm{VAL}(\Gamma^N_\lambda) + \frac{1}{2}\mathrm{VAL}(\Gamma^S_\lambda) \quad \forall S \subseteq N$$

**Note**: In the case $S = N$, $v_\lambda(N) = \mathrm{VAL}(\Gamma^N_\lambda)$ is the maximal expected payoff in the MDP with the single player $N$, where the pure stage actions are $A^N$ and the stage payoff is $\sum_{i\in N}\lambda_i g^i(z, a)$.

Our main result is as follows.

**Theorem 5.** *For every finite stochastic game there exists an NTU- value.*

The proof is presented in Appendix B.

## 9. NTU-value of the $r$-discounted game

We now define an NTU value of the $r$-discounted game $\Gamma_r$. The required steps are obviously simpler than in the undiscouted case,

We define the feasible set, $F_r$, as follows:

$$\begin{aligned} F_r := & \quad \{\gamma_r(\sigma)\colon \sigma \in \Sigma^N\} \\ = & \quad \mathrm{conv}\{\gamma_r(\sigma)\colon \sigma \in \Sigma^N_{s.p.}\}. \end{aligned} \tag{9}$$

**Note**: The equation says that $F_r$ is a convex polytope spanned by the expected payoffs of the finitely many pure stationary strategies. It is a simple analog of the first equation in Lemma 2.

Since every two-person zero-sum r-discounted stochastic game has a minmax value (Theorem 2), an NTU-value can be defined in the same way as for strategic games.
Let

$$\mathrm{Val}(\Gamma^S_{r,\lambda_r}) := \max_{\sigma \in \Sigma^S} \min_{\tau \in \Sigma^{N \setminus S}} \left( \sum_{i \in S} \lambda^i_r \gamma^i_r(\sigma, \tau) - \sum_{i \notin S} \lambda^i_r \gamma^i_r(\sigma, \tau) \right).$$

**Definition 6.** $\psi_r \in F_r$ is an NTU-value of the $r$-discounted stochastic game $\Gamma_r$ if $\exists \lambda_r \in \Delta$ such that $\varphi(v_{r,\lambda_r}) = \lambda_r * \psi$, where $v_{r,\lambda_r}$ is the coalitional game defined by

$$v_{r,\lambda_r}(S) := \frac{1}{2}\mathrm{Val}(\Gamma^N_{r,\lambda_r}) + \frac{1}{2}\mathrm{Val}(\Gamma^S_{r,\lambda_r}) \quad \forall S \subseteq N \tag{10}$$

**Note**: In the case $S = N$, $v_{r,\lambda_r}(N) = \mathrm{Val}(\Gamma^N_{r,\lambda_r}) = \max_{\sigma \in \Sigma^N} \sum_{i \in N} \lambda^i_r \gamma^i_r(\sigma)$

**Theorem 6.** *For every r-discounted stochastic game there exists an NTU-value.*

The proof proceeds in complete analogy with the proof in Appendix A for strategic games. In particular, it is easy to verify that properties $(i)$ and $(ii)$, required in that proof, are still valid.

## 10. Asymptotic Expansions

Recall that an *atomic formula* is an expression of the form $p > 0$ or $p = 0$, where p is a polynomial with integer coefficients in one or more variables; an *elementary formula* is an expression constructed in a finite number of steps from atomic formulae by means of conjunctions ($\wedge$), disjunctions ($\vee$), negations ($\sim$), and quantifiers of the form "there exists" ($\exists$) or "for all" ($\forall$). A variable is *free* in a formula if somewhere in the formula it is not modified by a quantifier $\exists$ or $\forall$. An *elementary sentence* is an elementary formula with no free variables.

**Lemma 4.** For fixed (N,Z,A) the statement of Theorem 6 is an elementary sentence.

The proof is given in Appendix D.

If we think of the variables as belonging to a certain ordered field, then a sentence is either true or false. For instance, the sentence $\forall x\ \exists y\ s.t.\ y^2 = x$ is false over the field of real numbers but true over the field of complex numbers.

An ordered field is said to be *real closed* if no proper algebraic extension is ordered. Tarski's principle states that an elementary sentence that is true over one real-closed field is true over every real-closed field. (See, e.g., [2].)

It is well known that the field of power series in a fractional power of $r$ (real Puiseux series) that converge for $r > 0$ sufficiently small, ordered according to the assumption that $r$ is "infinitesimal" (i.e., $r < a$ for any real number $a > 0$), is real-closed. (See, e,g, [2], or [14].)

Thus, given Theorem 6 and Lemma 4, Tarski's principle implies the following:

**Theorem 7.** *Fix (N,Z,A). For every* $1 > r > 0$ *there exist* $\psi_r \in \mathbb{R}^N$, $\lambda_r \in \mathbb{R}^N$ *and* $v_{r,\lambda_r} \in \mathbb{R}^{2^N}$ *satisfying the NTU-value conditions* (14) *to* (17)*, such that each one of these variables has an expansion of the form*

$$\Sigma_{k=0}^{\infty} \alpha_k r^{k/M} \tag{11}$$

*that converges for* $r > 0$ *sufficiently small.*

**Note**: The general form of an element of the field of real Puiseaux series is $\Sigma_{k=-K}^{\infty} \alpha_k r^{k/M}$. However, because the $\psi$, $\lambda$, and $v$ are bounded, K=0.

We now apply this result to derive an asymptotic version of Theorem 5.
Let $r \to 0$. In light of the asymptotic expansion (11), $\psi_r \to \psi_0$, $\lambda_r \to \lambda_0 \in \Delta$, and $v_{r,\lambda_r} \to v_{\lambda_0}$.
By Lemma 6 (in Appendix C), $\psi_0 \in F_0$. By Corollary 4, $v_{\lambda_0}$ is the uniform minmax value of $\Gamma_{\lambda_0}^S$ for all $S \subseteq N$. Thus, $\psi_0$, $\lambda_0$, and $v_{\lambda_0}$ satisfy the requirements of Definition 5; hence $\psi_0$ is an NTU-value of $\Gamma$.

**Theorem 8.** *Every finite stochastic game* $\Gamma$ *has an NTU-value that is the limit, as* $r \to 0$*, of NTU-values of the* $r$*-discounted games. Furthermore, these NTU-values, as well as their scaling factors and the associated minmax values and optimal strategies in the zero-sum scaled games, are real Puiseaux series converging to their counterparts in the game* $\Gamma$*.*

**Note**: The above provides an alternative proof for the existence of an NTU-value in stochastic games.

**Note**: An alternative proof of Theorem 7 is obtained by noting that for every fixed $(N, Z, A, g, p)$, the set of tuples $(r, \psi_r, \lambda_r, v_{r,\lambda_r})$ that satisfy the NTU-value conditions (14) to (17) is a semi-algebraic set, whose projection on the first $(r)$ -coordinate is $(0, 1)$. Therefore, there is a function $r \mapsto (\psi_r, \lambda_r, v_{r,\psi_r})$, such that each one of its coordinates has an expansion of the form (11). (See, [14]).

## 11. Discussion

The paper details the extension of the Harsanyi-Shapley-Nash cooperative solution for one-shot strategic games to finite stochastic games. The properties of a finite stochastic game that are used are: A) finitely many players, states, and actions, B) complete information, and C) perfect monitoring, i.e., current state and players' past actions are observable.

In the general model of a repeated game, which can be termed a *stochastic game with incomplete information and imperfect monitoring*, the stage payoff and the state transitions are as in a classical stochastic game, but the initial state is random, and each player receives a stochastic signal about players' previous stage actions and current state.

The result that for each fixed $1 > r > 0$ the $r$-discounted game has an NTU-value, as well as its proof, are both identical to those given here for the finite stochastic game with perfect monitoring. The existence of an NTU-value in the undiscounted case depends on the existence of a uniform value in the corresponding two-person zero-sum model. Note, however, that the existence of an NTU-value in the undiscounted game does not depend on the existence of equilibrium payoffs in the corresponding undiscounted games.

## 12. Appendix A: Existence of NTU-value in strategic games

**Theorem 1** *For every finite strategic game there exists an NTU-value.*

*Proof.* Recall that $F = \text{conv}\{g(a) \colon a \in A\}$. Let $F(\lambda) = \{\lambda * x : x \in F\}$ and $E(\lambda) = \big\{y \in F(\lambda) \bigm| \sum_{i \in N} y_i \text{ is maximal on } F(\lambda)\big\}$. We claim that

(i) $y_i \leq K\lambda_i \quad \forall y \in E(\lambda)$

(ii) $\varphi_i(v_\lambda) \geq -K\lambda_i \quad \forall \lambda \in \Delta$,

where $K := \max_{i\in N} \max_{a\in A} |g^i(a)|$ denotes the largest absolute value of a payoff in G.
To see (i), note that $|x_i| \le K\ \forall x \in F$; therefore $|y_i| \le K\lambda_i\ \forall y \in F(\lambda)$, and in particular $y_i \le K\lambda_i\ \forall y \in E(\lambda)$.
To see (ii), note that, by (6),

$$\begin{aligned}
& 2v_\lambda(S\cup i) - v_\lambda(N) \\
= \ & \max_{x\in X^{S\cup i}} \min_{y\in X^{N\setminus(S\cup i)}} \left(\sum_{j\in S\cup i} \lambda_j g^j(x,y) - \sum_{j\notin S\cup i} \lambda_j g^j(x,y)\right) \\
\ge \ & \max_{x\in X^{S}} \min_{y\in X^{N\setminus S}} \left(\sum_{j\in S\cup i} \lambda_j g^j(x,y) - \sum_{j\notin S\cup i} \lambda_j g^j(x,y)\right) \\
\ge \ & \max_{x\in X^{S}} \min_{y\in X^{N\setminus S}} \left(\sum_{j\in S} \lambda_j g^j(x,y) - \sum_{j\notin S} \lambda_j g^j(x,y)\right) - 2K\lambda_i \\
= \ & 2v_\lambda(S) - v_\lambda(N) - 2K\lambda_i
\end{aligned} \tag{12}$$

so that

$$v_\lambda(S\cup i) - v_\lambda(S) \ge -K\lambda_i \ \ \forall S \not\ni i. \tag{13}$$

Since $\varphi_i v_\lambda$ is an average of the marginal contributions $v_\lambda(S\cup i) - v_\lambda(S)$, this implies (ii).

We now define a correspondence $H : \Delta \to \mathbb{R}^N$ as follows:

$$H(\lambda) := \left\{ \lambda + \frac{\varphi(v_\lambda) - y}{2K} \;\middle|\; y \in E(\lambda) \right\}.$$

We wish to show that $H(\lambda) \subset \Delta$.
Let $z \in H(\lambda)$. Since the Shapley value is efficient, $\varphi(v_\lambda)$ lies in $E(\lambda)$, which implies that $\sum_{i\in N}(\varphi(v_\lambda) - y)_i = 0$ for any $y \in E(\lambda)$. Thus $\sum_{i\in N} z_i = \sum_{i\in N} \lambda_i = 1$.

It remains to show that $z_i \ge 0$. Indeed, by (ii) and (i),

$$z_i = \lambda_i + \frac{\varphi_i(v_\lambda) - y_i}{2K} \ge \lambda_i + \frac{-K\lambda_i - K\lambda_i}{2K} \ge \lambda_i - \lambda_i = 0$$

Rewriting

$$H(\lambda) = (\lambda + \frac{\varphi v_\lambda}{2K}) - \frac{1}{2K} E(\lambda)$$

and noting that $E(\lambda)$ is convex, we conclude that $H(\lambda)$ is convex for every $\lambda$.

The minmax value is continuous in the payoffs, and so $v_\lambda(S)$ is continuous in $\lambda$. Therefore – since the Shapley value of a coalitional game $v$ is linear in $v(S)_{S\subset N}$ – $\varphi(v_\lambda)$ is continuous in $\lambda$.

Clearly, the set-valued mapping $\lambda \to F(\lambda)$ is continuous, implying that the mapping $\lambda \to E(\lambda)$ is upper-semi-continuous. Therefore $H : \Delta \to \Delta$ is an upper-semi-continuous correspondence satisfying the conditions of the Kakutani fixed-point theorem.

Thus there exists a $\lambda_0$ such that $\lambda_0 \in H(\lambda_0)$, i.e., $\varphi(v_{\lambda_0}) = y_0$, where $y_0 \in E(\lambda_0)$. Let $\psi_0 \in F$ be such that $y_0 = \lambda_0 * \psi_0$. Then $\psi_0$ is an NTU-value of the game G.

□

## 13. Appendix B: Existence of NTU-value in stochastic games

**Theorem** 5
*For every finite stochastic game there exists an NTU-value.*

*Proof.* The proof is carried out in analogy with the proof of Theorem 1, with the following adjustments:

- The feasible set $F = \text{conv}\{g(a) \colon a \in A^N\}$ is replaced by $F_0 = \{\lim_{r\to 0} \gamma_r(\sigma) : \sigma \in \Sigma^N\}$.
- The coalitional game $v_\lambda$ is no longer defined by reference to the minmax value of the one-shot game between S and $N\backslash S$, but rather it is defined by reference to the uniform value of the stochastic game played between $S$ and $N\backslash S$.

The two properties of $F$ that are needed in the proof are that, for some constant $K$, $x_i \leq K\lambda_i$ for all $x \in F$, and that the mapping from $\lambda$ to $F(\lambda) = \{\lambda * x : x \in F\}$ is continuous in $\lambda$. These properties hold for $F_0$ as well. (See Lemma 3.)

The two properties of $v_\lambda$ that are needed in the proof are the continuity of $v_\lambda$ in $\lambda$ and inequality (13), namely:

$$v_\lambda(S \cup i) - v_\lambda(S) \geq -K\lambda_i \;\; \forall S \not\ni i.$$

But (13) can be proved in the same way as in the proof of Theorem 1, i.e., by means of the inequalities (12).

The first and last equation in (12) just state the definition of $v_\lambda$.

The second inequality says that, if we compare two two-person zero-sum games with the same payoffs, where in the first game player 1's (respectively, player 2's) strategy set is larger (respectively, smaller) than in the second game, then the value of the first game is greater than or equal to the value of the second game. But this is true for the minmax value of stochastic games just as well as it is true for the standard minmax value of matrix games.

The third inequality says that, if we compare two two-person zero-sum games with the same strategy sets, where the payoffs of the two games differ by at most $2\lambda_i\|g^i\|$, then the values of these games differ by at most $2\lambda_i\|g^i\|$. By Corollary 3, this holds in stochastic games just as well, when "payoffs" are replaced by "stage payoffs".

Finally, we note that the continuity of $v_\lambda$ is also a consequence of Corollary 3:

$$|v_\lambda(S) - v_{\lambda'}(S)| = |\mathrm{uVal}(\Gamma^S_\lambda) - \mathrm{uVal}(\Gamma^S_{\lambda'}|) \le \sum_{i=1}^N \|g^i\||\lambda_i - \lambda'_i|.$$

With these adjustments, the proof of Theorem 5 goes through in the same way as the proof of Theorem 1.

□

## 14. Appendix C: Stationary Strategies

**Lemma 5.** If $\sigma$ is a stationary strategy then

(i) $\lim_{k\to\infty} \gamma_k(\sigma)$ and $\lim_{r\to 0} \gamma_r(\sigma)$ exist and are equal.

(ii) $\gamma_r(\sigma)$ is a bounded rational function in $r$.

This result is well known (e.g., [4], [13], or [3]). For completeness, we provide a proof.

*Proof.* A stationary strategy, $\sigma_t = \sigma\ \forall t$, induces the same expected payoffs, $g_\sigma$, and the same transition probabilities, $P_\sigma$, at every stage, where $g_\sigma\colon Z \to \mathbb{R}^N$ is defined by

$$g_\sigma[z] = g(z, \sigma(z)) = \sum_{a\in A} \sigma(z)[a]g(z,a)$$

and $P\colon Z \to \Delta(Z)$ is defined by

$$P_\sigma(z)[z'] = p(z, \sigma(z))[z'] = \sum_{a\in A} \sigma(z)[a]p(z,a)[z'].$$

Since $P_\sigma$ is a Markov matrix, $||P_\sigma|| \le 1$ . As is well known, this implies that the sequence $\frac{1}{k}\sum_{t=1}^k P_\sigma^{t-1}$ converges, and therefore

$$\gamma_k(\sigma) = \frac{1}{k}\sum_{t=1}^k P_\sigma^{t-1} g_\sigma$$

converges as $k \to \infty$. But the convergence of $\gamma_k(\sigma)$ as $k \to \infty$ implies the convergence of $\gamma_r(\sigma)$ as $r \to 0$, to the same limit. (This follows from, e.g., Lemma 1.)

To prove (ii) note that, since $||P_\sigma|| \le 1$, the power series $(1-r)^t P_\sigma^t$ converges to $(I - (1 - r)P_\sigma)^{-1}$, so that

$$\gamma_r(\sigma) = \sum_{t=1}^\infty r(1-r)^{t-1} P_\sigma^{t-1} g_\sigma = r(I - (1-r)P_\sigma)^{-1} g_\sigma.$$

Thus $\gamma_r(\sigma)$ is a rational function in $r$. It is bounded by $\max_{z,a} |g^i(z,a)|$.

□

**Note**: Part (ii) provides yet another proof that $\gamma_r(\sigma)$ converges as $r \to 0$.

We now apply the lemma to provide a proof of Blackwell's theorem.

**Theorem** 3
In every MDP there exists a stationary strategy $\sigma^*$ such that $\sigma^*$ is optimal in the $r$-discounted MDP for all $0 < r < r_0$ for some $r_0 > 0$.

*Proof.* By Corollary 1, for any $0 < r < 1$ some pure stationary strategy is optimal in the $r$-discounted MDP. Thus, a pure stationary strategy that yields the highest expected payoff among the finitely many pure stationary strategies is optimal.

Since the expected payoffs of these strategies are rational functions, they can cross only finitely many times. It follows that one of them is maximal in an interval $[0, r_0]$, and so the corresponding pure stationary strategy is optimal in that interval.

□

**Lemma 6.** $\lim_{r\to 0} F_r = F_0$

*Proof.* Let $x_0 \in \lim_{r\to 0} F_r$. By (9),

$$x_0 = lim_{r\to 0} \sum_{m\in M} \mu_{r,m}\ \gamma_r(\eta_m),$$

where $\{\eta_m\}_{m\in M}$ are the finitely many pure stationary strategies, and where $\mu_{r,m} \geq 0$ and $\sum_{m\in M} \mu_{r,m} = 1$.
Let $r_n$ be a subsequence such that $\lim_{n\to\infty} \mu_{r_n,m} = \mu_{0,m}\ \ \forall m \in M$. Since $\eta_m$ is stationary, $\lim_{n\to\infty} \gamma_{r_m}(\eta_m)$ exists. (Lemma 5). Denoting this limit by $\gamma_0(\eta_m)$, we have

$$x_0 = \sum_{m\in M} \mu_{0,m}\ \gamma_0(\eta_m).$$

Let

$$\sigma_0 = \sum_{m\in M} \mu_{0,m}\ \eta_m.$$

Then $x_0 = \lim_{r\to 0} \gamma_r(\sigma_0) \in F_0$.

□

The definition:

**Definition 7.** $v \in \mathbb{R}$ is the *minmax value* of a two-person zero-sum stochastic game if $\forall \varepsilon > 0$, $\exists \sigma_\varepsilon, \tau_\varepsilon$ and $k_\varepsilon > 0$ s.t. $\forall \sigma, \tau$

(i) $\varepsilon + \gamma_k(\sigma_\varepsilon, \tau) \geq v \geq \gamma_k(\sigma, \tau_\varepsilon) - \varepsilon \quad \forall k > k_\varepsilon$.
(ii) $\varepsilon + E_{\sigma_\varepsilon,\tau} \liminf_{k\to\infty} \bar{g}_k \geq v \geq E_{\sigma,\tau_\varepsilon} \limsup_{k\to\infty} \bar{g}_k$,

where $\bar{g}_k := \frac{1}{k}\sum_{t=1}^k g_t$.

If we go that route we should say something like:
$v \in \mathbb{R}$ is the *uniform*, respectively, the *limiting-average*, *value* of a two-person zero-sum stochastic game if $\forall\varepsilon > 0$, $\exists\sigma_\varepsilon, \tau_\varepsilon$ and $k_\varepsilon > 0$ s.t. $\forall\sigma, \tau$ (i), respectively, (ii), holds.

**Note**: Obviously, if the value, respectively, the unform value or the limiting-average exists, then it is unique.
Such a modification will require further wording changes thereafter. E.g., uniform value has to be changed to the minmax value.
We can mention that, apriori, the uniform minmax value can exists when the limiting-average minmax value does not exists, and vice versa, and that both can exists and differ. However, in the finite stochastic game the value exists thus each one exists and the two are equal.

## 15. Appendix D: "NTU-value exists in $\Gamma_r$" is an elementary sentence

**Lemma** 4
The statement "for every r-discounted stochastic game there exists an NTU-value" is an elementary sentence.

*Proof.* Fix finite N, Z, and A. The statement may be written as follows:
$\forall(g,p)$ and $\forall\ 0 < r < 1$, $\exists\psi_r \in \mathbb{R}^N$, $\exists\lambda_r \in \mathbb{R}^N$, and $\exists v_{r,\lambda_r} \in \mathbb{R}^{2^N}$ s.t.

$$\psi_r \in F_r \tag{14}$$

$$\lambda_r \in \Delta \tag{15}$$

$$v_{r,\lambda_r}(S) := \frac{1}{2}\mathrm{Val}(\Gamma^N_{r,\lambda_r}) + \frac{1}{2}\mathrm{Val}(\Gamma^S_{r,\lambda_r}) \quad \forall S \subseteq N \tag{16}$$

and

$$\varphi(v_{r,\lambda_r}) = \lambda_r * \psi_r. \tag{17}$$

In this statement, the variables $g, p, r, \psi_r, \lambda_r$, and $v_{r,\lambda_r}$ are all modified by $\exists$ or $\forall$. So we must show that (14) - (17) are elementary formulas where these are the only free variables.
In the interest of brevity, we only show that (14) - (17) are elementary formulas. It is straightforward to verify that no variables but the ones listed above are free in any of these formulas.

We first consider (15). The statement that each coordinate in non-negative and the sum of the coordinates is 1, is obviously an elementary formula.

Next, we consider (17). This is an elementary formula because the Shapley value, $\varphi\colon \mathbb{R}^{2^N} \to \mathbb{R}^N$, being a linear function, can be expressed in the form

$$\varphi(v)_i = \sum_{S \subset N} c_i^S v(S),$$

where the $c_i^S$ are (rational) constants, independent of $v$.

Next, we consider (16). It is well known that, if $G$ is a one-shot two-person zero-sum game, then the statement $y = \mathrm{Val}(G)$ is an elementary formula. (See, e.g., [2] ). By (7), then, the statement $y = \mathrm{Val}(\Gamma_r)$, where $\Gamma_r$ is an r-discounted stochastic game, is also an elementary formula.

Finally, we consider (14). Obviously, (7) applies in the case of a stochastic r-discounted game with a single-player who has a single strategy, $\sigma$. Therefore the statement $y = \gamma_r(\sigma)$ is an elementary formula. Since $F_r$ is the convex hull of the finitely many $\gamma_r(\sigma)$ corresponding to pure stationary strategies, (14) is an elementary formula as well.

□

## References


[1] Aumann, R.J.. and M. Kurtz (1977), Power and Taxes,*Econometrica*, 45, 1137-1161.

[2] Bewley, T. and E. Kohlberg (1976), The Asymptotic Theory of Stochastic Games, *Mathematics of Operations Research*, 1, 197-208.

[3] Bewley, T. and E. Kohlberg (1978), On Stochastic Games with Stationary Optimal Strategies, *Mathematics of Operations Research*, 3, 104-125..

[4] Blackwell, D. (1962), Discrete Dynamic Programming, *The Annals of Mathematical Statistics*, 2, 724-738.

[5] Blackwell, D. and T.S. Ferguson (1968), The Big Match, *Annals of Mathematical Statistics*, 39, 159-168.

[6] Gilette, D. (1957), Stochastic Games with Zero Stop Probabilities, *Contributions to the Theory of Games*, 3, 179-187.

[7] Harsanyi, J. (1963), A Simplified Bargaining Model for the n-Person Cooperative Game, *International Economic Review*, 4, 194-220.

[8] Kohlberg, E. (1974), Repeated Games with Absorbing States, *Annals of Statistics*, 4, 194-220.

[9] Mertens, J.F. and A. Neyman (1981), Stochastic Games, *International Journal of Game Theory*, 10, 53-66.

[10] Mertens J.F., S. Sorin and S. Zamir (1994), Repeated Games, Core Discussion Paper 9420-9421-9422, Université Catholique de Louvain la Neuve, Belgium.

[11] Nash, J. (1950), The Bargaining Problem, *Econometrica*, 18, 155-162.

[12] Neyman, A. (1977), Values for Non-transferable Utility Games with a Continuum of Players, Technical Report No. 351, School of Operations Research and Industrial Engineering, Cornell University.

[13] Neyman, A. (2003), From Markov Chains to Stochastic Games, A. Neyman and S. Sorin (eds.), NATO ASI series 2003, Kluwer Academic Publishers, pp. 9-25.

[14] Neyman, A. (2003), Real Algebraic Tools in Stochastic Games, in *Stochastic Games and Applications*, A. Neyman and S. Sorin (eds.), NATO ASI series 2003, Kluwer Academic Publishers, pp. 58-75.

[15] Shapley, L. (1953), Stochastic Games, *Proceedings of the National Academy of Sciences , U.S.A.*

[16] Shapley, L. (1969), Utility and the Theory of Games, *Editions du Centre Nattional de la Recherche Scientifique*, 39, 251-263.