HARVARD | BUSINESS | SCHOOL

The Cooperative Solution of Stochastic Games

Elon Kohlberg
Abraham Neyman

Working Paper 15-071

The Cooperative Solution of Stochastic Games

Elon Kohlberg
Harvard Business School

Abraham Neyman
The Hebrew University of Jerusalem

Working Paper 15-071

The Cooperative Solution of Stochastic Games

Elon Kohlberg* and Abraham Neyman†

January 23, 2015

Abstract

Building on the work of Nash, Harsanyi, and Shapley, we define a cooperative solution for strategic games that takes account of both the competitive and the cooperative aspects of such games. We prove existence in the general (NTU) case and uniqueness in the TU case. Our main result is an extension of the definition and the existence and uniqueness theorems to stochastic games - discounted or undiscounted.

1 Introduction

Stochastic games, introduced by Shapley ([21]), may be described as follows. 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.

While the theory of stochastic games has been developed in many different directions, there has been practically no work on the interplay between stochastic games and cooperative game theory. Our purpose here is to take an initial step in this direction.

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

†Institute of Mathematics, and the Federmann 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.

Building on the work of Nash ([16]), Harsanyi ([9]), and Shapley ([22]), we define a cooperative solution for strategic games and prove an existence theorem. Our main result is an extension of the definition and the existence theorem to stochastic games.

The original idea appears in Nash ([16]), who pioneered the notion of a solution that takes account of both the competitive and the cooperative aspects of a strategic game. Nash defined such a solution for two-person games and proved an existence and uniquenss theorem.

The solution is derived by means of "bargaining with variable threats." In an initial competitive stage, each player declares a "threat" strategy, to be used if negotiations break down; the outcome resulting from deployment of these strategies constitutes a "disagreement point." In a subsequent cooperative stage, the players coordinate their strategies to achieve a Pareto optimal oucome, and share the gains over the disagreement point; the sharing is done in accordance with principles of fairness.

The extension of the Nash solution to n-player strategic games requires several ideas. These appear, more or less explicitly, in the work of Harsanyi ([9]), Shapley ([22]), Aumann and Kurtz ([3]), and Neyman ([17]). However, there does not seem to be a single comprehensive treatment. Thus we provide a formal definition and existence proof for what we believe is the simplest possible generalization of the Nash solution. We refer to it as *the cooperative solution of a strategic game.*

The first step is to consider an alternative view of the two-player case. Under the assumption of transferable utility (TU), the outcome of "bargaining with variable threats" can be described very simply. (The description that follows is due to Shapley ([23]). It also appears in Kalai and Kalai ([10]), who use it to define their "co-co-value.")

Let s denote the maximal sum of the players' payoffs in any entry of the payoff matrix, and let d be the minmax value of the zero-sum game constructed by taking the difference between player 1's and 2's payoffs. Then the Nash solution splits the amount s in such a way that the difference in payoffs is d. Specifically, the payoffs to players 1 and 2 are, respectively, $\frac{1}{2}s + \frac{1}{2}d$ and $\frac{1}{2}s - \frac{1}{2}d$. (Appendix A provides a simple example highlighting the insights that this solution affords.)

The procedure can be generalized to n-player TU games as follows. Let s denote the maximal sum of payoffs in any entry of the payoff matrix, and let d_S be the minmax value of the zero-sum game between a subset of players, S, and its complement, $N \setminus S$, where the players in each of these subsets collaborate as a single player, and where the payoff is the difference between the sum of payoffs to the players in S and the sum of payoffs to the players in $N \backslash S$.

The cooperative solution splits the amount s in such a way as to reflect the relative strengths of the subsets S. Specifically, the cooperative solution is the Shapley value of the[1] coalitional game v, where $v(N) = s$ and for $S \subsetneq N$, $v(S) - v(N \backslash S) = d_S$.

Since the above procedure is well defined and results in a single vector of payoffs, it follows that any n-person TU strategic game has a unique cooperative solution.

To get a sense of this solution, note that in a three-player game the payoff to player 1 is $\frac{1}{3}s + \frac{1}{3}[d_1 - \frac{1}{2}(d_2 + d_3)]$. (See Proposition 1.)

Next, we consider the more general case of non-transferable utility (NTU) games. Here, implementation of the above-described procedure is more challenging. While in the TU case, the objective of a set of players acting as a single player is, obviously, to maximize the sum of their payoffs, in the NTU case, the objective is unclear.

Still, it is possible to generalize the Nash solution to NTU games. The essential idea is Shapley's method of "lambda-transfers." Imagine that utility becomes transferable after multiplication by some scaling factors, $\lambda = (\lambda_1, \ldots, \lambda_n) \geq 0$. Compute the cooperative solution of the resulting TU game. If this cooperative solution is, in fact, feasible without actual transfers of utility, then it is defined as a cooperative solution of the NTU game.

It is easy to verify that, in two-person NTU games,"bargaining with variable threats" in fact coincides with the lambda-transfer method. Thus Nash's

[1] The use of the definite article is not entirely correct, as the provided equations do not determine the coalitional game. However, they do determine its Shapley value (Corollary 1).

([16]) original proof establishes existence and uniqueness of the cooperative solution in any two-person strategic game.

We prove that a cooperative solution exists in any n-person strategic game. In NTU games with more than two players there may well be multiple solutions [2]

In the central part of this paper, we extend these developments to stochastic games. Our main result is a definition and existence theorem for the cooperative value in TU or NTU stochastic games. The result applies to both discounted and undiscounted games.

In fact, we show that in stochastic games, just as in strategic games, the cooperative value satisfies the following. Existence and uniqueness in two-person games, existence and uniqueness in n-person TU games, and existence in n-person NTU games[3].

The structure of the paper is as follows.

In Section 2 we review the notions of a coalitional game and the Shapley value of such a game.

In Section 3 we define the cooperative solution for TU and NTU strategic games and provide existence and uniqueness theorems.

As the notion of the minmax value in undiscounted stochastic games is crucial for the cooperative solution in such games, we provide in Section 4 a self-contained review.

In Section 5 we define the cooperative solution for TU and NTU stochastic games and provide existence and uniqueness theorems.

In Section 6 we present asymptotic properties of this solution concept.

[2] Note the analogy with exchange economies. In the TU case there is a unique competitive equilibrium; in the NTU case there might be multiple equilibria.

[3] At first blush, existence theorem in the undiscounted case may seem surprising. After all, the analysis of n-person undiscounted stochastic games is notoriously difficult. In particular, there is no existence theorem for ε-Nash equilibria in such games. However, the generaliztion still goes through. The reason is that (as will become clear in the sequel) the notion of a cooperative value does not rely on Nash equilibrium analysis in the n-player stochastic game itself. Rather, it only makes use of minmax analysis in two-person (zero-sum) stochastic games, played betwen a coaltion of players and its complement.

2 The Shapley Value of Coalitional Games

A *coalitional game with transferable utility* ("coalitional game" for short) 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 [21] introduced the notion of a "value," or an a priori 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, φv.

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

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

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 $\mathcal{R}$. Thus, the allocation to player i is the weighted average of the marginal contributions, $v(S \cup i) - v(S)$, weighted according to the frequency with which i appears right after S in a random ordering of N.

From the formula, it is easy to see that the Shapley value has the following properties.

For all coalitional games $(N, v), (N, w)$,

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

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

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

We will also use three additional consequences of (1):

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

$$\varphi_i v = \sum_{\substack{S\subset N\\ S\ni i}} \frac{(s-1)!(n-s)!}{n!} \left(v(S) - v(N\backslash S)\right), \text{ where } \ s = |S|, \tag{5}$$

and

$$\varphi v_i = \frac{1}{n}\sum_{k=1}^{n} \frac{(k-1)!(n-k)!}{(n-1)!} \sum_{\substack{S\subset N\\ S\ni i, |S|=k}} d_S \ = \frac{1}{n}\sum_{k=1}^{n} d_{i,k}, \tag{6}$$

where $d_S := v(S) - v(N\backslash S)$, and $d_{i,k}$ denotes the average of the d_S over all k-player coalitions that include i.

Obviously, (5) implies:

Corollary 1. *The Shapley value, φv, is determined by the differences $v(S) - v(N\backslash S)$, $S\subset N$.*

Notes:

- Property (4), known as the *Milnor condition*, was proposed by Milnor [14] as a desirable property of a solution to a coalitional game. It says that the allocation to a player must lie between the smallest and the largest marginal contributions of that player. Cleraly, this holds for the Shapley value.

- Property (5) is well known. One way to see its validity is as follows. Let $S \ni i$. In an ordering in which i is the last element of S, the marginal contribution is $v(S) - v(S-i)$. In the reverse ordering the marginal contribution is $(v(N \setminus S \cup i) - (v(N \setminus S))$. Since the set of reverse orderings is the same as the set of orderings, we might as well replace each summand in the r.h.s. of (1) by the average marginal contributions in the ordering and its reverse. Thus

  $$\varphi_i v := \frac{1}{n!}\sum_{S\ni i}\sum_{\mathcal{R}:P_i^{\mathcal{R}}=S-i}(\frac{1}{2}\left((v(S)-v(N\setminus S)) + (v(N\setminus S\cup i) - v(S-i))\right).$$

  But as S ranges over the subsetst of N that include i, so does $N \setminus S \cup i$. Thus we have

  $$\varphi_i v := \frac{1}{n!}\sum_{S\ni i}(s-1)!(n-s)!(v(S)-v(N\setminus S)),$$

  which is the same as (5).

- Property (6) is an immediate consequence of (5).

- An alternative way to compute the Shapley value from the $d_S = v(S) - v(N \backslash S)$ is to choose a representative v, e.g.,

  $$v(\emptyset)=0, v(N)=d_N,\ \text{and for } \emptyset \neq S \subset N,\ v(S)=\frac{1}{2}d_S \ \text{ and } \ v(N\backslash S) = -\frac{1}{2}d_S, \tag{7}$$

  and then apply the general formula (1).

3 The Cooperative Solution of Strategic Games

3.1 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

- $A^i = A$ 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 randomizations by the players in S) and not for $\Delta(A^S)$ (correlated randomizations). Still, we adopt this notation for its compactness.

3.2 The cooperative solution of TU strategic games

In TU games it is assumed that players can make side payments; i.e., utility can be transferred from one player to another. Thus it is meaningful to consider the maximal sum of payoffs available to all the players:

$$v(N) := \max_{x \in X^N} \sum_{i \in N} g^i(x). \tag{8}$$

Note: In a single-person maximization there is no advantage in using randomized strategies. Thus $v(N) = \max_{a \in A^N} \sum_{i \in N} g^i(x)$. We use the formulation in (8) merely in order to conform with the case $S = N$ in (9).

The question, then, is how the amount $v(N)$ is split among the players. As discussed in the Introduction, we capture the strength of any coalition, S, by means of the minmax value of the zero-sum game played between S and its complement, where each of these coalitions acts as a single player:

$$v(S) - v(N \backslash S) := \max_{x \in X^S} \min_{y \in X^{N \backslash S}} \left(\sum_{i \in S} g^i(x,y) - \sum_{i \notin S} g^i(x,y) \right). \qquad (9)$$

We then apply the Shapley value. This may be justified by the view that the Shapley value is a fair allocation of $v(N)$ reflecting the strength of the various coalitions, $S \subset N$. (See, e.g., Young [24].[4])

Definition 1. *The cooperative solution of the TU strategic game G is the Shapley value, φv, where v is a coalitional game satisfying (9).*

By Corollary 1, any coalitional game satisfying (9) has the same Shapley value. Thus the above procedure is well defined:

Theorem 1. *Every TU strategic game has a unique cooperative solution.*

And, applying equation (5), we have:

Proposition 1. *Let G be a TU strategic game. Its cooperative solution, $\psi \in \mathbb{R}^N$, may be described as follows:*

$$\psi_i = \frac{1}{n} \sum_{k=1}^{n} d_{i,k},$$

where $d_{i,k}$ denotes the average of the

$$d_S := \max_{x \in X^S} \min_{y \in X^{N \backslash S}} \left(\sum_{i \in S} g^i(x,y) - \sum_{i \notin S} g^i(x,y) \right)$$

over all k-player coalitions S that include i.

[4] According to Young, a fair sharing rule must allocate to each player an amount that depends only on that player's marginal contributions. He shows that this requirement, along with standard conditions of efficiency and symmetry, characterizes the Shapley value.

3.3 The cooperative solution of NTU strategic games

In a general strategic game, the payoffs represent von Neumann–Morgenstern utilities of the outcomes. Therefore, two different games where the payoffs of each player differ by a positive factor are the same. Thus we cannot assume transferable utility.

Since in NTU games addition of payoffs is meaningless, the focus of attention no longer is the maximum sum of payoffs but rather the Pareto frontier of the feasible set,

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

which consists of all the payoff vectors that can be attained when the players correlate their strategies.

To define the cooperative solution of an NTU strategic game, we deploy the lambda-transfer method of Shapley ([22]).

Assume that utility becomes transferable after an appropriate multiplication by *scaling factors* $\lambda = (\lambda_1, \ldots, \lambda_n) \geq 0$, $\lambda \neq 0$. Then we can proceed in analogy with the TU case.

The total payoff available to all players is

$$v_\lambda(N) := \max_{x \in X^N} \sum_{i \in N} \lambda_i g^i(x), \tag{10}$$

and the relative strength of a coalition S is captured by

$$v_\lambda(S) - v_\lambda(N \backslash S) := \max_{x \in X^S} \min_{y \in X^{N \backslash S}} \left(\sum_{i \in S} \lambda^i g^i(x,y) - \sum_{i \notin S} \lambda^i g^i(x,y) \right). \tag{11}$$

Note: When $S = N$ this is the same formula as (10), considering that $N \setminus N = \emptyset$.

As in the TU case, we may define the cooperative value of the strategic game G by taking the Shapley value, φv_λ. However, there are two difficulties.

First, it is unclear how the scaling factors λ ought to be chosen. Second, to attain φv_λ the players might have to resort to transfers of utility, which are ruled out by the assumption of NTU.

The idea underlying the lambda-tranfer method is not to specify a single λ but rather to accept any λ for which the associated Shapley value can be implemented without actual transfers of utility. Thus we require that $\varphi(v_\lambda)$ be a λ-rescaling of an allocation in the feasible set.

In other words, using the notation

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

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

Note that, by the linearity of the Shapley value, for every vector λ 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 λ 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 2. *$\psi \in F = conv\{g(a)\colon a \in A^N\}$ is a cooperative solution of the strategic game G if $\exists \lambda \in \Delta$ such that $\varphi(v_\lambda) = \lambda * \psi$, where v_λ is a coalitional game satisfying (11).*

Theorem 2. *Every finite strategic game has a cooperative solution.*

This theorem is closely related to the results of Shapley [22] and Harsanyi [9]. A proof is provided in Appendix A. It is a special case of Neyman [17].

Notes:

- The manner in which certain exchange rates – those that allow implementation of a solution without actual side payments – arise from the data of the game, bears some similarity to the manner in which market clearing prices arise from the data of an exchange economy.

- The lambda-transfer method has been applied in other contexts, most notably in defining a Shapley value for NTU coalitional games.

- The philosophical underpinnings of the lambda-transfer method have generated lively discussion among game theorists, e.g., Aumann ([1] and [2]) and Roth ([20]).

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 3. *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*

- $p \colon Z \times A^N \to \Delta(Z)$ *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 σ along with the initial state z determine a probability distribution P^z_σ 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_σ.

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 Γ_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 Γ_r.

Note: In fact, Γ_r is a family of games, Γ^z_r, parameterized by the initial state. Similarly for Γ_k.

We denote by v_r, respectively v_k, the minmax value of Γ_r, respectively Γ_k.

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

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

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 3. *(Shapley 1953) Let* Γ_r *be a two-person zero-sum r-discounted stochastic game.*

- Γ_r *has a minmax value and stationary optimal strategies. Furthermore:*

- $(v[z])_{z \in Z}$ *is the minmax value of* Γ_r *with initial state z iff it is the (unique) solution of the equations*

$$v[z] = Val\, G_r[z, v] \ \ \forall z \in Z \tag{12}$$

  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* Γ_r*.*

We denote by v_r, respectively v_k, the minmax value of Γ_r, respectively Γ_k. (The existence of v_k is obvious, as Γ_k is a finite game.)

4.2 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 3 implies:

Corollary 2. *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 4. *(Blackwell 1962) In every MDP there exists a uniformly optimal pure stationary strategy. That is, there exists a pure stationary strategy σ^* such that*

(i) σ^* *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* σ^* *is ε-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_{σ^*} *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_{σ^*} a.e. because a stationary strategy induces fixed transition probabilities on the states, resulting in a Markov chain.

- As σ^* is ε-optimal in Γ_k, it follows that $\lim_{k\to\infty} v_k = \lim_{k\to\infty} E_{\sigma^*}\bar{g}_k = E_{\sigma^*} \lim_{k\to\infty} \bar{g}_k$ 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‘) σ^* is ε-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 ε-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 a 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 0.9 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 simultaneously (essentially) 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 σ^*; and the right inequalities indicate that no larger payoff can be guaranteed by any strategy.

4.3 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 σ_0, τ_0 for player 1 and 2, respectively, to be optimal, if there exist $r_0 > 0$ and $k_0 > 0$ such that, for all σ, τ,

(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 ε.)

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 ϵ.

This is illustrated by the game known as the Big Match (Gilette [8]) where, moreover, there are even no Markov strategies that satisfy either (i), or (ii), or (iii) within an ϵ ([7]).

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, guarantees only the maxmin in pure strategies (+o(1) as $r \to 0$) in the r^2-discounted Big Match.

However, upon reflection, it appears that if we wish to define the notion that a player can "guarantee" 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 ε, simultaneously in all (i) r-discounted games with r sufficiently small, (ii) k-stage games with k sufficiently large, and (iii) games where the valuation of a stream of payoffs lies between the lim inf and lim sup of the average payoff in the first k stages. It is *not* essential that this strategy be stationary or that it be independent of ε, as is the case in MDPs.

In other words, our requirement should be an analog of conditions (i'), (ii'), and (iii') above, but where the strategy σ^* may depend on ε and it need not be stationary (or even Markov).

Thus we define v to be the minmax value of the (undiscounted) 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 τ, τ_ε denote strategies of player 2.

Definition 5. *$v \in \mathbb{R}^Z$ 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 $v = \lim_{r\to 0} v_r = \lim_{k\to\infty} v_k$.
- As noted earlier, every MDP has a value.

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
$$\varepsilon + \gamma_w(\sigma_\varepsilon, \tau) \geq v \geq \gamma_w(\sigma, \tau_\varepsilon) - \varepsilon,$$
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 5. *(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 [7]. They showed that in the Big Match, for any $\varepsilon > 0$, there exist non-Markov strategies that satisfy (iii) within an ε. This was extended by Kohlberg [11] to a special class of stochastic games – repeated games with absorbing states. The general definition and existence theorem were provided by Mertens and Neyman [12].

- A priori there is no reason to rule out the possibility that the uniform value exists 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 3. *Let v_r (respectively, v_k) denote the minmax value of the r-discounted game (respectively, the k-stage game). Then $v = VAL(\Gamma)$ iff $v = \lim_{r \to 0} v_r$*
(respectively, $v = \lim_{k \to \infty} v_k$).

Corollary 4. *If $\Gamma = (N, Z, A, g, p)$ and $\Gamma' = (N, Z, A, g', p)$ then*

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

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 $\bar{g}_k$ and $\bar{g}'_k$ differ by at most the same amount; thus an optimal strategy in Γ_k guarantees $v_k - \|g - g'\|_\infty$ in Γ'_k, and vice versa, which implies that $\|v_k - v'_k\| \le \|g - g'\|_\infty$. Next, let $k \to \infty$ and apply the previous corollary.

Corollary 5. *Every MDP has a uniform value.*

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

5 The cooperative solution of stochastic games

We now proceed to define the cooperative solution of a stochastic game in analogy with the definition for strategic games.

Let Γ 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 Σ^N.

5.1 The cooperative solution of two-person stochastic games

The existence and uniqueness theorem of Nash ([16]) requires only the existence of a minmax value in two-person strategic games. Since a minmax value exists in stochastic games (Theorem 5), the same proof goes through.

Theorem 6. *Every two-person stochastic game, discounted or undiscounted, TU or NTU, has a unique cooperative solution.*

5.2 The cooperative solution of TU stochastic games

The existence and uniqueness of a cooperative solution in TU stochastic games goes through in the same way as for TU strategic games, with the following adjustments:

Equation 8 is replaced by

$$v(N) := \max_{\sigma \in \Sigma^N} \sum_{i \in N} g^i(x), \tag{13}$$

and equation 9 is replaced by

$$v(S) - v(N \backslash S) := \text{VAL}(\Gamma^S), \tag{14}$$

.

where Γ^S denotes 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

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

.

Thus we have:

Theorem 7. *Every TU stochastic game, discounted or undiscounted, has a unique cooperative solution.*

5.3 The cooperative solution of NTU discounted stochastic games

Consider the r-discounted game, Γ_r. We define the feasible set, F_r, as follows:

$$\begin{aligned} F_r &:= \{\gamma_r(\sigma) \colon \sigma \in \Sigma^N\} \\ &= \operatorname{conv}\{\gamma_r(\sigma) \colon \sigma \in \Sigma^N_{s.p.}\}. \end{aligned} \tag{15}$$

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 (17).

Since every two-person zero-sum r-discounted stochastic game has a minmax value (Theorem 3), a cooperative solution can be defined in the same way as for strategic games.

Let

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

Definition 6. *$\psi_r \in F_r$ is an NTU-value of the r-discounted stochastic game Γ_r if*

$\exists \lambda \in \Delta$ such that $\varphi(v_{r,\lambda}) = \lambda * \psi$, where $v_{r,\lambda}$ is a coalitional game satisfying

$$v_{r,\lambda}(S) - v_{r,\lambda}(N\setminus S) := \mathit{Val}(\Gamma^S_{r,\lambda}) \quad \forall S \subseteq N. \tag{16}$$

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

Theorem 8. *Every NTU discounted stochastic game has a cooperative solution.*

5.4 The cooperative solution of NTU undiscounted stochastic games

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.

$$\begin{aligned} F_0 &= conv\{x\colon \exists \sigma \in \Sigma^N_{s.p.} \ s.t. \ x = \lim_{r\to 0} \gamma_r(\sigma)\} \\ &= \{x\colon \exists \sigma \in \Sigma^N \ s.t. \ x = \lim_{k\to\infty} \gamma_k(\sigma)\} \\ &= conv\{x\colon \exists \sigma \in \Sigma^N_{s.p.} \ s.t. \ x = \lim_{k\to\infty} \gamma_k(\sigma)\}. \end{aligned} \tag{17}$$

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 $\exists \hat\sigma \in \Sigma^N$ that induces the same distribution on the plays of the game as the mixed strategy $\frac{1}{2}\sigma' + \frac{1}{2}\sigma''$. Thus $\gamma_r(\hat\sigma) = \gamma_r(\frac{1}{2}\sigma' + \frac{1}{2}\sigma'') = \frac{1}{2}\gamma_r(\sigma') + \frac{1}{2}\gamma_r(\sigma'')$ 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, $F_0 \supseteq \text{conv}\{x\colon \exists \sigma \in \Sigma^N_{s.p.} \text{ s.t. } x = \lim_{r\to 0} \gamma_r(\sigma)\}$. To prove the equality, assume $F_0 \ni x_0 \notin \text{conv}\{x\colon \exists \sigma \in \Sigma^N_{s.p.} \text{ s.t. } x = \lim_{r\to 0} \gamma_r(\sigma)\}$.

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 4 w.r.t. the MDP 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 σ 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 (17) 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 $y_i \le \lambda_i \|g^i\|$ $\forall i \in N$, and

(ii) the mapping $\lambda \to F_0(\lambda)$ is continuous.

We define the cooperative solution of the undiscounted stochastic game analogously to Definition 2 for strategic games.

Definition 7. *$\psi \in F_0$ is a cooperative solution of the stochastic game Γ if $\exists \lambda \in \Delta$ such that $\varphi(v_\lambda) = \lambda * \psi$, where v_λ is a coalitional game with*

$$v_\lambda(S) - v_\lambda(N\backslash S) := VAL(\Gamma^S_\lambda) \quad \forall S \subseteq N.$$

Note: In the case $S = N$, $v_\lambda(N) = \text{VAL}(\Gamma^N_\lambda)$ 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 $\sum_{i\in N}\lambda_i g^i(z,a)$.

Theorem 9. *Every NTU undiscounted stochastic game has a cooperative solution.*

The proof is presented in Appendix C.

In summary, we have:

Theorem 10. *Every finite stochastic game, discounted or undiscounted, TU or NTU, has a cooperative solution.*

6 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 E.

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, $\forall x > 0 \; \exists y \; s.t. \; y^2 = x$, is true over the field of real numbers but false over the field of rational 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., [4].)

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., [4], or [19].)

A generic element of this field is a series of the form

$$\Sigma_{k=K}^{\infty} \alpha_k r^{k/M},$$

where M is a positive integer, K is an integer, and $\alpha_k \in \mathbb{R}$, and where the series converges for $r > 0$ sufficiently small. Of course, if the series remains bounded as $r \to 0$ then $K = 0$.

Thus, given Theorem 8 and Lemma 4, Tarski's principle implies the following:

Theorem 11. *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 cooperative solution conditions (25) to (28), such that each coordinate of these variables has an asymptotic expansion (in a right neighborhood of 0) of the form*

$$\Sigma_{k=0}^{\infty} \alpha_k r^{k/M}. \tag{18}$$

More precisely, there exist real Puiseux series $\tilde{\psi^i}$, $\tilde{\lambda^i}$, *and* $\tilde{v}(S)$ *s.t. for any* $r > 0$ *sufficiently small and for all* $i \in N$ *and* $S \subset N$*, these series converge to* ψ_r^i, λ_r^i *and* $v_{r,\lambda_r}(S)$*, respectively.*

Notes:

- One may consult [4] for additional detail regarding the application of Tarski's principle for obtaining asymptotic solutions, as $r \to 0$, in r-discounted stochastic games.

- An alternative proof of Theorem 11 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 cooperative solution conditions (25) to (28) is a semialgebraic set, whose projection on the first coordinate (r) 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 (18). (See, [19].)

We now apply Theorem 11 to derive an asymptotic version of Theorem 10.

Let $r \to 0$. In light of the asymptotic expansion (18), $\psi_r \to \psi_0$, $\lambda_r \to \lambda_0 \in \Delta$, and $v_{r,\lambda_r} \to v_{\lambda_0}$.

By Lemma 6 (in Appendix C), $\psi_0 \in F_0$. By Corollary 4, v_{λ_0} is the uniform minmax value of $\Gamma_{\lambda_0}^S$ for all $S \subseteq N$. Thus, ψ_0, λ_0, and v_{λ_0} satisfy the requirements of Definition 5; hence ψ_0 is an NTU-value of Γ. In summary:

Theorem 12. *Every finite stochastic game* Γ *has a cooperative solution that is the limit, as* $r \to 0$*, of cooperative solutions of the* r*-discounted games. Furthermore, these cooperative solutions, as well as their scaling factors and the associated minmax values and optimal strategies in the zero-sum scaled games, are real Puiseux series converging to their counterparts in the game* Γ.

Note: The above derivation of Theorem 12 provides an alternative proof of the existence of a cooperative solution in undiscounted stochastic games.

7 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., the 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 classic stochastic game, but the initial state is random, and each player receives a stochastic signal about players' previous stage actions and the current state.

The result, namely, that for each fixed $1 > r > 0$ the r-discounted game has a cooperative solution, and its proof, are both identical to those given here for the finite stochastic game with perfect monitoring. The existence of a cooperative solution 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 a cooperative solution in the undiscounted game does not depend on the existence of equilibrium payoffs in the corresponding undiscounted games.

8 Appendix A: The Cooperative Solution in a Simple Example

Consider the two-person TU game

$$\begin{bmatrix} 2,1 & -1,-2 \\ -2,-1 & 1,2 \end{bmatrix}.$$

At first blush the game looks entirely symmetrical. The set of feasible payoffs (the convex hull of the four entries in the matrix) is symmetrical; and the maximum payoff that each player can guarantee is the same, namely, 0. (Player 1's and 2's minmax strategies are $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{2}{3}, \frac{1}{3})$, respectively.)

Thus one would expect the maximal sum of payoffs $s = 3$ to be shared equally, i.e., $(1.5, 1.5)$. However, the Nash analysis reveals a fundamental asymmetry. In the zero-sum game of differences

$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

the minmax value $d = 1$ indicates that player 1 has an advantage. Indeed, the Nash solution – $(2, 1)$ – reflects this advantage.

9 Appendix B: Existence of a cooperative solution in strategic games

Theorem 2 *Every finite strategic game has a cooperative solution.*

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

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

(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 $|x_i| \leq K \; \forall x \in F$; therefore $|y_i| \leq K\lambda_i \; \forall y \in F(\lambda)$, and in particular $y_i \leq K\lambda_i \; \forall y \in E(\lambda)$.

To see (ii), set $v(S)$ as in (7). Then

$$\begin{aligned}
& 2v_\lambda(S \cup i) \\
= & \max_{x \in X^{S \cup i}} \min_{y \in X^{N \setminus (S \cup i)}} \left(\sum_{j \in S \cup i} \lambda_j g^j(x,y) - \sum_{j \notin S \cup i} \lambda_j g^j(x,y) \right) \\
\geq & \max_{x \in X^{S}} \min_{y \in X^{N \setminus S}} \left(\sum_{j \in S \cup i} \lambda_j g^j(x,y) - \sum_{j \notin S \cup i} \lambda_j g^j(x,y) \right) \\
\geq & \max_{x \in X^{S}} \min_{y \in X^{N \setminus S}} \left(\sum_{j \in S} \lambda_j g^j(x,y) - \sum_{j \notin S} \lambda_j g^j(x,y) \right) - 2K\lambda_i \\
= & 2v_\lambda(S) - 2K\lambda_i, \qquad (19)
\end{aligned}$$

where the first and last equalities follow from (9). Thus

$$v_\lambda(S \cup i) - v_\lambda(S) \geq -K\lambda_i \;\; \forall S \not\ni i. \qquad (20)$$

By (4), 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 $z_i \geq 0$. Indeed, by (ii) and (i),

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

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 λ.

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

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 λ_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 ψ_0 is an NTU-value of the game G.

□

10 Appendix C: Existence of a cooperative solution in stochastic games

Theorem 9

Every NTU undiscounted stochastic game has a cooperative solution.

Proof. The proof is carried out in analogy with the proof of Theorem 2, 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_λ 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 minmax 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 λ to $F(\lambda) = \{\lambda * x : x \in F\}$ is continuous in λ. These properties hold for F_0 as well. (See Lemma 3.)

The two properties of v_λ that are needed in the proof are the continuity of v_λ in λ and inequality (20), namely:

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

But the validity of (20) for stochastic games can be proved in the same way as for one-shot games, i.e., by means of the inequalities (19). Specifically:

The first and last equations in (19) just state the definition of v_λ.

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 4, this holds in stochastic games just as well, when "payoffs" are replaced by "stage payoffs."

Finally, we note that the continuity of v_λ is also a consequence of Corollary 4:

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

With these adjustments, the proof of Theorem 10 goes through in the same way as the proof of Theorem 2.

□

11 Appendix D: Stationary Strategies

Lemma 5. *If σ 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., [5], [6], or [18]). For completeness, we provide a proof.

Proof. A stationary strategy, $\sigma_t = \sigma\ \forall t$, induces the same expected payoffs, g_σ, and the same transition probabilities, P_σ, 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_σ 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 $\sum_{t=1}^{\infty}(1-r)^tP_\sigma^t$ 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: An alternaitve proof of (i) can be obtained by applying the Hardy–Littlewood Tauberian theorem, which asserts that if $\alpha_t \geq 0,\ t = 1, 2, 3, \ldots$, then the convergence of $\sum_{t=1}^{\infty} r(1-r)^{t-1}\alpha_t$ as $r \to 0$, implies the convergence of $\frac{1}{k}\sum_{t=1}^{k} \alpha_t$ as $k \to \infty$, to the same limit. First, establish, as in (ii) above, that $\gamma_r(\sigma)$ is a bounded rational function in r. Thus $\lim_{r\to 0} \gamma_r(\sigma)$ exists. Assuming, w.l.o.g., that $g(z, a) \geq 0\ \forall(z, a)$, and applying the Hardy–Littlewood theorem to the sequence $\alpha_t := E^z_\sigma g(z_t, a_t)$, it follows that $\lim_{k\to\infty} \gamma_k(\sigma)$ exists and the limits are equal.

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

Proof of Theorem 4

Proof. By Corollary 2, 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 $[0, r_0]$; thus the corresponding pure stationary strategy, σ^*, is optimal in that interval. This proves part (i).

Now, let σ^* be as above. Then for $r < r_0,\ \gamma_r(\sigma^*) = v_r$. By Lemma 5, then,

$$\lim_{k\to\infty} \gamma_k(\sigma^*) = \lim_{r\to 0} \gamma_r(\sigma^*) = \lim_{r\to 0} v_r.$$

Since statment (ii) amounts to

$$\lim_{k\to\infty} \gamma_k(\sigma^*) = \lim_{k\to\infty} v_k,$$

we must show that

$$\lim_{r\to 0} v_r = \lim_{k\to\infty} v_k. \tag{21}$$

First, by Theorem 3,

$$v_{\frac{1}{k}} = \text{Val } G_{\frac{1}{k}}\ [v_{\frac{1}{k}}], \tag{22}$$

where $G_r[v] = (G_r[z, v])_z$ is the one-shot game defined in (12).

Next, by backwards-induction,

$$v_k = \text{Val } G_{\frac{1}{k}}[v_{k-1}]. \tag{23}$$

Finally, since $\gamma_r(\sigma^*)$ is a bounded rational function in r, there is an interval, $[0, r']$, where $v_r = \gamma_r(\sigma^*)$ is continuously differentiable. Thus, for some $c \in \mathbb{R}$,

$$\|v_{\frac{1}{k}} - v_{\frac{1}{k-1}}\| \le c\, \frac{1}{k(k-1)}, \quad \forall k > K' := \frac{1}{r'}. \tag{24}$$

From (22), (23), and (24),

$$\|v_{\frac{1}{k}} - v_k)\| \le (1 - \frac{1}{k})\ \|v_{\frac{1}{k}} - v_{k-1}\| \le (1 - \frac{1}{k})\ \|v_{\frac{1}{k-1}} - v_{k-1}\| + ck^{-2}.$$

Multiplying by k and adding up for $k = K' + 1, \dots, K$, we have

$$K\ |v_{\frac{1}{K}} - v_K| \le K'\ |v_{\frac{1}{K'}} - v_{K'}| + c\,\lg(K).$$

Thus $v_{\frac{1}{K}} - v_K \to 0$ as $K \to \infty$, which implies (21). This completes the proof of *(ii)*.

The first part of *(iii)* is proved in the note following the statement of Blackwell's theorem. The second part is 4) in Proposition 3 of Neyman [18]. For a proof see pp. 21–22 there.

□

Lemma 6. *If $x_0 = \lim_{r\to 0} x_r$, where $x_r \in F_r$, then $x_0 \in F_0$.*

Proof. By (15),

$$x_r = \sum_{m\in M} \mu_{r,m}\, \gamma_r(\eta_m)$$

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 η_m is stationary, $\lim_{n\to\infty} \gamma_{r_n}(\eta_m)$ exists. (Lemma 5.) Thus

$$x_0 = \lim_{r\to 0} x_r = \sum_{m\in M} \mu_{0,m} \lim_{r\to 0} \gamma_r(\eta_m) = \lim_{r\to 0} \gamma_r(\sigma_0),$$

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

12 Appendix E: "A cooperative solution exists in Γ_r" is an elementary sentence

Lemma 4

The statement "for every r-discounted stochastic game there exists a cooperative solution" 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.

$$\psi_r \in F_r \tag{25}$$

$$\lambda_r \in \Delta \tag{26}$$

$$v_{r,\lambda_r}(S) - v_{r,\lambda_r}(N \backslash S) := \mathrm{Val}(\Gamma^S_{r,\lambda_r}) \quad \forall S \subseteq N \tag{27}$$

and

$$\varphi(v_{r,\lambda_r}) = \lambda_r * \psi_r. \tag{28}$$

In this statement, the variables $g, p, r, \psi_r, \lambda_r$, and v_{r,λ_r} are all modified by $\exists$ or $\forall$. Thus we must show that (25)–(28) are elementary formulae where these are the only free variables.

In the interest of brevity, we show only that (25)–(28) are elementary formulae. It is straightforward to verify that no variables but the ones listed above are free in any of these formulae.

We first consider (26). The statement that "each coordinate is non-negative and the sum of the coordinates is 1," is obviously an elementary formula.

Next, we consider (28). 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 (27). 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., [4].) By (12), then, the statement $y = \mathrm{Val}(\Gamma_r)$, where Γ_r is an r-discounted stochastic game, is also an elementary formula.

Finally, we consider (25). Obviously, (12) applies in the case of a stochastic r-discounted game with a single-player who has a single strategy, σ. 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, (25) is an elementary formula as well.

□

References

[1] Aumann, R.J. (1985), On the Non-transferable Utility Value: A Comment on the Roth–Shafer Examples, *Econometrica*, 53, 667–677.

[2] Aumann, R.J. (1986), On the Non-transferable Utility Value: Rejoinder, *Econometrica*, 54, 985–989.

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

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

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

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

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

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

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

[10] Kalai, A. and E. Kalai (2013), Cooperation in Strategic Games Revisited, *Quarterly Journal of Economics*, 128, 917–966.

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

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

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

[14] Milnor, J. (1952), Reasonable Outcomes for n-Person Games, *The Rand Corporation*, 18, 196.

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

[16] Nash, J. (1953), Two-Person Cooperative Games, *Econometrica*, 21, 128–140.

[17] 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.

[18] 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.

[19] 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.

[20] Roth, A. (1985), On the Non-transferable Utility Value: A Reply to Aumann, *Econometrica*, 54, 981–984.

[21] Shapley, L. (1953), Stochastic Games, *Proceedings of the National Academy of Sciences, U.S.A.*, 39, 1095–1100.

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

[23] Shapley, L. (1984), Mathematics 147 Game Theory, UCLA Department of Mathematics, 1984, 1987, 1988, 1990.

[24] Young, H.P. (1988), Individual Contribution and Just Compensation, in *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, A. Roth (ed.), Cambridge University Press, pp. 267–278.