

An invitation to play

Wieslaw Zielonka

▶ **To cite this version:**

Wieslaw Zielonka. An invitation to play. Mathematical Foundations of Computer Science 2005, 2005, Gdansk, Poland. pp.58-70, 10.1007/11549345_7 . hal-00160437

HAL Id: hal-00160437
https://hal.science/hal-00160437

Submitted on 6 Jul 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An invitation to play*

Wiesław Zielonka

LIAFA, case 7014, Université Paris 7 and CNRS
2, Place Jussieu, 75251 Paris Cedex 05, France
zielonka@liafa.jussieu.fr

Abstract. Parity games and their subclasses and variants pop up in various contexts: μ-calculus, tree automata, program verification [3, 1, 8]. Such games provide only binary information indicating the winning player. However, in classical games theory [12] the emphasis is rather on how much we win or lose. Can we incorporate the information about the profits and losses into parity games?

1 Games

Our games oppose two players, player 1 and player 2. At each moment the game is in some state s and the player controlling s chooses an action available at s which results in issuing an immediate reward r and changing the state to a new one s'. Both the reward and the new state depend deterministically on the executed action, i.e. we can assume without loss of generality that the set of actions A is just a subset of $S \times \Re \times S$, where S is the set of all states and $\Re$ is a set of (immediate) rewards. If $a = (s_1, r, s_2) \in A$ then the state $s_1 = \text{source}(a)$ is the *source* of the action a indicating the state where a is available, $s_2 = \text{target}(a)$ is the *target* state where the game moves upon the execution of a and finally $r = \text{reward}(a) \in \Re$ is the reward associated with a.

The set S of states is partitioned onto two sets, the set S_1 of states controlled by player 1 and the set S_2 of states controlled by player 2. For each state s the set $A(s) = \{a \in A \mid \text{source}(a) = s\}$ is the set of actions *available* at s and we assume that this set is always non-empty for each state s.

The tuple $\mathcal{A} = (S_1, S_2, A)$ satisfying the conditions above is called an *arena* over the set $\Re$ of rewards. Unless otherwise stated, we assume always that an "arena" means in fact a finite arena, i.e. an arena with finite state and action spaces.

A *history* in arena $\mathcal{A}$ is a finite or an infinite sequence $h = a_1 a_2 \ldots$ of actions such that $\forall i, \text{target}(a_i) = \text{source}(a_{i+1})$. The source of the first action a_0 is the source, source(h), of history h. If h is finite then the target of the last action is the target, target(h), of h.

It is convenient to assume that for each state s there is an empty history $\mathbf{1}_s$ with the source and the target s.

* This research was supported by European Research Training Network: Games and Automata for Synthesis and Validation and ACI Sécurité Informatique 2003-22 VERSYDIS

We start the play by putting a token at some initial state s_1 and the players play by moving the token from state to state: at each stage if the token is at a state $s \in S_i$ controlled by player i than player i chooses an action $a \in A(s)$ available at s and moves the token to the state target(a).

Starting from an initial state s_1, the infinite sequence of actions $p = a_1 a_2 \dots$ executed by the players is called a *play* in the arena $\mathcal{A}$, i.e. *plays* are just infinite histories in $\mathcal{A}$.

Upon the termination of a play p player 1 receives from player 2 a payoff. In this paper we assume that the payoff depends uniquely on the infinite sequence of rewards occurring in the play p.

An infinite sequence $r = r_1 r_2 \dots$ of elements of $\Re$ is said to be finitely generated if there exists a finite subset X of $\Re$ such that all r_i belong to X.

A *payoff mapping* u over $\Re$ maps finitely generated infinite sequences of rewards $r = r_1 r_2 \dots$ into $\mathbb{R}$, $u : r \mapsto u(r) \in \mathbb{R}$. Since we are concerned only with plays over finite arenas we do not need to specify what is the payoff for those infinite reward sequences which are not finitely generated.

A *game* over $\Re$ is just a couple $\mathbf{G} = (\mathcal{A}, u)$ consisting of an arena and a payoff mapping. A play $p = a_0 a_1 \dots$ in the game $\mathbf{G}$ is a play in the underlying arena. Upon completing p player 1 receives from player 2 the amount $u(\text{reward}(p))$, where $\text{reward}(p) := \text{reward}(a_1), \text{reward}(a_2), \dots$ is the sequence of rewards occurring in p. To avoid clutter we abuse the notation and we write systematically $u(p)$ to denote $u(\text{reward}(p))$ (this can be seen as an extension of payoff mapping to plays).

Parity games. For parity games the set of rewards $\Re$ is the set $\mathbb{N}$ of non-negative integers and following the tradition we call the elements of $\Re = \mathbb{N}$ priorities rather than rewards.

For any infinite finitely generated sequence of priorities $n = n_1 n_2 \dots$ let

$$\text{priority}(n) = \limsup_{i \to \infty} n_i \tag{1}$$

be the maximal priority occurring infinitely often in n. The payoff mapping in the parity games is given by

$$u(n) = \begin{cases} 1 & \text{if priority}(n) \text{ is odd,} \\ 0 & \text{if priority}(n) \text{ is even.} \end{cases} \tag{2}$$

Two remarks are in order. Usually in parity games we speak about the winning and the losing player, however it is clear that this is equivalent to the binary payoff formulation given above and we prefer payoffs since subsequently we will be interested in profits or losses and not just in the mere information who wins. Secondly, in parity games we usually attach priorities to states not to actions but this has no influence on the game analysis and game theoretists prefer to associate rewards with actions [11].

Discounted and mean-payoff games. Let us compare briefly parity games with other similar games studied in game theory rather than in computer science. Two types of games are particularly popular, discounted and mean-payoff games [11, 4]. In both these games $\Re = \mathbb{R}$, i.e. the rewards are real numbers.

In *mean-payoff* games the payoff for an infinite sequence of real numbers $r = r_1 r_2 \dots$ is calculated through the formula $\overline{\text{mean}}(r) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r_i$. Instead of taking lim sup it is possible to consider the games with the payoff $\underline{\text{mean}}(r) = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r_i$.

In the case of *discounted games* player 1 receives from 2 the amount $\text{disc}_\lambda(r) = (1-\lambda) \sum_{i=0}^{\infty} \lambda^i r_i$, where $\lambda \in (0,1)$ is a discount factor.

The striking difference between the parity games on the one hand and the mean-payoff or the discounted games on the other hand is that in the later the emphasis is put on the amount of profit/loss while for the parity games the information is just binary, indicating the winner without any attempt to quantify his profit. Obviously for games inspired by economic applications to be able to quantify the profit is essential, after all, the difference between winning or losing 10$ is hardly noticeable (and both events in themselves are of little interest) while the difference between winning 10$ and winning $10^6$$ is formidable and of great interest to the player.

Can parity games be adapted to provide a pertinent information about the player's profits/losses instead of just a plain indication who wins? It turns out that in fact several such extensions are possible for parity games and moreover these games preserve the most appealing property of parity games: the existence of optimal memoryless strategies for both players.

1.1 Strategies

A strategy of a player is his plan of action, it tells him which action to take when it is his turn to move. The choice of the action to be executed can depend on the whole sequence of previous moves. Thus a *strategy* for player 1 is a mapping

$$\sigma : \{h \mid h \text{ a finite history with target}(h) \in S_1\} \longrightarrow A \tag{3}$$

such that if $s = \text{target}(h)$ then $\sigma(h) \in A(s)$.

A strategy σ of player 1 is said to be *positional* or *memoryless* if the chosen action depends only on the last state in the history. It is convenient to view a positional strategy as a mapping

$$\sigma : S_1 \to A \tag{4}$$

such that $\sigma(s) \in A(s)$, $\forall s \in S_1$.

Strategies and positional strategies for player 2 are defined in the similar way with S_2 replacing S_1.

In the sequel, σ and τ, possibly with subscripts or superscripts, will always denote strategies for players 1 and 2 respectively.

A finite or infinite history $h = a_1 a_2 \dots$ is said to be *consistent* with a strategy σ of player 1 if for each i such that $\text{target}(a_i) \in S_1$, $a_{i+1} = \sigma(a_0 \dots a_i)$. Moreover,

if $s = \text{source}(a_1) \in S_1$ then we require that $a_1 = \sigma(\mathbf{1}_s)$ (recall that $\mathbf{1}_s$ is a special play of zero length with the source and target s). The consitency with strategies of player 2 is defined similarly.

Given a pair of strategies σ and τ for both players and a state s, there exists in arena $\mathcal{A}$ a unique infinite play p, denoted $p_{\mathcal{A}}(s, \sigma, \tau)$, consistent with σ and τ and such that $s = \text{source}(p)$.

Strategies $\sigma^\sharp$ and $\tau^\sharp$ of players 1 and 2 are *optimal* in the game $\mathbf{G} = (\mathcal{A}, u)$ if for any state $s \in S$ and any strategies σ and τ

$$u(p_{\mathcal{A}}(s, \sigma, \tau^\sharp)) \leq u(p_{\mathcal{A}}(s, \sigma^\sharp, \tau^\sharp)) \leq u(p_{\mathcal{A}}(s, \sigma^\sharp, \tau)) \ . \tag{5}$$

Thus if both strategies are optimal the players do not have any incentive to change them unilaterally.

Note that for zero sum games that we consider here, where the profit of one player is equal to the loss of his adversary, we have the exchangeability property for optimal strategies: for any other pair of optimal strategies $\tau^\ddagger$, $\sigma^\ddagger$, the couples $(\tau^\ddagger, \sigma^\sharp)$ and $(\tau^\sharp, \sigma^\ddagger)$ are also optimal and $u(p_{\mathcal{A}}(s, \sigma^\sharp, \tau^\sharp)) = u(p_{\mathcal{A}}(s, \sigma^\ddagger, \tau^\ddagger))$; this last quantity is called the value of the game $\mathbf{G} = (\mathcal{A}, u)$ at the state s.

The basic problem of game theory is to determine for a given payoff mapping u if for every game $\mathbf{G} = (\mathcal{A}, u)$ both players have optimal strategies.

In computer science we prefer positional strategies since they are particularly easy to implement. For this reason the question that we ask in this paper for every payoff u is whether for each game $\mathbf{G} = (\mathcal{A}, u)$ over a finite arena $\mathcal{A}$ both players have positional optimal strategies.

2 From Parity Games to Games with Profits

2.1 Simple Priority Games.

The simplest adjustment of parity games enabling any real-valued payoff consists in associating with each priority a real number by means of a mapping $\alpha : \mathbb{N} \to \mathbb{R}$, we call α a priority valuation. Let $n = n_1 n_2 \ldots$ be any finitely generated infinite sequence of elements of $\mathbb{N}$.

Then the payoff mapping of *simple priority games* is given by

$$u_\alpha(n) = \alpha(\text{priority}(n)), \tag{6}$$

where priority(n) is defined as in (1). Clearly for different priority valuations α we have different simple priority games, in particular for α that maps even numbers to 0 and odd numbers to 1 we recover the parity game. In fact simple priority games are still very close to parity games. Let $\alpha(\mathbb{N}) = \{x_1 < \ldots < x_k\}$ be all priority values taken in the increasing order[1]. Then to establish if player 1 has a strategy allowing him to win at least x_i in the game with the priority valuation α we solve the game with the binary priority valuation β_i defined by $\beta_i(l) = 1$ if $\alpha(l) \geq x_i$ and $\beta_i(l) = 0$ if $\alpha(l) < x_i$. Games with binary valuations are obviously

[1] We can assume without loss of genarality that $\alpha(\mathbb{N})$ is finite.

equivalent to parity games thus both players have optimal positional strategies $\sigma_i^\sharp, \tau_i^\sharp$ in the game with the valuation β_i (in fact this is true even for infinite arenas [3, 10]). These strategies can used to build optimal positional strategies $\sigma^\sharp, \tau^\sharp$ in the game with the valuation α. For a given state $s \in S$ define the rank of s to be the maximal l such that the strategy β_l of player 1 allows him to win 1 in the binary priority game with the valuation β_l when the initial state is s. Then, for $s \in S_1$, we set $\sigma^\sharp(s) = \sigma_l^\sharp(s)$ while for $s \in S_2$ we set $\tau^\sharp(s) = \tau_{l+1}^\sharp(s)$, where l is the rank of s. Clearly the strategies $\sigma^\sharp$ and $\tau^\sharp$ are positional. Moreover, it is easy to see that in $(\mathcal{A}, u_\alpha)$ for plays starting at a state s with the rank l the strategy $\sigma^\sharp$ assures for player 1 that he will win at least x_l while the strategy $\tau^\sharp$ assures for player 2 that he will pay no more than x_l. This proves the optimality of strategies $\sigma^\sharp$ and $\tau^\sharp$ (also for infinite arenas).

2.2 Mean-payoff Priority Games.

To generalize yet further our games set $\Re = \mathbb{N} \times \mathbb{R}$ as the set of rewards. Each couple $(n, r) \in \Re$ consists now of a non-negative priority n and a real valued reward $r \in \mathbb{R}$. For an infinite finitely generated reward sequence $x = (n_1, r_1), (n_2, r_2), \ldots$ we calculate now the payoff in the following way. Let $n = \text{priority}(n_1 n_2 \ldots)$ be the maximal priority appearing infinitely often in x and let $x(n) = (n_{i_1}, r_{i_1}), (n_{i_2}, r_{i_2}), \ldots$ be the subsequence of x consisting of the elements with priority n, $n = n_{i_1} = n_{i_2} = \cdots$. Then

$$\overline{\text{mean}}(x) = \limsup_{k \to \infty} \frac{r_{i_1} + \cdots + r_{i_k}}{k} \tag{7}$$

defines the payoff for *mean-payoff priority games*. Thus, intuitively, we calculate here mean-payoff of rewards but limited to the subsequence of the maximal priority occurring infinitely often. Note that if there is only one priority then this payoff mapping reduces to the payoff of mean-payoff games (and for this reason we keep the same name). But, on the other hand, if we limit ourselves to reward sequences such that $n_i = n_j$ implies $r_i = r_j$ for all i, j, i.e. to sequences where the reward is constant for each priority, then $\overline{\text{mean}}$ reduces to a simple priority payoff of Sect. 2.1 with an appropriate priority valuation. Thus mean-payoff priority games combine the principal characteristics of mean-payoff and parity games.

Are these games positional?

If the arena is controlled by player 1, i.e. $S = S_1$, then player 1 has an obvious optimal positional strategy that can be found in the following way (we do not pretend that the method given below is the most efficient one). First note that for any play of the form $p = xy^\omega$, where x is a finite history, $y = a_1 a_2 \ldots a_k$ is a simple cycle[2] in the arena $\mathcal{A}$ and $y^\omega = yy \ldots$ is the infinite concatenation of y, we can calculate $\overline{\text{mean}}(p)$ in the following way:
let for $1 \leq i \leq k$, $\text{reward}(a_i) = (n_i, r_i)$, let $l = \max\{n_i \mid 1 \leq i \leq k\}$ be the

[2] That means that $\text{source}(y) = \text{target}(y)$ and $\text{source}(a_i) \neq \text{source}(a_j)$ for $1 \leq i < j \leq k$.

maximal priority occurring in y, and let $M = \{i \mid 1 \leq i \leq k \quad \text{and} \quad n_i = l\}$ be the occurrences of l in y, then $\overline{\text{mean}}(p) = \frac{1}{|M|} \sum_{m \in M} r_m$.

Let y be a simple cycle such that $\overline{\text{mean}}(y^\omega)$ is maximal. It is easy to see that for any other play p in $\mathcal{A}$, $\overline{\text{mean}}(p) \leq \overline{\text{mean}}(y^\omega)$. Thus to maximize his gain player 1 should arrive at this cycle y, which can be done with a positional strategy, and then he should turn round y forever which is obviously positional. If there are states in $\mathcal{A}$ from which the cycle y of the maximal payoff is not accessible then in the subarena consisting of such states we repeat the procedure described above.

For arenas controlled by player 2 (which means that $S = S_2$) the optimal positional strategy of player 2 can be found in the similar way by finding the simple cycle minimizing the payoff.

The main result of Sect. 3 (Theorem 3) states that the existence of optimal positional strategies for one-player games implies the existence of optimal positional strategies for two-player games and thus we can conclude

Proposition 1. *For all priority mean-payoff games over finite arenas both players have optimal positional strategies.*

This result above was first established by other methods in [7].

2.3 Weighted Reward Games.

Yet another extension of parity games can be obtained in the following way. Suppose that for an infinite finitely generated sequence of priorities $n = n_1 n_2 \ldots$, the priorities n_e and n_o are respectively the greatest even and the greatest odd priority occurring infinitely often in n. Then player 1 wins the parity game iff the quantity $n_o - n_e$ is positive. However, intuitively, $n_o - n_e$ gives us a more detailed information of how much the winning player outperforms the losing player in parity games, and we can as well consider the game where $n_o - n_e$ is the payoff obtained by player 1, i.e. the game where player 1 tries now to maximize this value. It is convenient then to replace even priorities by their negatives, i.e. consider finitely generated sequences $m_1 m_2 \ldots$ of integers and then the payoff for player 1 takes the form $\limsup_{i \to \infty} m_i + \liminf_{i \to \infty} m_i$. However, we can then go a step further and take as the set of rewards the set $\mathbb{R}$ of real numbers and for a fixed parameter $\lambda \in [0; 1]$ and a finitely generated sequence $r_1 r_2 \ldots$ of real numbers define a weighted reward payoff mapping

$$u_\lambda^{\text{wr}}(r_1 r_2 \ldots) = (1 - \lambda) \liminf_{i \to \infty} r_i + \lambda \limsup_{i \to \infty} r_i \ . \tag{8}$$

Note that for $\lambda = 1$ the definition above gives just the classical gambling payoff [2, 9] while $\lambda = 1/2$ can be used, as explained above, to generalize parity games. In the same way as for mean-payoff priority games one can verify that weighted reward one-player games have optimal positional strategies which implies by Theorem 3 the following result obtained first in [6]:

Proposition 2. *For all weighted reward games over finite arenas both players have optimal positional strategies.*

3 From One-Player Games to Two-Player Games.

It turns our that to assure that a payoff mapping u allows optimal positional strategies for all two-player games it suffices to verify whether one-player games with payoff u have optimal positional strategies. In fact a similar result holds also for perfect information stochastic games [5] and the proof below is just an adaptation of the one of [5].

This result is useful in practice since, as we have seen, the verification if a given payoff mapping admits optimal positional strategies can be trivial for one-person games but can require a bit of dexterity for two-person games.

An arena $\mathcal{A} = (S_1, S_2, A)$ is said to be *controlled by player* i if for each state $s \in S_j$ controlled by his adversary j, $j \neq i$ and $i, j \in \{1, 2\}$, there is only one action $a \in A$ with source s. Thus essentially the adversary player j has never any choice, in particular he has only one strategy and this strategy is positional. In this case we can as well put all the states of j under the control of player i and remove player j altogether from our game. *A one-player arena* is just an arena controlled by one of the two players and a *one-player game* is a game on a one-player arena. Note that in one-player games it suffices to exhibit an optimal strategy for the controlling player since the unique strategy of his adversary is trivial.

Theorem 3. *Let u be a payoff mapping over a set $\Re$ of rewards. If for each finite one-player arena $\mathcal{A}$ over $\Re$ the player controlling $\mathcal{A}$ has an optimal positional strategy in the game $\mathbf{G} = (\mathcal{A}, u)$ then for all two-person games over finite arenas with payoff u both players have optimal positional strategies.*

Proof. Suppose that u satisfies the conditions of the theorem. In the sequel whenever we speak about games over arenas $\mathcal{A}$ the payoff u is tacitly assumed.

For any arena $\mathcal{A} = (S_1, S_2, A)$ we call the value $|A| - |S|$ the *rank* of $\mathcal{A}$ ($|X|$ denotes the cardinality of X). Since for each state there is at least one available action the rank is always non-negative. If the rank is 0 then for each state s there is exactly one available action and therefore each player has only one possible strategy and these strategies are positional and optimal.

We shall continue the proof of Theorem 3 by induction over the rank value.

Let $\mathcal{A} = (S_1, S_2, A)$ be an arena with rank $k > 0$ and suppose that both players have optimal positional strategies for all games over the arenas with the ranks smaller than k. We shall construct a pair of optimal strategies $\sigma^\sharp, \tau^\sharp$ for the game over $\mathcal{A}$, the strategy $\sigma^\sharp$ of player 1 will be positional but the strategy $\tau^\sharp$ of player 2 will use some finite memory. In the next step we shall show that also player 2 has an optimal positional strategy.

If one of the players controls $\mathcal{A}$ then both of them have optimal positional strategies and there is nothing to do.

Thus we can assume that there exists a state $x \in S_1$ controlled by player 1 such that the set $A(x) = \{a \in A \mid \text{source}(a) = s\}$ of actions available at x contains more than one element. Let us fix such a state x which we shall call the *pivot*. We fix also a partition of the set $A(x)$ onto two non-empty sets $A_L(x)$ and $A_R(x)$, $A(x) = A_L(x) \cup A_R(x)$, $A_L(x) \cap A_R(x) = \emptyset$.

We define two subarenas $\mathcal{A}_L$ and $\mathcal{A}_R$ of $\mathcal{A}$ which we call respectively the left and the right (sub)arena. In $\mathcal{A}_L$ and $\mathcal{A}_R$ we keep the same states as in $\mathcal{A}$. Also the actions with the source in the states $s \neq x$ are the same in $\mathcal{A}_L$, $\mathcal{A}_R$ and $\mathcal{A}$. The only difference concerns the actions with the source x, in the left arena $\mathcal{A}_L$ we keep only the actions of $A_L(x)$ while in the right arena only the actions of $A_R(x)$ removing all the other actions with source x.

Since the ranks of the arenas $\mathcal{A}_L$ and $\mathcal{A}_R$ are smaller than the rank of $\mathcal{A}$, by the induction hypothesis, there exist optimal positional strategies $\sigma_L^\sharp$ and $\tau_L^\sharp$ in the game $(\mathcal{A}_L, u)$ and optimal positional strategies $\sigma_R^\sharp$ and $\tau_R^\sharp$ in the game $(\mathcal{A}_R, u)$.

We pretend that one of the two strategies $\sigma_L^\sharp$ or $\sigma_R^\sharp$ is also optimal for player 1 in the initial game over $\mathcal{A}$. The situation is more complicated for player 2, usually neither $\tau_L^\sharp$ nor $\tau_R^\sharp$ is optimal for him in the game over $\mathcal{A}$. However, it turns out to be possible to intertwine in some way the strategies $\tau_L^\sharp$ and $\tau_R^\sharp$ to obtain an optimal strategy for player 2 on $\mathcal{A}$.

Using the arena $\mathcal{A}_L$ and the strategy $\tau_L^\sharp$ of player 2 we construct an arena $\mathcal{A}_L[\tau_L^\sharp]$ that has the same states as $\mathcal{A}_L$ but we restrict the actions available to player 2: for each state $s \in S_2$ controlled by 2 we leave in $\mathcal{A}_L[\tau_L^\sharp]$ only one action with the source s, namely the action $\tau_L^\sharp(s)$ provided by the strategy $\tau_L^\sharp$. We do not restrict the moves of player 1, he can take exactly the same actions as in $\mathcal{A}_L$.

In a similar way we construct from the arena $\mathcal{A}_R$ and the optimal strategy $\tau_R^\sharp$ of player 2 in the game on $\mathcal{A}_R$ an arena $\mathcal{A}_R[\tau_R^\sharp]$ by restricting the actions player 2 to those that are provided by the strategy $\tau_R^\sharp$.

Notice that arenas $\mathcal{A}_L[\tau_L^\sharp]$ and $\mathcal{A}_R[\tau_R^\sharp]$ are controlled by player 1.

Next we rename in $\mathcal{A}_L[\tau_L^\sharp]$ and $\mathcal{A}_R[\tau_R^\sharp]$ all the states that are different from the pivot state x.

Let

$$U = S \setminus \{x\} \tag{9}$$

be the set of states that are different from the pivot x. Let U_L, U_R be two disjoint copies of the set U and let

$$S_L := U_L \cup \{x\} \quad \text{and} \quad S_R := U_R \cup \{x\} \ . \tag{10}$$

For a state $s \in U$ its left and right copy are denoted respectively s_L and s_R. It is convenient to assume that the pivot x is the only state that is a copy of itself, i.e. $x_L = x = x_R$. By π_L we shall denote the natural bijections

$$\pi_L : S \to S_L \quad \text{and} \quad \pi_R : S \to S_R \tag{11}$$

$\pi_L : s \mapsto s_L = \pi_L(s)$ and $\pi_R : s \mapsto s_R = \pi_R(s)$, for all $s \in S$. The renaming mappings π_L and π_R are extended in a natural way to actions

$$\pi_L((s,r,t)) = (\pi_L(s), r, \pi_L(t)) \quad \text{and} \quad \pi_R((s,r,t)) = (\pi_R(s), r, \pi_R(t)) \tag{12}$$

for actions (s,r,t) respectively on the left and the right subarena.

The arenas obtained from $\mathcal{A}_L[\tau_L^\sharp]$ and $\mathcal{A}_R[\tau_R^\sharp]$ by applying the corresponding renaming mappings are denoted $\pi_L(\mathcal{A}_L[\tau_L^\sharp])$ and $\pi_R(\mathcal{A}_R[\tau_R^\sharp])$. Note that $\pi_L(\mathcal{A}_L[\tau_L^\sharp])$ and $\pi_R(\mathcal{A}_R[\tau_R^\sharp])$ have only one common state x and the only common actions are eventually the actions of the form (x, r, x) if such actions with source and target x exist in $\mathcal{A}$. Finally we construct the arena $\mathcal{A}_{LR} = \pi_L(\mathcal{A}_L[\tau_L^\sharp]) \cup \pi_R(\mathcal{A}_R[\tau_R^\sharp])$, where the union means that we take simply the union of state sets and the union of action sets of $\pi_L(\mathcal{A}_L[\tau_L^\sharp])$ and $\pi_R(\mathcal{A}_R[\tau_R^\sharp])$. Let us note that since the only state common to $\pi_L(\mathcal{A}_L[\tau_L^\sharp])$ and $\pi_R(\mathcal{A}_R[\tau_R^\sharp])$ is the pivot x the arena $\mathcal{A}_{LR}$ can be seen informally as the arena obtained from $\mathcal{A}_L[\tau_L^\sharp]$ and $\mathcal{A}_R[\tau_R^\sharp]$ by gluing them together at x.

Obviously, $\mathcal{A}_{LR}$ is a one-player arena controlled by player 1. Intuitively, for each state of $\mathcal{A}_{LR}$ controlled by player 1 he has at his disposition the same actions as in $\mathcal{A}$. On the other hand, player 2 is compelled to use either the strategy $\tau_L^\sharp$ or the strategy $\tau_R^\sharp$ depending on whether the current position is in the left or in the right subarena of $\mathcal{A}_{LR}$. Each time the pivot x is visited player 1 can choose if he prefers to play till the next visit to x against the strategy $\tau_L^\sharp$ or against the strategy $\tau_R^\sharp$ by choosing either a left or a right action at x.

Example 4. Figure 1 illustrates different stages of the construction of $\mathcal{A}_{LR}$. To avoid clutter the rewards associated with actions are omitted. The states controlled by players 1 and 2 are represented respectively by circles and squares. The pivot state x has three outgoing actions and we fix the following left/right partition of $A(x)$: $A_L(x) = \{(x, s^2)\}$, $A_R(x) = \{(x, s^1), (x, s^3)\}$. Suppose now that for the state s^4 the optimal positional strategy $\tau_L^\sharp$ for player 2 in $(\mathcal{A}_L, u)$ chooses the action (s^4, s^1) while in $(\mathcal{A}_R, u)$ the optimal positional strategy $\tau_R^\sharp$ for the same player chooses the action (s^4, s^3) (for the other state s^2 both strategies choose the only available action (s^2, s^4)). The bottom part of Fig. 1 presents the resulting arenas $\mathcal{A}_L[\tau_L^\sharp]$ and $\mathcal{A}_R[\tau_R^\sharp]$. Finally, the upper left part of Fig. 1 shows the arena $\mathcal{A}_{LR}$. Note that in $\mathcal{A}_{LR}$ at the pivot state x player 1 has again three available actions, as in the initial arena $\mathcal{A}$.

Since the game $(\mathcal{A}_{LR}, u)$ is controlled by player 1 he has in this game an optimal positional strategy $\sigma_{LR}^\sharp$. Now let us look which action is chosen by $\sigma_{LR}^\sharp$ at the pivot state, we can have either $\mathrm{target}(\sigma_{LR}^\sharp(x)) \in S_L$ or $\mathrm{target}(\sigma_{LR}^\sharp(x)) \in S_R$. Exchanging if necessary "left" and "right", we can assume without loss of generality that

$$\mathrm{target}(\sigma_{LR}^\sharp(x)) \in S_L \ . \tag{13}$$

Under condition (13) it turns out that the strategy

$$\sigma^\sharp := \sigma_L^\sharp \tag{14}$$

is optimal for player 1 in the game $(\mathcal{A}, u)$

It remains to define an optimal strategy $\tau^\sharp$ for player 2 on $\mathcal{A}$. Let h be a finite history in $\mathcal{A}$ with $\mathrm{target}(h) \in S_2$, then

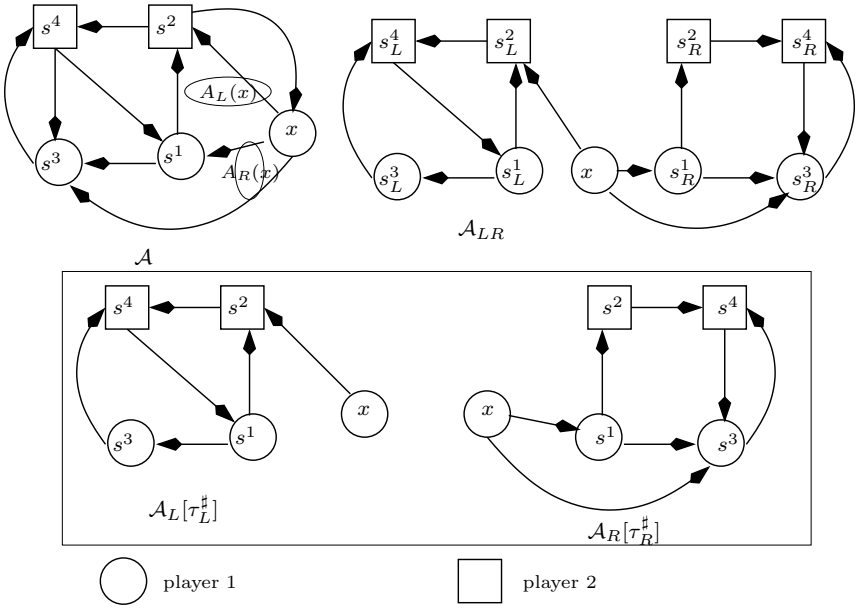


Fig. 1. Construction of $\mathcal{A}_{LR}$.

$$\tau^\sharp(h) = \begin{cases} \tau^\sharp_L(\mathrm{target}(h)) & \text{if either } h \text{ does not contain any action with source } x \\ & \text{or the last such action belongs to } A_L(x), \\ \tau^\sharp_R(\mathrm{target}(h)) & \text{if } h \text{ contains at least one action with source } x \\ & \text{and the last such action belongs to } A_R(x). \end{cases} \tag{15}$$

Thus player 2 applies in $\tau^\sharp$ either the strategy $\tau^\sharp_L$ or $\tau^\sharp_R$ and which of the two strategies is chosen depends on the action taken by player 1 at the last passage through the pivot state. To implement the strategy $\tau^\sharp$ one needs a memory, albeit a finite memory taking two values L and R is sufficient. The initial memory value is L. Every time the play traverses the pivot state x player 2 observes the action taken by player 1 and updates his memory either to L or to R depending on whether this action belongs to $A_L(x)$ or to $A_R(x)$.

Up to the next visit to x player 2 uses either the strategy $\tau^\sharp_L$ or $\tau^\sharp_R$ depending on the memory value.

We shall show that strategies $\sigma^\sharp$ and $\tau^\sharp$ defined by (13) and (15) are optimal in the game over $\mathcal{A}$.

Let $s \in S$ and consider the play $p_{\mathcal{A}}(s, \sigma^\sharp, \tau^\sharp)$. From (13) it follows that player 1 chooses during this play at each passage through x the same left hand side action from $A_L(x)$, however in this case player 2 plays all the time using the

strategy $\tau_L^\sharp$, thus

$$p_{\mathcal{A}}(s, \sigma^\sharp, \tau^\sharp) = p_{\mathcal{A}}(s, \sigma_L^\sharp, \tau_L^\sharp) = p_{\mathcal{A}_L}(s, \sigma_L^\sharp, \tau_L^\sharp) \ . \tag{16}$$

Let τ be any strategy of player 2 on $\mathcal{A}$ and let τ' be the restriction of this strategy to histories in $\mathcal{A}_L$. Obviously, τ' is a valid strategy on $\mathcal{A}_L$ and $p_{\mathcal{A}_L}(s, \sigma_L^\sharp, \tau') = p_{\mathcal{A}}(s, \sigma_L^\sharp, \tau) = p_{\mathcal{A}}(s, \sigma^\sharp, \tau)$. On the other hand, $u(p_{\mathcal{A}_L}(s, \sigma_L^\sharp, \tau_L^\sharp)) \leq u(p_{\mathcal{A}_L}(s, \sigma_L^\sharp, \tau'))$ by optimality of $\sigma_L^\sharp$, $\tau_L^\sharp$ on $\mathcal{A}_L$. These facts and (16) imply

$$u(p_{\mathcal{A}}(s, \sigma^\sharp, \tau^\sharp)) \leq u(p_{\mathcal{A}}(s, \sigma^\sharp, \tau)) \ . \tag{17}$$

Now let σ be any strategy of player 1 on $\mathcal{A}$. This strategy can be transformed to a strategy σ_{LR} on $\mathcal{A}_{LR}$ in the following way.

Let

$$\pi : S_L \cup S_R \to S$$

be the mapping from the states of $\mathcal{A}_{LR}$ to the states of $\mathcal{A}$ such that $\pi(s_L) = \pi(s_R) = s$ for all $s \in S$. This mapping can be extended to actions by putting $\pi((y', r, y'')) = (\pi(y'), r, \pi(y''))$ for any states $y', y'' \in S_L \cup S_R$ and next to finite and infinite histories, for a history $h = a_1 a_2 \dots$ in $\mathcal{A}_{LR}$, $\pi(h) = \pi(a_1)\pi(a_2)\dots$ is a history in $\mathcal{A}$. Now for any history $h = a_1 \dots a_n$ in $\mathcal{A}_{LR}$ with the target $y := \text{target}(h)$ controlled by player 1 we define

$$\sigma_{LR}(h) = \begin{cases} \pi_L(\sigma(\pi(h))) & \text{if target}(h) \in S_L = U_L \cup \{x\}, \\ \pi_L(\sigma(\pi(h))) & \text{if target}(h) \in U_R, \end{cases} \tag{18}$$

where π_L and π_R were defined in (11). Thus, intuitively, when playing according to σ_{LR} player 1 takes the projection $\pi(h)$ of the history h onto $\mathcal{A}$, applies the strategy σ which gives him an action $(s, r, t) := \sigma(\pi(h))$ in $\mathcal{A}$. The target state of h, target(h), is either the state $s_L \in U_L \cup \{x\}$ or the state $s_R \in U_R$. In the first case player 1 executes the action $\pi_L((s, r, t)) = (s_L, r, t_L)$, in the second case he executes $\pi_R((s, r, t)) = (s_R, r, t_R)$.

Let $p_{\mathcal{A}_{LR}}(s_L, \sigma_{LR}, \cdot)$ be the play in $\mathcal{A}_{LR}$ starting at a left hand side state $s_L \in S_L$ and consistent with σ_{LR} (we left out here the strategy of player 2 since he has only one strategy on $\mathcal{A}_{LR}$ and therefore it is useless to specify it explicitly). From the construction of $\mathcal{A}_{LR}$ and definitions (18) and (15) of σ_{LR} and $\tau^\sharp$ it follows that for any strategy σ of player 1 on $\mathcal{A}$

$$p_{\mathcal{A}}(s, \sigma, \tau^\sharp) = \pi(p_{\mathcal{A}_{LR}}(s_L, \sigma_{LR}, \cdot)), \tag{19}$$

Since $\sigma_{LR}^\sharp$ is an optimal positional strategy for player 1 in $(\mathcal{A}_{LR}, u)$ we have

$$u(p_{\mathcal{A}_{LR}}(s_L, \sigma_{LR}, \cdot)) \leq u(p_{\mathcal{A}_{LR}}(s_L, \sigma_{LR}^\sharp, \cdot)). \tag{20}$$

The play $p_{\mathcal{A}_{LR}}(s_L, \sigma_{LR}^\sharp, \cdot)$ starts in the left hand side state s_L and is consistent with the strategy $\sigma_{LR}^\sharp$ that, according to (13), chooses at the pivot state x a left hand side action, therefore this play traverses uniquely the left hand side

states S_L. We can define for player 1 a positional strategy $\pi \circ \sigma^\sharp_{LR} \circ \pi_L$ on $\mathcal{A}_L$ that corresponds to the left hand side part of the strategy $\sigma^\sharp_{LR}$: for $s \in S_1$, $\pi \circ \sigma^\sharp_{LR} \circ \pi_L(s) = \pi(\sigma^\sharp_{LR}(s_L))$. That we have defined in this way a valid strategy for player 1 on $\mathcal{A}_L$ is guaranteed by (13). Since player 2 is constrained in $\mathcal{A}_{LR}$ to play on S_L according to the strategy $\tau^\sharp_L$ we can see that applying the strategy $\sigma^\sharp_{LR}$ for a play starting at s_L in $\mathcal{A}_{LR}$ gives, modulo the renaming, the same result as applying the strategies $\pi \circ \sigma^\sharp_{LR} \circ \pi_L$ and $\tau^\sharp_L$ in $\mathcal{A}_L$ for a play staring at s, formally

$$\pi(p_{\mathcal{A}_{LR}}(s_L, \sigma^\sharp_{LR}, \cdot)) = p_{\mathcal{A}_L}(s, \pi \circ \sigma^\sharp_{LR} \circ \pi_L, \tau^\sharp_L). \tag{21}$$

Eq. (19) and (21) imply the equality of corresponding rewards sequences, i.e. also the equality of corresponding payoffs while by optimality $\sigma^\sharp_L$ and $\tau^\sharp_L$ in $\mathcal{A}_L$ we have $u(p_{\mathcal{A}_L}(s, \pi \circ \sigma^\sharp_{LR} \circ \pi_L, \tau^\sharp_L)) \leq u(p_{\mathcal{A}_L}(s, \sigma^\sharp_L, \tau^\sharp_L))$.

Putting together the last inequality and (16), (19), (20), (21) we deduce

$$u(p_{\mathcal{A}}(s, \sigma, \tau^\sharp)) \leq u(p_{\mathcal{A}}(s, \sigma^\sharp, \tau^\sharp).$$

This and (17) imply the optimality of strategies $\sigma^\sharp$ and $\tau^\sharp$ in the game $(\mathcal{A}, u)$.

Our problem is that the optimal strategy of player 2 constructed above is not positional, however, the remedy is simple. Consider the game with the payoff mapping $-u$ and where the roles of players 1 and 2 are permuted, i.e. it is player 1 that pays to player 2 the amount $-u(p)$ after a play p. Thus player 2 wants to maximize the payment while player 2 tries to minimize it. In the new game choose as the pivot a state controlled by player 2 with at least two available actions, then the construction above repeated in this new setting will provide optimal strategies $\sigma^\ddagger$ and $\tau^\ddagger$ for players 1 and 2 in the new game with $\tau^\ddagger$ beeing positional. However optimal strategies in the new and the old games are the same thus $\tau^\ddagger$ is an optimal positional strategy for player 2 in $(\mathcal{A}, u)$. By exchangeability property for optimal strategies, $\sigma^\sharp$ and $\tau^\ddagger$ constitute a pair of optimal positional strategies in the game $(\mathcal{A}, u)$. □

4 Final Remarks

In Sect. 3 and in Theorem 3 instead of payoff mappings we could use, without any substantial modification, preference relations [12] over infinite reward sequences. Such a relation $\preceq$ is a binary complete transitive relation (where "complete" means that $a \preceq b$ or $b \preceq a$ for all a, b in the domain of $\preceq$). Obviously each payoff mapping u defines a preference relation $\preceq_u$, for infinite finitely generated sequences of rewards r and r', $r \preceq_u r'$ iff $u(r) \leq u(r')$. Although, at least in principle, preference relations can be represented by real valued payoffs, this representation is not always natural and therefore it may be advantageous to reformulate Sect. 3 and trade payoffs for preference relations.

4.1 Nash equlibria

Suppose that we have a finite set $\{1, \ldots, N\}$ of players and the set of states is partitioned onto N disjoint sets $S = S_1 \cup \ldots \cup S_N$, S_i beeing the states controlled

by player i. Again, if the current state is s then the player controlling s chooses and executes an action available at s. Now each player i has his own payoff mapping u_i that gives for each infinite finitely generated sequence of rewards r the payoff $u_i(r)$ of player i. A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_N)$ is an N-tuple of strategies, where σ_i is a strategy of player i. Fixing a strategy profile σ and an initial state $s \in S$ we have exactly one play $p_{\mathcal{A}}(s, \sigma)$ starting at s and consistent with all strategies σ. For a strategy profile σ and a strategy σ'_i of player i by (σ^{-i}, σ'_i) we denote the strategy profile obtained from the profile σ by replacing σ_i by σ'_i.

A strategy profile σ is in *Nash equilibrium* if for each i, $1 \leq i \leq N$, and each strategy σ'_i of player i, $u_i(p_{\mathcal{A}}(s, (\sigma^{-i}, \sigma'_i))) \leq u_i(p_{\mathcal{A}}(s, \sigma))$. From the result of Sect. 3, using the trigger strategy described in [12], we can deduce that

Proposition 5. *Suppose that for all i, $1 \leq i \leq N$, for one-player games over finite arenas with payoffs u_i and $-u_i$ there exist optimal positional strategies. Then for each N-person game with payoff profile $(u_1, \ldots, u_N)$ over a finite arena there exists a Nash equilibrium profile σ where the strategy σ_i of each player i is a finite memory strategy.*

Acknowledgments. During the preparation of this paper we have profited from numerous conversations with Hugo Gimbert.

References

1. André Arnold and Damian Niwiński. *Rudiments of μ-calculus*, volume 146 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 2001.
2. Lester E. Dubins and Leaonard J. Savage. *Inequalities for Stochastic Processes (How to gamble if you must)*. Dover Publications, 2nd edition, 1976.
3. E.A. Emerson and C. Jutla. Tree automata, μ-calculus and determinacy. In *FOCS'91*, pages 368–377. IEEE Computer Society Press, 1991.
4. Jerzy Filar and Koos Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.
5. Hugo Gimbert and Wiesław Zielonka. From Markov decision processes to perfect information stochastic games. in preparation.
6. Hugo Gimbert and Wiesław Zielonka. When can you play positionally? In *Mathematical Foundations of Computer Science 2004*, volume 3153 of *LNCS*, pages 686–697. Springer, 2004.
7. Hugo Gimbert and Wiesław Zielonka. Games where you can play optimally without any memory. In *CONCUR 2005*, LNCS. Springer, 2005. to appear.
8. Erich Grädel, Wolfgang Thomas, and Thomas Wilke, editors. *Automata, Logics, and Infinite Games*, volume 2500 of *LNCS*. Springer, 2002.
9. Ashok P. Maitra and William D. Sudderth. *Discrete Gambling and Stochastic Games*. Springer, 1996.
10. A.W. Mostowski. Games with forbidden positions. Technical Report 78, Uniwersytet Gdański, Instytut Matematyki, 1991.
11. Abraham Neyman and Sylvain Sorin, editors. *Stochastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Physical Sciences*. Kluwer Academic Publishers, 2004.
12. Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, 2002.