Applying Blackwell optimality: priority mean-payoff games as limits of multi-discounted game

Hugo Gimbert, Wieslaw Zielonka

Applying Blackwell Optimality: Priority Mean-Payoff Games as Limits of Multi-Discounted Games

Hugo Gimbert[1]
Wiesław Zielonka[2*]

[1] LIX, École Polytechnique, Palaiseau, France

[2] Université Denis Diderot and CNRS, LIAFA
case 7014, 75205 Paris Cedex 13, France

gimbert@lix.polytechnique.fr zielonka@liafa.jussieu.fr

Abstract

We define and examine priority mean-payoff games — a natural extension of parity games. By adapting the notion of Blackwell optimality borrowed from the theory of Markov decision processes we show that priority mean-payoff games can be seen as a limit of special multi-discounted games.

1 Introduction

One of the major achievements of the theory of stochastic games is the result of Mertens and Neyman [15] showing that the values of mean-payoff games are the limits of the values of discounted games. Since the limit of the discounted payoff is related to Abel summability while the mean-payoff is related to Cesàro summability of infinite series, and classical abelian and tauberian theorems establish tight links between these two summability methods, the result of Mertens and Neyman, although technically very difficult, comes with no surprise.

In computer science similar games appeared with the work of Gurevich and Harrington [12] (games with Muller condition) and Emerson and Jutla [5] and Mostowski [16] (parity games).

However discounted and mean-payoff games also seem very different from Muller/parity games. The former, inspired by economic applications, are games with real valued payments, the latter, motivated by logics and automata theory, have only two outcomes, the player can win or lose.

The theory of parity games was developed independently from the theory of discounted/mean-payoff games [11] even though it was noted by Jur-

* Partially supported by the french ANR-SETI project AVERISS

dziński [14] that deterministic parity games on finite arenas can be reduced to mean-payoff games[1].

Recently de Alfaro, Henzinger and Majumdar [3] presented results that indicate that it is possible to obtain parity games as an appropriate limit of multi-discounted games. In fact, the authors of [3] use the language of the μ-calculus rather than games, but as the links between μ-calculus and parity games are well-known since the advent [5], it is natural to wonder how discounted μ-calculus from [3] can be reflected in games.

The aim of this paper is to examine in detail the links between discounted and parity games suggested by [3]. In our study we use the tools and methods that are typical for classical game theory but nearly never used for parity games. We want to persuade the reader that such tools, conceived for games inspired by economic applications, can be successfully applied to games that come from computer science.

As a by-product we obtain a new class of games — priority mean-payoff games — that generalise in a very natural way parity games but contrary to the latter allow to quantify the gains and losses of the players.

The paper is organised as follows.

In Section 2 we introduce the general framework of deterministic zero-sum infinite games used in the paper, we define optimal strategies, game values and introduce positional (i.e. memoryless) strategies.

In Section 3 we present discounted games. Contrary to classical game theory where there is usually only one discount factor, for us it is crucial to work with multi-discounted games where the discount factor can vary from state to state.

Section 4 is devoted to the main class of games examined in this paper — priority mean-payoff games. We show that for these games both players have optimal positional strategies (on finite arenas).

In classical game theory there is a substantial effort to refine the notion of optimal strategies. To this end Blackwell [2] defined a new notion of optimality that allowed him a fine-grained classification of optimal strategies for mean-payoff games. In Section 5 we adapt the notion of Blackwell optimality to our setting. We use Blackwell optimality to show that in some strong sense priority mean-payoff games are a limit of a special class of multi-discounted games.

The last Section 6 discusses briefly some other applications of Blackwell optimality.

Since the aim of this paper is not only to present new results but also to familiarize the computer science community with methods of classical game theory we have decided to make this paper totally self-contained. We present

[1] But this reduction seems to be proper for deterministic games and not possible for perfect information stochastic games.

all proofs, even the well-known proof of positionality of discounted games[2]. For the same reason we also decided to limit ourselves to deterministic games. Similar results can be proved for perfect information stochastic games [10, 9] but the proofs become much more involved. We think that the deterministic case is still of interest and has the advantage of beeing accessible through elementary methods.

The present paper is an extended and improved version of [8].

2 Games

An *arena* is a tuple $\mathcal{A} = (S_1, S_2, A)$, where S_1 and S_2 are the sets of *states* that are controlled respectively by player 1 and player 2, A is the set of *actions*.

By $S = S_1 \cup S_2$ we denote the set of all states. Then $A \subseteq S \times S$, i.e. each action $a = (s', s'') \in A$ is a couple composed of the *source state* $\mathrm{source}(a) = s'$ and the *target state* $\mathrm{target}(a) = s''$. In other words, an arena is just as a directed graph with the set of vertices S partitioned onto S_1 and S_2 with A as the set of edges.

An action a is said to be *available* at state s if $\mathrm{source}(a) = s$ and the set of all actions available at s is denoted by $A(s)$.

We consider only arenas where the set of states is finite and such that for each state s the set $A(s)$ of available actions is non-empty.

A *path* in arena $\mathcal{A}$ is a finite or infinite sequence $p = s_0 s_1 s_2 \ldots$ of states such that $\forall i$, $(s_i, s_{i+1}) \in A$. The first state is the source of the p, $\mathrm{source}(p) = s_0$, if p is finite then the last state is the target of p, $\mathrm{target}(p)$.

Two players 1 and 2 play on $\mathcal{A}$ in the following way. If the current state s is controlled by player $P \in \{1, 2\}$, i.e. $s \in S_P$, then player P chooses an action $a \in A(s)$ available at s, this action is executed and the system goes to the state $\mathrm{target}(a)$.

Starting from an initial state s_0, the infinite sequence of consecutive moves of both players yields an infinite sequence $p = s_0 s_1 s_2 \ldots$ of visited states. Such sequences are called *plays*, thus plays in this game are just infinite paths in the underlying arena $\mathcal{A}$.

We shall also use the term “a finite play” as a synonym of “a finite path” but “play” without any qualifier will always denote an infinite play/path.

A *payoff mapping*

$$u : S^{\omega} \to \mathbb{R} \tag{1.1}$$

maps infinite sequences of states to real numbers. The interpretation is that at the end of a play p player 1 receives from player 2 the *payoff* $u(p)$ (if

[2] But this can be partially justified since we need positionality of multi-discounted games while in the literature usually simple discounted games are treated. We should admit however that passing from discounted to multi-discounted games needs only minor obvious modifications.

$u(p) < 0$ then it is rather player 2 that receives from player 1 the amount $|u(p)|$).

A *game* is couple $(\mathcal{A}, u)$ composed of an arena and a payoff mapping.

The obvious aim of player 1 (the maximizer) in such a game is to maximize the received payment, the aim of player 2 (the minimizer) is opposite, he wants to minimize the payment paid to his adversary.

A strategy of a player P is his plan of action that tells him which action to take when the game is at a state $s \in S_P$. The choice of the action can depend on the whole past sequence of moves.

Therefore a *strategy* for player 1 is a mapping

$$\sigma : \{p \mid p \text{ a finite play with } \mathrm{target}(p) \in S_1\} \longrightarrow S \tag{1.2}$$

such that for each finite play p with $s = \mathrm{target}(p) \in S_1$, $(s, \sigma(p)) \in A(s)$.

Strategy σ of player 1 is said to be *positional* if for every state $s \in S_1$ and every finite play p with $\mathrm{target}(p) = s$, $\sigma(p) = \sigma(s)$. Thus the action chosen by a positional strategy depends only on the current state, previously visited states are irrelevant. Therefore a positional strategy of player 1 can be identified with a mapping

$$\sigma : S_1 \to S \tag{1.3}$$

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

A finite or infinite play $p = s_0 s_1 \ldots$ is said to be *consistent* with a strategy σ of player 1 if, for each $i \in \mathbb{N}$ such that $s_i \in S_1$, we have $(s_i, \sigma(s_0 \ldots s_{i-1})) \in A$.

Strategies, positional strategies and consistent plays are defined in the analogous way for player 2 with S_2 replacing S_1.

In the sequel Σ and $\mathcal{T}$ will stand for the set of strategies for player 1 and player 2 while Σ_p and $\mathcal{T}_p$ are the corresponding sets of positional strategies.

The letters σ and τ, with subscripts or superscripts if necessary, will be used to denote strategies of player 1 and player 2 respectively.

Given a pair of strategies $\sigma \in \Sigma$ and $\tau \in \mathcal{T}$ and an initial state s, there exists a unique infinite play in arena $\mathcal{A}$, denoted $p(s, \sigma, \tau)$, consistent with σ and τ and such that $s = \mathrm{source}(p(s, \sigma, \tau))$.

Definition 2.1. Strategies $\sigma^\sharp \in \Sigma$ and $\tau^\sharp \in \mathcal{T}$ are *optimal* in the game $(\mathcal{A}, u)$ if

$$\begin{aligned}&\forall s \in S, \forall \sigma \in \Sigma, \forall \tau \in \mathcal{T},\\ &\qquad u(p(s, \sigma, \tau^\sharp)) \leq u(p(s, \sigma^\sharp, \tau^\sharp)) \leq u(p(s, \sigma^\sharp, \tau)) \ .\end{aligned} \tag{1.4}$$

Thus if strategies $\sigma^\sharp$ and $\tau^\sharp$ are optimal then the players do not have any incentive to change them unilaterally: player 1 cannot increase his gain by switching to another strategy σ while player 2 cannot decrease his losses by switching to another strategy τ.

In other words if player 2 plays according to $\tau^\sharp$ then the best response of player 1 is to play with $\sigma^\sharp$, no other strategy can do better for him. Conversely, if player 1 plays according to $\sigma^\sharp$ then the best response of player 2 is to play according to $\tau^\sharp$ as no other strategy does better to limit his losses.

We say that a payoff mapping u *admits optimal positional strategies* if for all games $(\mathcal{A}, u)$ over *finite arenas* there exist optimal positional strategies for both players. We should emphasize that the property defined above is a property of the payoff mapping and not a property of a particular game, we require that both players have optimal positional strategies for *all possible games* over finite arenas.

It is important to note that zero-sum games that we consider here, i.e. the games where the gain of one player is equal to the loss of his adversary, satisfy the exchangeability property for optimal strategies:

for any two pairs of optimal strategies $(\sigma^\sharp, \tau^\sharp)$ and $(\sigma^\star, \tau^\star)$, the pairs $(\sigma^\star, \tau^\sharp)$ and $(\sigma^\sharp, \tau^\star)$ are also optimal and, moreover,

$$u(p(s, \sigma^\sharp, \tau^\sharp)) = u(p(s, \sigma^\star, \tau^\star)) \ \ ,$$

i.e. the value of $u(p(s, \sigma^\sharp, \tau^\sharp))$ is independent of the choice of the optimal strategies — this is *the value of the game* $(\mathcal{A}, u)$ *at state* s.

We end this general introduction with two simple lemmas.

Lemma 2.2. Let u be a payoff mapping admitting optimal positional strategies for both players.

(A) Suppose that $\sigma \in \Sigma$ is any strategy while $\tau^\sharp \in \mathcal{T}_p$ is positional. Then there exists a positional strategy $\sigma^\sharp \in \Sigma_p$ such that

$$\forall s \in S, \quad u(p(s, \sigma, \tau^\sharp)) \leq u(p(s, \sigma^\sharp, \tau^\sharp)) \ \ . \tag{1.5}$$

(B) Similarly, if $\tau \in \mathcal{T}$ is any strategy and $\sigma^\sharp \in \Sigma_p$ a positional strategy then there exists a positional strategy $\tau^\sharp \in \mathcal{T}_p$ such that

$$\forall s \in S, \quad u(p(s, \sigma^\sharp, \tau^\sharp)) \leq u(p(s, \sigma^\sharp, \tau)) \ \ .$$

Proof. We prove (A), the proof of (B) is similar. Take any strategies $\sigma \in \Sigma$ and $\tau^\sharp \in \mathcal{T}_p$. Let $\mathcal{A}'$ be a subarena of $\mathcal{A}$ obtained by restricting the actions of player 2 to the actions given by the strategy $\tau^\sharp$, i.e. in $\mathcal{A}'$ the only possible strategy for player 2 is the strategy $\tau^\sharp$. The actions of player 1 are not restricted, i.e. in $\mathcal{A}'$ player 1 has the same available actions as in $\mathcal{A}$, in particular σ is a valid strategy of player 1 on $\mathcal{A}'$. Since u admits optimal positional strategies, player 1 has an optimal positional strategy $\sigma^\sharp$ on $\mathcal{A}'$. But (1.5) is just the optimality condition of $\sigma^\sharp$ on $\mathcal{A}'$. Q.E.D.

Lemma 2.3. Suppose that the payoff mapping u admits optimal positional strategies. Let $\sigma^\sharp \in \Sigma_p$ and $\tau^\sharp \in \mathcal{T}_p$ be positional strategies such that

$$\forall s \in S, \forall \sigma \in \Sigma_p, \forall \tau \in \mathcal{T}_p,$$
$$u(p(s,\sigma,\tau^\sharp)) \leq u(p(s,\sigma^\sharp,\tau^\sharp) \leq u(p(s,\sigma^\sharp,\tau)) \ , \tag{1.6}$$

i.e. $\sigma^\sharp$ and $\tau^\sharp$ are optimal in the class of positional strategies. Then $\sigma^\sharp$ and $\tau^\sharp$ are optimal in the class of all strategies.

Proof. Suppose that

$$\exists \tau \in \mathcal{T}, \quad u(p(s,\sigma^\sharp,\tau)) < u(p(s,\sigma^\sharp,\tau^\sharp)) \ . \tag{1.7}$$

By Lemma 2.2 (B) there exists a positional strategy $\tau^\star \in \mathcal{T}_p$ such that $u(p(s,\sigma^\sharp,\tau^\star)) \leq u(p(s,\sigma^\sharp,\tau)) < u(p(s,\sigma^\sharp,\tau^\sharp))$, contradicting (1.6). Thus $\forall \tau \in \mathcal{T}, u(p(s,\sigma^\sharp,\tau^\sharp)) \leq u(p(s,\sigma^\sharp,\tau))$. The left hand side of (1.4) can be proved in a similar way. Q.E.D.

3 Discounted Games

Discounted games where introduced by Shapley [19] who proved that stochastic discounted games admit stationary optimal strategies. Our exposition follows very closely the original approach of [19] and that of [17]. Nevertheless we present a complete proof for the sake of completeness.

Arenas for discounted games are equipped with two mappings defined on the set S of states: the *discount mapping*

$$\lambda : S \longrightarrow [0,1)$$

associates with each state s a discount factor $\lambda(s) \in [0,1)$ and the *reward mapping*

$$r : S \longrightarrow \mathbb{R} \tag{1.8}$$

maps each state s to a real valued reward $r(s)$.

The payoff mapping

$$u_\lambda : S^\omega \longrightarrow \mathbb{R}$$

for discounted games is defined in the following way: for each play $p = s_0 s_1 s_2 \ldots \in S^\omega$

$$\begin{aligned} u_\lambda(p) &= (1-\lambda(s_0))r(s_0) + \lambda(s_0)(1-\lambda(s_1))r(s_1) + \lambda(s_0)\lambda(s_1)(1-\lambda(s_2))r(s_2) + \ldots \\ &= \sum_{i=0}^{\infty} \lambda(s_0)\ldots\lambda(s_{i-1})(1-\lambda(s_i))r(s_i) \ . \end{aligned} \tag{1.9}$$

Usually when discounted games are considered it is assumed that there is only one discount factor, i.e. that there exists $\lambda \in [0,1)$ such that $\lambda(s) = \lambda$ for all $s \in S$. But for us it is essential that discount factors depend on the state.

It is difficult to give an intuitively convincing interpretation of (1.9) if we use this payoff mapping to evaluate infinite games. However, there is a natural interpretation of (1.9) in terms of stopping games, in fact this is the original interpretation given by Shapley[19].

In *stopping games* the nature introduces an element of uncertainty. Suppose that at a stage i a state s_i is visited. Then, before the player controlling s_i is allowed to execute an action, a (biased) coin is tossed to decide if the game stops or if it will continue. The probability that the game stops is $1 - \lambda(s_i)$ (thus $\lambda(s_i)$ gives the probability that the game continues). Let us note immediately that since we have assumed that $0 \leq \lambda(s) < 1$ for all $s \in S$, the stopping probabilities are strictly positive therefore the game actually stops with probability 1 after a finite number of steps.

If the game stops at s_i then player 1 receives from player 2 the payment $r(s_i)$. This ends the game, there is no other payment in the future.

If the game does not stop at s_i then there is no payment at this stage and the player controlling the state s_i is allowed to choose an action to execute[3].

Now note that $\lambda(s_0) \dots \lambda(s_{i-1})(1 - \lambda(s_i))$ is the probability that the game have not stopped at any of the states $s_0, \dots, s_{i-1}$ but it does stop at state s_i. Since this event results in the payment $r(s_i)$ received by player 1, Eq. (1.9) gives in fact the *payoff expectation* for a play $s_0 s_1 s_2 \dots$.

Shapley [19] proved[4] that

Theorem 3.1 (Shapley)**.** Discounted games $(\mathcal{A}, u_\lambda)$ over finite arenas admit optimal positional strategies for both players.

Proof. Let $\mathbb{R}^S$ be the vector space consisting of mappings from S to $\mathbb{R}$. For $f \in \mathbb{R}^S$, set $||f|| = \sup_{s \in S} |f(s)|$. Since S is finite $|| \cdot ||$ is a norm for which $\mathbb{R}^S$ is complete. Consider an operator $\Psi : \mathbb{R}^S \longrightarrow \mathbb{R}^S$, for $f \in \mathbb{R}^S$ and $s \in S$,

$$\Psi[f](s) = \begin{cases} \max_{(s,s') \in A(s)} (1 - \lambda(s)) r(s) + \lambda(s) f(s') & \text{if } s \in S_1 \\ \min_{(s,s') \in A(s)} (1 - \lambda(s)) r(s) + \lambda(s) f(s') & \text{if } s \in S_2 \end{cases} .$$

$\Psi[f](s)$ can be seen as the value of a one shot game that gives the payoff $(1 - \lambda(s)) r(s) + \lambda(s) f(s')$ if the player controlling the state s choses an action $(s, s') \in A(s)$.

[3] More precisely, if the nature does not stop the game then the player controlling the current state *is obliged* to execute an action, players cannot stop the game by themselves.

[4] In fact, Shapley considered a much larger class of stochastic games.

We can immediately note that Ψ is monotone, if $f \geq g$ then $\Psi[f] \geq \Psi[g]$, where $f \geq g$ means that $f(s) \geq g(s)$ for all states $s \in S$.

Moreover, for any positive constant c and $f \in \mathbb{R}^S$

$$\Psi[f] - c\lambda\mathbf{1} \leq \Psi[f - c \cdot \mathbf{1}] \quad \text{and} \quad \Psi[f + c \cdot \mathbf{1}] \leq \Psi[f] + c\lambda\mathbf{1} \ , \tag{1.10}$$

where $\mathbf{1}$ is the constant mapping, $\mathbf{1}(s) = 1$ for each state s, and $\lambda = \sup_{s \in S} \lambda(s)$.

Therefore, since

$$f - ||f - g|| \cdot \mathbf{1} \leq g \leq f + ||f - g|| \cdot \mathbf{1} \ ,$$

we get

$$\Psi[f] - \lambda||f - g|| \cdot \mathbf{1} \leq \Psi[g] \leq \Psi[f] + \lambda||f - g|| \cdot \mathbf{1} \ ,$$

implying

$$||\Psi[f] - \Psi[g]|| \leq \lambda||f - g|| \ .$$

By Banach contraction principle, Ψ has a unique fixed point $w \in \mathbb{R}^S$, $\Psi[w] = w$. From the definition of Ψ we can see that this unique fixed point satisfies the inequalities

$$\forall s \in S_1, \forall (s, s') \in A(s), \quad w(s) \geq (1 - \lambda(s))r(s) + \lambda(s)w(s') \tag{1.11}$$

and

$$\forall s \in S_2, \forall (s, s') \in A(s), \quad w(s) \leq (1 - \lambda(s))r(s) + \lambda(s)w(s') \ . \tag{1.12}$$

Moreover, for each $s \in S$ there is an action $\xi(s) = (s, s') \in A(s)$ such that

$$w(s) = (1 - \lambda(s))r(s) + \lambda(s)w(s') \ . \tag{1.13}$$

We set $\sigma^\sharp(s) = \xi(s)$ for $s \in S_1$ and $\tau^\sharp(s) = \xi(s)$ for $s \in S_2$ and we show that $\sigma^\sharp$ and $\tau^\sharp$ are optimal for player 1 and 2. Suppose that player 1 plays according to the strategy $\sigma^\sharp$ while player 2 according to some strategy τ. Let $p(s_0, \sigma^\sharp, \tau) = s_0 s_1 s_2 \ldots$. Then, using (1.12) and (1.13), we get by induction on k that

$$w(s_0) \leq \sum_{i=0}^{k} \lambda(s_0) \ldots \lambda(s_{i-1})(1 - \lambda(s_i))r(s_i) + \lambda(s_0) \ldots \lambda(s_k)w(s_{k+1}) \ .$$

Tending k to infinity we get

$$w(s_0) \leq u_\lambda(p(s_0, \sigma^\sharp, \tau)) \ .$$

In a similar way we can establish that for any strategy σ of player 1,

$$w(s_0) \geq u_\lambda(p(s_0, \sigma, \tau^\sharp))$$

and, finally, that

$$w(s_0) = u_\lambda(p(s_0, \sigma^\sharp, \tau^\sharp)) \ ,$$

proving the optimality of $\sigma^\sharp$ and $\tau^\sharp$. Q.E.D.

4 Priority mean-payoff games

In mean-payoff games the players try to optimize (maximize/minimize) the mean value of the payoff received at each stage. In such games the *reward mapping*

$$r : S \longrightarrow \mathbb{R} \tag{1.14}$$

gives, for each state s, the payoff received by player 1 when s is visited. The payoff of an infinite play is defined as the mean value of daily payments:

$$u_m(s_0 s_1 s_2 \ldots) = \limsup_k \frac{1}{k+1} \sum_{i=0}^{k} r(s_i) \ , \tag{1.15}$$

where we take lim sup rather than the simple limit since the latter may not exist. As proved by Ehrenfeucht and Mycielski [4], such games admit optimal positional strategies; other proofs can be found for example in [1, 7].

We slightly generalize mean-payoff games by equipping arenas with a new mapping

$$w : S \longrightarrow \mathbb{R}_+ \tag{1.16}$$

associating with each state s a *strictly positive* real number $w(s)$, the *weight* of s. We can interpret $w(s)$ as the amount of time spent at state s each time when s is visited. In this setting $r(s)$ should be seen as the payoff by a time unit when s is visited, thus the mean payoff received by player 1 is

$$u_m(s_0 s_1 s_2 \ldots) = \limsup_k \frac{\sum_{i=0}^{k} w(s_i) r(s_i)}{\sum_{i=0}^{k} w(s_i)} \ . \tag{1.17}$$

Note that in the special case when the weights are all equal to 1, the weighted mean value (1.17) reduces to (1.15).

As a final ingredient we add to our arena a *priority mapping*

$$\pi : S \longrightarrow \mathbb{Z}_+ \tag{1.18}$$

giving a positive integer *priority* $\pi(s)$ of each state s.

We define the *priority* of a play $p = s_0 s_1 s_2 \ldots$ as the *smallest* priority appearing infinitely often in the sequence $\pi(s_0)\pi(s_1)\pi(s_2)\ldots$ of priorities visited in p:

$$\pi(p) = \liminf_i \pi(s_i) \ . \tag{1.19}$$

For any priority a, let $\mathbf{1}_a : S \longrightarrow \{0, 1\}$ be the indicator function of the set $\{s \in S \mid \pi(s) = a\}$, i.e.

$$\mathbf{1}_a(s) = \begin{cases} 1 & \text{if } \pi(s) = a \\ 0 & \text{otherwise.} \end{cases} \tag{1.20}$$

Then the priority mean payoff of a play $p = s_0 s_1 s_2 \dots$ is defined as

$$u_{\mathrm{pm}}(p) = \limsup_k \frac{\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s) \cdot w(s_i) \cdot r(s_i)}{\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s_i) \cdot w(s_i)} \quad . \tag{1.21}$$

In other words, to calculate priority mean payoff $u_{\mathrm{pm}}(p)$ we take weighted mean payoff but with the weights of all states having priorities different from $\pi(p)$ shrunk to 0. (Let us note that the denominator $\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s_i) \cdot w(s_i)$ is different from 0 for k large enough, in fact it tends to infinity since $\mathbf{1}_{\pi(p)}(s_i) = 1$ for infinitely many i. For small k the numerator and the denominator can be equal to 0 and then, to avoid all misunderstanding, it is convenient to assume that the indefinite value 0/0 is equal to $-\infty$.)

Suppose that for all states s,

- $w(s) = 1$ and
- $r(s)$ is 0 if $\pi(s)$ is even, and $r(s)$ is 1 if $\pi(s)$ is odd.

Then the payoff obtained by player 1 for any play p is either 1 if $\pi(p)$ is odd, or 0 if $\pi(p)$ is even. If we interpret the payoff 1 as the victory of player 1, and payoff 0 as his defeat then such a game is just the usual parity game [5, 11].

It turns out that

Theorem 4.1. For any arena $\mathcal{A}$ the priority mean-payoff game $(\mathcal{A}, u_{\mathrm{pm}})$ admits optimal positional strategies for both players.

There are many possible ways to prove Theorem 4.1, for example by adapting the proofs of positionality of mean payoff games from [4] and [1] or by verifying that u_{pm} satisfies sufficient positionality conditions given in [7]. Below we give a complete proof based mainly on ideas from [7, 20].

A payoff mapping is said to be *prefix independent* if for each play p and for each factorization $p = xy$ with x finite we have $u(p) = u(y)$, i.e. the payoff does not depend on finite prefixes of a play. The reader can readily persuade herself that the priority mean payoff mapping u_{pm} is prefix independent.

Lemma 4.2. Let u be a prefix-independent payoff mapping such that both players have optimal positional strategies $\sigma^\sharp$ and $\tau^\sharp$ in the game $(\mathcal{A}, u)$. Let $\mathrm{val}(s) = p(s, \sigma^\sharp, \tau^\sharp)$, $s \in S$, be the game value for an initial state s.

For any action $(s, t) \in A$,

(1) if $s \in S_1$ then $\mathrm{val}(s) \geq \mathrm{val}(t)$,

(2) if $s \in S_2$ then $\mathrm{val}(s) \leq \mathrm{val}(t)$,

(3) if $s \in S_1$ and $\sigma^\sharp(s) = t$ then $\mathrm{val}(s) = \mathrm{val}(t)$,

(4) if $s \in S_2$ and $\tau^\sharp(s) = t$ then $\mathrm{val}(s) = \mathrm{val}(t)$.

Proof. (1) This is quite obvious. If $s \in S_1$, $(s,t) \in A$ and $\mathrm{val}(s) < \mathrm{val}(t)$ then for a play starting at s player 1 could secure for himself at least $\mathrm{val}(t)$ by executing first the action (s,t) and next playing with his optimal strategy. But this contradicts the definition of $\mathrm{val}(s)$ since from s player 2 has a strategy that limits his losses to $\mathrm{val}(s)$.

The proof of (2) is obviously similar.

(3) We know by (1) that if $s \in S_1$ and $\sigma^\sharp(s) = t$ then $\mathrm{val}(s) \geq \mathrm{val}(t)$. This inequality cannot be strict since from t player 2 can play in such a way that his loss does not exceed $\mathrm{val}(t)$.

(4) is dual to (1).

Q.E.D.

Proof of Theorem 4.1. We define the size of an arena $\mathcal{A}$ to be the difference $|A| - |S|$ of the number of actions and the number of states and we carry the proof by induction on the size of $\mathcal{A}$. Note that since for each state there is at least one available action the size of each arena is ≥ 0.

If for each state there is only one available action then the number of actions is equal to the number of states, the size of $\mathcal{A}$ is 0, and each player has just one possible strategy, both these strategies are positional and, obviously, optimal.

Suppose that both players have optimal positional strategies for arenas of size $< k$ and let $\mathcal{A}$ be of size k, $k \geq 1$.

Then there exists a state with at least two available actions. Let us fix such a state t, we call it the *pivot*. We assume that t is controlled by player 1

$$t \in S_1 \tag{1.22}$$

(the case when it is controlled by player 2 is symmetric).

Let $A(t) = A_L(t) \cup A_R(t)$ be a partition of the set $A(t)$ of actions available at t onto two disjoint non-empty sets. Let $\mathcal{A}_L$ and $\mathcal{A}_R$ be two arenas, we call them left and right arenas, both of them having the same states as $\mathcal{A}$, the same reward, weight and priority mappings and the same available actions for all states different from t. For the pivot state t, $\mathcal{A}_L$ and $\mathcal{A}_R$ have respectively $A_L(t)$ and $A_R(t)$ as the sets of available actions. Thus, since $\mathcal{A}_L$ and $\mathcal{A}_R$ have less actions than $\mathcal{A}$, their size is smaller than the size of $\mathcal{A}$ and, by induction hypothesis, both players have optimal positional strategies: $(\sigma_L^\sharp, \tau_L^\sharp)$ on $\mathcal{A}_L$ and $(\sigma_R^\sharp, \tau_R^\sharp)$ on $\mathcal{A}_R$.

We set $\mathrm{val}_L(s) = u_{\mathrm{pm}}(p(s, \sigma_L^\sharp, \tau_L^\sharp))$ and $\mathrm{val}_R(s) = u_{\mathrm{pm}}(p(s, \sigma_R^\sharp, \tau_R^\sharp))$ to be the values of a state s respectively in the left and the right arena.

Without loss of generality we can assume that for the pivot state t

$$\mathrm{val}_L(t) \leq \mathrm{val}_R(t) \ . \tag{1.23}$$

We show that this implies that

$$\forall s \in S, \quad \mathrm{val}_L(s) \leq \mathrm{val}_R(s) \ . \tag{1.24}$$

Suppose the contrary, i.e. that the set

$$X = \{s \in S \mid \mathrm{val}_L(s) > \mathrm{val}_R(s)\}$$

is non-empty. We define a positional strategy σ^* for player 1

$$\sigma^*(s) = \begin{cases} \sigma^\sharp_L(s) & \text{if } s \in X \cap S_1 \\ \sigma^\sharp_R(s) & \text{if } s \in (S \setminus X) \cap S_1. \end{cases} \tag{1.25}$$

Note that, since the pivot state t does not belong to X, for $s \in X \cap S_1$, $\sigma^\sharp_L(s)$ is valid action for player 1 not only in $\mathcal{A}_L$ but also in $\mathcal{A}_R$, therefore the strategy σ^* defined above is a valid positional strategy on the arena $\mathcal{A}_R$.

We claim that

For games on $\mathcal{A}_R$ starting at a state $s_0 \in X$ strategy σ^* guarantees that player 1 wins at least $\mathrm{val}_L(s_0)$ (against any strategy of player 2). (1.26)

Suppose that we start a game on $\mathcal{A}_R$ at a state s_0 and player 1 plays according to σ^* while player 2 uses any strategy τ. Let

$$p(s_0, \sigma^*, \tau) = s_0 s_1 s_2 \ldots \tag{1.27}$$

be the resulting play. We define

$$\forall s \in S, \quad \mathrm{val}(s) = \begin{cases} \mathrm{val}_L(s) & \text{for } s \in X, \\ \mathrm{val}_R(s) & \text{for } s \in S \setminus X. \end{cases} \tag{1.28}$$

We shall show that the sequence $\mathrm{val}(s_0), \mathrm{val}(s_1), \mathrm{val}(s_2), \ldots$ is non-decreasing,

$$\forall i, \quad \mathrm{val}(s_i) \leq \mathrm{val}(s_{i+1}) \ . \tag{1.29}$$

Since strategies $\sigma^\sharp_L$ and $\sigma^\sharp_R$ are optimal in $\mathcal{A}_L$ and $\mathcal{A}_R$, Lemma 4.2 and (1.28) imply that for all i

$$\mathrm{val}(s_i) = \mathrm{val}_L(s_i) \leq \mathrm{val}_L(s_{i+1}) \quad \text{if } s_i \in X, \tag{1.30}$$

and

$$\mathrm{val}(s_i) = \mathrm{val}_R(s_i) \leq \mathrm{val}_R(s_{i+1}) \quad \text{if } s_i \in S \setminus X. \tag{1.31}$$

To prove (1.29) there are four cases to examine:

(1) Suppose that s_i and s_{i+1} belong to X. Then $\mathrm{val}(s_{i+1}) = \mathrm{val}_L(s_{i+1})$ and (1.29) follows from (1.30).

(2) Suppose that s_i and s_{i+1} belong to $S \setminus X$. Then $\mathrm{val}(s_{i+1}) = \mathrm{val}_R(s_{i+1})$ and now (1.29) follows from (1.31).

(3) Let $s_i \in X$ and $s_{i+1} \in S \setminus X$. Then (1.29) follows from (1.30) and from the fact that $\mathrm{val}_L(s_{i+1}) \le \mathrm{val}_R(s_{i+1}) = \mathrm{val}(s_{i+1})$.

(4) Let $s_i \in S \setminus X$ and $s_{i+1} \in X$. Then $\mathrm{val}_R(s_{i+1}) < \mathrm{val}_L(s_{i+1}) = \mathrm{val}(s_{i+1})$, which, by (1.31), implies (1.29). Note that in this case we have the strict inequality $\mathrm{val}(s_i) < \mathrm{val}(s_{i+1})$.

This terminates the proof of (1.29).

Since the set $\{\mathrm{val}(s) \mid s \in S\}$ is finite (1.29) implies that the sequence $\mathrm{val}(s_i), i = 0, 1, \ldots,$ ultimately constant. But examining the case (4) above we have established that each passage from $S \setminus X$ to X strictly increases the value of val. Thus from some stage n onward all states s_i, $i \ge n$, are either in X or in $S \setminus X$. Therefore, according to (1.25), from the stage n onward player 1 always plays either $\sigma_L^\sharp$ or $\sigma_R^\sharp$ and the optimality of both strategies assures that he wins at least $\mathrm{val}(s_n)$, i.e.

$$u_{\mathrm{pm}}(p(s_0, \sigma^*, \tau)) = u_{\mathrm{pm}}(s_0 s_1 \ldots) = u_{\mathrm{pm}}(s_n s_{n+1} s_{n+2} \ldots) \ge \mathrm{val}(s_n) \ge \mathrm{val}(s_0).$$

In particular, if $s_0 \in X$ then using strategy σ^* player 1 secures for himself the payoff of at least $\mathrm{val}(s_0) = \mathrm{val}_L(s_0)$ against any strategy of player 2, which proves (1.26). On the other hand, the optimality of $\tau_R^\sharp$ implies that player 2 can limit his losses to $\mathrm{val}_R(s_0)$ by using strategy $\tau_R^\sharp$. But how player 1 can win at least $\mathrm{val}_L(s_0)$ while player 2 loses no more than $\mathrm{val}_R(s_0)$ if $\mathrm{val}_L(s_0) > \mathrm{val}_R(s_0)$ for $s_0 \in X$? We conclude that the set X is empty and (1.24) holds.

Now our aim is to prove that (1.23) implies that the strategy $\sigma_R^\sharp$ is optimal for player 1 not only in $\mathcal{A}_R$ but also for games on the arena $\mathcal{A}$. Clearly player 1 can secure for himself the payoff of at least $\mathrm{val}_R(s)$ by playing according to $\sigma_R^\sharp$ on $\mathcal{A}$. We should show that he cannot do better. To this end we exhibit a strategy $\tau^\sharp$ for player 2 that limits the losses of player 2 to $\mathrm{val}_R(s)$ on the arena $\mathcal{A}$.

At each stage player 2 will use either his positional strategy $\tau_L^\sharp$ optimal in $\mathcal{A}_L$ or strategy $\tau_R^\sharp$ optimal in $\mathcal{A}_R$. However, in general neither of these strategies is optimal for him in $\mathcal{A}$ and thus it is not a good idea for him to stick to one of these strategies permanently, he should rather adapt his strategy to the moves of his adversary. To implement the strategy $\tau^\sharp$ player 2 will need one bit of memory (the strategy $\tau^\sharp$ we construct here is *not* positional). He uses this memory to remember if at the last passage through the pivot state t player 1 took an action of $A_L(t)$ or an action of $A_R(t)$. In the former case player 2 plays using the strategy $\tau^\sharp_L$, in the latter case he plays using the strategy $\tau^\sharp_R$. In the periods between two passages through t player 2 does not change his strategy, he sticks either to $\tau^\sharp_L$ or to $\tau^\sharp_R$, he switches from one of these strategies to the other only when compelled by the action taken by player 1 during the last visit at the pivot state[5]. It remains to specify which strategy player 2 uses until the first passage through t and we assume that it is the strategy $\tau^\sharp_R$.

Let $s_0 \in S$ be an initial state and let σ be some, not necessarily positional, strategy of 1 for playing on $\mathcal{A}$. Let

$$p(s_0, \sigma, \tau^\sharp) = s_0 s_1 s_2 \ldots \tag{1.32}$$

be the resulting play. Our aim is to show that

$$u_{\text{pm}}(p(s_0, \sigma, \tau^\sharp)) \le \text{val}_R(s_0) \ . \tag{1.33}$$

If $p(s_0, \sigma, \tau^\sharp)$ never goes through t then $p(s_0, \sigma, \tau^\sharp)$ is in fact a play in $\mathcal{A}_R$ consistent with $\tau^\sharp_R$ which immediately implies (1.33).

Suppose now that $p(s_0, \sigma, \tau^\sharp)$ goes through t and let k be the first stage such that $s_k = t$. Then the initial history $s_0 s_1 \ldots s_k$ is consistent with $\tau^\sharp_R$ which, by Lemma 4.2, implies that

$$\text{val}_R(t) \le \text{val}_R(s_0) \ . \tag{1.34}$$

If there exists a stage n such that $s_n = t$ and player 2 does not change his strategy after this stage[6], i.e. he plays from the stage n onward either $\tau^\sharp_L$ or $\tau^\sharp_R$ then the suffix play $s_n s_{n+1} \ldots$ is consistent with one of these strategies implying that either $u_{\text{pm}}(s_n s_{n+1} \ldots) \le \text{val}_L(t)$ or $u_{\text{pm}}(s_n s_{n+1} \ldots) \le \text{val}_R(t)$. But $u_{\text{pm}}(s_n s_{n+1} \ldots) = u_{\text{pm}}(p(s_0, \sigma, \tau^\sharp))$ and thus (1.34) and (1.23) imply (1.33).

The last case to consider is when player 2 switches infinitely often between $\tau^\sharp_R$ and $\tau^\sharp_L$.

In the sequel we say that a non-empty sequence of states z contains only actions of $\mathcal{A}_R$ if for each factorization $z = z' s' s'' z''$ with $s', s'' \in S$, (s', s'') is an action of $\mathcal{A}_R$. (Obviously, there is in a similar definition for $\mathcal{A}_L$.)

[5] Note the intuition behind the strategy $\tau^\sharp$: If at the last passage through the pivot state t player 1 took an action of $A_L(t)$ then, at least until the next visit to t, the play is like the one in the game $\mathcal{A}_L$ (all actions taken by the players are actions of $\mathcal{A}_L$) and then it seems reasonable for player 2 to respond with his optimal strategy on $\mathcal{A}_L$. On the other hand, if at the last passage through t player 1 took an action of $A_R(t)$ then from this moment onward until the next visit to t we play like in $\mathcal{A}_R$ and then player 2 will respond with his optimal strategy on $\mathcal{A}_R$.

[6] In particular this happens if $p(s_0, \sigma, \tau^\sharp)$ goes finitely often through t.

Since now we consider the case when the play $p(s_0, \sigma, \tau^\sharp)$ contains infinitely many actions of $A_L(t)$ and infinitely many actions of $A_R(t)$ there exists a unique infinite factorization

$$p(s_0, \sigma, \tau^\sharp) = x_0 x_1 x_2 x_3 \dots \quad , \tag{1.35}$$

such that

- each x_i, $i \geq 1$, is non-empty and begins with the pivot state t,
- each path $x_{2i}t$, $i = 0, 1, 2, \dots$ contains only actions of $\mathcal{A}_R$ while
- each path $x_{2i+1}t$ contains only actions of $\mathcal{A}_L$.

(Intuitively, we have factorized the play $p(s_0, \sigma, \tau^\sharp)$ according to the strategy used by player 2.)

Let us note that the conditions above imply that

$$x_R = x_2 x_4 x_6 \dots \quad \text{and} \quad x_L = x_1 x_3 x_5 \dots \quad . \tag{1.36}$$

are infinite paths respectively in $\mathcal{A}_R$ and $\mathcal{A}_L$.

Moreover x_R is a play consistent with $\tau_R^\sharp$ while x_L is consistent with $\tau_L^\sharp$. By optimality of strategies $\tau_R^\sharp$, $\tau_L^\sharp$,

$$u_{\text{pm}}(x_R) \leq \text{val}_R(t) \quad \text{and} \quad u_{\text{pm}}(x_L) \leq \text{val}_L(t) \quad . \tag{1.37}$$

It is easy to see that path priorities satisfy $\pi(x_R) \geq \pi(p(s_0, \sigma, \tau^\sharp))$ and $\pi(x_L) \geq \pi(p(s_0, \sigma, \tau^\sharp))$ and at most one of these inequalities is strict.

(1) If $\pi(x_R) > \pi(p(s_0, \sigma, \tau^\sharp))$ and $\pi(x_L) = \pi(p(s_0, \sigma, \tau^\sharp))$ then there exists m such that all states in the suffix $x_{2m} x_{2m+2} x_{2m+4} \dots$ of x_R have priorities greater than $\pi(p(s_0, \sigma, \tau^\sharp))$ and do not contribute to the payoff $u_{\text{pm}}(x_{2m} x_{2m+1} x_{2m+2} x_{2m+3} \dots)$.

This and the prefix-independence property of u_{pm} imply

$$\begin{aligned} u_{\text{pm}}(p(s_0, \sigma, \tau^\sharp)) &= u_{\text{pm}}(x_{2m} x_{2m+1} x_{2m+2} x_{2m+3} \dots) = \\ & u_{\text{pm}}(x_{2m+1} x_{2m+3} \dots) = u_{\text{pm}}(x_L) \leq \text{val}_L(s_0) \leq \text{val}_R(s_0), \end{aligned}$$

where the first inequality follows from the fact that x_L is consistent with the optimal strategy $\tau_L^\sharp$.

(2) If $\pi(x_L) > \pi(p(s_0, \sigma, \tau^\sharp))$ and $\pi(x_R) = \pi(p(s_0, \sigma, \tau^\sharp))$ then we get in a similar way $u_{\text{pm}}(p(s_0, \sigma, \tau^\sharp)) = u_{\text{pm}}(x_R) \leq \text{val}_R(s_0)$.

(3) Let $a = \pi(x_R) = \pi(p(s_0, \sigma, \tau^\sharp)) = \pi(x_L)$. For a sequence $t_0 t_1 \dots t_l$ of states we define

$$F_a(t_0 \dots t_l) = \sum_{i=1}^{l} \mathbf{1}_a(t_i) \cdot w(t_i) \cdot r(t_i)$$

and

$$G_a(t_0 \dots t_l) = \sum_{i=1}^{l} \mathbf{1}_a(t_i) \cdot r(t_i),$$

where $\mathbf{1}_a$ is defined in (1.20). Thus for an infinite path p, $u_{\mathrm{pm}}(p) = \limsup_i F_a(p_i)/G_a(p_i)$, where p_i is the prefix of length i of p.

Take any $\epsilon > 0$. Eq. (1.37) implies that for all sufficiently long prefixes y_L of x_L, $F_a(y_L)/G_a(y_L) \leq \mathrm{val}_L(t) + \epsilon \leq \mathrm{val}_R(t) + \epsilon$ and similarly for all sufficiently long prefixes y_R of x_R, $F_a(y_R)/G_a(y_R) \leq \mathrm{val}_R(t) + \epsilon$. Then we also have

$$\frac{F_a(y_R) + F_a(y_L)}{G_a(y_R) + G_a(y_L)} \leq \mathrm{val}_R(t) + \epsilon \ . \tag{1.38}$$

If y is a proper prefix of the infinite path $x_1x_2x_3\dots$ then

$$y = x_1x_2\dots x_{2i-1}x'_{2i}x'_{2i+1} \ ,$$

where

- either x'_{2i} is a prefix of x_{2i} and x'_{2i+1} is empty or
- $x'_{2i} = x_{2i}$ and x'_{2i+1} is a prefix of x_{2i+1}

(and x_i are as in factorization (1.35)). Then $y_R = x_2x_4\dots x'_{2i}$ is a prefix of x_R while $y_L = x_1x_3\dots x_{2i-1}x'_{2i+1}$ is a prefix of x_L. If the length of y tends to ∞ then the lengths of y_R and y_L tend to ∞. Since $G_a(y) = G_a(y_R) + G_a(y_L)$ and $F_a(y) = F_a(y_R) + G_a(y_L)$ Eq. (1.38) implies that $G_a(y)/F_a(y) \leq \mathrm{val}_R(t) + \epsilon$. Since the last inequality holds for all sufficiently long finite prefixes of $x_1x_2x_3\dots$ we get that $u_{\mathrm{pm}}(p(s_0, \sigma, \tau^\sharp)) = u_{\mathrm{pm}}(x_1x_2x_3\dots) \leq \mathrm{val}_R(s_0) + \epsilon$. As this is true for all $\epsilon > 0$ we have in fact $u_{\mathrm{pm}}(p(s_0, \sigma, \tau^\sharp)) \leq \mathrm{val}_R(s_0)$.

This terminates the proof that if player 2 plays according to strategy $\tau^\sharp$ then his losses do not extend $\mathrm{val}_R(s_0)$.

We can conclude that strategies $\sigma^\sharp_R$ and $\tau^\sharp$ are optimal on $\mathcal{A}$ and for each initial state s the value of a game on $\mathcal{A}$ is the same as in $\mathcal{A}_R$.

Note however that while player 1 can use his optimal positional strategy $\sigma^\sharp_R$ to play optimally on $\mathcal{A}$ the situation is more complicated for player 2. The optimal strategy that we have constructed for him is not positional and certainly if we pick some of his optimal positional strategies on $\mathcal{A}_R$ then we cannot guarantee that it will remain optimal on $\mathcal{A}$.

To obtain an optimal positional strategy for player 2 we proceed as follows:

If for each state $s \in S_2$ controlled by player 2 there is only one available action then player 2 has only one strategy ($\tau^\sharp_R = \tau^\sharp_L$). Thus in this case player 2 needs no memory.

If there exists a state $t \in S_2$ with at least two available actions then we take this state as the pivot and by the same reasoning as previously we find a pair of optimal strategies $(\sigma^*, \tau^\sharp)$ such that $\tau^\sharp$ is positional while σ^* may need one bit of memory to be implemented.

By exchangeability property of optimal strategies we can conclude that $(\sigma^\sharp, \tau^\sharp)$ is a couple of optimal positional strategies.

Q.E.D.

5 Blackwell optimality

Let us return to discounted games. In this section we examine what happens if, for all states s, the discount factors $\lambda(s)$ tend to 1 or, equivalently, the stopping probabilites tend to 0.

When all discount factors are equal and tend to 1 with the same rate then the value of discounted game tends to the value of a simple mean-payoff game, this is a classical result examined extensively by many authors in the context of stochastic games, see [6] and the references therein.

What happens however if discount factors tend to 1 with different rates for different states? To examine this limit we assume in the sequel that arenas for discounted games are equipped not only with a reward mapping $r : S \longrightarrow \mathbb{R}$ but also with a priority mapping $\pi : S \longrightarrow \mathbb{Z}_+$ and a weight mapping $w : S \longrightarrow (0, 1]$, exactly as for priority mean-payoff games of Section 4.

Let us take $\beta \in (0, 1]$ and assume that the stopping probability of each state s is equal to $w(s)\beta^{\pi(s)}$, i.e. the discount factor is

$$\lambda(s) = 1 - w(s)\beta^{\pi(s)} \quad . \tag{1.39}$$

Note that with these discount factors, for two states s and s', $\pi(s) < \pi(s')$ iff $1 - \lambda(s') = o(1 - \lambda(s))$ for $\beta \downarrow 0$.

If (1.39) holds then the payoff mapping (1.9) can be rewritten in the following way, for a play $p = s_0 s_1 s_2 \ldots$,

$$u_\beta(p) = \sum_{i=0}^{\infty} (1 - w(s_0)\beta^{\pi(s_0)}) \ldots (1 - w(s_{i-1})\beta^{\pi(s_{i-1})})\beta^{\pi(s_i)} w(s_i) r(s_i) \quad . \tag{1.40}$$

Let us fix a finite arena $\mathcal{A}$. Obviously, it depends on the parameter β which positional strategies are optimal in the games with payoff (1.40). It is remarkable that for β sufficiently close to 0 the optimality of positional strategies does not depend on β any more. This phenomenon was discovered, in the framework of Markov decision processes, by David Blackwell [2] and is now known under the name of Blackwell optimality.

We shall say that positional strategies $(\sigma^\sharp, \tau^\sharp) \in \Sigma \times \mathcal{T}$ are *β-optimal* if they are optimal in the discounted game $(\mathcal{A}, u_\beta)$.

Definition 5.1. Strategies $(\sigma^\sharp, \tau^\sharp) \in \Sigma \times \mathcal{T}$ are Blackwell optimal in a game $(\mathcal{A}, u_\beta)$ if they are β-optimal for all β in an interval $0 < \beta < \beta_0$ for some constant $\beta_0 > 0$ (β_0 depends on the arena $\mathcal{A}$).

Theorem 5.2. (a) For each arena $\mathcal{A}$ there exists $0 < \beta_0 < 1$ such that if $\sigma^\sharp, \tau^\sharp$ are β-optimal positional strategies for players 1 and 2 for some $\beta \in (0, \beta_0)$ then they are β-optimal for all $\beta \in (0, \beta_0)$, i.e. they are Blackwell optimal.

(b) If $\sigma^\sharp, \tau^\sharp$ are positional Blackwell optimal strategies then they are also optimal for the priority mean-payoff game $(\mathcal{A}, u_{\mathrm{pm}})$.

(c) For each state s, $\lim_{\beta \downarrow 0} \mathrm{val}(\mathcal{A}, s, u_\beta) = \mathrm{val}(\mathcal{A}, s, u_{\mathrm{pm}})$, where $\mathrm{val}(\mathcal{A}, s, u_\beta)$ and $\mathrm{val}(\mathcal{A}, s, u_{\mathrm{pm}})$ are the values of, respectively, the β-discounted game and the priority mean-payoff game.

The remaining part of this section is devoted to the proof of Theorem 5.2.

Lemma 5.3. Let p be an ultimately periodic infinite sequence of states. Then $u_\beta(p)$ is a rational function[7] of β and

$$\lim_{\beta \downarrow 0} u_\beta(p) = u_{\mathrm{pm}}(p) \ . \tag{1.41}$$

Proof. First of all we need to extend the definition (1.40) to finite sequences of states, if $x = s_0 s_1 \dots s_l$ then $u_{\mathrm{pm}}(x)$ is defined like in (1.40) but with the sum taken from 0 to l.

Let $p = xy^\omega$ be an ultimately periodic sequence of states, where x, y are finite sequences of states, y non-empty. Directly from (1.40) we obtain that, for $x = s_0 \dots s_l$,

$$u_\beta(p) = u_\beta(x) + (1 - w(s_0)\beta^{\pi(s_0)}) \dots (1 - w(s_l)\beta^{\pi(s_l)}) u_\beta(y^\omega) \ . \tag{1.42}$$

For any polynomial $f(\beta) = \sum_{i=0}^{l} a_i \beta^i$ the *order*[8] of f is the *smallest* j such that $a_j \neq 0$. By definition the order of the zero polynomial is $+\infty$.

Now note that $u_\beta(x)$ is just a polynomial of β of order strictly greater than 0, which implies that $\lim_{\beta \downarrow 0} u_\beta(x) = 0$. Thus $\lim_{\beta \downarrow 0} u_\beta(p) = \lim_{\beta \downarrow 0} u_\beta(y^\omega)$. On the other hand, $u_{\mathrm{pm}}(p) = u_{\mathrm{pm}}(y^\omega)$. Therefore it suffices to prove that

$$\lim_{\beta \downarrow 0} u_\beta(y^\omega) = u_{\mathrm{pm}}(y^\omega) \ . \tag{1.43}$$

[7] The quotient of two polynomials.

[8] Not to be confounded with the degre of f which is te greatest j such that $a_j \neq 0$.

Suppose that $y = t_0 t_1 \dots t_k$, $t_i \in S$. Then

$$u_\beta(y^\omega) = u_\beta(y) \sum_{i=0}^{\infty} [(1 - w(t_0)\beta^{\pi(t_0)}) \cdots (1 - w(t_k)\beta^{\pi(t_k)})]^i = \frac{u_\beta(y)}{1 - (1 - w(t_0)\beta^{\pi(t_0)}) \cdots (1 - w(t_k)\beta^{\pi(t_k)})} \quad . \tag{1.44}$$

Let $a = \min\{\pi(t_i) \mid 0 \le i \le k\}$ be the priority of y, $L = \{l \mid 0 \le l \le k \quad \text{and} \quad \pi(t_l) = a\}$. Now it suffices to observe that the right hand side of (1.44) can be rewritten as

$$u_\beta(y^\omega) = \frac{\sum_{l \in L} w(t_l) r(t_l) \beta^a + f(\beta)}{\sum_{l \in L} w(t_l) \beta^a + g(\beta)} \quad ,$$

where f and g are polynomials of order greater than a. Therefore

$$\lim_{\beta \downarrow 0} u_\beta(y^\omega) = \frac{\sum_{l \in L} w(t_l) r(t_l)}{\sum_{l \in L} w(t_l)} \quad . \tag{1.45}$$

However, the right hand side of (1.45) is the value of $u_{\mathrm{pm}}(y^\omega)$. Q.E.D.

Proof of Theorem 5.2. The proof of condition (a) given below follows very closely the one given in [13] for Markov decision processes.

Take a sequence (β_n), $\beta_n \in (0,1]$, such that $\lim_{n\to\infty} \beta_n = 0$. Since for each β_n there is at least one pair of β_n-optimal positional strategies and there are only finitely many positional strategies for a finite arena $\mathcal{A}$, passing to a subsequence of (β_n) if necessary, we can assume that there exists a pair of positional strategies $(\sigma^\sharp, \tau^\sharp)$ that are β_n-optimal for all β_n.

We claim that there exists $\beta_0 > 0$ such that $(\sigma^\sharp, \tau^\sharp)$ are β-optimal for all $0 < \beta < \beta_0$.

Suppose the contrary. Then there exists a state s and a sequence (γ_m), $\gamma_m \in (0,1]$, such that $\lim_{m\to\infty} \gamma_m = 0$ and either $\sigma^\sharp$ or $\tau^\sharp$ is not γ_m-optimal. Therefore

(i) either for each m player 1 has a strategy $\sigma^\star_m$ such that $u_{\gamma_m}(p(s, \sigma^\sharp, \tau^\sharp)) < u_{\gamma_m}(p(s, \sigma^\star_m, \tau^\sharp))$,

(ii) or for each m player 2 has a strategy $\tau^\star_n$ such that $u_{\gamma_m}(p(s, \sigma^\sharp, \tau^\star_m)) < u_{\gamma_m}(p(s, \sigma^\sharp, \tau^\sharp))$.

Due to Lemma 2.2, all the strategies $\sigma^\star_m$ and $\tau^\star_m$ can be chosen to be positional and since the number of positional strategies is finite, taking a subsequence of (γ_m) if necessary, we can assume that

(1) either there exist a state s, a positional strategy $\sigma^\star \in \Sigma_p$ and a sequence (γ_m), $\gamma_m \downarrow 0$, such that

$$u_\beta(p(s,\sigma^\sharp,\tau^\sharp)) < u_\beta(p(s,\sigma^\star,\tau^\sharp)) \quad \text{for all } \beta = \gamma_1, \gamma_2, \dots\ , \tag{1.46}$$

(2) or there exist a state s, a positional strategy $\tau^\star \in \mathcal{T}_p$ and a sequence (γ_m), $\gamma_m \downarrow 0$, such that

$$u_\beta(p(s,\sigma^\sharp,\tau^\star)) < u_\beta(p(s,\sigma^\sharp,\tau^\sharp)) \quad \text{for all } \beta = \gamma_1, \gamma_2, \dots\ . \tag{1.47}$$

Suppose that (1.46) holds.

The choice of $(\sigma^\sharp,\tau^\sharp)$ guarantees that

$$u_\beta(p(s,\sigma^\star,\tau^\sharp)) \le u_\beta(p(s,\sigma^\sharp,\tau^\sharp)) \quad \text{for all } \beta = \beta_1, \beta_2, \dots\ . \tag{1.48}$$

Consider the function

$$f(\beta) = u_\beta(p(s,\sigma^\star,\tau^\sharp)) - u_\beta(p(s,\sigma^\sharp,\tau^\sharp)). \tag{1.49}$$

By Lemma 5.3, for $0 < \beta < 1$, $f(\beta)$ is a rational function of β. But from (1.46) and (1.48) we can deduce that when β tends to 0 then $f(\beta) \le 0$ infinitely often and $f(\beta) > 0$ infinitely often. This is possible for a rational function f only if this function is identicaly equal to 0, contradicting (1.46). In a similar way we can prove that (1.47) entails a contradiction. We conclude that $\sigma^\sharp$ and $\tau^\sharp$ are Blackwell optimal.

To prove condition (b) of Theorem 5.2 suppose the contrary, i.e. that there are positional Blackwell optimal strategies $(\sigma^\sharp,\tau^\sharp)$ that are not optimal for the priority mean-payoff game. This means that there exists a state s such that either

$$u_{\mathrm{pm}}(p(s,\sigma^\sharp,\tau^\sharp)) < u_{\mathrm{pm}}(p(s,\sigma,\tau^\sharp)) \tag{1.50}$$

for some strategy σ of player 1 or

$$u_{\mathrm{pm}}(p(s,\sigma^\sharp,\tau)) < u_{\mathrm{pm}}(p(s,\sigma^\sharp,\tau^\sharp)) \tag{1.51}$$

for some strategy τ of player 2. Since priority mean-payoff games have optimal positional strategies, by Lemma 2.2, we can assume without loss of generality that σ and τ are positional. Suppose that (1.50) holds. As $\sigma, \sigma^\sharp, \tau^\sharp$ are positional the plays $p(s,\sigma^\sharp,\tau^\sharp)$ and $p(s,\sigma,\tau^\sharp)$ are ultimately periodic, by Lemma 5.3, we get

$$\lim_{\beta\downarrow 0} u_\beta(p(s,\sigma^\sharp,\tau^\sharp)) = u_{\mathrm{pm}}(p(s,\sigma^\sharp,\tau^\sharp)) < u_{\mathrm{pm}}(p(s,\sigma,\tau^\sharp)) = \lim_{\beta\downarrow 0} u_\beta(p(s,\sigma,\tau^\sharp)). \tag{1.52}$$

However inequality (1.52) implies that there exists $0 < \beta_0$ such that

$$\forall \beta < \beta_0, \quad u_\beta(p(s, \sigma^\sharp, \tau^\sharp)) < u_\beta(p(s, \sigma, \tau^\sharp)) \ ,$$

in contradiction with the Blackwell optimality of $(\sigma^\sharp, \tau^\sharp)$. Similar reasoning shows that also (1.51) contradicts the Blackwell optimality of $(\sigma^\sharp, \tau^\sharp)$.

This also shows that

$$\lim_{\beta \downarrow 0} \operatorname{val}(\mathcal{A}, s, u_\beta) = \lim_{\beta \downarrow 0} u_\beta(p(s, \sigma^\sharp, \tau^\sharp)) = u_{\mathrm{pm}}(p(s, \sigma^\sharp, \tau^\sharp)) = \operatorname{val}(\mathcal{A}, s, u_{\mathrm{pm}}),$$

i.e. condition (c) of Theorem 5.2 holds as well. Q.E.D.

Let us note that there is another known link between parity and discounted games: Jurdziński [14] has shown how parity games can be reduced to mean-payoff games and it is well-known that the value of mean-payoff games is a limit of the value of discounted games, see [15] or [21] for the particular case of deterministic games. However, the reduction of [14] does not seem to extend to priority mean-payoff games and, more significantly, it also fails for perfect information stochastic games. Note also that [21] concentrates only on value approximation and the issue of Blackwell optimality of strategies in not touched at all.

6 Final remarks

6.1 Interpretation of infinite games

In real life all systems have a finite life span: computer systems become obsolete, economic environment changes. Therefore it is reasonable to ask if infinite games are pertinent as models of such systems. This question is discussed for example in [18].

If there exists a family of payoff mappings u_n such that $u_n : S^n \longrightarrow \mathbb{R}$ is defined for paths of length n (n-stage payoff) and the payoff $u(s_0 s_1 \ldots)$ for an infinite play is a limit of $u_n(s_0 s_1 \ldots s_{n-1})$ when the number of stages n tends to ∞ then we can say that infinite games are just approximations of finite games where the length of the game is very large or not precisely known. This interpretation is quite reasonable for simple mean-payoff games for example, where the payoff for infinite plays is a limit of n stage mean-payoff. However such an interpretation fails for priority mean-payoff games and for parity games where no appropriate n-stage payoff mappings exist.

However the stopping (or discounted) games offer another attractive probabilistic interpretation of priority mean-payoff games. For sufficiently small β if we consider a stopping game with the stopping probabilities $w(s)\beta^{\pi(s)}$ for each state s then Theorem 5.2 states that optimal positional strategies for the stopping game are optimal for the priority mean-payoff game. Moreover, the value of the stopping game tends to the value of the priority mean-payoff game when β tends to 0. And the stopping game is a finite game but in a probabilistic rather than deterministic sense, such a game stops with probability 1. Thus we can interpret infinite priority mean-payoff games as an approximation of stopping games where the stopping probabilities are very small. We can also see that smaller priorities are more significant since the corresponding stopping probabilities are much greater: $w(s)\beta^{\pi(s)} = o(w(t)\beta^{\pi(t)})$ if $\pi(s) > \pi(t)$.

6.2 Refining the notion of optimal strategies for priority mean-payoff games

Optimal strategies for parity games (and generally for priority mean-payoff games) are under-selective. To illustrate this problem let us consider the game of Figure 6.2.

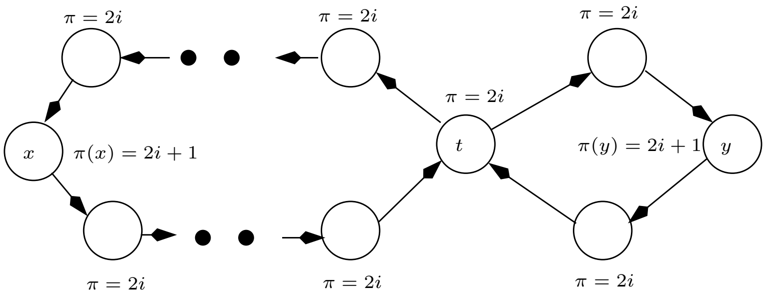

FIGURE 1. The left and the right loop contain one state, x and y respectively, with priority $2i + 1$, all the other states have priority $2i$. The weight of all states is 1. The reward for x and for y is 1 and 0 for all the other states. This game is in fact a parity (Büchi) game, player 1 gets payoff 1 if one of the states $\{x, y\}$ is visited infinitely often and 0 otherwise.

For this game all strategies of player 1 guarantee him the payment 1. Suppose however that the left loop contains $2^{1000000}$ states while the right loop only 3 states. Then, intuitively, it seems that the positional strategy choosing always the small right loop is much more advantageous for player 1 than the positional strategy choosing always the big left loop. But with the traditional definition of optimality for parity games one strategy is as good as the other.

On the other hand, Blackwell optimality clearly distinguishes both strategies, the discounted payoff associated with the right loop is strictly greater than the payoff for the left loop.

Let us note that under-selectiveness of simple mean-payoff games originally motivated the introduction of the Blackwell's optimality criterion [2].

Indeed, the infinite sequence of rewards $100, 0, 0, 0, 0, \ldots$ gives, at the limit, the mean-payoff 0, the same as an infinite sequence of 0. However it is clear that we prefer to get once 100 even if it is followed by an infinite sequence of 0 than to get 0 all the time.

6.3 Evaluating β_0.

Theorem 5.2 is purely existential and does not provide any evaluation of the constant β_0 appearing there. However it is not difficult to give an elementary estimation for β_0, at least for deterministic games considered in this paper. We do not do it here since the bound for β_0 obtained this way does not seem to be particularly enlightening.

The preceding subsection discussing the meaning of the Blackwell optimality raises the question what is the complexity of finding Blackwell optimal strategies. This question remains open. Note that if we can find efficiently Blackwell optimal strategies then we can obviously find efficiently optimal strategies for priority mean-payoff games and, in particular, for parity games. But the existence of a polynomial time algorithm solving parity games is a well-known open problem.

References

[1] H. Björklund, S. Sandberg, and S. Vorobyov. Memoryless determinacy of parity and mean payoff games: a simple proof. *Theor. Computer Science*, 310:365–378, 2004.

[2] D. Blackwell. Discrete dynamic programming. *Annals of Mathematical Statistics*, 33:719–726, 1962.

[3] L. de Alfaro, T. A. Henzinger, and Rupak Majumdar. Discounting the future in systems theory. In *ICALP 2003*, volume 2719 of *LNCS*, pages 1022–1037. Springer, 2003.

[4] A. Ehrenfeucht and J. Mycielski. Positional strategies for mean payoff games. *Intern. J. of Game Theory*, 8:109–113, 1979.

[5] E.A. Emerson and C. Jutla. Tree automata, μ-calculus and determinacy. In *FOCS'91*, pages 368–377. IEEE Computer Society Press, 1991.

[6] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.

[7] H. Gimbert and W. Zielonka. When can you play positionally? In *Mathematical Foundations of Computer Science 2004*, volume 3153 of *LNCS*, pages 686–697. Springer, 2004.

[8] H. Gimbert and W. Zielonka. Deterministic priority mean-payoff games as limits of discounted games. In *ICALP 2006*, volume 4052, part II of *LNCS*, pages 312–323. Springer, 2006.

[9] H. Gimbert and W. Zielonka. Limits of multi-discounted Markov decision processes. In *LICS 2007*, pages 89–98. IEEE Computer Society Press, 2007.

[10] H. Gimbert and W. Zielonka. Perfect information stochastic priority games. In *ICALP 2007*, volume 4596 of *LNCS*, pages 850–861. Springer, 2007.

[11] E. Grädel, W. Thomas, and T. Wilke, editors. *Automata, Logics, and Infinite Games*, volume 2500 of *LNCS*. Springer, 2002.

[12] Y. Gurevich and L. Harrington. Trees, automata and games. In *Proc. 17th Symp. on the Theory of Comp.*, pages 60–65. IEEE Computer Society Press, 1982.

[13] A. Hordijk and A.A. Yushkevich. Blackwell optimality. In E.A. Feinberg and A. Schwartz, editors, *Handbook of Markov Decision Processes*, chapter 8. Kluwer, 2002.

[14] M. Jurdziński. Deciding the winner in parity games is in UP $\cap$ co-UP. *Information Processing Letters*, 68(3):119–124, 1998.

[15] J.F. Mertens and A. Neyman. Stochastic games. *International Journal of Game Theory*, 10:53–56, 1981.

[16] A.W. Mostowski. Games with forbidden positions. Technical Report 78, Uniwersytet Gdański, Instytut Matematyki, 1991.

[17] A. Neyman. From Markov chains to stochastic games. In A. Neyman and S. Sorin, editors, *Stochastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Physical Sciences*, pages 397–415. Kluwer Academic Publishers, 2003.

[18] M. J. Osborne and A. Rubinstein. *A Course in Game Theory.* The MIT Press, 2002.

[19] L. S. Shapley. Stochastic games. *Proceedings Nat. Acad. of Science USA*, 39:1095–1100, 1953.

[20] W. Zielonka. An invitation to play. In *Mathematical Foundations of Computer Science 2005*, volume 3618 of *LNCS*, pages 58–70. Springer, 2005.

[21] Uri Zwick and Mike Paterson. The complexity of mean payoff games on graphs. *Theor. Computer Science*, 158(1-2):343–359, 1996.