Available online at www.sciencedirect.com
SCIENCE DIRECT®
Electronic Notes in Theoretical Computer Science
ELSEVIER Electronic Notes in Theoretical Computer Science 119 (2005) 3–9
www.elsevier.com/locate/entcs

Discounting Infinite Games But How and Why? [1]

Hugo Gimbert[2] and Wiesław Zielonka[3]

LIAFA, Université Denis Diderot Paris 7
case 7014, 2 place Jussieu
75251 Paris Cedex 05, France

Abstract

In a recent paper de Alfaro, Henzinger and Majumdar [8] observed that discounting successive payments, the procedure that is employed in the classical stochastic game theory since the seminal paper of Shapley [16], is also pertinent in the context of much more recent theory of stochastic parity games [7,6,5] which were proposed as a tool for verification of probabilistic systems.
We show that, surprisingly perhaps, the particular discounting used in [8] is in fact very close to the original ideas of Shapley. This observation allows to realize that the specific discounting of [8] suffers in fact from some needless restrictions. We advocate that dropping the constraints imposed in [8] leads to a more general and elegant theory that includes parity and mean payoff games as particular limit cases.

Keywords: parity games, discounting games

1 Stochastic Games

The proper framework for our presentation are stochastic games introduced by Shapley [16].

Such games are played by two players [4] : the player 0 and the player 1. We

[1] This research was supported by European Research Training Network: Games and Automata for Synthesis and Validation
[2] Email: hugo@liafa.jussieu.fr
[3] Email: zielonka@liafa.jussieu.fr
[4] We consider here exclusively two players' zero sum games even if some definitions can obviously be stated in the broader framework of many players non zero sum games.

doi:10.1016/j.entcs.2004.07.005

are given a finite set [5] of *states* S, for each state $s \in S$ we have two finite sets of *actions* : $A(s)$ – the actions of player 0 and $B(s)$ the set of actions of player 1. If the system is at the state $s \in S$ both players choose simultaneously and independently actions $a \in A(s)$ and $b \in B(s)$ respectively and the system goes to a new state s' with the probability $p(s' \mid s, a, b)$ that, as we can see, depends on the current state and the chosen actions. We suppose that the conditional probabilities are correctly and consistently defined, i.e., $0 \leq p(s' \mid s, a, b) \leq 1$ and $\sum_{s' \in S} p(s' \mid s, a, b) = 1$.

A *play* in such a game is an infinite sequence

$$p = (s_0, a_0, b_0), (s_1, a_1, b_1), (s_2, a_2, b_2), \ldots$$

of triples (s_i, a_i, b_i) belonging to the set

$$T = \{(s, a, b) \mid s \in S \text{ and } a \in A(s), b \in B(s) \}$$

whose elements will be called *transitions*. Intuitively, the play p describes the sequence of the visited states and the actions chosen by both players at each stage i of the game.

A payoff mapping u maps each possible play p to a real number $u(p)$ — the payment received by player 0 from player 1 resulting from the play p. The obvious aim of 0 is to play in a way that maximizes his gain while player 1 tries to minimize his loss. Both players use strategies, that indicate how they should play at each stage of a game, i.e., which available action will be chosen. In general the choice of the next action can depend on the past history and can be probabilistic in nature, i.e., strategies provide a conditional probability distribution over the actions that are available at the current stage, see any of the following textbooks and monographs [18,10,19,17] for a formal definition. Fixing the strategies σ of player 0 and τ of player 1 and an initial state s yields a unique probability measure $\mu_{s,\sigma,\tau}$ over the Borel sets of plays starting at s. Now we can state more formally that the aim of player 0 is to choose, if possible, a strategy maximizing his expected payment

$$\mathbb{E}_{s,\sigma,\tau}(u) = \int u(p)\mu_{s,\sigma,\tau}(dp)$$

where the integral is taken over the set of all plays p starting at s (we assume tacitly that u is integrable).

Varying the payment mapping u we obtain different classes of stochastic games.

[5] Finiteness of the state space is not really necessary.

We say that a game starting at s has a *value* for a payment map u if

$$\sup_\sigma \inf_\tau \mathbb{E}_{s,\sigma,\tau}(u) = \inf_\tau \sup_\sigma \mathbb{E}_{s,\sigma,\tau}(u),$$

where σ and τ range over all strategies of both players. The equation above means that both players have ε-optimal strategies.

There are two simpler but important subclasses of stochastic games.

In *perfect information stochastic games* the set S of states is partitioned onto the sets S_0 and S_1 of states of player 0 and player 1 respectively. For the states of S_i belonging to player i the set of actions available to his adversary contains just one element. Such games allow a description simpler than that of general stochastic games. We can assume that with each state s there is associated a finite set $A(s)$ of actions. When we are at the state s the owner i of s ($s \in S_i$) choses an action $a \in A(s)$ to execute and the execution of a leads to a new state $s' \in S$ with a fixed probability $p(s' \mid s, a)$. Again we assume that $p(\cdot \mid s, a, b)$ is a fixed conditional probability distribution with $\sum_{s' \in S} p(s' \mid s, a) = 1$ for all s and $a \in A(s)$.

Yet even simpler class of games is composed of *deterministic games*. This are perfect information games where for each state s and each action $a \in A(s)$ there is one state s' such that $p(s' \mid s, a) = 1$, i.e., the choice of a determines unambiguously the next state.

2 Pieces of the Puzzle

Different types of game models share the same framework described in the previous section and differ only by their payoff mappings.

2.1 Mean-payoff and Discounted Games

Let us suppose that for each transition $(s, a, b) \in T$ we have a real number $r(s, a, b)$ — a one day payoff.

In the *mean payoff games* we look at the long run mean value of one day payoffs. For a play $p = (s_0, a_0, b_0), (s_1, a_1, b_1), (s_2, a_2, b_2), \ldots$ let $\mathbf{r} = (r_i)_{i=0}^\infty$ be the sequence of corresponding one day rewards ($r_i = r(s_i, a_i, b_i)$) and $\sigma_n(\mathbf{r}) = \frac{1}{n+1} \sum_{i=0}^n r_i$ their mean value over the first $n + 1$ days. Since the limit $\lim_{n\to\infty} \sigma_n(\mathbf{r})$ need not exist we consider either upper or lower limits:

$$\limsup_{n\to\infty} \sigma_n(\mathbf{r}) = \lim_{n\to\infty} \sup_{i\geq n} \sigma_i(\mathbf{r}) \quad \text{and} \quad \liminf_{n\to\infty} \sigma_n(\mathbf{r}) = \lim_{n\to\infty} \inf_{i\geq n} \sigma_i(\mathbf{r})$$

and take one of them as the payment $u_{\mathrm{mean}}(p)$ corresponding to the play p.

In *discounted games* we fix a discount factor $\lambda \in (0;1)$ and the payoff of the play $p = (s_0, a_0, b_0), (s_1, a_1, b_1), (s_2, a_2, b_2), \dots$ is given by

$$u_\lambda(p) = (1-\lambda)\sum_{i=0}^{\infty} r_i \lambda^i, \quad \text{where } r_i = r(s_i, a_i, b_i). \tag{1}$$

Shapley [16] showed that discounted games have values and that both players have optimal positional strategies. Bewley and Kohlberg [1,2,3] showed that the limit of the value of a discounted game exists as $\lambda \nearrow 1$. Subsequently Mertens and Neyman [12] proved that this limit gives in fact the value of the mean payoff stochastic game. The proofs of the results of Bewley, Kohlberg, Mertens and Neyman are difficult [10].

2.2 Parity Games and How to Discount Them

Let us recall that parity games were first defined in the framework of deterministic games in Emerson and Jutla [9] (and Mostowski[13]) with two applications: the complementation problem for automata over infinite trees and modal μ-calculus.

Stochastic parity games were introduced and investigated in detail in a series of paper by de Alfaro et al. [6,5,7]. The last of these papers proves, by means of μ-calculus, that such games have a value. In stochastic parity games the one day rewards, that are sometimes called colors in this setting, are non negative integers, $r(s, a, b) \in \mathbb{N}$, $(s, a, b) \in T$, and the payoff of the play $p = (s_0, a_0, b_0), (s_1, a_1, b_1), (s_2, a_2, b_2), \dots$ is given by

$$u_{\text{parity}}(p) = (\limsup r_i) \mod 2, \quad \text{where } r_i = r(s_i, a_i, b_i).$$

Thus $u_{\text{parity}}(p)$ is either 0 or 1 depending on the parity of the maximal one day payoff visited infinitely often [6]. Although there are many apparent similarities between mean payoff and parity games, especially in the deterministic case, [4,15], the exact relation between these two types of games remained elusive. In part at least this was due to the absence of appropriate *discounted parity games*. In fact, one of the most striking features of mean payoff games is the possibility of approximating them by discounted games. Therefore it seems that, unless we find discounted counterpart for parity games, the analogies between parity and mean payoff games should be considered as superfluous.

[6] Usually, one day payoffs or colors used in parity games are associated with states, i.e., it is assumed that each state $s \in S$ is colored by $r(s) \in \mathbb{N}$. We prefer to color the "transitions" in order to allow a more uniform setting for parity/mean payoff/discounted games.

It turns out, however, that discounted version of parity games has already been discovered by de Alfaro, Henzinger and Majumdar [8]. More exactly, the paper [8] deals mainly with discounted μ-calculus and the references to games are cursory and concern only the simplest cases like reachability, safety and Büchi games. Probably this is the reason why the potential residing in this approach was not fully exploited up to now. In fact it seems rather improbable to unify mean payoff and parity games through μ-calculus. Stochastic mean payoff games need much more sophisticated tools and their theory is related to the theory of nonexpansive mappings [14], that, in general, may have no fixed points. Therefore game theory seems to offer a broader perspective than μ-calculus. The first task is then to translate discounted μ-calculus to games, i.e., to provide an appropriate payoff mapping for infinite plays. There is no need to provide the corresponding formula immediately since we can realize quickly that the payoff mapping obtained by this translation is just a very special case of the discounted payoff considered originally by Shapley [16].

More precisely, Shapley [16] considered total payoff stochastic games where for each state $s \in S$ there is a fixed probability $\alpha(s)$ that the game stops when visiting s. As it is well-known (and can be seen easily) the expected total payoff under the stopping condition is the same as the expected payoff for infinite non stopping games where each passage though a state s results in discounting all subsequent one day payoffs by the factor $\lambda(s) = 1 - \alpha(s)$. Thus Shapley games can be seen as games where the payoff of an infinite play $p = (s_0, a_0, b_0), (s_1, a_1, b_1), (s_2, a_2, b_2), \dots$ is given by

$$u_{\text{Shapley}}(p) = \sum_{n=0}^{\infty} \lambda_0 \dots \lambda_n r_n, \quad \text{where } \forall i \in \mathbb{N}, \lambda_i = \lambda(s_i) \text{ and } r_i = r(s_i, a_i, b_i). \tag{2}$$

The idea of many different discount factors was abandoned in all subsequent papers and textbooks relating discounted games since it turned out essentially useless and many discount factors add only unnecessary clutter. It is the formula (1), with one discount factor, that is universally applied.

However, as observed in [8], many different discount factors are essential for discounting parity games. But the formula (2) is inappropriate when we want to investigate the limit payment as various discount factors tend to 1. To this end we should first amend (2) and add to it supplementary factors of the form $(1 - \lambda_i)$. This yields our final *multi-discount payment* mapping.

Let λ be a mapping that for each transition (s, a, b) gives a discount factor $\lambda(s, a, b) \in (0; 1)$ (which can be different for different transitions) and let $r(s, a, b) \in \mathbb{R}$ be, as previously, the corresponding one day reward. For a play $p = (s_0, a_0, b_0), (s_1, a_1, b_1), (s_2, a_2, b_2), \dots$ we set $\forall i \in \mathbb{N}$, $\lambda_i = \lambda(s_i, a_i, b_i)$ and

$r_i = r(s_i, a_i, b_i)$. Then *multi-discount payoff* for p is given by

$$u_{\mathrm{multi}}(p) = \sum_{n=0}^{\infty} (1 - \lambda_n) \lambda_0 \cdots \lambda_{n-1} r_n. \tag{3}$$

The discounted μ-calculus of de Alfaro et al.[8] corresponds in fact to the payoff (3) with the additional constraint:

(A) the one day rewards $r(s, a, b)$ take only the values 0 and 1.

Yet another restriction appears in [8] when the limit of the multi-discount payoff is considered with various discount factors tending to 1. To explain it precisely we should change first the semantics of discounted factors. Instead of supposing that λ maps the transitions to fixed real numbers from the interval $(0; 1)$ we shall assume that there is a finite set Λ of *discount variables* or parameters and that λ is a mapping from the set of transitions into the set Λ of variables, with different transitions that can be mapped to the same variable. Then the multi-discount payment $u_{\mathrm{multi}}(p)$ can be viewed as a function of the variables Λ and we can examine what happens if the variables of Λ tend to 1. When investigating such limits [8] imposes an additional condition restricting syntactically the occurrences of discount factors in μ-calculus formulas. Roughly speaking, in the game framework this restriction translates to the following condition:

(B) if two transitions are mapped to the same discount variable then the one day rewards for these transitions are also equal, i.e.,
for all $(s', a', b'), (s, a, b) \in T$,
if $\lambda(s', a', b') = \lambda(s, a, b) \in \Lambda$ then $r(s', a', b') = r(s, a, b)$.

It turns out, however, that the most interesting things happen precisely when we relax either (A) or (B) or both these restrictions and examine (3) when discount variables tend to 1 in some order. This leads in the limit to several new games that generalize either parity or mean payoff games or both of them. This approach turns out to be very fruitful, we can profit largely from the accumulated knowledge concerning classical stochastic games [10] to establish effortlessly results about parity games and their extensions. On the other hand, this method suggests also how to define "prioritized" versions of classical stochastic games which can be of some interest for game theory. This subject is developed extensively in [11].

References

[1] T. Bewley and E. Kohlberg. The asymptotic solution of a recursive equation arising in stochastic games. *Mathematics of Operations Research*, 1:321–336, 1976.

[2] T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Mathematics of Operations Research*, 1:197–208, 1976.

[3] T. Bewley and E. Kohlberg. On stochastic games with stationary optimal strategies. *Mathematics of Operations Research*, 3:104–125, 1978.

[4] Henrik Björklund, Sven Sandberg, and Sergei Vorobyov. Memoryless determinacy of parity and mean payoff games: a simple proof. *TCS*, 310:365–378, 2004.

[5] L. de Alfaro and T.A. Henzinger. Concurrent ω-regular games. In *LICS'00*, pages 142–154. IEEE Computer Society Press, 2000.

[6] L. de Alfaro, T.A. Henzinger, and O. Kupferman. Concurrent reachability games. In *FOCS'98*, pages 564–575. IEEE Computer Society Press, 1998.

[7] L. de Alfaro and R. Majumdar. Quantitative solution to omega-regular games. In *STOC'01*, pages 675–683. ACM Press, 2001. final version to appear in Journal of Computer and System Sciences.

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

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

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

[11] Hugo Gimbert and Wiesław Zielonka. Mean-payoff and parity games: two faces of the coin. in preparation.

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

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

[14] Abraham Neyman. Stochastic games and nonexpansive maps. In A. Neyman and S. Sorin, editors, *Stohastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Pphysical Sciences*, pages 397–415. Kluwer Academic Publishers, 2003.

[15] Anuj Puri. *Theory of hybrid systems and discrete event systems*. PhD thesis, EECS, University of California, 1995.

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

[17] Sylvain Sorin. *A First Course on Zero-Sum Repeated Games*. Springer, 2002.

[18] Sylvain Sorin. Classification and basic tools. In A. Neyman and S. Sorin, editors, *Stohastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Pphysical Sciences*, pages 27–36. Kluwer Academic Publishers, 2003.

[19] O.J. Vrieze. *Stochastic games with finite state and action spaces*. CWI Tract. Centrum voor Wiskunde en Informatica, 1987.