A Class of Markov Decision Processes with Pure and Stationary Optimal Strategies *

Hugo Gimbert
CNRS, LABRI, Bordeaux, France
hugo.gimbert@labri.fr

January 15, 2008

1 Introduction

We are interested in the existence of pure and stationary optimal strategies in Markov decision processes. We restrict to Markov decision processes with finitely many states and actions and infinite duration.

In a Markov decision process, each state is labelled by an immediate payoff and each infinite history generates a stream of immediate payoffs.

The final payoff associated with an infinite history may be computed in several ways. For example, the final payoff may be the discounted sum [12] of all immediate payoffs or their average value[3]. Also some authors considered the supremum or the infimum limit of the stream of immediate payoffs [10]. Even more exotic is the so-called "parity condition", a prominent tool in automata theory and logic [8].

Surprisingly, these five kinds of Markov decision processes share a common property: they admit pure and stationary optimal strategies. A natural question is the following: what do the discounted, mean, limsup, liminf and parity payoff functions have in common, which explains this surprising fact?

In this paper, we introduce the class of *prefix-independent* and *submixing* payoff functions, and we prove that all Markov decision processes equipped with such payoff functions admit pure and stationary optimal strategies.

This result partially solves our problem, since the parity, limsup and mean-payoff functions are prefix-independent and submixing.

Our result has several interesting consequences. First, it unifies and shortens disparate proofs of existence of pure and stationary optimal strategies for Markov decision processes equipped with the parity [2], limsup [10] and mean [1, 11] payoff function. Second, it allows us to generate a bunch of new examples of pure and stationary payoff functions.

*This research was supported by Instytut Informatyki of Warsaw University, and European Research Training Network: Games and Automata for Synthesis and Validation.

Plan. This paper is organized as follows. In section 1, we introduce notions of Markov decision process, payoff function and optimal strategy. In section 2, we state our main result : if a payoff function is prefix-independent and submixing, this guarantees the existence of pure and stationary strategies (cf. Theorem 1). In section 3, we show that our main result unifies various disparate proofs. In section 4, we present new examples of pure and stationary payoff functions. In section 5, we prove Theorem 1.

1 Markov decision processes

Let $\mathbf{S}$ be a finite set. The set of finite (resp. infinite) sequences on $\mathbf{S}$ is denoted $\mathbf{S}^*$ (resp. $\mathbf{S}^\omega$). A *probability distribution* on $\mathbf{S}$ is a function $\delta : \mathbf{S} \to \mathbb{R}$ such that $\forall s \in \mathbf{S},\ 0 \leq \delta(s) \leq 1$ and $\sum_{s\in\mathbf{S}} \delta(s) = 1$. The set of probability distributions on $\mathbf{S}$ is denoted $\mathcal{D}(\mathbf{S})$.

Definition 1. *A Markov decision process i a tuple $\mathcal{M} = (\mathbf{S}, \mathbf{A}, (\mathbf{A}(s))_{s\in\mathbf{S}}, p)$, with $\mathbf{S}$ a finite set of states, $\mathbf{A}$ a finite set of actions, for each state $s \in \mathbf{S}$, a set $\mathbf{A}(s) \subseteq \mathbf{A}$ of actions available in s, transition probabilities $p : \mathbf{S} \times \mathbf{A} \to \mathcal{D}(\mathbf{S})$.*

When the current state of the chain is s, then the controller chooses an available action $a \in \mathbf{A}(s)$, and the new state is t with probability $p(t|s, a)$.

A triple $(s, a, t) \in \mathbf{S} \times \mathbf{A} \times \mathbf{S}$ such that $a \in \mathbf{A}(s)$ and $p(t|s, a) > 0$ is called a transition.

A *history* in $\mathcal{M}$ is an infinite sequence $h = s_0a_1s_1 \cdots \in \mathbf{S}(\mathbf{AS})^\omega$ such that for each n, (s_n, a_{n+1}, s_{n+1}) is a transition. State s_0 is called the source of h. The set of histories with source s is denoted $\mathbf{P}^\omega_{\mathcal{M},s}$. A *finite history* in $\mathcal{M}$ is a finite sequence $h = s_0a_1 \cdots a_{n-1}s_n \in \mathbf{S}(\mathbf{AS})^*$ such that for each n, (s_n, a_{n+1}, s_{n+1}) is a transition. s_0 is the source of h and s_n its target. The set of finite histories (resp. of finite histories with source s) is denoted $\mathbf{P}^*_\mathcal{M}$ (resp. $\mathbf{P}^*_{\mathcal{M},s}$).

A *strategy* in $\mathcal{M}$ is a function $\sigma : \mathbf{P}^*_\mathcal{M} \to \mathcal{D}(\mathbf{A})$ such that for any finite history $h \in \mathbf{P}^*_\mathcal{M}$ with target $t \in \mathbf{S}$, the distribution $\sigma(h)$ puts non-zero probabilities only on actions that are available in t, i.e. $(\sigma(h)(a) > 0) \implies (a \in \mathbf{A}(t))$. The set of strategies in $\mathcal{M}$ is denoted $\Sigma_\mathcal{M}$.

As explained in the introduction, certain types of strategies are of particular interest, such as *pure* and *stationary* strategies. A strategy is stationary when choices only depend on the current state, and not on the past finite history. A strategy is pure if no lottery is used to choose actions.

Definition 2. *A strategy $\sigma \in \Sigma_\mathcal{M}$ is said to be* pure *if for every finite history h and action a, either $\sigma(h)(a) = 0$ or $\sigma(h)(a) = 1$. A strategy σ is said to be* stationary *if for every finite history $h \in \mathbf{P}^*_\mathcal{M}$ with target t, $\sigma(h) = \sigma(t)$.*

We use the following random variables on the set $\mathbf{P}^\omega_{\mathcal{M},s}$ of infinite histories.

For $n \in \mathbb{N}$, and $t \in \mathbf{S}$,

$$\begin{array}{ll} S_n(s_0 a_1 s_1 \cdots) = s_n & \text{the } (n+1)\text{-th state,} \\ A_n(s_0 a_1 s_1 \cdots) = a_n & \text{the } n\text{-th action,} \\ H_n = S_0 A_1 \cdots A_n S_n & \text{the finite history of the first } n \text{ stages,} \\ N_t = |\{n > 0 : S_n = t\}| \in \mathbb{N} \cup \{+\infty\} & \text{the number of visits to state } t. \end{array} \quad (1)$$

With every initial state s and strategy σ is associated a probability measure $\mathbb{P}_s^\sigma$ on $\mathbf{P}_{\mathcal{M},s}^\omega$, equipped with the σ-algebra generated by random variables $S_0, A_1, S_1, \ldots$. We abuse the notation: for every finite history $h \in \mathbf{P}_{\mathcal{M},s}^*$ of length $n+1 \in \mathbb{N}$ with target $t \in \mathbf{S}$, and for every $a \in \mathbf{A}(t)$, the event $\{S_0 A_1 \cdots A_n S_n = h\}$ is simply denoted h and the event $\{S_0 A_1 \cdots A_n S_n A_{n+1} = ha\}$ is denoted ha. With this notation, $\mathbb{P}_s^\sigma$ is the only probability measure on $\mathbf{P}_{\mathcal{M},s}^\omega$ such that:

$$\mathbb{P}_s^\sigma(ha \mid h) = \sigma(h)(a) \ , \quad (2)$$

$$\mathbb{P}_s^\sigma(har \mid ha) = p(r|t,a) \ . \quad (3)$$

1.1 Payoff functions

After an infinite history, the controller gets some payoff. There are various ways for computing this payoff.

Mean payoff. In a mean-payoff Markov decision process, each transition (s,a,t) of is labeled with a daily payoff $r(s,a,t) \in \mathbb{R}$. An history $s_0 a_1 s_1 \cdots$ gives rise to a sequence $r_0 r_1 \cdots$ of daily payoffs, where $r_n = r(s_n, a_{n+1}, s_{n+1})$. The final payoff is:

$$\phi_{\text{mean}}(r_0 r_1 \cdots) = \limsup_{n \in \mathbb{N}} \frac{1}{n+1} \sum_{i=0}^{n} r_i \ . \quad (4)$$

The mean-payoff function has been introduced by Gillette [3] and is used to evaluate average performance.

Discounted payoff. Each transition (s,a,t) is labeled not only with a daily payoff $r(s,a,t) \in \mathbb{R}$ but also with a discount factor $0 \leq \lambda(s,a,t) < 1$. The final payoff associated with a sequence $(r_0, \lambda_0)(r_1, \lambda_1) \cdots \in (\mathbb{R} \times [0,1[)^\omega$ of daily payoffs and discount factors is:

$$\phi_{\text{disc}}((r_0, \lambda_0)(r_1, \lambda_1) \cdots) = r_0 + \lambda_0 r_1 + \lambda_0 \lambda_1 r_2 + \cdots \ . \quad (5)$$

The discounted payoff has been introduced by Shapley [12].

Parity payoff. The parity condition is used in automata theory and in the study of certain temporal logic [8]. Each transition (s, a, t) is labeled with an integer $c(s, a, t) \in \{0, \dots, d\}$, called a *priority*. The controller receives payoff 1 if the highest priority seen infinitely often is odd, and 0 otherwise. For $c_0 c_1 \cdots \in \{0, \dots, d\}^\omega$,

$$\phi_{\text{par}}(c_0 c_1 \cdots) = \begin{cases} 0 \text{ if } \limsup_n c_n \text{ is even,} \\ 1 \text{ otherwise.} \end{cases} \tag{6}$$

General payoff functions. Observe that in the examples we gave above, the transitions were labeled with various types of data: real numbers for the mean-payoff, couple of real numbers for the discounted payoff and integers for the parity payoff.

We wish to treat those examples in a unified framework. For this reason, we consider now that each Markov decision process $\mathcal{M}$ comes together with a finite set of colours $\mathbf{C}$ and a mapping $\text{col} : \mathbf{S} \times \mathbf{A} \times \mathbf{S} \to \mathbf{C}$, which colors transitions. In the case of the mean payoff, transitions are coloured with real numbers hence $\mathbf{C} \subseteq \mathbb{R}$, whereas in the case of the discounted payoff colours are couples $\mathbf{C} \subseteq \mathbb{R} \times [0, 1[$ and for the parity game colours are integers $\mathbf{C} = \{0, \dots, d\}$. Each infinite history $h = s_0 a_1 s_1 \cdots$ generates an infinite sequence of colours $\text{col}(h) = \text{col}(s_0, a_1, s_1)\,\text{col}(s_1, a_2, s_2) \cdots$, and the final payoff is computed by a payoff function:

Definition 3. *Let $\mathbf{C}$ be a finite set. A payoff function on $\mathbf{C}$ is a function $\phi : \mathbf{C}^\omega \to \mathbb{R}$ which is bounded and Borel-measurable..*

After an infinite history h, the controller receives payoff $\phi(\text{col}(h))$.

1.2 Values and optimal strategies in Markov decision processes

Let $\mathcal{M}$ be a Markov decision process with payoff function ϕ. After history h, the controller receives payoff $\phi(\text{col}(h)) \in \mathbb{R}$. We extend the definition domain of ϕ to $\mathbf{P}^\omega_{\mathcal{M},s}$:

$$\forall h \in \mathbf{P}^\omega_{\mathcal{M},s}, \quad \phi(h) = \phi(\text{col}(h)) \ .$$

The expected value of ϕ under the probability $\mathbb{P}^\sigma_s$ is called the *expected payoff* of the controller and is denoted $\mathbb{E}^\sigma_s[\phi]$. It is well-defined because ϕ is measurable and bounded. The *value of a state s* is the maximal expected payoff that the controller can get :

$$\text{val}(\mathcal{M})(s) = \sup_{\sigma \in \Sigma_\mathcal{M}} \mathbb{E}^\sigma_s[\phi] \ .$$

A strategy σ is said to be *optimal* in $\mathcal{M}$ if for any state $s \in \mathbf{S}$,

$$\mathbb{E}^\sigma_s[\phi] = \text{val}(\mathcal{M})(s) \ .$$

2 A sufficient condition for the existence of pure and stationary optimal strategies

Our result is based on the notion of prefix-independent and stationary payoff functions.

Definition 4. *Let ϕ be a payoff function on $\mathbf{C}^\omega$. We say that ϕ is* prefix-independent *if for any finite word $u \in \mathbf{C}^*$ and infinite word $v \in \mathbf{C}^\omega$, $\phi(uv) = \phi(v)$. We say that ϕ is* submixing *if for any sequence of finite non-empty words $u_0, v_0, u_1, v_1, \ldots \in \mathbf{C}^*$,*

$$\phi(u_0 v_0 u_1 v_1 \cdots) \leq \max\{\ \phi(u_0 u_1 \cdots)\ ,\ \phi(v_0 v_1 \cdots)\ \}\ .$$

The notion of prefix-independence is classical. The submixing property is close to the notions of *fairly-mixing* payoff functions introduced in [5] and of *concave* winning conditions introduced in [9]. We are now ready to state our main result.

Theorem 1. *Let ϕ be a prefix-independent and submixing payoff function. Then in any Markov decision process $\mathcal{M}$ with payoff function ϕ, there exists pure and stationary optimal strategies.*

The proof of Theorem 1 is postponed to Section 5.

3 Unification of classical results

We now show how Theorem 1 unifies several proofs of the existence of pure and stationary optimal strategies in the parity [2], the limsup and liminf [10] and the mean-payoff [1, 11] Markov decision processes.

The parity, mean, limsup and liminf payoff functions are denoted respectively $\phi_{\mathrm{par}}, \phi_{\mathrm{mean}}, \phi_{\mathrm{lsup}}$ and ϕ_{linf}. Both ϕ_{par} and ϕ_{mean} have already been defined in subsection 1.1. ϕ_{lsup} and ϕ_{linf} are defined as follows. Let $\mathbf{C} \subseteq \mathbb{R}$ be a finite set of real numbers, and $c_0 c_1 \cdots \in \mathbf{C}^\omega$. Then

$$\phi_{\mathrm{lsup}}(c_0 c_1 \cdots) = \limsup_n c_n$$
$$\phi_{\mathrm{linf}}(c_0 c_1 \cdots) = \liminf_n c_n\ .$$

The four payoff functions $\phi_{\mathrm{par}}, \phi_{\mathrm{mean}}, \phi_{\mathrm{lsup}}$ and ϕ_{linf} are very different. Indeed, ϕ_{lsup} may measure peak performances of a system, ϕ_{linf} its worst performances, and ϕ_{mean} its average performances. The function ϕ_{par} may encode logical specifications, expressed in MSO or LTL for example [8].

Proposition 1. *The payoff functions ϕ_{lsup}, ϕ_{linf}, ϕ_{par} and ϕ_{mean} are submixing.*

Proof. Let $\mathbf{C} \subseteq \mathbb{R}$ be a finite set of real numbers and $u_0, v_0, u_1, v_1, \ldots \in \mathbf{C}^*$ be a sequence of finite non-empty words on $\mathbf{C}$. Define $u = u_0 u_1 \cdots \in \mathbf{C}^\omega$,

$v = v_0 v_1 \cdots \in \mathbf{C}^\omega$ and $w = u_0 v_0 u_1 v_1 \cdots \in \mathbf{C}^\omega$. The following elementary fact immediately implies that ϕ_{lsup} is submixing.

$$\phi_{\text{lsup}}(w) = \max\{\phi_{\text{lsup}}(u), \phi_{\text{lsup}}(v)\} \ . \tag{7}$$

In a similar way, ϕ_{linf} is submixing since

$$\phi_{\text{linf}}(w) = \min\{\phi_{\text{linf}}(u), \phi_{\text{linf}}(v)\} \ . \tag{8}$$

Now suppose that $\mathbf{C} = \{0, \ldots, d\}$ is a finite set of integers and consider function ϕ_{par}. Remember that $\phi_{\text{par}}(w)$ equals 1 if $\phi_{\text{lsup}}(w)$ is odd and 0 if $\phi_{\text{lsup}}(w)$ is even. Then using (7) we get that if $\phi_{\text{par}}(w)$ has value 1 then it is the case of either $\phi_{\text{par}}(u)$ or $\phi_{\text{par}}(v)$. It proves that ϕ_{par} is also submixing.

Now let us consider function ϕ_{mean}. A proof that ϕ_{mean} is submixing already appeared in [5], and we reproduce it here, updating the notations. Again $\mathbf{C} \subseteq \mathbb{R}$ is a finite set of real numbers. Let $c_0, c_1, \ldots \in \mathbf{C}$ be the sequence of letters of $\mathbf{C}$ such that $w = (c_i)_{i \in \mathbb{N}}$. Since word w is a shuffle of words u and v, there exists a partition (I_0, I_1) of $\mathbb{N}$ such that $u = (c_i)_{i \in I_0}$ and $v = (c_i)_{i \in I_1}$. For any $n \in \mathbb{N}$, let $I_0^n = I_0 \cap \{0, \ldots, n\}$ and $I_1^n = I_1 \cap \{0, \ldots, n\}$. Then for $n \in \mathbb{N}$,

$$\begin{aligned}\frac{1}{n+1}\sum_{i=0}^{n} c_i &= \frac{|I_0^n|}{n+1}\left(\frac{1}{|I_0^n|}\sum_{i \in I_0^n} c_i\right) + \frac{|I_1^n|}{n+1}\left(\frac{1}{|I_1^n|}\sum_{i \in I_1^n} c_i\right) \\ &\leq \max\left\{\frac{1}{|I_0^n|}\sum_{i \in I_0^n} c_i, \frac{1}{|I_1^n|}\sum_{i \in I_1^n} c_i\right\} \ .\end{aligned}$$

The inequality holds since $\frac{|I_0^n|}{n+1} + \frac{|I_1^n|}{n+1} = 1$. Taking the superior limit of this inequality, we obtain $\phi_{\text{mean}}(w) \leq \max\{\phi_{\text{mean}}(u), \phi_{\text{mean}}(v)\}$. It proves that ϕ_{mean} is submixing. □

Since ϕ_{lsup}, ϕ_{linf}, ϕ_{par} and ϕ_{mean} are clearly prefix-independent, Proposition 1 and Theorem 1 imply that those four payoff functions are pure and stationary. Hence, we unify and simplify existing proofs of [2, 10] and [1, 11]. In particular, we use only elementary tools for the mean-payoff function, whereas [1] uses martingale theory and relies on other papers, and [11] uses a reduction to discounted games, as well as analytical tools.

4 Generating new examples of pure and stationary payoff functions.

We now present three different techniques for generating new examples of Markov decision processes with pure and stationary optimal strategies.

4.1 Mixing with the liminf payoff

In last section was presented the limsup payoff function, which intuitively measure the peak performances of a system and the liminf payoff function, which intuitively measures its worst performances. The *compromise payoff* function is used when the controller wants to achieve a trade-off between good peak performances and not too bad worst performances. It was introduced in [5]. We fix a factor $\lambda \in [0,1]$, a finite set $\mathbf{C} \subseteq \mathbb{R}$ and for $u \in \mathbf{C}^\omega$, we define

$$\phi^\lambda_{\text{comp}}(u) = \lambda \cdot \phi_{\text{lsup}}(u) + (1-\lambda) \cdot \phi_{\text{linf}}(u) \ .$$

The fact that $\phi^\lambda_{\text{comp}}$ is submixing is a corollary of the following proposition.

Proposition 2. *Let $\mathbf{C} \subseteq \mathbb{R}$, $0 \leq \lambda \leq 1$ and ϕ be a payoff function on $\mathbf{C}$. Suppose that ϕ is prefix-independent and submixing. Then the payoff function*

$$\lambda \cdot \phi + (1-\lambda) \cdot \phi_{linf} \tag{9}$$

is also prefix-independent and submixing.

The proof is straightforward, using (8) above. According to Theorem 1 and Proposition 1, any payoff function defined by equation (9), where ϕ is either $\phi_{\text{mean}}, \phi_{\text{par}}$ or ϕ_{lsup}, is pure and stationary.

4.2 The approximation operator

Consider an increasing function $f : \mathbb{R} \to \mathbb{R}$ and a payoff function $\phi : \mathbf{C}^\omega \to \mathbb{R}$. Then their composition $f \circ \phi$ is also a payoff function and moreover, if ϕ guarantees the existence of pure and stationary optimal strategies then $f \circ \phi$ also does. Indeed, a strategy optimal for an Markov decision process $\mathcal{M}$ with payoff function ϕ is also optimal for the Markov decision process $\mathcal{M}$ with payoff function $f \circ \phi$. In fact, it is straightforward to check that composition by f conserves the prefix-independent and submixing properties.

An example is the threshold function $f = \mathbf{1}_{\geq 0}$ which associates 0 with strictly negative real numbers and 1 with positive number. Then $f \circ \phi$ indicates whether the performance evaluated by ϕ reaches the critical value of 0.

4.3 The hierarchical product

Now we define a binary operator between payoff functions, which stabilizes the family of prefix-independent and submixing payoff functions. We call this operator the *hierarchical product*.

Let ϕ_0, ϕ_1 be two payoff functions on sets of colours $\mathbf{C}_0$ and $\mathbf{C}_1$ respectively. We do not require $\mathbf{C}_0$ and $\mathbf{C}_1$ to be identical nor disjoints.

The hierarchical product $\phi_0 \rhd \phi_1$ of ϕ_0 and ϕ_1 is a payoff function on the set of colours $\mathbf{C}_0 \cup \mathbf{C}_1$ and is defined as follows. Let $u = c_0 c_1 \cdots \in (\mathbf{C}_0 \cup \mathbf{C}_1)^\omega$ and u_0 and u_1 the two projections of u on $\mathbf{C}_0$ and $\mathbf{C}_1$ respectively. Then

$$(\phi_0 \rhd \phi_1)(u) = \begin{cases} \phi_0(u_0) & \text{if } u_0 \text{ is infinite,} \\ \phi_1(u_1) & \text{otherwise.} \end{cases}$$

This definition makes sense : although each word u_0 and u_1 can be either finite or infinite, at least one of them must be infinite.

The hierarchical product conserves the prefix-independence and submixing properties.

Proposition 3. *Let ϕ_0 and ϕ_1 be two payoff functions. If ϕ_0 and ϕ_1 are prefix-independent and submixing, then $\phi_0 \rhd \phi_1$ also is.*

Now we give two examples of the use of the hierarchical product. For $e \in \mathbb{N}$, let 0_e and 1_e be the payoff functions defined on the one-letter alphabet $\{e\}$ and constant equal to 0 and 1 respectively. Let d be an odd number, and ϕ_{par} be the parity payoff function on $\{0, \dots, d\}$. Then

$$\phi_{\text{par}} = 1_d \rhd 0_{d-1} \rhd \cdots \rhd 1_1 \rhd 0_0 \ .$$

Another example of hierarchical product is given in, which deals with Also hierarchical products of *mean-payoff function* have been considered, the corresponding Markov decision processes are tightly linked with discounted Markov decision processes [6, 7].

5 Proof of Theorem 1.

The proof of Theorem 1 is organized as follows. In the first subsection, we establish two useful elementary lemmas. Then in subsection 5.2, we establish a property about Markov chains. In subsection 5.3, we establish that the expected value of histories that never reach their initial state is no more than the value of that state. Then in subsection 5.4, we introduce the notion of a split of an arena. Basic properties of the split operation are described in Proposition 4, and Theorem 4 shows how one can simulate a strategy in an arena with strategies in the split of that arena. Theorem 5 and 6 are the key results to show that value of a state in an arena is no more than its maximal value in splits of the arena, i.e. Corollary 1. End of proof of Theorem 1 is given in subsection 5.7.

5.1 Preliminary lemmas

In the proof of Theorem 1, we will often use the following lemmas. Recall that we abuse the notation and for every finite play h of length $n+1$, we denote also by h the event $\{S_0 A_1 S_1 \cdots S_n = h\}$.

First Lemma is called the shifting lemma.

Lemma 1 (shifting lemma)**.** *Let $\mathcal{M}$ be a Markov decision process, $s, t \in \mathbf{S}$ some states, $h \in \mathbf{P}^*_{\mathcal{M},s}$ a finite history with source s and target t, σ a strategy in $\mathcal{M}$, and X a real valued random variable such that $\sup X \neq +\infty$ or $\inf X \neq -\infty$. Then*

$$\mathbb{E}^\sigma_s[X \mid h] = \mathbb{E}^{\sigma[h]}_t[\ X[h]\ ], \tag{10}$$

where $\sigma[h]$ is the strategy defined as $\sigma[h](s_0 a_1 \cdots s_n) = \sigma(h a_1 \cdots s_n)$ and $X[h]$ is the random variable defined by $X[h](s_0 a_1 s_1 \cdots) = X(h a_1 s_1 \cdots)$.

Proof. First prove the Lemma when X is the indicator function of an event h', with h' a finite history. Then deduce that it holds for any random variable X. □

The following lemma will also be very useful.

Lemma 2. *Let $\mathcal{M}$ be an Markov decision process, s a state of $\mathcal{M}$, $E \subseteq \mathbf{P}^{\omega}_{\mathcal{M},s}$ an event and σ and τ two strategies. Let us suppose that σ and τ coincide on E, in the sense that for all finite history $h \in \mathbf{P}^{*}_{\mathcal{M},s}$,*

$$(h \text{ is a prefix of an history in } E) \implies (\sigma(h) = \tau(h)) \ .$$

Then for any event F,

$$\mathbb{P}^{\sigma}_{s}(F \mid E) = \mathbb{P}^{\tau}_{s}(F \mid E) \ . \tag{11}$$

Proof. Start with proving:

$$\mathbb{P}^{\sigma}_{s}(E) = \mathbb{P}^{\tau}_{s}(E) \tag{12}$$

This is easy to prove when E is the event h, where h is a finite play. Since these events generate all events, (12) holds in general. Since σ and τ coincide on F, we obtain 11.

□

5.2 About Markov chains

Second step consists in proving theorem 2, which establishes a property of Markov chains. A Markov decision process $\mathcal{M}$ is a Markov chain when $\forall s \in \mathbf{S}, |\mathbf{A}(s)| = 1$. In that case, there is a unique strategy σ in $\mathcal{M}$. The probability measure on $\mathbf{P}^{\omega}_{\mathcal{M},s}$ associated with that unique strategy is denoted $\mathbb{P}_s$ instead of $\mathbb{P}^{\sigma}_{s}$.

Theorem 2. *Let $\mathcal{M}$ be an Markov decision process with payoff function ϕ. Suppose that $\mathcal{M}$ is a Markov chain and ϕ is prefix-independent. Let s be a recurrent state of $\mathcal{M}$. Then*

$$\mathbb{P}_{s}(\phi > \mathrm{val}(\mathcal{M})(s)) = 0 \ . \tag{13}$$

Proof. Let E be the event $E = \{\phi > \mathrm{val}(G, s)\}$. Event E is a tail-event hence according to the 0-1 law, E has probability either 0 or 1. Suppose for a second that $\mathbb{P}_{s}(E) = 1$ and find a contradiction. Then $\mathbb{P}_{s}(\phi > \mathrm{val}(\mathcal{M}, s)) = 1$, hence $\mathbb{E}_{s}[\phi] > \mathrm{val}(\mathcal{M}, s)$, which contradicts the definition of $\mathrm{val}(\mathcal{M}, s)$. We deduce that $\mathbb{P}_{s}(E) = 0$ which gives (13) and achieves the proof of this theorem. □

5.3 Histories that never reach again their initial state

Consider the definition of N_s given by equation (1). The event $\{N_s = 0\}$ is the set of histories that never reach again s after the first stage. The following theorem states a property about the expected value of those histories.

Theorem 3. *Let $\mathcal{M}$ be a Markov decision process with payoff function ϕ, s a state of $\mathcal{M}$ and σ a strategy. Suppose that ϕ is prefix-independent. Then*

$$\mathbb{E}_s^\sigma[\phi \mid N_s = 0] \leq \mathrm{val}(\mathcal{M})(s). \tag{14}$$

Proof. Let $f : \mathbf{P}^*_{\mathcal{M},s} \to \mathbf{P}^*_{\mathcal{M},s}$ be the mapping that "forget cycles on s", defined by:

$$f(s_0 a_1 \cdots s_n) = s_k a_{k+1} \cdots s_n, \text{ where } k = \max\{i \mid s_i = s\} \ .$$

Let τ the strategy that consists in forgetting the cycles on s, and apply σ. Formally τ is defined by $\tau(h) = \sigma(f(h))$. We are going to show that:

$$\mathbb{E}_s^\sigma[\phi \mid N_s = 0] = \mathbb{E}_s^\tau[\phi], \tag{15}$$

which implies immediately (14), by definition of the value of a state. Even if (15) may seem obvious, we proof it for the sake of completeness.

We suppose that

$$e = \mathbb{P}_s^\sigma(N_s = 0) > 0 \ , \tag{16}$$

otherwise (14) is not defined, and there is nothing to prove. Then clearly:

$$\mathbb{P}_s^\tau(N_s = \infty) = 0. \tag{17}$$

Define last_s, the last date where history reaches s:

$$\mathrm{last}_s = \sup\{n \in \mathbb{N},\ S_n = s\}.$$

Then $\{N_s = \infty\} = \{\mathrm{last}_s = \infty\}$, hence (17) implies $\mathbb{P}_s^\tau(\mathrm{last}_s < \infty) = 1$, and

$$\begin{aligned}\mathbb{E}_s^\tau[\phi] &= \sum_{n\in\mathbb{N}} \mathbb{E}_s^\tau[\phi \mid \mathrm{last}_s = n] \cdot \mathbb{P}_s^\tau(\mathrm{last}_s = n). \\ &= \sum_{n\in\mathbb{N}} \sum_{h\in\mathbf{P}^*_{\mathcal{M},s}} \mathbb{E}_s^\tau[\phi \mid \mathrm{last}_s = n, H_n = h] \cdot \mathbb{P}_s^\tau(\mathrm{last}_s = n, H_n = h). \end{aligned} \tag{18}$$

Let $n \in \mathbb{N}$ and $h \in \mathbf{P}^*_{\mathcal{M},s}$ such that $\mathbb{P}_{\tau,s}(\mathrm{last}_s = n, H_n = h) > 0$. Then

$$\begin{aligned}\mathbb{E}_s^\tau[\phi \mid \mathrm{last}_s = n, H_n = h] &= \mathbb{E}_s^{\tau[h]}[\phi \mid \mathrm{last}_s = 0] \\ &= \mathbb{E}_s^\tau[\phi \mid \mathrm{last}_s = 0] \\ &= \mathbb{E}_s^\sigma[\phi \mid \mathrm{last}_s = 0]. \end{aligned} \tag{19}$$

The first equality is obtained using the shifting lemma and the prefix-independence of ϕ. The second equality comes from the fact that since $\mathbb{P}_{\tau,s}(\mathrm{last}_s = N, H_N = h) > 0$, h is s and by definition of τ, $\tau[h] = \tau$. The third equality comes from the fact that τ and σ coincide on the set $\{\mathrm{last}_s = 0\}$, and applying the lemma 2.

Eventually, (19) and (18) give $\mathbb{E}_s^\tau[\phi] = \mathbb{E}_s^\sigma[\phi | \, \mathrm{last}_s = 0]$. Since $\{N_s = 0\} = \{\mathrm{last}_s = 0\}$, we get

$$\mathbb{E}_s^\tau[\phi] = \mathbb{E}_s^\sigma[\phi | N_s = 0] \ . \tag{20}$$

By definition of the value of a state, $\mathrm{val}(\mathcal{M})(s) \geq \mathbb{E}_s^\tau[\phi]$, which together with (20) gives (14) and achieves the proof of this theorem. □

5.4 Submixing payoff functions and split of a Markov decision process

The proof of 1 is by induction on the number of actions in the Markov decision process. For that purpose, we introduce the notion of split of an Markov decision process, and associated projections.

Definition 5. *Let $\mathcal{M}$ be an Markov decision process and $s \in \mathbf{S}$ a state such that $|\mathbf{A}(s)| > 1$. Let $(\mathbf{A}_0(s), \mathbf{A}_1(s))$ a partition of $\mathbf{A}(s)$ in two non-empty sets.*

Let $\mathcal{M}_0 = (\mathbf{S}, \mathbf{A}_0, (\mathbf{A}_0(s))_{s\in\mathbf{S}}, p, \mathrm{col})$ be the Markov decision process obtained from
$\mathcal{M} = (\mathbf{S}, \mathbf{A}, (\mathbf{A}(s))_{s\in\mathbf{S}}, p, \mathrm{col})$ in the following way. We restrict the set of actions available in s to $\mathbf{A}_0(s)$. For $t \neq s$, nothing changes, i.e. $\mathbf{A}_0(t) = \mathbf{A}(t)$. The transition probabilities p and the colouring mapping col *do not change. Let $\mathcal{M}_1$ be the Markov decision process obtained symmetrically, restricting the set of actions available in s to $\mathbf{A}_1(s)$.*

Then $(\mathcal{M}_0, \mathcal{M}_1)$ is called a split *of $\mathcal{M}$ on s.*

Now consider a split $(\mathcal{M}_0, \mathcal{M}_1)$ of a Markov decision process $\mathcal{M}$ on a state s. There exists a natural projection (π_0, π_1) from finite histories $h \in \mathbf{P}^*_{\mathcal{M},s}$ to couples of finite histories $(h_0, h_1) \in \mathbf{P}^*_{\mathcal{M}_0,s} \times \mathbf{P}^*_{\mathcal{M}_1,s}$. Let us describe informally this projection.

Consider a finite history $h \in \mathbf{P}^*_{\mathcal{M},s}$. Then h factorizes in a unique way in a sequence

$$h = h_0 h_1 \cdots h_k h_{k+1} \ , \tag{21}$$

such that

- for $0 \leq i \leq k$, h_i is a simple cycle on s,
- h_{k+1} is a finite history with source s, which does not reach s again.

For any $0 \leq i \leq k+1$, the source of h_i is s hence the first action a_i in h_i is available in s, i.e. $a_i \in \mathbf{A}(s)$. Since $(\mathbf{A}_0(s), \mathbf{A}_1(s))$ is a partition of $\mathbf{A}(s)$, we have either $a_i \in \mathbf{A}_0(s)$ or $a_i \in \mathbf{A}_1(s)$. Then $\pi_0(h)$ is obtained by deleting from the factorization (21) of h every simple cycle h_i which first action a_i is in $\mathbf{A}_1(s)$. Symmetrically, $\pi_1(h)$ is obtained by erasing every simple cycle h_i such that $a_i \in \mathbf{A}_0(s)$.

Let us formalize this construction in an inductive way. First we define inductively the *mode* of a play. For $h \in \mathbf{P}^*_{\mathcal{M},s}$, $a \in \mathbf{A}(h)$ and $t \in \mathbf{S}$

$$\mathrm{mode}(hat) = \begin{cases} \mathrm{mode}(h) \text{ if the target of } h \text{ is not } s. \\ 0 \text{ if the target of } h \text{ is } s \text{ and } a \in \mathbf{A}_0(s) \\ 1 \text{ if the target of } h \text{ is } s \text{ and } a \in \mathbf{A}_1(s) \end{cases} \tag{22}$$

For $i \in \{0, 1\}$, the projection π_i is defined by $\pi_i(s) = s$, and for $h \in \mathbf{P}^*_{\mathcal{M},s}$, $a \in \mathbf{A}(h)$ and $t \in \mathbf{S}$,

$$\pi_i(hat) = \begin{cases} \pi_i(h)at & \text{if } \mathrm{mode}(hat) = i \\ \pi_i(h) & \text{if } \mathrm{mode}(hat) = 1 - i. \end{cases} \tag{23}$$

The definition domain of π_0 and π_1 naturally extends to $\mathbf{P}^{\omega}_{\mathcal{M},s}$, in the following way. Let $h = s_0 a_1 s_1 \cdots \in \mathbf{P}^{\omega}_{\mathcal{M},s}$ be an infinite history, and for every $n \in \mathbb{N}$, let $h_n = s_0 a_1 \cdots s_n$. Then for every $n \in \mathbb{N}$, $\pi_0(h_n)$ is a prefix of $\pi_0(h_{n+1})$. If the sequence $(\pi_0(h_n))_{n\in\mathbb{N}}$ is stationary equal to some finite word $h' \in \mathbf{P}^{*}_{\mathcal{M}_0,s}$, then we define $\pi_0(h) = h'$. Otherwise, the sequence $(\pi_0(h_n))_{n\in\mathbb{N}}$ has a limit $h' \in \mathbf{P}^{\omega}_{\mathcal{M}_0,s}$, and we define $\pi_0(h) = h'$. Let us define the random variables:

Definition 6. *The two random variables*

$$\Pi_0 = \pi_0(S_0 A_1 S_1 \cdots) \text{ with values in } \mathbf{P}^{*}_{\mathcal{M}_0,s} \cup \mathbf{P}^{\omega}_{\mathcal{M}_0,s}$$
$$\Pi_1 = \pi_1(S_0 A_1 S_1 \cdots) \text{ with values in } \mathbf{P}^{*}_{\mathcal{M}_1,s} \cup \mathbf{P}^{\omega}_{\mathcal{M}_1,s}$$

are called the projections associated with the split $(\mathcal{M}_0, \mathcal{M}_1)$.

Useful properties of Π_0 and Π_1 are summarized in the following proposition.

Proposition 4. *Let* $\mathcal{M}$ *be a Markov decision process,* s, t *states of* $\mathcal{M}$, $(\mathcal{M}_0, \mathcal{M}_1)$ *a split of* $\mathcal{M}$ *on* s, *and* Π_0 *and* Π_1 *the projections associated with that split.*

- *Let* $h_0 \in \mathbf{P}^{*}_{\mathcal{M}_0,s}$ *be a finite history in* $\mathcal{M}_0$, *with source* s *and target* t, *and* $a \in \mathbf{A}_0(t)$. *Let* $\sqsubseteq$ *be the prefix order relation on finite and infinite words. Then*
$$\forall r \in \mathbf{S}, \quad \mathbb{P}^{\sigma}_{s}(h_0 a r \sqsubseteq \Pi_0 \mid h_0 a \sqsubseteq \Pi_0) = p(r|t,a). \tag{24}$$
- *Let* $x \in \mathbb{R}$ *and* ϕ *be a prefix-independent submixing payoff function. Then*
$$\begin{aligned}\{N_s = \infty \text{ and } \phi > x\} \subseteq \\ \{\Pi_0 \text{ is infinite and } N_s(\Pi_0) = \infty \text{ and } \phi(\Pi_0) > x\} \\ \bigcup \{\Pi_1 \text{ is infinite and } N_s(\Pi_1) = \infty \text{ and } \phi(\Pi_1) > x\}\,.\end{aligned} \tag{25}$$

Proof. We first prove (24). Let π_0 and π_1 be the functions defined by (23) and (22) above. Remark that their definition show that they are both $\sqsubseteq$-increasing. For $h \in \mathbf{P}^{*}_{\mathcal{M},s}$ we denote the event $\{h \sqsubseteq S_0 A_1 S_1 \cdots\}$ as $\mathcal{O}_h$. Fix a finite history h_0 in $\mathcal{M}_0$, an action $a \in \mathbf{A}_0(s)$ and define:

$$Y = \{h \in \mathbf{P}^{*}_{\mathcal{M},s} \mid \pi_0(h) = h_0 \text{ and } \exists r \in \mathbf{S}, \pi_0(har) = h_0 a r\}.$$

Let us start with proving

$$\forall r \in \mathbf{S}, \quad \{h_0 a r \sqsubseteq \pi_0\} = \bigcup_{h\in Y} \mathcal{O}_{har}\ . \tag{26}$$

We start with inclusion $\subseteq$. Let $r \in \mathbf{S}$, $h \in Y$ and $l \in \mathbf{P}^{\omega}_{\mathcal{M},s}$ such that $har \sqsubseteq l$. Since $h \in Y$, and by definition (22) and (23), we deduce that $\forall r, \mathrm{mode}(har) = 0$ and $\pi_0(har) = h_0 a r$. Since π_0 is $\sqsubseteq$-increasing, and $har \sqsubseteq l$, we get $\pi_0(har) \sqsubseteq \pi_0(l)$, hence $h_0 a r \sqsubseteq \pi_0(l)$ thus $l \in \{h_0 a r \sqsubseteq \Pi_0\}$. It proves inclusion $\subseteq$ of (26).

Let us prove now inclusion $\supseteq$ of (26). Let $r \in \mathbf{S}$ and $l \in \{h_0 a r \sqsubseteq \Pi_0\}$. Then $h_0 a r \sqsubseteq \pi_0(l)$. Rewrite l as $l = s_0 a_1 s_1 \cdots$. Since Π_0 is $\sqsubseteq$-increasing,

$\exists n \in \mathbb{N}$ s.t. $h_0 ar \sqsubseteq \pi_0(s_0 \cdots s_{n-1} a_n s_n)$ and $h_0 ar \not\sqsubseteq \pi_0(s_0 \cdots s_{n-1})$. Define $h = s_0 a_1 \cdots s_{n-1}$, then last equation rewrites as $h_0 ar \not\sqsubseteq \pi_0(h)$ and $h_0 ar \sqsubseteq \pi_0(h a_n s_n)$. According to definition (23) of π_0, it necessarily means that $h_0 = \pi_0(h)$ and $h_0 ar = \pi_0(h) a_n s_n$. Hence $h \in Y$, $a_n = a$ and $s_n = r$, thus $har \sqsubseteq l$ and $l \in \cup_{h \in Y} \{har\}$. It achieves to prove (26).

Let X the prefix-free closure of Y, i.e.

$$X = \{h \in Y \mid \nexists h' \in Y \text{ s.t. } h' \neq h \text{ and } h' \sqsubseteq h\} \ .$$

Then

$$\forall r \in \mathbf{S}, \quad \bigcup_{h \in Y} \mathcal{O}_{har} = \bigcup_{h \in X} \mathcal{O}_{har},$$

and the second union is in fact a disjoint union. Hence, according to (26),

$$\forall r \in \mathbf{S}, \quad (\mathcal{O}_{har})_{h \in X} \text{ is a partition of } \{h_0 ar \sqsubseteq \pi_0\} \ , \tag{27}$$

$$\text{and} \quad (\mathcal{O}_{ha})_{h \in X} \text{ is a partition of } \{h_0 a \sqsubseteq \pi_0\} \ . \tag{28}$$

From (27), we get for $r \in \mathbf{S}$,

$$\begin{aligned}
\mathbb{P}_s^\sigma(h_0 ar \sqsubseteq \pi_0) &= \sum_{h \in X} \mathbb{P}_s^\sigma(\mathcal{O}_{har}) \\
&= \sum_{h \in X} p(r|t,a) \cdot \mathbb{P}_s^\sigma(\mathcal{O}_{ha}) && \text{from (3)} \\
&= p(r|t,a) \cdot \sum_{h \in X} \mathbb{P}_s^\sigma(\mathcal{O}_{ha}) \\
&= p(r|t,a) \cdot \mathbb{P}_s^\sigma(h_0 a \sqsubseteq \pi_0) && \text{from (28)}.
\end{aligned}$$

It achieves the proof of (24).

Now let us prove (25). Let ϕ be a prefix-independent submixing payoff function, and $x \in \mathbb{R}$. Let $h \in \{N_s = +\infty \text{ and } \phi > x\}$.

Suppose first that $\pi_1(h)$ is a finite word. Then according to (23), the set $\{h' \in \mathbf{P}^*_{\mathcal{M},s} \mid h' \sqsubseteq h \text{ and } \mathrm{mode}(h') = 1\}$ is finite. According to (23) again, it implies that h and $\pi_0(h)$ are identical, except for a finite prefix. Since ϕ is prefix-independent, it implies $\phi(h) = \phi(\pi_0(h))$. Moreover, since $N_s(h) = +\infty$, we have $N_s(\pi_0(h)) = +\infty$. This two last facts prove (25) in the case where $\pi_1(h)$ is finite.

The case where $\pi_0(h)$ is finite is symmetrical.

Let us suppose now that both $\pi_0(h)$ and $\pi_1(h)$ are infinite. We prove that there exists $u_0, v_0, u_1, v_1 \in (\mathbf{SA})^*$ such that

$$\begin{aligned}
h &= u_0 v_0 u_1 v_1 \cdots \\
\pi_0(h) &= u_0 u_1 u_2 \cdots \\
\pi_1(h) &= v_0 v_1 v_2 \cdots \ .
\end{aligned} \tag{29}$$

Write $h = s_0 a_1 s_1 \cdots$. Let

$$\{n_0 < n_1 < \ldots\} = \{n > 0 \mid \mathrm{mode}(s_0 a_1 \cdots s_n) = 0 \text{ and } \mathrm{mode}(s_0 a_1 \cdots s_{n+1}) = 1\} \ ,$$
$$\{m_0 < m_1 < \ldots\} = \{m > 0 \mid \mathrm{mode}(s_0 a_1 \cdots s_m) = 1 \text{ and } \mathrm{mode}(s_0 a_1 \cdots s_{m+1}) = 0\} \ .$$

Then, by definition (22),

$$\forall i \in \mathbb{N}, \quad s_{n_i} = s_{m_i} = s \ . \tag{30}$$

Without loss of generality suppose $a_1 \in \mathbf{A}_0(s)$. Then by (22), $\mathrm{mode}(s_0 a_1 s_1) = 0$ hence $0 < n_0 < m_0 < n_1 < \cdots$. Define $u_0 = s_0 a_1 \cdots s_{n_0 - 1} a_{n_0}$, for $i \in \mathbb{N}$ define $v_i = s_{n_i} \cdots a_{m_i}$ and for $i \in \mathbb{N}$ define $u_{i+1} = s_{m_i} \cdots a_{n_{i+1}}$. Then by (23) we get (29).

Since ϕ is submixing, (29) implies $\phi(h) \leq \max\{\phi(\pi_0(h)), \phi(\pi_1(h)\}$. Since $\phi(h) > x$ we deduce $x < \max\{\phi(\pi_0(h)), \pi_1(h)\}$, i.e.

$$(\phi(\pi_0(h)) < x) \text{ or } (\phi(\pi_1(h)) < x). \tag{31}$$

Moreover, by (30) and (29), histories $\pi_0(h)$ and $\pi_1(h)$ reaches infinitely often s, hence $N_s(\pi_0(h)) = N_s(\pi_1(h)) = +\infty$. This last fact together with (31) implies (25) which achieves this proof. □

The following theorem shows that any strategy σ in $\mathcal{M}$ can be simulated by a strategy σ_0 in $\mathcal{A}_0$, in a way that for any Π_0-measurable event E in $\mathcal{M}$, the probability of E under σ in $\mathcal{A}$ is less than the probability of $\Pi_0(E)$ under σ_0 in $\mathcal{A}_0$.

Theorem 4. *Let $\mathcal{M}$ be a Markov decision process, σ a strategy in $\mathcal{M}$, s a state of $\mathcal{M}$ such that $|\mathbf{A}(s)| \geq 2$, $(\mathcal{M}_0, \mathcal{M}_1)$ a split of $\mathcal{M}$ on s, and Π_0 he associated projection. Then there exists a strategy σ_0 in $\mathcal{M}_0$ such that for any event $E_0 \subseteq \mathbf{P}^{\omega}_{\mathcal{M}_0, s}$,*

$$\mathbb{P}_{\sigma, s}(\pi_0 \in E_0) \leq \mathbb{P}_{\sigma_0, s}(E_0). \tag{32}$$

Proof. The symbol $\sqsubseteq$ denotes the prefix ordering on finite and infinite words. For two words u, v, we write $u \sqsubset v$ if u is a strict prefix of v i.e. if $u \sqsubseteq v$ and $u \neq v$.

For any state $t \neq s$, let us choose in an arbitrary way an action $a_t \in \mathbf{A}(t)$, and let us also choose an action $a_s \in \mathbf{A}_0(s)$. For any $h \in \mathbf{P}^*_{\mathcal{M}_0, s}$ with target t and for any action $a \in \mathbf{A}(t)$, we define

$$\sigma_0(h)(a) = \begin{cases} \mathbb{P}^{\sigma}_s(ha \sqsubseteq \Pi_0 \mid h \sqsubset \Pi_0) & \text{if } \mathbb{P}^{\sigma}_s(h \sqsubset \Pi_0) > 0 \\ 1 & \text{if } \mathbb{P}^{\sigma}_s(h \sqsubset \Pi_0) = 0 \text{ and } a = a_t \\ 0 & \text{if } \mathbb{P}^{\sigma}_s(h \sqsubset \Pi_0) = 0 \text{ and } a \neq a_t \end{cases}$$

Then σ_0 is a strategy in $\mathcal{M}_0$ since by definition of $\sqsubset$,

$$\mathbb{P}^{\sigma}_s(h \sqsubset \Pi_0) = \sum_{a \in \mathbf{A}(t)} \mathbb{P}^{\sigma}_s(ha \sqsubseteq \Pi_0) \ .$$

We first show (32) in the particular case where there exists $h_0 \in \mathbf{P}^*_{\mathcal{M}_0,s}$ such that $E_0 = \{l \in \mathbf{P}^\omega_{\mathcal{M}_0,s} \mid h \sqsubseteq l\}$. Remember that we abuse the notation and write simply $E_0 = h$. With this notation, we wish to prove that:

$$\forall h' \in \mathbf{P}^*_{\mathcal{M}_0,s}, \quad \mathbb{P}^\sigma_s(h' \sqsubseteq \Pi_0) \leq \mathbb{P}^{\sigma_0}_s(h'). \tag{33}$$

We prove (33) inductively. If $h' = s$ then since Π_0 has values in $\mathbf{P}^*_{\mathcal{M},s} \cup \mathbf{P}^\omega_{\mathcal{M},s}$, we get $\mathbb{P}^\sigma_s(s \sqsubseteq \pi_0) = 1 = \mathbb{P}^{\sigma_0}_s(s)$. Now let us suppose that (33) is proved for some finite history $h \in \mathbf{P}^*_{\mathcal{M}_0,s}$. Let t be the target of h and $a \in \mathbf{A}_0(t)$, and let us prove that (33) holds for $h' = hat$. First case is $\mathbb{P}^\sigma_s(h \sqsubset \Pi_0) = 0$, then a fortiori $\mathbb{P}^\sigma_s(har \sqsubseteq \Pi_0) = 0$, and (33) holds for $h' = hat$. Now let us suppose $\mathbb{P}^\sigma_s(h \sqsubset \Pi_0) \neq 0$. Then,

$$\begin{aligned}
\mathbb{P}^\sigma_s(har \sqsubseteq \Pi_0) &= p(r|t,a) \cdot \mathbb{P}^\sigma_s(ha \sqsubseteq \Pi_0) \\
&= p(r|t,a) \cdot \mathbb{P}^\sigma_s(ha \sqsubseteq \Pi_0 \mid h \sqsubset \Pi_0) \cdot \mathbb{P}^\sigma_s(h \sqsubset \Pi_0) \\
&= p(r|t,a) \cdot \sigma_0(h)(a) \cdot \mathbb{P}^\sigma_s(h \sqsubset \Pi_0) \\
&\leq p(r|t,a) \cdot \sigma_0(h)(a) \cdot \mathbb{P}^\sigma_s(h \sqsubseteq \Pi_0) \\
&\leq p(r|t,a) \cdot \sigma_0(h)(a) \cdot \mathbb{P}^{\sigma_0}_s(h) \\
&= \mathbb{P}^{\sigma_0}_s(har).
\end{aligned}$$

The first equality comes from (24), and the third is by definition of σ_0. The last inequality is by induction hypothesis and the last equality by (2) and (3). It achieves the proof of equality (33).

Let us achieve the proof of Theorem 4. Let $\mathcal{E}$ be the collection of events $E_0 \subseteq \mathbf{P}^\omega_{\mathcal{M}_0,s}$ such that (32) holds. Then observe that $\mathcal{E}$ is stable by enumerable disjoint unions and enumerable increasing unions. According to (33), $\mathcal{E}$ contains all the events $(\mathcal{O}_{h_0})_{h_0 \in \mathbf{P}^*_{\mathcal{M}_0,s}}$. Since $\mathcal{E}$ is stable by enumerable disjoint unions, it contains the collection $\{\bigcup_{h_0 \in H_0} \mathcal{O}_{h_0} \mid H_0 \subseteq \mathbf{P}^*_{\mathcal{M}_0,s}\}$. This last collection is a Boolean algebra. Since $\mathcal{E}$ is stable by enumerable increasing union, it implies that $\mathcal{E}$ contains the σ-field generated by $(\mathcal{O}_{h_0})_{h_0 \in \mathbf{P}^*_{\mathcal{M}_0,s}}$, i.e. all measurable sets of $\mathbf{P}^\omega_{\mathcal{M}_0,s}$. This achieves this proof. □

5.5 Histories that never come back in their initial state.

We deduce from theorem 3 the following result.

Theorem 5. *Let $\mathcal{M}$ be a Markov decision process with payoff function ϕ, s a state, σ a strategy and $(\mathcal{M}_0, \mathcal{M}_1)$ a split of $\mathcal{M}$ on s. Let us suppose that ϕ is prefix-independent. Then*

$$\mathbb{E}^\sigma_s[\phi \mid N_s < \infty] \leq \max\{\mathrm{val}(\mathcal{M}_0)(s), \mathrm{val}(\mathcal{M}_1)(s)\}. \tag{34}$$

Proof. Let us define $v_0 = \mathrm{val}(\mathcal{M}_0, \phi)$ and $v_1 = \mathrm{val}(\mathcal{M}_1)(\phi)$. For any action $a \in \mathbf{A}(s)$ we denote σ_a the strategy in $\mathcal{M}$ defined for $h \in \mathbf{P}^*_{\mathcal{M},s}$ by:

$$\begin{cases} \sigma_a(h) = \sigma(h) \text{ if the target of } h \text{ is not } s \\ \sigma_a(h) \text{ chooses action } a \text{ with probability 1 otherwise.} \end{cases}$$

Remark that the strategy σ_a always chooses the same action when plays reaches state s, hence it is a strategy either in $\mathcal{M}_0$ or in $\mathcal{M}_1$. From Theorem 3, we deduce

$$\forall a \in \mathbf{A}(s), \quad \mathbb{E}_s^{\sigma_a}[\phi \mid N_s = 0] \leq \max\{v_0, v_1\}. \tag{35}$$

Since σ and σ_a coincide on $\{N_s = 0, A_1 = a\}$, lemma 2 implies :

$$\begin{aligned}\mathbb{E}_s^{\sigma}[\phi \mid A_1 = a, N_s = 0] &= \mathbb{E}_s^{\sigma_a}[\phi \mid A_1 = a, N_s = 0] \\ &= \mathbb{E}_s^{\sigma_a}[\phi \mid N_s = 0] \ ,\end{aligned}$$

where the last equality holds since by definition of σ_a, $\mathbb{P}_s^{\sigma_a}(A_1 = a) = 1$. Together with (35), we get $\mathbb{E}_s^{\sigma}[\phi \mid A_1 = a, N_s = 0] \leq \max\{v_0, v_1\}$, whatever be action a and strategy σ. It implies :

$$\forall \sigma \in \Sigma_{\mathcal{M}}, \quad \mathbb{E}_s^{\sigma}[\phi \mid N_s = 0] \leq \max\{v_0, v_1\}.$$

Conditioning on the last moment where history reaches s, and using the shifting lemma and the prefix-independence of ϕ, this last equation implies :

$$\mathbb{E}_s^{\sigma}[\phi \mid N_s < \infty] \leq \max\{v_0, v_1\}.$$

It achieves the proof of Theorem 5. □

5.6 Histories that infinitely often reach their initial state.

The following theorem shows that if an history reaches infinitely often its initial state, then its value is no more than the value of that state.

Theorem 6. *Let $\mathcal{M}$ be a Markov decision process with payoff function ϕ, s a state and σ a strategy. Suppose that ϕ is prefix-independent and submixing. Then*

$$\mathbb{P}_s^{\sigma}(\phi > \mathrm{val}(\mathcal{M})(s) \mid N_s = \infty) = 0. \tag{36}$$

Moreover, suppose that $|\mathbf{A}(s)| \geq 2$ and let $(\mathcal{M}_0, \mathcal{M}_1)$ be a split of $\mathcal{M}$ on s. Then

$$\mathbb{P}_s^{\sigma}\left(\phi > \max\{\mathrm{val}(\mathcal{M}_0)(s), \mathrm{val}(\mathcal{M})(s)\} \mid N_s = \infty\right) = 0. \tag{37}$$

Proof. We prove that theorem by induction on $N(\mathcal{M}) = \sum_{s \in \mathbf{S}}(|\mathbf{A}(s)| - 1)$.

If $N(\mathcal{M}) = 0$ then $\mathcal{M}$ is a Markov chain. In that case, $\mathbb{P}_s^{\sigma}(N_s = \infty) > 0$ iff s is a recurrent state iff $\mathbb{P}_s^{\sigma}(N_s = \infty) = 1$. Hence (36) is a direct consequence of Theorem 2. Moreover, since $N(\mathcal{M}) = 0$, then $\forall s, |\mathbf{A}(s)| = 1$ and we do not need to prove (37).

Now let us suppose that $N(\mathcal{M}) > 0$ and that Theorem 6 is proved for any $\mathcal{M}'$ such that $N(\mathcal{M}') < N(\mathcal{M})$. We first prove (37). Let s be a state, σ a strategy, suppose that $|\mathbf{A}(s)| > 2$ and let $(\mathcal{M}_0, \mathcal{M}_1)$ be a split of $\mathcal{M}$ on s. Let $\mathcal{M}_0 = (\mathcal{M}_0, \phi)$, $\mathcal{M}_1 = (\mathcal{M}_1, \phi)$, $v_0 = \mathrm{val}(\mathcal{M}_0, \phi)$, $v_1 = \mathrm{val}(\mathcal{M}_1, \phi)$, and Π_0, Π_1 the associated projections. Let

$$\begin{aligned}E_0 &= \{h_0 \in \mathbf{P}^{\omega}_{\mathcal{M}_0, s} \mid \phi(h_0) > v_0 \text{ and } N_s(h_0) = +\infty\} \\ E &= \{h \in \mathbf{P}^{\omega}_{\mathcal{M}, s} \mid \pi_0(h) \in E_0\} \ .\end{aligned}$$

We start with proving that

$$\mathbb{P}_s^\sigma(E) = 0. \tag{38}$$

From Theorem 4, there exists a strategy σ_0 in $\mathcal{M}_0$ such that $\mathbb{P}_s^\sigma(\Pi_0 \in E_0) \leq \mathbb{P}_s^{\sigma_0}(E_0)$. Hence

$$\begin{aligned}\mathbb{P}_{\sigma,s}(E) &= \mathbb{P}_{\sigma,s}(\Pi_0 \in E_0) \leq \mathbb{P}_{\sigma_0,s}(E_0) \\ &= \mathbb{P}_{\sigma_0,s}(\phi > v_0 \text{ and } N_s = +\infty) \\ &= 0,\end{aligned}$$

where this last equality holds by induction hypothesis, since $N(\mathcal{M}_0) < N(\mathcal{M})$. Hence we have shown (38) and by symmetry, we obtain for $i \in \{0, 1\}$,

$$\mathbb{P}_s^\sigma(\Pi_i \text{ is infinite and } N_s(\Pi_i) = \infty \text{ and } \phi(\Pi_i) > v_i) = 0 \ .$$

Now consider (25) of Proposition 4, with $x = \max\{v_0, v_1\}$. Together with the last equation, it gives (37).

Now we prove that (36) holds.

First we show that (36) holds for any state s such that $|\mathbf{A}(s)| \geq 2$. Every strategy in $\mathcal{M}_0$ or $\mathcal{M}_1$ is also a strategy in $\mathcal{M}$, hence $\text{val}(\mathcal{M})(s) \geq \max\{v_0, v_1\}$ and we deduce from (37) that $\mathbb{P}_s^\sigma\ (\phi > \text{val}(\mathcal{M})(s) \mid N_s = \infty) = 0$. Hence the set

$$T = \{s \in \mathbf{S} \mid \forall \sigma \in \Sigma_\mathcal{M},\ \ \mathbb{P}_s^\sigma(\phi > \text{val}(\mathcal{M})(s) \text{ and } N_s = \infty) = 0\} \tag{39}$$

contains any state $s \in \mathbf{S}$ such that $|\mathbf{A}(s)| \geq 2$. Hence (36) holds for any s such that $|\mathbf{A}(s)| \geq 2$. Let $U = \mathbf{S} \backslash T$. We have proved that :

$$\forall s \in U,\ \ |\mathbf{A}(s)| = 1 \ . \tag{40}$$

For achieving the proof of (36) we must prove that $T = \mathbf{S}$, i.e. $U = \emptyset$. Suppose the contrary, and let us search a contradiction. If $U \neq \emptyset$, then the set

$$W = \{s \in U \mid \text{val}(\mathcal{M})(s) = \min_{t \in U} \text{val}(\mathcal{M})(t)\}$$

is not empty and contains a state $s \in W$. According to (40), there exists a unique action a available in s.

Now we show that $\forall t \in \mathbf{S}$ such that $p(t|s, a) > 0$,

$$\text{if } t \in U \text{ then } \ \text{val}(\mathcal{M})(t) \geq \text{val}(\mathcal{M})(s) \tag{41}$$

$$\text{if } t \in T \text{ then } \ \text{val}(\mathcal{M})(t) > \text{val}(\mathcal{M})(s) \ . \tag{42}$$

The case where $t \in U$ is clear since we choose s with minimal value in U. Now let $t \in T$ such that $p(t|s, a) > 0$ and let us prove 42. Since $s \in U$, $s \notin T$ and by definition of T,

$$\exists \sigma \in \Sigma_\mathcal{M} \quad \text{s.t. } \mathbb{P}_s^\sigma(\phi > \text{val}(\mathcal{M})(s) \text{ and } N_s = \infty) > 0 \ . \tag{43}$$

Now remark that since $p(t|s, a) > 0$ we have $\mathbb{P}_s^\sigma(N_t = \infty \mid N_s = \infty) = 1$. Together with (43), it implies

$$\mathbb{P}_s^\sigma(\phi > \text{val}(\mathcal{M})(s) \text{ and } N_t = \infty) > 0 \ .$$

Conditioning this probability on the first moment the history reaches state t, we deduce that there exists a finite history $h \in \mathbf{P}^*_{\mathcal{M},s}$ with source s and target t such that

$$\mathbb{P}^\sigma_s(\phi > \mathrm{val}(\mathcal{M})(s) \text{ and } N_t = \infty \mid h) > 0 \ .$$

Since ϕ is prefix-independent, and according to the shifting lemma it implies :

$$\mathbb{P}^{\sigma[h]}_t(\phi > \mathrm{val}(\mathcal{M})(s) \text{ and } N_t = \infty) > 0.$$

Since $t \in T$, the definition of T implies :

$$\mathbb{P}^{\sigma[h]}_t(\phi > \mathrm{val}(\mathcal{M})(t) \text{ and } N_t = \infty) = 0 \ .$$

Those two last equations imply $\mathrm{val}(\mathcal{M})(t) > \mathrm{val}(\mathcal{M})(s)$, which achieves the proof of (42).

Now we are close to get the contradiction we are looking for. Since ϕ is prefix-independent, we deduce :

$$\mathrm{val}(\mathcal{M})(s) = \sum_{t : p(t|s,a)>0} p(t|s,a) \cdot \mathrm{val}(\mathcal{M})(t).$$

Together with (42) we get :

$$\forall t \in \mathbf{S}, \quad (p(t|s,a) > 0) \implies (t \in U) \text{ and } (\mathrm{val}(\mathcal{M})(t) = \mathrm{val}(\mathcal{M})(s)) \ .$$

This last equation holds for any $s \in W$. Thus any transition with source in W has target in W. It implies that any history in $\mathcal{M}$ with source in W stays in W with probability 1, hence the restriction $\mathcal{M}[W]$ of $\mathcal{M}$ to the set of states W is an Markov decision process, and

$$\forall s \in W, \quad \mathrm{val}(\mathcal{M})(s) = \mathrm{val}(\mathcal{M}[W])(s) \tag{44}$$

Let $\mathcal{M}[W] = (\mathcal{M}[W], \phi)$. By definition of U, there exists a strategy σ in $\mathcal{M}[W]$ such that $\mathbb{P}_{\sigma,s}(\phi > \mathrm{val}(\mathcal{M})(s) \text{ and } N_s = \infty) > 0$ and together with (44) we get

$$\mathbb{P}^\sigma_s(\phi > \mathrm{val}(\mathcal{M}[W])(s) \text{ and } N_s = \infty) > 0 \ .$$

Since $W \subseteq U$ and according to (40), $\mathcal{M}[W]$ is a Markov chain. According to the last equation, $\mathbb{P}_s N_s = \infty > 0$ hence s is a recurrent state. Hence the last equation contradicts Theorem 2.

Finally we get a contradiction, hence $U = \emptyset$. This achieves the proof of Theorem 6. □

5.7 Proof of Theorem 1

The above results aggregate as follows.

Corollary 1. *Let $\mathcal{M}$ be a Markov decision process with payoff function ϕ, s a state of $\mathcal{M}$ and $(\mathcal{M}_0, \mathcal{M}_1)$ a split of $\mathcal{M}$ on s. Suppose that ϕ is prefix-independent and submixing. Then*

$$\mathrm{val}(\mathcal{M})(s) = \max\{\mathrm{val}(\mathcal{M}_0)(s), \mathrm{val}(\mathcal{M}_1)(s)\}. \tag{45}$$

Proof. Let $\sigma \in \Sigma_{\mathcal{M}}$. Then

$$\begin{aligned}\mathbb{E}_s^\sigma[\phi] &= \mathbb{E}_s^\sigma[\phi \mid N_s < \infty] \cdot \mathbb{P}_s^\sigma(N_s < \infty) + \mathbb{E}_s^\sigma[\phi \mid N_s = \infty] \cdot \mathbb{P}_s^\sigma(N_s = \infty) \\ &\leq \max\{\mathrm{val}(\mathcal{M}_0)(s), \mathrm{val}(\mathcal{M}_1)(s)\} \cdot (\mathbb{P}_s^\sigma(N_s < \infty) + \mathbb{P}_s^\sigma(N_s = \infty)) \\ &= \max\{\mathrm{val}(\mathcal{M}_0)(s), \mathrm{val}(\mathcal{M}_1)(s)\}\end{aligned}$$

The second inequality is a consequence of Theorems 6 and 5. Since it is true for any strategy σ, we get :

$$\mathrm{val}(\mathcal{M})(s) \leq \max\{\mathrm{val}(\mathcal{M}_0)(s), \mathrm{val}(\mathcal{M}_1)(s)\} \ .$$

To conclude, notice that a strategy for the Markov decision process $\mathcal{M}_0$ or $\mathcal{M}_1$ is also a strategy for the Markov decision process $\mathcal{M}$, hence $\mathrm{val}(\mathcal{M}_0)(s) \leq \mathrm{val}(\mathcal{M})(s)$ and $\mathrm{val}(\mathcal{M}_1)(s) \leq \mathrm{val}(\mathcal{M})(s)$. □ □

Now we can achieve the proof of Theorem 1.

Proof. of Theorem 1. We prove Theorem 1 by induction on $N(\mathcal{M}) = \sum_{s \in \mathbf{S}}(|\mathbf{A}(s)| - 1)$.

First case is the case where $N(\mathcal{M}) = 0$, i.e. $\mathcal{M}$ is a Markov chain. In that case, there exists a unique strategy, which is necessarily optimal and positional.

Now let us suppose that $N(\mathcal{M}) = \sum_{s \in \mathbf{S}}(|\mathbf{A}(s)| - 1) > 0$ and that Theorem 1 is proved for any $\mathcal{M}'$ such that $N(\mathcal{M}') < N(\mathcal{M})$. Since $N(\mathcal{M}) > 0$, there exists a state s of $\mathcal{M}$ such that $|\mathbf{A}(s)| \geq 2$. Let $(\mathcal{M}_0, \mathcal{M}_1)$ be a split of $\mathcal{M}$ on s. Without loss of generality, we can suppose that :

$$\mathrm{val}(\mathcal{M}_0)(s) \geq \mathrm{val}(\mathcal{M}_1)(s) \ , \tag{46}$$

and according to corollary 1, we deduce :

$$\mathrm{val}(\mathcal{M}_0)(s) = \mathrm{val}(\mathcal{M})(s) \ . \tag{47}$$

By inductive hypothesis, there exists a positional strategy σ_0 optimal for the Markov decision process $\mathcal{M}_0$. We are going to prove that σ_0 is also optimal for the Markov decision process $\mathcal{M}$. Let $\sigma \in \Sigma_{\mathcal{M}}$ and $t \in \mathbf{S}$. Then

$$\begin{aligned}\mathbb{E}_t^\sigma[\phi] = \mathbb{E}_t^\sigma[\phi \mid \exists n, S_n = s] \cdot \mathbb{P}_t^\sigma(\exists n, S_n = s) + \\ \mathbb{E}_t^\sigma[\phi \mid \forall n, S_n \neq s] \cdot \mathbb{P}_t^\sigma(\forall n, S_n \neq s).\end{aligned} \tag{48}$$

Let $\tau \in \Sigma_{\mathcal{M}}$ defined as follows :

$$\tau(s_0 a_1 \cdots s_n) = \begin{cases} \sigma(s_0 a_1 \cdots s_n) \text{ if } \forall 0 \leq i \leq n, s_i \neq s, \\ \sigma_0(s_n) \text{ otherwise.} \end{cases}$$

Then we have the three following equalities. First, since σ and τ coincide on the event $\{\forall n, S_n \neq s\}$, lemma 2 implies :

$$\mathbb{E}_t^\sigma[\phi \mid \forall n, S_n \neq s] \cdot \mathbb{P}_t^\sigma(\forall n, S_n \neq s) = \mathbb{E}_t^\tau[\phi \mid \forall n, S_n \neq s] \cdot \mathbb{P}_t^\tau(\forall n, S_n \neq s). \tag{49}$$

Second, by definition of τ,

$$\mathbb{P}_t^\sigma(\exists n, S_n = s) = \mathbb{P}_t^\tau(\exists n, S_n = s) \ . \tag{50}$$

And finally,

$$\begin{aligned}
\mathbb{E}_t^\sigma[\phi \mid \exists n, S_n = s] &\leq \mathrm{val}(\mathcal{M})(s) && (51)\\
&= \mathrm{val}(\mathcal{M}_0)(s) && (52)\\
&= \mathbb{E}_s^{\sigma_0}[\phi] && (53)\\
&= \mathbb{E}_t^\tau[\phi \mid \exists n, S_n = s] \ , && (54)\\
& && (55)
\end{aligned}$$

where the inequality comes from the shifting lemma and the prefix-independence of ϕ, the first equality from (46), the second from the fact that σ_0 is optimal in $\mathcal{M}_0$ and the third by the shifting lemma again.

Finally, (50), (49) and (54) together with (48) prove that

$$\forall t \in \mathbf{S}, \quad \forall \sigma \in \Sigma_{\mathcal{M}}, \quad \mathbb{E}_t^\sigma[\phi] \leq \mathbb{E}_t^\tau[\phi] \ . \tag{56}$$

By definition, τ always chooses an action of $\mathbf{A}_0(s)$ when the history has target s, hence τ is a strategy in $\mathcal{M}_0$. Since σ_0 is optimal in $\mathcal{M}_0$ hence we get :

$$\forall t \in \mathbf{S}, \quad \mathbb{E}_t^\tau[\phi] \leq \mathbb{E}_t^{\sigma_0}[\phi] \ .$$

Since σ_0 is also a strategy in $\mathcal{M}$, this last equation together with (56) proves that σ_0 is optimal in $\mathcal{M}$. Since σ_0 is positional, it achieves the proof of the inductive step, and of Theorem 1. □

6 Conclusion

In that paper, we have introduced the class of prefix-independent and submixing payoff functions, and we proved that they guarantee the existence of pure and stationary optimal strategies. Moreover, we have defined three operators on payoff functions, that can be used to generate new examples of Markov decision processes with pure and stationary optimal strategies.

Most of the results of this paper can be extended to the broader framework of two-player zero-sum stochastic games with perfect information [4].

To conclude, let us formulate the following conjecture about pure and stationary payoff functions. "Let ϕ be a prefix-independent payoff function. Suppose that in every non-stochastic one player game with payoff function ϕ, there exists pure and stationary optimal strategies. Then the same holds in every Markov decision process with payoff function ϕ.

Acknowledgments

I would like to thank Wiesław Zielonka for numerous discussions about payoff games on Markov decision processes, and for careful proof-reading of this paper.

References

[1] K.-J. Bierth. An expected average reward criterion. *Stochastic Processes and Applications*, 26:133–140, 1987.

[2] C. Courcoubetis and M. Yannakakis. Markov decision processes and regular events. In *ICALP'90*, volume 443 of *Lecture Notes in Computer Science*, pages 336–349. Springer, 1990.

[3] D. Gillette. Stochastic games with zero stop probabilities. In *Contribution to the Theory of Games III, Annals of Mathematics Studies*, volume 39, pages 179–187, 1957.

[4] H. Gimbert. *Jeux Positionnels*. PhD thesis, Université Denis Diderot, Paris, 2006.

[5] H. Gimbert and W. Zielonka. When can you play positionally? In *Proc. of MFCS'04*, volume 3153 of *LNCS*, pages 686–697. Springer, 2004.

[6] H. Gimbert and W. Zielonka. Limits of multidiscounted markov decision processes. In *Proceedings of LICS'07*, 2007.

[7] H. Gimbert and W. Zielonka. Perfect information stochastic priority games. In *Proceedings of ICALP'07*, 2007.

[8] E. Graedel, W. Thomas, and T. Wilke. *Automata, Logics and Infinite Games*, volume 2500 of *Lecture Notes in Computer Science*. Springer, 2002.

[9] E. Kopczyński. Half-positional determinacy of infinite games. In *Proceedings of ICALP'06*, volume 4052 of *Lecture Notes in Computer Science*. Springer, 2006.

[10] A.P. Maitra and W.D. Sudderth. *Discrete gambling and stochastic games*. Springer-Verlag, 1996.

[11] A. Neyman and S. Sorin. *Stochastic games and applications*. Kluwer Academic Publishers, 2003.

[12] L. S. Shapley. Stochastic games. In *Proceedings of the National Academy of Science USA*, volume 39, pages 1095–1100, 1953.