



HAL
open science

Can a Weaker Player Win? Adaptive Play in Repeated Games

Jonatha ANSELMi, Bruno Gaujal

► **To cite this version:**

Jonatha ANSELMi, Bruno Gaujal. Can a Weaker Player Win? Adaptive Play in Repeated Games. 2026. ⟨hal-05513121v1⟩

HAL Id: hal-05513121

<https://hal.science/hal-05513121v1>

Preprint submitted on 16 Feb 2026 (v1), last revised 18 Mar 2026 (v2)

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

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



Distributed under a Creative Commons CC BY-SA 4.0 - Attribution - ShareAlike - International License

Can a Weaker Player Win? Adaptive Play in Repeated Games

Jonatha ANSELMi and Bruno GAUJAL

{jonatha.anselmi,bruno.gaujal}@inria.fr

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France.

Abstract

Consider a two-player game repeated N times. Player 1 can choose between two styles (for interpretability, *offensive* and *defensive*), whereas Player 2 uses a single fixed style. Let $X_N := \#wins - \#losses$ for Player 1 after N games, and define the match gain as $\mathbb{E}[\text{sign}(X_N)]$, with $\text{sign}(0) = 0$. We assume Player 1 is weaker in the sense that each pure style is losing in expectation. Our objective is to identify under which parameter regimes Player 1 can nevertheless achieve a positive gain under an optimal adaptive policy.

Using dynamic programming, we solve the finite-horizon control problem and numerically identify parameter regimes in which the optimal gain is strictly positive at some horizon N^* . We also derive structural conditions guaranteeing that g_N^* is always negative, and regimes (notably with fair (D)) where g_N^* is nonnegative for all N and can be strictly positive for every $N \geq 2$. We then characterize the asymptotic behavior as $N \rightarrow \infty$ for a weak player. In the *safe* case, where the defensive style induces a sure draw, the limiting gain varies continuously with the parameters and may take any value in $[0, 1]$. In the *non-safe* case, the limiting gain converges to -1 when both styles are strictly losing, and to 0 when (D) is fair (and non-safe).

Keywords: Sequential decision making; repeated games; dynamic programming; Parrondo's paradox

1. Introduction

1.1. A challenge between two chess players

We start with a simple yet striking situation illustrating how adaptivity can overturn a per-game disadvantage. Two chess players face each other repeatedly, so empirical win-loss frequencies are well established. Player 1 can choose between two styles, (O) (offense) and (D) (defense), whereas Player 2 uses a single fixed style. Under either choice of Player 1, Player 2 is more likely to win. The game outcomes are:

- If Player 1 plays (O), she wins with probability 0.45 and loses with probability 0.55 (no draws).
- If Player 1 plays (D), she wins with probability 0.10, loses with probability 0.15, and draws with probability 0.75.

Player 1's one-game score is $+1$ for a win (W) and -1 for a loss (L), and her one-game expected gain is

$$g_1 := \mathbb{P}(W) - \mathbb{P}(L).$$

Thus Player 1 is weaker under both pure styles: $g_1^{(O)} = -0.10$ and $g_1^{(D)} = -0.05$. Despite this, Player 1 proposes a two-game match and we ask:

Should one bet on Player 1 or on Player 2?

Any non-adaptive two-game plan, i.e., ((O), (O)), ((O), (D)), ((D), (O)), or ((D), (D)), has strictly negative expected gain. In contrast, Player 1 can achieve a positive expected gain with an adaptive rule: play (O) in game 1; in game 2 play (D) if game 1 was a win, and (O) otherwise. A direct enumeration of the possible two-game outcomes yields the expected gain

$$\begin{aligned} g_2 &:= \mathbb{P}(W-W) + \mathbb{P}(W-\text{Draw}) - \mathbb{P}(L-L) \\ &= 0.45 \cdot 0.10 + 0.45 \cdot 0.75 - 0.55 \cdot 0.55 \\ &= 0.08, \end{aligned}$$

where outcomes with a draw in game 1 are impossible because game 1 is played under (O). This example illustrates a general mechanism: even if each pure style is losing in expectation, a state-dependent choice of styles over a finite horizon can yield a positive expected gain. In this paper, we study an idealized model capturing this effect.

1.2. Model and addressed problem

We formalize the preceding vignette as a finite-horizon repeated game between two players of unequal

strength. Player 2 plays a single fixed style throughout. At each round, Player 1 chooses one of two styles: (O) (offense) or (D) (defense). Conditional on the style chosen by Player 1 at round n , the round outcome is drawn independently of the past. Under (O), the outcome distribution is

$$\mathbf{p} = (p_w, p_d, p_\ell), \quad p_w + p_d + p_\ell = 1,$$

where p_w (resp. p_d, p_ℓ) is the probability that Player 1 wins (resp. draws, loses). Under (D), the distribution is

$$\mathbf{q} = (q_w, q_d, q_\ell), \quad q_w + q_d + q_\ell = 1.$$

We call (D) *defensive* when draws are more likely under (D) than under (O), i.e.,

$$q_d \geq p_d.$$

In the remainder, we use the following terminology.

Definition 1. *Player 1 is weak if each pure style is losing in expectation, i.e.,*

$$p_w \leq p_\ell \quad \text{and} \quad q_w \leq q_\ell,$$

and strictly weak if both inequalities are strict.

Definition 2. *The defense (D) is safe if $q_d = 1$ and fair but non-safe if $q_w = q_\ell$ and $q_d < 1$.*

The example in Section 1.1 exhibits the core effect: despite strict weakness under both styles, an adaptive choice over two rounds yields a positive expected gain.

The match is played over N games, and the outcome is evaluated *globally* rather than game-by-game. A natural objective is the sign of the final score difference: Player 1 “wins the match” if the number of wins exceeds the number of losses, “loses” otherwise, and draws if the two are equal. Under this criterion, the relevant question is not whether Player 1 can outperform Player 2 in expectation in a single game (it cannot), but whether adaptive switching between styles over time can bias the *distribution* of the final score enough to yield a positive expected match gain.

Let W_n and L_n denote the number of wins and losses of Player 1 after n games and let the *score* be $X_n := W_n - L_n$. We evaluate the expected *gain* at the end of the match via

$$g_N := \mathbb{E}[\text{sign}(X_N) \mid X_0 = 0], \quad \text{sign}(0) = 0,$$

so that $g_N > 0$ means Player 1 has a positive expected advantage under the win/lose/draw match criterion. The central question is whether, and for which parameter

regimes, Player 1 can achieve $g_N > 0$ by selecting (O) or (D) adaptively as a function of the match state (e.g., the current score difference), even though both pure styles are unfavorable.

This viewpoint places the interaction in the scope of finite-horizon sequential decision-making: at each round, Player 1 observes the current match state (e.g., the score difference so far) and selects a style that trades off immediate expected loss against variance and future opportunities. The key phenomenon is that, even when each pure style is unfavorable in isolation, a state-dependent mixture can exploit the finite horizon and the nonlinear match criterion to produce regimes where the weaker player has a positive chance—and even a positive expected gain—of winning the overall challenge.

1.3. Contribution

In this paper, we are interested in the game above from the standpoint of Player 1 (the weak player). We aim at understanding under which conditions her gain is non-negative.

- First, we clarify the role of the defensive play parameter q_d . In the *safe* regime $q_d = 1$, we analyze the adaptive state-dependent policy *catenaccio* (cat) and obtain an explicit limit for its gain as $N \rightarrow \infty$ (Theorem 1). In contrast, when (D) is fair but non-safe ($q_w = q_\ell > 0$ and $q_d < 1$), the same policy becomes asymptotically detrimental and its gain converges to a negative value (Theorem 2).
- Then, we consider the *optimal* adaptive policy. We formulate the problem as a finite-horizon control problem on the score state X_n and solve it via Bellman optimality equation. A numerical solution shows that g_N^* can be non-monotone in N and that strict positivity may occur only at specific horizons N^* (see Figure 2). We complement these computations with structural results: (i) we provide simple sufficient conditions on $(p_w, p_\ell, q_w, q_\ell)$ under which g_N^* is *uniformly negative* for all N , and identify a parameter regime (when (D) is fair) where g_N^* is nonnegative (Proposition 2).
- Finally, we characterize the asymptotic behavior of the optimal gain as $N \rightarrow \infty$ (Theorem 3). This is the main technical difficulty and contribution. Outside the safe defense case, the limit is *quantized*: for weak players it converges either to -1 (both styles strictly losing) or to 0 (defense is fair but non-safe defense). In the safe case, the limiting gain can take any value

in $[0, 1]$ as a function of the parameters, and the theorem also shows that a slight refinement of *catenaccio*, i.e., (cat^+) , is essentially optimal in that case.

Our proofs combine random-walk arguments, finite-horizon dynamic programming, and asymptotic martingale estimates.

2. Related Work

The phenomenon introduced in Section 1.1, i.e., combining or alternating between individually unfavorable modes of play can improve performance, is closely related to *Parrondo's paradox* [4, 6], originally introduced in the context of gambling games. In its canonical formulation, two Markovian games (or Markov chains) that are losing in isolation (negative drift) can yield a positive long-run average gain (equivalently, positive asymptotic drift) when alternated or mixed. This occurs because the per-step drift is state-dependent: alternating between games changes the occupation frequencies over states (e.g., capital modulo a period) and can therefore reverse the average drift.

Parrondo's paradox has found applications across disciplines, including physics, biology, and finance; see [5] for a survey. The effect also extends beyond the two-game setting: variants with three or more individually losing games can still exhibit a winning outcome under suitable alternation schemes [1]. Beyond random or periodic mixing, finite-horizon optimal sequencing and adaptive selection rules have also been investigated, typically via backward induction/dynamic programming [2, 1].

Our model shares this high-level mechanism, i.e., state-dependent switching between two styles. In this sense, our setting is Parrondo-like: two individually losing modes can yield a positive outcome under switching, but here the effect is driven by finite-horizon adaptive control under a nonlinear match criterion, together with the safe/non-safe dichotomy of the defensive style. In fact, in this paper we distinguish between the *safe* (score-freezing) from *non-safe* defense style of play. This yields a dichotomy that drives the long-run behavior: it enables a complete asymptotic classification for weak players and identifies regimes where *catenaccio*-type policies are essentially optimal. In addition, in contrast with the works above, our focus is not algorithmic. We are interested in structural conditions on the input parameters under which a weaker player can (or cannot) achieve a positive gain.

3. Random walks, (un)certainity, and *catenaccio*

If Player 1 commits to a fixed style (say (D)), the match score evolves as a one-dimensional (lazy) random walk and the probability that Player 1 wins an N -game match under (D) is (see, e.g., [3])

$$\mathbb{P}(X_N > 0) = \sum_{\substack{i > j \geq 0 \\ i+j \leq N}} \binom{N}{i} \binom{N-i}{j} q_w^i q_\ell^j q_d^{N-i-j}, \quad (1)$$

where i (resp. j) counts wins (resp. losses) and the remaining $N-i-j$ rounds are draws. Therefore, if $q_w < q_\ell$ then $\mathbb{P}(X_N > 0) < 1/2$, the gain is negative for all N , $g_N^{(D)} < 0$ and goes to -1 as N goes to infinity.

On the other hand, with two available styles, a natural policy used by weaker teams in sports is to pursue variance early and then “protect the lead” once ahead. We model this as the *catenaccio* policy:

*Play (O) until the score becomes positive,
then play (D) thereafter.*

Formally, let $\tau := \inf\{n \geq 0 : X_n = +1\}$ and play (O) up to time τ and (D) afterwards.

3.1. Safe defense: a positive asymptotic gain

When the defensive style (D) is safe, the expected gain of Player 1 under the *catenaccio* policy converges to $2p_w/p_\ell - 1$ as N goes to infinity. Thus, it is positive if and only if $p_w \geq p_\ell/2$. This is stated in the next result.

Theorem 1. *If $q_d = 1$, then $g_N^{(\text{cat})} \xrightarrow{N \rightarrow \infty} 2p_w/p_\ell - 1$.*

Proof. Under (O), the score evolves as a (possibly lazy) random walk with increments in $\{-1, 0, +1\}$ and drift $p_w - p_\ell \leq 0$. Let $\mathcal{W} := \{\exists n \geq 0 : X_n = +1\}$ be the event that the walk ever hits $+1$. Classical ruin/hitting-probability theory yields (see [3, p. 272]), $\mathbb{P}(\mathcal{W}) = p_w/p_\ell$, with $p_w \leq p_\ell$, for the non-lazy case; the same formula holds here since laziness only inserts zero-steps. Under *catenaccio* with safe (D), hitting $+1$ at any time implies $X_N = +1$ for all subsequent horizons, whereas on \mathcal{W}^c the score never becomes positive, hence $\text{sign}(X_N) = -1$ for all N . Therefore,

$$\lim_{N \rightarrow \infty} g_N^{(\text{cat})} = (+1)\mathbb{P}(\mathcal{W}) + (-1)\mathbb{P}(\mathcal{W}^c) = 2\frac{p_w}{p_\ell} - 1.$$

This concludes the proof. \square

3.2. Fair but non-safe defense

Now, assume that the defensive style (D) is fair but non-safe, meaning $q_w = q_\ell > 0$. Intuitively, once catenaccio switches to (D), the score continues to fluctuate and eventually forgets the early advantage. The next result shows that the expected gain of Player 1 under the catenaccio policy always becomes negative.

Theorem 2. *If $q_w = q_\ell > 0$, $g_N^{(\text{cat})} \xrightarrow{N \rightarrow \infty} p_w/p_\ell - 1 < 0$.*

Proof. Let \mathcal{W} be as in the proof of Theorem 1; ruin theory again gives $\mathbb{P}(\mathcal{W}) = p_w/p_\ell$. On \mathcal{W}^c , the score never reaches +1, so catenaccio never switches and (O) is used at every round. Since $p_w < p_\ell$, the law of large numbers implies $X_N \rightarrow -\infty$ almost surely on \mathcal{W}^c , hence $\text{sign}(X_N) \rightarrow -1$.

On \mathcal{W} , let $K := \inf\{n \geq 0 : X_n = +1\}$ be the first hitting time of +1. By definition of catenaccio, for all $n \geq 1$,

$$X_{K+n} = 1 + \sum_{i=1}^n B_i,$$

where $(B_i)_{i \geq 1}$ are i.i.d. with $\mathbb{P}(B_i = +1) = q_w$, $\mathbb{P}(B_i = 0) = q_d$, and $\mathbb{P}(B_i = -1) = q_\ell = q_w$. In particular, B_i is mean-zero and symmetric, so $X_{K+n} - 1$ is a centered (lazy) symmetric walk. By the central limit theorem (or a local limit theorem),

$$\mathbb{P}(X_{K+n} > 0 \mid K) \rightarrow \frac{1}{2}, \quad \mathbb{P}(X_{K+n} < 0 \mid K) \rightarrow \frac{1}{2},$$

and therefore $\mathbb{E}[\text{sign}(X_{K+n}) \mid K] \rightarrow 0$. Since $K < \infty$ on \mathcal{W} , this yields $\mathbb{E}[\text{sign}(X_N) \mid \mathcal{W}] \rightarrow 0$ as $N \rightarrow \infty$. Combining the two cases,

$$\begin{aligned} \lim_{N \rightarrow \infty} g_N^{(\text{cat})} &= \lim_{N \rightarrow \infty} \mathbb{E}[\text{sign}(X_N) \mid \mathcal{W}] \mathbb{P}(\mathcal{W}) + \\ &\quad \lim_{N \rightarrow \infty} \mathbb{E}[\text{sign}(X_N) \mid \mathcal{W}^c] \mathbb{P}(\mathcal{W}^c) \\ &= 0 \cdot \frac{p_w}{p_\ell} + (-1) \left(1 - \frac{p_w}{p_\ell}\right), \end{aligned}$$

as desired. \square

Therefore, when defense is merely fair but non-safe, the “post-switch” dynamics asymptotically erase the advantage, so catenaccio only yields transient finite-horizon benefits.

4. Beyond catenaccio: optimal policies and dynamic programming

The *catenaccio* policy provides a useful benchmark, but it is generally not optimal. In particular, when (D) is fair, we will see that the *optimal* policy can yield a

nonnegative gain for every horizon and may be strictly positive for finite N even when Player 1 is weak under both pure styles. This section addresses the decision problem:

Given \mathbf{p} and \mathbf{q} , does there exist a horizon N such that Player 1 can achieve $g_N(\mathbf{p}, \mathbf{q}) > 0$ under an optimal adaptive policy?

4.1. Computing an optimal policy

For a fixed horizon N , the problem is a finite-horizon Markov decision process on the score X_n . We encode the match outcome through the terminal value function

$$V_N^N(x) = \begin{cases} +1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

For $0 \leq n \leq N - 1$, let $V_n^N(x)$ denote the optimal value at time n when the current score is x . The Bellman optimality recursion reads: for all $n \geq 0$ and $x \in \{-n, \dots, n\}$,

$$V_n^N(x) = \max\{p_w V_{n+1}^N(x+1) + p_d V_{n+1}^N(x) + p_\ell V_{n+1}^N(x-1), \\ q_w V_{n+1}^N(x+1) + q_d V_{n+1}^N(x) + q_\ell V_{n+1}^N(x-1)\}. \quad (2)$$

The optimal gain is the value at the origin:

$$g_N^* = V_0^N(0).$$

In some cases, we mark the dependence on \mathbf{p} and \mathbf{q} , by denoting the gain $g_N^*(\mathbf{p}, \mathbf{q})$. When the context is clear, we simply denote the gain as g_N^* . The backward computation induced by (2) is illustrated in Figure 1. The lattice points (n, x) shown in red or green need not be evaluated. Green nodes have value +1 (from such states Player 1 cannot lose, regardless of future actions), whereas red nodes have value -1 (Player 1 cannot win). Exploiting these forced regions eliminates roughly half of the states from the backward recursion.

4.2. Properties of the optimal gain

Numerical solutions of (2) reveal two non-intuitive effects: (i) g_N^* can be *non-monotone* in N ; (ii) the optimal action is highly horizon-dependent. Figure 2 illustrates this phenomenon for strictly losing parameters $p_w = 0.43$, $p_d = 0$, $p_\ell = 0.57$ and $q_w = 0.06$, $q_d = 0.86$, $q_\ell = 0.08$, where the optimal gain is positive at $N^* = 4$ but not for all nearby horizons.

Identifying an optimal horizon N^* as a function of $(p_w, p_d, p_\ell, q_w, q_d, q_\ell)$ appears delicate because N^* can be large when both styles are only mildly losing.

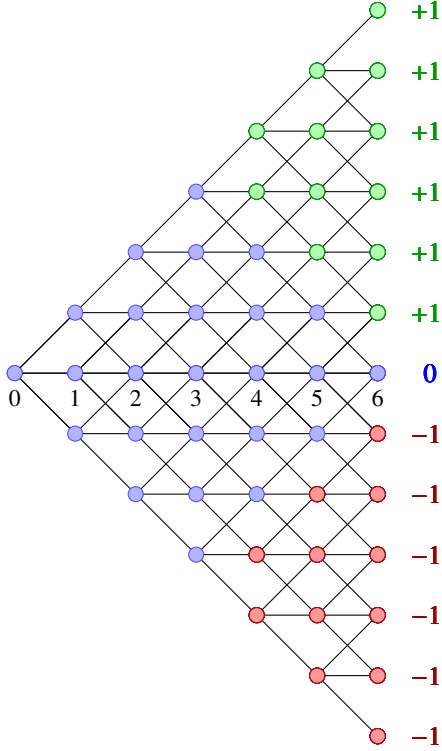


Figure 1: Backward computation of $g_N^* = V_0^N(0)$ for $N = 6$. Terminal values are $\text{sign}(X_6)$. Green (resp. red) nodes are states from which a match win (resp. loss) is already forced, hence their values are +1 (resp. -1).

For example, with $(p_w, p_d, p_\ell) = (0.49, 0, 0.51)$ and $(q_w, q_d, q_\ell) = (0.02, 0.95, 0.03)$, we obtain $N^* = 32$ and $g_{32}^* \approx 0.453$. It is not difficult to show that N^* can grow to infinity by making (D) safer and safer (see Section 4.3).

Therefore, deciding whether a weak player can achieve $g_N^* > 0$ for some horizon N is non-trivial. When draws are rare, odd horizons often yield a smaller gain than the preceding even horizon. The next proposition formalizes this no-draw disadvantage. Note that this is not true in general: when $q_d > 0$ and $p_d = 0$, it is possible to construct an example where $g_{2N+1}^* > g_{2N}^*$ (e.g., by taking $\mathbf{p} = (0.4, 0, 0.6)$ and $\mathbf{q} = (0.15, 0.7, 0.15)$).

Proposition 1. *If $p_d = q_d = 0$, then $g_{2N+1}^* \leq g_{2N}^*$.*

Proof. Let us consider a $(2N + 1)$ -game match. With $p_d = q_d = 0$, the score parity matches the time index: after $2N$ games, X_{2N} is even. If $X_{2N} \geq 2$ (resp. $X_{2N} \leq -2$), the final outcome is forced regardless of the last action, hence $V_{2N}^{2N+1}(X_{2N}) = +1$ (resp. -1). If $X_{2N} = 0$, the last game yields expected value $\max(p_w - p_\ell, q_w -$

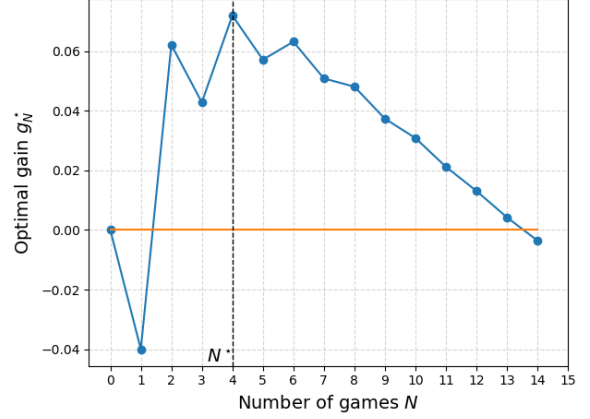


Figure 2: Optimal gain g_N^* under the strictly losing parameters $(p_w, p_d, p_\ell) = (0.43, 0, 0.57)$ and $(q_w, q_d, q_\ell) = (0.06, 0.84, 0.10)$.

$q_\ell) < 0$ under weakness, so $V_{2N}^{2N+1}(0) < 0$. In all cases,

$$V_{2N}^{2N+1}(x) \leq V_{2N}^{2N}(x) \quad \forall x \in \{-2N, \dots, 2N\},$$

since $V_{2N}^{2N}(x) = \text{sign}(x)$.

Let π^* be optimal for horizon $2N+1$. Then

$$\begin{aligned} g_{2N+1}^* &= \mathbb{E}_{\pi^*} \left[V_{2N+1}^{2N+1}(X_{2N+1}) \mid X_0 = 0 \right] \\ &= \mathbb{E}_{\pi^*} \left[V_{2N}^{2N+1}(X_{2N}) \mid X_0 = 0 \right] \\ &\leq \mathbb{E}_{\pi^*} \left[V_{2N}^{2N}(X_{2N}) \mid X_0 = 0 \right] \leq g_{2N}^*. \end{aligned}$$

This concludes the proof. \square

Let us recall that with no loss of generality we assume $p_d \leq q_d$ (draws are more likely when playing (D)). Note that this implies $q_w \leq p_\ell$ in the weak player case.

Let us consider two defensive strategies with different probabilities, denoted (D) and (D)' with \mathbf{q} and $\mathbf{q}' = (q'_w, q'_d, q'_\ell)$, respectively. We say that (D)' dominates (D) if the corresponding probabilities are stochastically ordered: $\mathbf{q} \leq_{st} \mathbf{q}'$; this simply means $q_w \leq q'_w$ and $q_\ell \geq q'_\ell$. It should be clear that the gain under (D)' is better than the gain under (D).

Lemma 1. *If $\mathbf{q} \leq_{st} \mathbf{q}'$ then $g_N^*(\mathbf{p}, \mathbf{q}) \leq g_N^*(\mathbf{p}, \mathbf{q}')$.*

Proof. The proof holds by backward induction on n . We show that for all x , the respective values satisfy $V_n^N(x) \leq V_n^{N'}(x)$, $V_n^N(x)$ and $V_n^{N'}(x)$ are non-decreasing in x . When $n = N$, $V_N^N(x) = V_N^{N'}(x)$ and they are both non-decreasing by definition. The Bellman equation at n for the original probabilities \mathbf{p}, \mathbf{q} is:

$$V_n^N(x) = \max \left\{ p_w V_{n+1}^N(x+1) + p_d V_{n+1}^N(x) + p_\ell V_{n+1}^N(x-1), \right. \\ \left. q_w V_{n+1}^N(x+1) + q_d V_{n+1}^N(x) + q_\ell V_{n+1}^N(x-1) \right\}$$

$$\begin{aligned}
&\leq \max \left\{ p_w V_{n+1}^N(x+1) + p_d V_{n+1}^N(x) + p_\ell V_{n+1}^N(x-1), \right. \\
&\quad \left. q'_w V_{n+1}^N(x+1) + q'_d V_{n+1}^N(x) + q'_\ell V_{n+1}^N(x-1) \right\} \\
&\leq \max \left\{ p_w V_{n+1}^{N'}(x+1) + p_d V_{n+1}^{N'}(x) + p_\ell V_{n+1}^{N'}(x-1), \right. \\
&\quad \left. q'_w V_{n+1}^{N'}(x+1) + q'_d V_{n+1}^{N'}(x) + q'_\ell V_{n+1}^{N'}(x-1) \right\} = V_n^{N'}(x).
\end{aligned}$$

The first inequality comes from $V_{n+1}^N(x)$ being non-decreasing in x by induction, so that $V_{n+1}^N(x-1) \leq V_{n+1}^N(x) \leq V_{n+1}^N(x+1)$. Using $\mathbf{q} \leq_{st} \mathbf{q}'$ implies the first inequality. The second inequality comes from induction $V_{n+1}^N(x-1) \leq V_{n+1}^{N'}(x-1)$, $V_{n+1}^N(x) \leq V_{n+1}^{N'}(x)$, $V_{n+1}^N(x+1) \leq V_{n+1}^{N'}(x+1)$. The fact that $V_n^N(x)$ and $V_n^{N'}(x)$ are non-decreasing in x comes from the monotonicity of the Bellman operator and because $V_n^N(x)$ and $V_n^{N'}(x)$ are non-decreasing in x . This concludes the induction. Finally, $g_N^*(\mathbf{p}, \mathbf{q}) = V_0^N(0) \leq V_0^{N'}(0) = g_N^*(\mathbf{p}, \mathbf{q}')$. \square

In the following, we use the notation $x \vee y := \max(x, y)$ and $x \wedge y := \min(x, y)$.

Proposition 2. *Consider a weak player. Then*

- (i) *If $q_\ell \geq p_w$, then for all $N \geq 1$, $g_N^* \leq 0$*
- (ii) *If $q_w = q_\ell$ ((D) is fair) then the gain is non-negative: $g_N^* \geq 0$. If in addition $q_\ell = q_w < p_w$ ((D) does not dominate (O)), then $g_N^* > 0$, for all $N \geq 2$.*

Proof. Proof of (i). Let us define (D)' with probabilities $q'_w := q_w \vee p_w$, $q'_\ell := q_\ell \wedge p_\ell$, $q'_d := 1 - q'_w - q'_\ell$. By definition, \mathbf{q}' dominates \mathbf{q} . Therefore, by using Lemma 1, $g_N^*(\mathbf{p}, \mathbf{q}) \leq g_N^*(\mathbf{p}, \mathbf{q}')$. By construction, \mathbf{q}' also dominates \mathbf{p} so $g_N^*(\mathbf{p}, \mathbf{q}') \leq g_N^*(\mathbf{q}', \mathbf{q}')$.

Now in the case where the offensive and defensive choices are the same, the gain is easier to analyse. The gain of (D)' in one game is equal to $g_1^{(D)'} = q'_w - q'_\ell = q_w \vee p_w - q_\ell \wedge p_\ell \leq 0$ because $q_\ell \geq p_w$ and $q_w \leq q_\ell$. For any $N \geq 1$, the probability $\mathbb{P}(X_N > 0)$ is given by Eq. (1) (replacing \mathbf{q} by \mathbf{q}'). This is always smaller than $1/2$, and goes to 0 as N goes to infinity. So the gain is negative and goes to -1 .

Proof of (ii). Since (D) is fair, playing (D) all the time will give a zero gain, hence $g_N^* \geq 0$. In addition, a case analysis shows that when $q_w = q_\ell < p_w$, the optimal policy uses (O) after game $N-1$ in state -1 so that $g_N^* > 0$ since $g_N^* = \mathbb{E}_{\pi^*}(V_{N-1}^N(X_{N-1}) | X_0 = 0)$ is strictly increasing in $V_{N-1}^N(-1)$ and since score -1 can be reached at step $N-1$ with positive probability. This last point also shows that in general, if no playing style dominates the other, then the gain of the optimal policy is strictly greater than the gain obtained by playing a fixed style. \square

The previous proposition has interesting consequences. In particular, if a draw is not possible ($p_d = p_\ell = 0$), then the gain is always negative for a strictly weak player. It can also be understood on the positive side: When the player is strictly weak, the only configuration where her gain has a chance to be positive is when $q_w < q_\ell \leq p_w < p_\ell$, although this is not sufficient to guarantee a positive gain.

4.3. Asymptotics

Unlike the finite-horizon case, the asymptotic behavior as $N \rightarrow \infty$ admits a complete classification under the weak-player assumption. If both (O) and (D) are strictly losing, the match is eventually lost with overwhelming probability and $g_N^* \rightarrow -1$. If (D) is fair but non-safe, then $g_N^* \geq 0$ for every N (by always playing (D)), yet $g_N^* \rightarrow 0$. Finally, if (D) is safe ($q_d = 1$), then g_N^* is non-decreasing in N and converges to the value induced by (essentially) *catenaccio*.

These results are formally given in the next theorem, which uses *catenaccio+* (denoted (cat+)), a slight refinement of (cat). *Catenaccio+* is defined as follows. It coincides with *catenaccio*, except in one case: if after $N-1$ games the score is zero, $X_{N-1} = 0$ (so the last game determines whether Player 1 wins or loses), then *catenaccio+* selects the better of (O) and (D). Specifically, it chooses (O) if $p_w - p_\ell > q_w - q_\ell$, and (D) otherwise. This differs from *catenaccio*, which plays (D) if the score has ever been positive, and (O) if the score has remained non-positive. Clearly, *catenaccio+* is (weakly) better than *catenaccio*.

Theorem 3. *Consider a weak player.*

- (i) *If $p_w < p_\ell$ and $q_w < q_\ell$, then $g_N^* \xrightarrow{N \rightarrow \infty} -1$.*
- (ii) *If $p_w < p_\ell$ and $q_w = q_\ell > 0$, then $\forall N \geq 1, g_N^* \geq 0$ and $g_N^* \xrightarrow{N \rightarrow \infty} 0$.*
- (iii) *If $p_w < p_\ell$ and $q_d = 1$, then $g_N^* = 0 \vee \max_{n \leq N} g_n^{(\text{cat}+)}$ and $g_N^* \xrightarrow{N \rightarrow \infty} 0 \vee \left(2 \frac{p_w}{p_\ell} - 1\right)$.*

Proof. (i). Let π be an arbitrary policy (defined for all $n \in \mathbb{N}$ and $x \in \mathbb{Z}$, with $\pi(x, n) \in \{(O), (D)\}$), and let $(X_n)_{n \geq 0}$ be the resulting score process with natural filtration $(\mathcal{F}_n)_{n \geq 0}$. Set $\delta := p_\ell - p_w \wedge q_\ell - q_w > 0$. By definition of the game, whatever action is taken at round n ,

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n - \delta.$$

Define $Y_n := X_n + \delta n$. Then (Y_n) is a supermartingale:

$$\mathbb{E}[Y_{n+1} | \mathcal{F}_n] \leq Y_n$$

with bounded increments as

$$|Y_{n+1} - Y_n| = |(X_{n+1} - X_n) + \delta| \leq 1 + \delta, \quad Y_0 = 0.$$

Therefore, Azuma–Hoeffding for supermartingales yields [7, Corollary 2.20]

$$\mathbb{P}(Y_n > h) \leq \exp\left(-\frac{h^2}{2n(1+\delta)^2}\right),$$

for any $h > 0$. Taking $h = (\delta n)^{2/3}$ and returning to X_n gives

$$\mathbb{P}(X_n > (\delta n)^{2/3} - \delta n) \leq \exp\left(-\frac{(\delta n)^{4/3}}{2(1+\delta)^2}\right). \quad (3)$$

Since $(\delta n)^{2/3} - \delta n \rightarrow -\infty$, it follows that $\mathbb{P}(X_n < 0) \rightarrow 1$, hence $g_n^\pi \rightarrow -1$ as $n \rightarrow \infty$.

(ii). Clearly $g_N^* \geq 0$ for all N , since the stationary policy that plays (D) at every step yields gain 0.

To prove $g_N^* \rightarrow 0$, fix an arbitrary policy π and let X_n be the induced score process. Here,

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n.$$

Thus (X_n) is a supermartingale with increments bounded by 1. However, Azuma–Hoeffding is not sufficient to conclude that $\mathbb{P}(X_n > 0)$ falls below $1/2$.

Let us refine the construction and let K_n be the number of times π uses (O) in the first n steps. Here, the idea is to decompose

$$X_n = \sum_{i=1}^{K_n} O_i + \sum_{j=1}^{n-K_n} D_j,$$

where $(O_i)_{i \in \mathbb{N}}$ are i.i.d. with values $+1$ w.p. p_w , 0 w.p. p_d , and -1 w.p. p_ℓ , and $(D_j)_{j \in \mathbb{N}}$ are i.i.d. with values $+1$ w.p. q_w , 0 w.p. q_d , and -1 w.p. $q_\ell = q_w$. Since π depends on (n, X_n) , K_n is a random variable depending on the realized samples $\{O_i, D_j : i, j \leq n\}$.

Consider the event $\mathcal{S} := \{(O_i)_{i \in \mathbb{N}}, (D_j)_{j \in \mathbb{N}} \text{ s.t. } K_n = o(\sqrt{n})\}$. First, let us assume first that \mathcal{S} holds. Since $-1 \leq O_i \leq 1$,

$$\sum_{j=1}^{n-K_n} D_j - K_n \leq X_n \leq \sum_{j=1}^{n-K_n} D_j + K_n. \quad (4)$$

Using also $-1 \leq D_j \leq 1$ yields

$$\sum_{j=1}^n D_j - 2K_n \leq X_n \leq \sum_{j=1}^n D_j + 2K_n. \quad (5)$$

By the central limit theorem,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n D_j \Rightarrow \mathcal{N}(0, q_w).$$

Dividing (5) by \sqrt{n} and using $K_n/\sqrt{n} \rightarrow 0$ on \mathcal{S} , we obtain

$$\mathbb{P}(X_n > 0) \rightarrow \frac{1}{2}, \quad \mathbb{P}(X_n < 0) \rightarrow \frac{1}{2},$$

hence $g_n^\pi \rightarrow 0$ on \mathcal{S} .

Assume now that \mathcal{S}^c holds. Then K_n is of order \sqrt{n} or larger, and $\sum_{i=1}^{K_n} O_i$ has negative drift. Using the same Azuma–Hoeffding argument as in (3) together with a union bound yields $\mathbb{P}(\sum_{i=1}^{K_n} O_i > 0) \rightarrow 0$. Indeed, for n large enough $K_n \geq \sqrt{n}$, and with $h = K_n^{2/3}$,

$$\begin{aligned} & \mathbb{P}\left(\sum_{i=1}^{K_n} O_i > K_n^{2/3} - (p_\ell - p_w)K_n\right) \\ & \leq \sum_{k=\sqrt{n}}^n \mathbb{P}\left(\sum_{i=1}^k O_i > k^{2/3} - (p_\ell - p_w)k\right) \\ & \leq \sum_{k=\sqrt{n}}^n \exp\left(-\frac{k^{4/3}}{2k}\right) \leq n \exp\left(-\frac{n^{1/6}}{2}\right). \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{K_n} O_i > 0\right) & \leq \mathbb{P}\left(\sum_{i=1}^{K_n} O_i > n^{1/3} - (p_\ell - p_w)\sqrt{n}\right) \\ & \leq \mathbb{P}\left(\sum_{i=1}^{K_n} O_i > K_n^{2/3} - (p_\ell - p_w)K_n\right) \\ & \leq n \exp\left(-\frac{n^{1/6}}{2}\right), \end{aligned}$$

which vanishes as $n \rightarrow \infty$. By symmetry, $\mathbb{P}(\sum_{j=1}^{n-K_n} D_j > 0) \leq 1/2$, hence for large n we have $\mathbb{P}(X_n > 0) < 1/2$ and therefore $g_n^\pi < 0$ on \mathcal{S}^c .

Combining the cases \mathcal{S} and \mathcal{S}^c gives

$$\limsup_{n \rightarrow \infty} g_n^\pi \leq 0$$

for every policy π . Since $g_N^* \geq 0$ for all N , it follows that $g_N^* \rightarrow 0$ as $N \rightarrow \infty$.

(iii). First, $g_{N+1}^* \geq g_N^*$ when $N \geq 1$ because one can use the following policy for $N+1$ steps: Use (D) at the first step and that use the optimal policy for the remaining N steps. Since (D) is safe, the first game does not change the score. The gain of this policy is the gain of the optimal policy for N games.

Now, let us prove that

$$g_N^* = 0 \vee \max_{n \leq N} g_n^{(\text{cat}+)}. \quad (6)$$

The inequality $g_N^* \geq 0 \vee \max_{n \leq N} g_n^{(\text{cat}+)}$ is immediate: the optimal policy dominates the stationary (D) policy (gain 0) and, for each $n \leq N$, the policy that plays (D) for the first $N - n$ rounds and then plays catenaccio+ for n rounds.

We now show $g_N^* \leq 0 \vee \max_{n \leq N} g_n^{(\text{cat}+)}$ by induction on N . For $N = 1$, catenaccio+ is optimal. Fix $N \geq 2$ and consider the score after the first round under an optimal policy. If the optimal policy plays (D) at step 1, then $X_1 = 0$ and the gain reduces to g_{N-1}^* , which is bounded by $0 \vee \max_{n \leq N-1} g_n^{(\text{cat}+)}$ by the induction hypothesis. If the optimal policy plays (O) at step 1, we distinguish the cases $X_1 \in \{-1, 0, 1\}$:

- If $X_1 = 1$, then playing (D) thereafter guarantees a match win; this coincides with catenaccio+, hence the conditional gain is bounded by $0 \vee \max_{n \leq N} g_n^{(\text{cat}+)}$: $(g_N^* | X_1 = 1) = +1 = (g_N^{(\text{cat}+)}) | X_1 = 1) \leq 0 \vee \max_{n \leq N} (g_n^{(\text{cat}+)}) | X_1 = 1)$
- If $X_1 = 0$, we reduce to horizon $N - 1$ and apply the induction hypothesis: $(g_N^* | X_1 = 0) = g_{N-1}^* \leq 0 \vee \max_{n \leq N-1} g_n^{(\text{cat}+)}$
- If $X_1 = -1$ then, there are two sub-cases: if score 0 is not reached, then the gain is $-1 \leq \max_{n \leq N} g_n^{(\text{cat}+)}$. If score 0 is reached at some step K , then we are back to the problem with $N - K$ games and induction applies. The gain $(g_N^* | X_1 = -1) = g_{N-K}^* \leq 0 \vee \max_{n \leq N-K} g_n^{(\text{cat}+)}$

Finally, (cat+) differs from (cat) only on the event $\{X_{N-1} = 0\}$, whose probability under either policy vanishes as $N \rightarrow \infty$; hence $\lim_{N \rightarrow \infty} g_N^{(\text{cat}+)} = \lim_{N \rightarrow \infty} g_N^{(\text{cat})}$. By Theorem 1, $\lim_{N \rightarrow \infty} g_N^{(\text{cat})} = 2p_w/p_\ell - 1$, and the coincidence of the (nonnegative) limits of g_N^* and $g_N^{(\text{cat}+)}$ follows directly from (6). \square

Theorem 3 yields a complete asymptotic classification for a weak player: the asymptotic gain is always in $\{-1, 0\}$ outside the safe regime, whereas safe defense yields a continuum of limits in $[0, 1]$ through the parameter p_w/p_ℓ . Point (iii) also formalizes that (cat)/(cat+) is essentially optimal when (D) is safe, but becomes asymptotically suboptimal as soon as (D) is merely fair and non-safe.

5. Conclusion

This paper identifies regimes in which an apparently weaker player can nevertheless attain a positive expected match outcome by adapting between two styles of play. The optimal advantage can be non-monotone in the number of games, and positivity may occur only at specific horizons. Finally, we provide a complete long-run classification for weak players.

Extending these insights to richer match structures, e.g., multiple outcome levels and multi-point scoring rules, is a natural direction for future work.

6. Second conclusion

To conclude on a sociological note, several behavioral studies suggest that the defeat of an apparently stronger player may generate frustration that can escalate into aggressive behavior. In such situations, the wiser strategy for the other player may be the one suggested by C-3PO to R2-D2 during a holographic chess game against Chewbacca: “Let the Wookiee win.”

Acknowledgments

This work has been partially supported by the MIAI cluster ANR-23-IACL-0006.

References

- [1] P. Arena, S. Fazzino, L. Fortuna, and P. Maniscalco. Game theory and non-linear dynamics: the parrondo paradox case study. *Chaos, Solitons & Fractals*, 17(2):545–555, 2003.
- [2] L. Dinis. Optimal sequence for Parrondo games. *Physical Review E*, 77(2):021124, 2008.
- [3] W. Feller. *An introduction to probability theory and its applications, Vol. 1*. John Wiley and Sons, 3rd edition edition, 1968.
- [4] G. Harmer and D. Abbott. Losing strategies can win by parrondo’s paradox. *Nature*, 402(864), 1999.
- [5] J. W. Lai and K. H. Cheong. Parrondo’s paradox from classical to quantum: A review. *Nonlinear Dynamics*, 100(1):849–861, 2020.
- [6] J. M. R. Parrondo, G. P. Harmer, and D. Abbott. New paradoxical games based on brownian ratchets. *Physical Review Letters*, 85(24):5226–5229, 2000.
- [7] M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.