# THE EXACT MINIMAX REGRET OF THE ADVERSARIAL TWO-ARMED BANDIT AT SMALL HORIZONS, WITH AND WITHOUT A SWITCHING COST

KOREA SUPERINTELLIGENCE LABS

Abstract. We compute exactly the finite-horizon minimax expected regret of the two-armed adversarial bandit with $\{0,1\}$ losses, an oblivious adversary, bandit feedback and regret measured against the best fixed arm: $V^{\mathrm{B}}(2,T) = \frac{1}{2}, \frac{1}{2}, \frac{3}{4}, \frac{33}{40}$ for $T = 1,2,3,4$, and $\frac{15}{16} \le V^{\mathrm{B}}(2,5) \le \frac{53}{40}$. Under full information the same game has Cover's value $\lceil T/2 \rceil \binom{T}{\lfloor T/2 \rfloor} 2^{-T}$, so bandit feedback is free for $T \le 3$ and first costs at $T = 4$, where the exact price of the missing observation is $\frac{33}{40} - \frac{3}{4} = \frac{3}{40}$. We then add the unit switching cost of Dekel, Ding, Koren and Peres and obtain $V^{\mathrm{sw}}(2,T) = \frac{3}{2}, 2, \frac{7}{3}, \frac{31}{12}, \frac{167}{60}$ for $T \le 5$, which is $H_T + \frac{1}{2}$ at each of these five horizons, and the full-information values $\frac{3}{2}, 2, \frac{9}{4}$ for $T \le 3$: with a switching cost, bandit feedback already costs at $T = 3$, one horizon earlier than without it. The harmonic pattern is a fit and not a law: an explicit twelve-point adversary shows $V^{\mathrm{sw}}(2,7) \ge \frac{373}{120} > H_7 + \frac{1}{2}$. Every value is an attained minimum over all behaviour strategies with probabilities in an arbitrary linearly ordered field, certified on both sides by explicit finite data; the certificate calculus and all finite checks are verified in Lean 4.

## 1. Introduction

The adversarial (non-stochastic) multi-armed bandit of Auer, Cesa-Bianchi, Freund and Schapire [2] is a central object of online learning, and its minimax regret is known up to constant factors: $\Theta(\sqrt{TK})$ for $K$ arms and $T$ rounds [2, 3, 7, 13]. What is not known, for any $T$ and any $K$, is its *value*. Abernethy and Warmuth, posing the open problem *Minimax Games with Bandits* at COLT 2009 [1], wrote that many online decision games "have been analyzed completely, with efficient solutions in each case. Yet, to our knowledge, no such minimax solution has been given for any decision game in the bandit setting." Among the questions they list is the stopping rule: "In the hedge setting we used a bound $k$ on the loss of the best action, and it may be reasonable to keep this restriction. Yet perhaps a bound on the time of the game would give an easier analysis." This paper takes that route in its smallest instance: two arms, $\{0,1\}$ losses, an oblivious adversary, regret against the best fixed arm, and a fixed horizon $T \le 4$ (with a bracket at $T = 5$). We solve this fixed-horizon variant at small horizons; we do not answer the open problem, whose stopping rule and action count are different.

The full-information side of the same game is classical: for two experts and binary outcomes Cover [8] determined the exact finite-horizon value in 1965 (see [12, 10], and the discussion after Proposition 4.3 for what his paper prints); in our normalisation it is $V^{\mathrm{F}}(2,T) = \lceil T/2 \rceil \binom{T}{\lfloor T/2 \rfloor} 2^{-T}$. A 2026 remark of Banerjee, Deo-Campo Vuong and Kleinberg on "the rarity of exact minimax solutions in online algorithms – results where an algorithm's performance perfectly matches the theoretical value of the game on input sequences of bounded length" lists prediction with two or three experts, unconstrained linear optimisation, ski rental and paging [6]; neither bandit feedback nor switching costs appear in that list, and we found no printed finite value for either game studied here (Section 6).

**Results.** Write $V^{\mathrm{B}}(2,T)$ and $V^{\mathrm{F}}(2,T)$ for the minimax expected regret of the two-armed $\{0,1\}$-loss game at horizon $T$ under bandit feedback and under full information (Section 2 defines both exactly). Our first results are the entries of Table 1.

So bandit feedback is free below horizon 4 and first costs at $T = 4$, by exactly $\frac{3}{40}$ (Theorems 4.1 and 4.4). The rate literature cannot make this statement: comparing $\sqrt{TK \log K}$ with $\sqrt{T \log K}$ says that bandit feedback eventually costs a factor of order $\sqrt{K}$, and nothing about where the cost begins. At $T = 4$, $K = 2$ the constants printed by Audibert and Bubeck [3] for rewards in $[0,1]$ (a game that contains ours) leave the value between $\frac{1}{20}\sqrt{TK} = 0.141$ — the lower bound they state over $K$-tuples of distributions on $[0,1]$, which they note bounds the adversarial case as well — and $100\sqrt{TK} = 282.8$, their Theorem 4 at $q = 3$, whose supremum is over all adversarial environments. The exact value is 0.825.

| $T$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $V^{\mathrm{B}}(2,T)$ | 1/2 | 1/2 | 3/4 | **33/40** | [15/16, 53/40] |
| $V^{\mathrm{F}}(2,T)$ | 1/2 | 1/2 | 3/4 | 3/4 | 15/16 |

TABLE 1. The plain game. Every entry except the bracket is an attained minimum, certified on both sides.

The second group of results concerns the same game with the *unit switching cost* of Dekel, Ding, Koren and Peres [9]: the learner pays 1 on every round on which its arm differs from that of the previous round, the first round always counting as a switch. They prove an $\widetilde{\Omega}(k^{1/3}T^{2/3})$ lower bound which, combined with the $O(k^{1/3}T^{2/3})$ upper bound they attribute to Arora, Dekel and Tewari [4], makes the minimax regret of that game $\widetilde{\Theta}(k^{1/3}T^{2/3})$ under bandit feedback against $\Theta(\sqrt{T\log k})$ under full information — "the first example that exhibits (even for constant $k$) a clear gap between the asymptotic difficulty, as $T$ grows, of online learning with bandit and full feedback." Table 2 gives the exact values at small horizons.

| $T$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $V^{\mathrm{sw}}(2,T)$ (bandit) | 3/2 | 2 | 7/3 | 31/12 | 167/60 | $\geq 59/20$ | $\geq 373/120$ |
| $V^{\mathrm{swF}}(2,T)$ (full info.) | 3/2 | 2 | **9/4** | – | – | – | – |
| $H_T + \frac{1}{2}$ | 3/2 | 2 | 7/3 | 31/12 | 167/60 | 59/20 | 433/140 |

TABLE 2. The game with a unit switching cost; $H_T = \sum_{j\leq T} 1/j$. The entries at $T = 6, 7$ are one-sided lower bounds, and $433/140 < 373/120$.

Three things are visible in the table and proved below. The five values $V^{\mathrm{sw}}(2,T)$, $T \leq 5$, are exact and equal the harmonic numbers shifted by $\frac{1}{2}$ (Theorem 5.2). With a switching cost, bandit feedback already costs at $T = 3$: $V^{\mathrm{swF}}(2,3) = \frac{9}{4} < \frac{7}{3} = V^{\mathrm{sw}}(2,3)$, a price of $\frac{1}{12}$, while the two feedback models agree at $T \leq 2$ (Theorem 5.4); the onset of the bandit-versus-full-information gap thus moves from $T = 4$ to $T = 3$, the finite-horizon shadow of the asymptotic gap of [9]. And the harmonic pattern does not continue: a twelve-point adversary distribution guarantees $\frac{373}{120} > H_7 + \frac{1}{2}$ against every horizon-7 learner (Theorem 5.6). The same construction gives exactly $H_6 + \frac{1}{2}$ at $T = 6$, so we do not claim that $T = 7$ is the first failure.

Each value is pinned by a *two-sided certificate*: an explicit behaviour strategy whose worst case over all $4^T$ loss matrices is the value, and an explicit finitely supported adversary distribution against which no behaviour strategy whatsoever does better. The second half rests on a backward-induction lemma (Section 3) valid for every behaviour strategy with probabilities in any linearly ordered commutative ring, so the values are attained minima over an infinite strategy space, not outputs of a search. The certificates were found by linear programming over the learner's sequence form; how they were found plays no role in their correctness. Every theorem, proposition and lemma below has been checked in Lean 4 with Mathlib (Section 7).

**What is not claimed.** "First" in "bandit feedback first costs at $T = 4$" is relative to the conventions of Section 2: two arms, $\{0,1\}$ losses, an oblivious adversary, regret against the best fixed arm, a fixed horizon, and, for the switching-cost game, the opening convention of [9]. Lattimore and Szepesvári state as an exercise that, without switching costs, the minimax regret over $\{0,1\}$ losses equals that over $[0,1]$ losses ([13], Exercise 36.13, "Binary is the worst case"); we cite that exercise and neither prove nor use it, and with a switching cost we claim only that the binary value is at most the $[0,1]$ value. The inequalities $V^{\mathrm{F}} \leq V^{\mathrm{B}}$ and $V^{\mathrm{swF}} \leq V^{\mathrm{sw}}$ are true but are not used: both values in every comparison are pinned independently. Finally, the stochastic symmetric two-armed Bernoulli bandit, whose finite-horizon minimax values go back to Fabius and van Zwet [11], is a different game (a parametric family of reward distributions and pseudo-regret, not $4^T$ adversarial loss matrices), and its values do not compare with ours.

## 2. The games

2.1. **Loss matrices, learners, and the two feedback models.** There are two arms, $a \in \{0,1\}$, and a horizon $T \in \mathbb{N}$. An oblivious adversary fixes, before the first round, a *loss matrix* $\ell = (\ell_1, \ldots, \ell_T)$ with

$\ell_t = (\ell_t(0), \ell_t(1)) \in \{0,1\}^2$; $\mathcal{L}_T$ is the set of all $4^T$ loss matrices, and $\mathrm{m}(\ell) = \min_a \sum_t \ell_t(a)$ is the loss of the best fixed arm in hindsight. The adversary may randomise over $\mathcal{L}_T$; since the expected regret of a fixed learner is affine in the adversary's distribution, the supremum over distributions is the maximum over the $4^T$ deterministic matrices, and only the latter appears in the definitions (the certificate lemma of Section 3 covers arbitrary finite mixtures).

At round $t$ the learner plays $I_t \in \{0,1\}$ at random, as a function of what it has observed and of private randomness. Under *bandit feedback* it observes $\ell_t(I_t)$ and nothing else; under *full information* it observes the pair $\ell_t$. The learner has perfect recall, so by Kuhn's theorem nothing is lost by describing it through a *behaviour strategy*, one probability per observed history. Fix a commutative ring $\mathbb{K}$ with a linear order making it a strictly ordered ring (in the theorems $\mathbb{K}$ is a linearly ordered field, $\mathbb{R}$ in particular). A behaviour strategy of depth 0 is a leaf; one of depth $T+1$ is a node $(p;\, c_{00}, c_{01}, c_{10}, c_{11})$ with $p \in \mathbb{K}$, $0 \le p \le 1$, and four behaviour strategies of depth $T$ as children: $p$ is the probability of playing arm 0 now, and $c_{ab}$ the continuation after playing arm $a$ and observing the bit $b$. Let $\mathcal{S}_T(\mathbb{K})$ be the set of behaviour strategies of depth $T$. The same trees serve both feedback models; only the rule selecting the child differs. For $\ell = (\ell_1, \ell')$ and $S = (p;\, c_{00}, c_{01}, c_{10}, c_{11}) \in \mathcal{S}_{T+1}(\mathbb{K})$ the expected losses are

$$C^{\mathrm{B}}_\ell(S) = p\big(\ell_1(0) + C^{\mathrm{B}}_{\ell'}(c_{0,\ell_1(0)})\big) + (1-p)\big(\ell_1(1) + C^{\mathrm{B}}_{\ell'}(c_{1,\ell_1(1)})\big),$$
$$C^{\mathrm{F}}_\ell(S) = p\,\ell_1(0) + (1-p)\,\ell_1(1) + C^{\mathrm{F}}_{\ell'}(c_{\ell_1(0),\ell_1(1)}),$$

with $C^{\mathrm{B}}_{()} = C^{\mathrm{F}}_{()} = 0$ at the empty matrix. Under bandit feedback the continuation depends on the arm played and the bit observed; under full information on the observed pair alone. The *regrets* are $R^{\mathrm{B}}_\ell(S) = C^{\mathrm{B}}_\ell(S) - \mathrm{m}(\ell)$ and $R^{\mathrm{F}}_\ell(S) = C^{\mathrm{F}}_\ell(S) - \mathrm{m}(\ell)$.

2.2. **The minimax value as an attained minimum.** Let

$$\mathcal{W}^{\mathrm{B}}_T(\mathbb{K}) = \big\{ v \in \mathbb{K} :\ \exists S \in \mathcal{S}_T(\mathbb{K})\ \ \forall \ell \in \mathcal{L}_T,\ R^{\mathrm{B}}_\ell(S) \le v \big\}$$

be the set of worst-case regrets a horizon-$T$ bandit learner can guarantee, and $\mathcal{W}^{\mathrm{F}}_T(\mathbb{K})$ the same with $R^{\mathrm{F}}$. The *minimax value* $V^{\mathrm{B}}(2,T)$ is the least element of $\mathcal{W}^{\mathrm{B}}_T(\mathbb{K})$; each theorem below asserts that this least element exists (the infimum is attained) and names it, for an arbitrary linearly ordered field $\mathbb{K}$.

*Remark* 2.1 (Why the full-information learner ignores its own past actions). The trees above let a full-information learner condition on the loss history only. That is without loss for this payoff: the expected loss $\sum_t \mathbb{E}[\ell_t(I_t)]$ depends only on the per-round marginals $\Pr(I_t = 0 \mid \ell)$, and any strategy that also looks at its own past actions induces such marginals. Under bandit feedback the same reduction is unavailable, because the learner's own action determines what it observes; that asymmetry is what Theorem 4.4 measures. The reduction is an argument in prose; the formal theorems are about the trees as defined.

2.3. **The switching cost.** Dekel, Ding, Koren and Peres [9] define the regret of a player who pays a unit cost for each change of action by their equation (1),

$$R = \sum_{t=1}^{T} \big(\ell_t(X_t) + \mathbf{1}_{X_t \ne X_{t-1}}\big) - \min_{x \in [k]} \sum_{t=1}^{T} \ell_t(x),$$

and fix the opening convention on the same page: "To make the loss on the first round well-defined, we set $X_0 = 0$ (so the first action always counts as a switch)." Their $X_0 = 0$ lies outside the action set $[k] = \{1, \dots, k\}$; we reproduce exactly that convention. Let $x \in \{\varnothing, 0, 1\}$ be the previously played arm, $\varnothing$ before round 1, and $\sigma(x,a) = 0$ if $x = a$, $\sigma(x,a) = 1$ otherwise, so that $\sigma(\varnothing, a) = 1$ for both arms. For $S = (p;\, c_{00}, c_{01}, c_{10}, c_{11})$ and $\ell = (\ell_1, \ell')$ put

$$C^{\mathrm{sw}}_\ell(S; x) = p\big(\ell_1(0) + \sigma(x,0) + C^{\mathrm{sw}}_{\ell'}(c_{0,\ell_1(0)}; 0)\big) + (1-p)\big(\ell_1(1) + \sigma(x,1) + C^{\mathrm{sw}}_{\ell'}(c_{1,\ell_1(1)}; 1)\big),$$

with $C^{\mathrm{sw}}_{()} = 0$, and $R^{\mathrm{sw}}_\ell(S) = C^{\mathrm{sw}}_\ell(S; \varnothing) - \mathrm{m}(\ell)$: this is equation (1) with the expectation over the learner's randomness taken. The strategy trees are unchanged – under bandit feedback the switching cost changes the payoff, not the information structure – and $\mathcal{W}^{\mathrm{sw}}_T(\mathbb{K})$, $V^{\mathrm{sw}}(2,T)$ are defined from $R^{\mathrm{sw}}$ as before.

Under full information the switching cost *does* change the information structure: the reduction of Remark 2.1 fails, because the switching term contributes $\Pr(I_t \ne I_{t-1} \mid \ell)$, a function of the joint law of consecutive actions and not of the two marginals. A full-information learner that pays for switching must

therefore be allowed to condition on its own past actions. Its behaviour strategies are the *8-ary* trees $\mathcal{S}_T^8(\mathbb{K})$: a node is $(p;\,(c_{a,r})_{a\in\{0,1\},\,r\in\{0,1\}^2})$ with $0\le p\le 1$, $c_{a,r}$ being the continuation after playing $a$ and observing the pair $r$, and

$$\begin{aligned} C_\ell^{\mathrm{swF}}(S;x) = p\big(\ell_1(0)+\sigma(x,0)+C_{\ell'}^{\mathrm{swF}}(c_{0,\ell_1};0)\big) \\ +(1-p)\big(\ell_1(1)+\sigma(x,1)+C_{\ell'}^{\mathrm{swF}}(c_{1,\ell_1};1)\big), \end{aligned}$$

$R_\ell^{\mathrm{swF}}(S)=C_\ell^{\mathrm{swF}}(S;\varnothing)-\mathrm{m}(\ell)$, with $\mathcal{W}_T^{\mathrm{swF}}(\mathbb{K})$ and $V^{\mathrm{swF}}(2,T)$ as before. Using the 4-ary full-information trees here would understate the full-information value and could manufacture a separation; the 8-ary class is the one over which Theorem 5.4 quantifies.

## 3. Certificates

3.1. **Backward induction against a mixture.** A *mixture* of horizon $n$ is a finite list $M=((w_i,\ell^i))_{i=1}^s$ with weights $w_i\in\mathbb{K}$, $w_i\ge 0$, and $\ell^i\in\mathcal{L}_n$ (repetitions allowed). Write $\mathrm{tot}(M)=\sum_i w_i$ and $\mu_a(M)=\sum_i w_i\,\ell_1^i(a)$, the weight the mixture puts on "arm $a$ loses in the current round". For $a,b\in\{0,1\}$ let $M_{a,b}$ be the sub-mixture of those $(w_i,\ell^i)$ with $\ell_1^i(a)=b$, with the first round of each matrix removed; for $r\in\{0,1\}^2$ let $M_r$ be the sub-mixture of those with $\ell_1^i=r$, first round removed. Define $\beta_0^{\mathrm{B}}=\beta_0^{\mathrm{F}}=0$, $\beta_0^{\mathrm{sw}}(M;x)=\beta_0^{\mathrm{swF}}(M;x)=0$ and

$$\begin{aligned} \beta_{n+1}^{\mathrm{B}}(M) &= \min_{a\in\{0,1\}}\Big(\mu_a(M)+\beta_n^{\mathrm{B}}(M_{a,1})+\beta_n^{\mathrm{B}}(M_{a,0})\Big), \\ \beta_{n+1}^{\mathrm{F}}(M) &= \min_{a\in\{0,1\}}\mu_a(M)+\sum_{r\in\{0,1\}^2}\beta_n^{\mathrm{F}}(M_r), \\ \beta_{n+1}^{\mathrm{sw}}(M;x) &= \min_a\Big(\sigma(x,a)\,\mathrm{tot}(M)+\mu_a(M)+\beta_n^{\mathrm{sw}}(M_{a,1};a)+\beta_n^{\mathrm{sw}}(M_{a,0};a)\Big), \\ \beta_{n+1}^{\mathrm{swF}}(M;x) &= \min_a\Big(\sigma(x,a)\,\mathrm{tot}(M)+\mu_a(M)+\sum_{r\in\{0,1\}^2}\beta_n^{\mathrm{swF}}(M_r;a)\Big). \end{aligned}$$

With a switching cost the recursion carries the previous arm and each branch gains one additive term, the switch penalty times the weight of the sub-mixture still alive.

**Lemma 3.1** (Backward induction bounds every behaviour strategy)**.** *Let $\mathbb{K}$ be a commutative ring with a linear order making it a strictly ordered ring, let $M=((w_i,\ell^i))$ be a mixture of horizon $n$ with all $w_i\ge 0$, and let $S\in\mathcal{S}_n(\mathbb{K})$. Then*

$$\beta_n^{\mathrm{B}}(M)\le\sum_i w_i\,C_{\ell^i}^{\mathrm{B}}(S)\qquad\text{and}\qquad\beta_n^{\mathrm{F}}(M)\le\sum_i w_i\,C_{\ell^i}^{\mathrm{F}}(S).$$

**Lemma 3.2** (The same with a switching cost)**.** *Under the hypotheses of Lemma 3.1, for every $x\in\{\varnothing,0,1\}$,*

$$\begin{aligned} \beta_n^{\mathrm{sw}}(M;x)&\le\sum_i w_i\,C_{\ell^i}^{\mathrm{sw}}(S;x)\qquad(S\in\mathcal{S}_n(\mathbb{K})), \\ \beta_n^{\mathrm{swF}}(M;x)&\le\sum_i w_i\,C_{\ell^i}^{\mathrm{swF}}(S;x)\qquad(S\in\mathcal{S}_n^8(\mathbb{K})). \end{aligned}$$

*Proof of both lemmas.* Induction on $n$; at $n=0$ both sides vanish. At a node $S=(p;\,c_{ab})$ the weighted cost splits by the first round:

$$\sum_i w_i\,C_{\ell^i}^{\mathrm{B}}(S)=p\,A_0+(1-p)\,A_1,\qquad A_a=\mu_a(M)+\sum_{b\in\{0,1\}}\ \sum_{i:\ \ell_1^i(a)=b} w_i\,C_{\ell^{i\prime}}^{\mathrm{B}}(c_{ab}),$$

where $\ell^{i\prime}$ is $\ell^i$ without its first round. Since $0\le p\le 1$, $pA_0+(1-p)A_1\ge\min(A_0,A_1)$, and the induction hypothesis on the four sub-mixtures $M_{a,b}$ gives $A_a\ge\mu_a(M)+\beta_n^{\mathrm{B}}(M_{a,1})+\beta_n^{\mathrm{B}}(M_{a,0})$. For full information the continuation does not depend on $a$ and the sum over the four $M_r$ factors out of the minimum. With a switching cost $A_a$ gains the term $\sigma(x,a)\,\mathrm{tot}(M)$, the same for every $i$, and the continuations are evaluated with previous arm $a$. No finiteness assumption on $S$ is made and no computation is involved. □

3.2. **Two-sided certificates.** For a mixture $M$ of horizon $n$ set

$$\mathrm{val}^{\mathrm{B}}(M) = \beta_n^{\mathrm{B}}(M) - \sum_i w_i \, \mathrm{m}(\ell^i),$$

the regret the mixture guarantees itself against every bandit learner, and likewise $\mathrm{val}^{\mathrm{F}}$, $\mathrm{val}^{\mathrm{sw}}$ (with $x = \varnothing$) and $\mathrm{val}^{\mathrm{swF}}$.

**Proposition 3.3** (A rational certificate pins the value over every ordered field)**.** *Let $n \geq 0$.*

(1) *If $M$ is a mixture of horizon $n$ with nonnegative weights summing to 1 and $S \in \mathcal{S}_n(\mathbb{K})$ satisfies $R^{\mathrm{B}}_{\ell^i}(S) \leq w$ for every $i$, then $\mathrm{val}^{\mathrm{B}}(M) \leq w$; the same holds for* F, sw *and* swF *(with $S \in \mathcal{S}_n^8(\mathbb{K})$ in the last case).*

(2) *Let $v \in \mathbb{Q}$. Suppose $S \in \mathcal{S}_n(\mathbb{Q})$ satisfies $R^{\mathrm{B}}_{\ell}(S) \leq v$ for all $\ell \in \mathcal{L}_n$, and $M$ is a mixture of horizon $n$ with rational weights, nonnegative and summing to 1, every $\ell^i \in \mathcal{L}_n$, and $\mathrm{val}^{\mathrm{B}}(M) = v$. Then for every linearly ordered field $\mathbb{K}$, $v$ is the least element of $\mathcal{W}_n^{\mathrm{B}}(\mathbb{K})$. The same holds for* F, sw *and* swF.

*Proof.* (1) Sum $R^{\mathrm{B}}_{\ell^i}(S) \leq w$ with weights $w_i$ and apply Lemma 3.1. (2) The embedding $\mathbb{Q} \to \mathbb{K}$ commutes with every recursion above and preserves $0 \leq p \leq 1$, so the image of $S$ shows $v \in \mathcal{W}_n^{\mathrm{B}}(\mathbb{K})$, and for $w \in \mathcal{W}_n^{\mathrm{B}}(\mathbb{K})$ witnessed by $S' \in \mathcal{S}_n(\mathbb{K})$, part (1) for the image of $M$ and $S'$ gives $v \leq w$. The other cases use Lemma 3.2. □

So a value is established by a rational learner tree, a rational adversary mixture, and four finite computations: the tree is complete of the right depth with all probabilities in $[0, 1]$; its regret against each of the $4^n$ matrices is at most $v$; the mixture is a probability distribution on $\mathcal{L}_n$; and its backward-induction value is exactly $v$. These four checks are the only computational content of the results below.

3.3. **How the certificates were found.** The learner's strategy space is a treeplex, so the game is a linear program in sequence form: one variable per learner sequence, one flow equality per information set, one inequality per adversary matrix – 170 variables, 85 equalities and up to 256 inequalities at $T = 4$; 682, 341 and up to 1024 at $T = 5$. Adversary rows were generated on demand (solve the restricted program, evaluate the plan against all $4^T$ matrices, add the most violated rows, repeat). For the plain game the simplex ran in exact rational arithmetic and the $T = 4$ bandit program converged in six rounds. For the switching-cost game a floating-point simplex supplied a warm start and the small restricted program was re-solved exactly; at $T = 5$ the learner's 341 probabilities were obtained by rounding the floating-point solution to denominators at most 360, and their exact worst case over the 1024 matrices is exactly $\frac{167}{60}$. Every certificate was re-verified by independently written code along a different path (forward evaluation of the tree against every matrix, backward induction on the mixture); at $T = 2$ the switching-cost value was also recomputed in normal form over all $2^5 = 32$ deterministic bandit learners and 16 matrices, as a check of the sequence-form encoding and of Kuhn's theorem in this instance. The full-information column of Table 1 was also reproduced by an unrelated dynamic program on the pair (rounds left, loss difference), which matches Cover's closed form for every $T \leq 14$ and supplies the $T = 5$ full-information certificate. The $T = 1$ switching-cost value is a hand computation: the learner plays arm 0 with probability $p$ and pays 1 for the opening switch, its regret is $1 + p$ against $(1, 0)$ and $2 - p$ against $(0, 1)$, and $\max(1 + p, 2 - p)$ is minimised at $p = \frac{1}{2}$ with value $\frac{3}{2}$.

## 4. The plain game

**Theorem 4.1** ($V^{\mathrm{B}}(2, 4) = 33/40$)**.** *For every linearly ordered field $\mathbb{K}$, $\frac{33}{40}$ is the least element of $\mathcal{W}_4^{\mathrm{B}}(\mathbb{K})$: some horizon-4 behaviour strategy has regret at most $\frac{33}{40}$ against every one of the 256 loss matrices, and no behaviour strategy whatsoever has worst-case regret below $\frac{33}{40}$.*

*Certificate.* The learner is a complete 4-ary tree of depth 4 with 85 node probabilities; the least common multiple of their denominators is 150480 (among them $\frac{3}{19}$, $\frac{10}{19}$, $\frac{8}{11}$, $\frac{3}{16}$, $\frac{13}{16}$), and its regret against each of the 256 matrices is at most $\frac{33}{40}$, with equality on some matrix. The adversary is a distribution of support 20 on $\mathcal{L}_4$, with weights of denominator dividing 80, whose value $\mathrm{val}^{\mathrm{B}}$ is exactly $\frac{33}{40}$. Both checks are exhaustive finite computations; Proposition 3.3 does the rest. □

**Proposition 4.2** (Horizons 1, 2, 3)**.** *For every linearly ordered field $\mathbb{K}$, the least elements of $\mathcal{W}_1^{\mathrm{B}}(\mathbb{K})$, $\mathcal{W}_2^{\mathrm{B}}(\mathbb{K})$, $\mathcal{W}_3^{\mathrm{B}}(\mathbb{K})$ are $\frac{1}{2}$, $\frac{1}{2}$, $\frac{3}{4}$, and the least elements of $\mathcal{W}_T^{\mathrm{F}}(\mathbb{R})$ for $T = 1, 2, 3$ are the same three numbers.*

*Certificate.* Learner trees of 1, 5 and 21 probabilities and adversary distributions of support 2, 2 and 6, in each feedback model. At $T = 1$ the learner flips a fair coin and the adversary plays $(1, 0)$ or $(0, 1)$ with probability $\frac{1}{2}$ each. The non-obvious half at $T = 3$ is a bandit learner matching the full-information value $\frac{3}{4}$. □

**Proposition 4.3** (Cover's column, re-derived)**.** *For every linearly ordered field $\mathbb{K}$ and $T = 1, \dots, 5$, the least element of $\mathcal{W}_T^{\mathrm{F}}(\mathbb{K})$ is $\frac{1}{2}, \frac{1}{2}, \frac{3}{4}, \frac{3}{4}, \frac{15}{16}$ respectively, that is, $\lceil T/2 \rceil \binom{T}{\lfloor T/2 \rfloor} 2^{-T}$.*

This is Cover's 1965 value for two experts [8], in the form Gravin, Peres and Sivan describe as "exactly half the expected distance travelled by a simple random walk in $T$ steps" [12]; the binomial expression is an elementary rewriting of that description and is not claimed as new. The proposition is the control column against which Theorem 4.4 compares, re-derived from the definition of the game and certified two-sidedly at each of the five horizons (adversary supports $2, 2, 6, 6, 20$; learner trees of $1, 5, 21, 85, 341$ probabilities).

What Cover's paper prints, read from the offprint of pp. 263–272, is a characterisation of the scores achievable by a sequential predictor of a binary sequence of length $T$: a score $\hat{s}(k/T)$ depending only on the number $k$ of ones is achievable exactly when $2^{-T} \sum_k \binom{T}{k} \hat{s}(k/T) = \frac{1}{2}$ together with a Lipschitz condition (his VII, p. 269), and one satisfying the second condition but not the first is achievable uniformly with the error $\varepsilon_T = \frac{1}{2} - 2^{-T} \sum_k \binom{T}{k} \hat{s}(k/T)$, which he displays on p. 270 for the envelope $\hat{s}(\eta) = \max\{\eta, 1 - \eta\}$, the score of the better of the two constant predictors. Since a score is $1 -$ (error rate), $-T\varepsilon_T$ is the minimax number of extra mistakes; it agrees with $\lceil T/2 \rceil \binom{T}{\lfloor T/2 \rfloor} 2^{-T}$ at every horizon $T \le 14$ we evaluated. Cover's adversary plays only rounds on which exactly one of the two predictors errs, whereas $\mathcal{L}_T$ also contains $(0, 0)$ and $(1, 1)$; Proposition 4.3 is proved over the larger $\mathcal{L}_T$, and the values coincide.

**Theorem 4.4** (Bandit feedback first costs at $T = 4$)**.** *For $T = 1, 2, 3$ the least elements of $\mathcal{W}_T^{\mathrm{B}}(\mathbb{R})$ and $\mathcal{W}_T^{\mathrm{F}}(\mathbb{R})$ coincide ($\frac{1}{2}, \frac{1}{2}, \frac{3}{4}$). At $T = 4$ the least element of $\mathcal{W}_4^{\mathrm{B}}(\mathbb{R})$ is $\frac{33}{40}$, that of $\mathcal{W}_4^{\mathrm{F}}(\mathbb{R})$ is $\frac{3}{4}$, $\frac{3}{4} < \frac{33}{40}$, and $\frac{33}{40} - \frac{3}{4} = \frac{3}{40}$.*

*Proof.* Theorem 4.1, Propositions 4.2 and 4.3, and arithmetic. □

**Proposition 4.5** (Horizon 5, bracketed)**.** *$\frac{53}{40} \in \mathcal{W}_5^{\mathrm{B}}(\mathbb{R})$, and every $v \in \mathcal{W}_5^{\mathrm{B}}(\mathbb{R})$ satisfies $v \ge \frac{15}{16}$; in particular every horizon-5 bandit strategy has worst-case regret strictly above $V^{\mathrm{B}}(2, 4) = \frac{33}{40}$.*

*Certificate.* The upper end is the learner that plays the horizon-4 optimum of Theorem 4.1 for four rounds and then a fair coin, evaluated against all 1024 matrices. The lower end is the optimal full-information adversary of Proposition 4.3 at $T = 5$ (support 20), read under bandit feedback: its bandit backward-induction value is at least $\frac{15}{16}$, and Lemma 3.1 applies. The exact value of $V^{\mathrm{B}}(2, 5)$ is not determined here (Section 8); two other cheap adversary mixtures tried for the lower end also gave exactly $\frac{15}{16}$. □

*Remark* 4.6 (Controls). The formal development also records that $\frac{3}{4}$ and $\frac{5}{6}$ are *not* least elements of $\mathcal{W}_4^{\mathrm{B}}(\mathbb{R})$, that $\frac{1}{2}$ is not the least element of $\mathcal{W}_3^{\mathrm{B}}(\mathbb{R})$, that the horizon-4 learner certificate is tight, that the constant strategy "always arm 0" is a legitimate strategy with worst case 4, and that the horizon-4 adversary certificate has support exactly 20; these guard against a misplaced quantifier making the theorems vacuous.

## 5. The game with a switching cost

**Lemma 5.1** (The opening penalty is deterministic)**.** *For every commutative ring $\mathbb{K}$, every $\ell_1 = (l_0, l_1) \in \{0, 1\}^2$, every $\ell'$, and every node $S = (p; c_{00}, c_{01}, c_{10}, c_{11})$,*

$$C^{\mathrm{sw}}_{(\ell_1, \ell')}(S; \varnothing) = 1 + \Big( p\big(l_0 + C^{\mathrm{sw}}_{\ell'}(c_{0,l_0}; 0)\big) + (1 - p)\big(l_1 + C^{\mathrm{sw}}_{\ell'}(c_{1,l_1}; 1)\big) \Big).$$

*Proof.* $\sigma(\varnothing, 0) = \sigma(\varnothing, 1) = 1$ and $p + (1 - p) = 1$. □

Consequently the variant that charges only the switches from round 2 on has value $V^{\mathrm{sw}}(2, T) - 1$ wherever $V^{\mathrm{sw}}(2, T)$ is known; that shifted set of guarantees is not itself defined in the formal development, and the shift is read off the lemma rather than stated as a theorem. The convention $X_0 = 0 \notin [k]$ is the stated definition of [9] (page 3 of the arXiv v2 PDF); in the paragraph that immediately follows the statement of their Theorem 2, at the head of their Section 4 and before its proof begins (page 9), the same paper writes "recall that we arbitrarily set $X_0 = 1$", which *is* in the action set. Asymptotically the difference is $O(1)$; at $T \le 5$ it is the whole third digit, and Remark 5.8 records what the other readings give.

**Theorem 5.2** ($V^{\mathrm{sw}}(2,T)$ for $T \le 5$)**.** *For every linearly ordered field $\mathbb{K}$ and $T = 1,2,3,4,5$, the least element of $\mathcal{W}_T^{\mathrm{sw}}(\mathbb{K})$ exists and equals $\frac{3}{2},\ 2,\ \frac{7}{3},\ \frac{31}{12},\ \frac{167}{60}$ respectively.*

*Certificate.* Learner trees of $1, 5, 21, 85, 341$ probabilities, the largest single denominator being $2, 2, 4, 24, 350$ at the five horizons, each with worst case over the $4^T$ matrices exactly the stated value (the $T = 4, 5$ certificates are tight). Adversary distributions of support $2, 2, 4, 6, 8$; at $T = 5$ the eight matrices are

$$\big((1,1)^j, (1,0)^{5-j}\big) \text{ and } \big((1,1)^j, (0,1)^{5-j}\big), \qquad j = 0,1,2,3,$$

with weight $\frac{1}{5}, \frac{1}{20}, \frac{1}{12}, \frac{1}{6}$ on each of the two matrices with commitment time $j$, where $(1,1)^j$ means $j$ rounds on which both arms lose and $(1,0)^{5-j}$ means $5-j$ rounds on which only arm 0 loses; its value $\mathrm{val}^{\mathrm{sw}}$ is exactly $\frac{167}{60}$. Proposition 3.3 transports both halves. □

**Proposition 5.3** (The price of the switching cost)**.** *With both minimax values attained (Theorems 4.1 and 5.2, Proposition 4.2),*

$$V^{\mathrm{sw}}(2,T) - V^{\mathrm{B}}(2,T) = 1,\ \tfrac{3}{2},\ \tfrac{19}{12},\ \tfrac{211}{120} \qquad (T = 1,2,3,4),$$

*and net of the deterministic opening penalty of Lemma 5.1 the price is $0, \frac{1}{2}, \frac{7}{12}, \frac{91}{120}$: the switching cost is free at $T = 1$, where there is no second round to switch into, and strictly positive from $T = 2$ on. At $T = 5$, where $V^{\mathrm{B}}(2,5)$ is only bracketed (Proposition 4.5), $\frac{167}{60} - \frac{53}{40} = \frac{35}{24} \le V^{\mathrm{sw}}(2,5) - V^{\mathrm{B}}(2,5) \le \frac{167}{60} - \frac{15}{16} = \frac{443}{240}$.*

*Proof.* Subtraction of attained values; at $T = 5$, Theorem 5.2 with Proposition 4.5. □

**Theorem 5.4** (With a switching cost, bandit feedback first costs at $T = 3$)**.** *For every linearly ordered field $\mathbb{K}$ the least elements of $\mathcal{W}_1^{\mathrm{swF}}(\mathbb{K})$, $\mathcal{W}_2^{\mathrm{swF}}(\mathbb{K})$, $\mathcal{W}_3^{\mathrm{swF}}(\mathbb{K})$ – the minimax values of the switching-cost game under full information, over all 8-ary behaviour strategies – exist and equal $\frac{3}{2},\ 2,\ \frac{9}{4}$. Hence the bandit and full-information values agree at $T = 1, 2$, and at $T = 3$*

$$V^{\mathrm{swF}}(2,3) = \tfrac{9}{4} < \tfrac{7}{3} = V^{\mathrm{sw}}(2,3), \qquad \tfrac{7}{3} - \tfrac{9}{4} = \tfrac{1}{12}.$$

*At the same horizon the plain game has $V^{\mathrm{B}}(2,3) = V^{\mathrm{F}}(2,3) = \frac{3}{4}$: adding the switching cost moves the onset of the bandit-versus-full-information gap from $T = 4$ to $T = 3$.*

*Certificate.* Full-information learner trees (8-ary) of $1, 9, 73$ probabilities with worst cases $\frac{3}{2}, 2, \frac{9}{4}$ (tight at $T = 3$), and adversary distributions of support $2, 2, 4$ with $\mathrm{val}^{\mathrm{swF}}$ equal to the same numbers; Lemma 3.2 for $\mathcal{S}^8$ and Proposition 3.3. The comparison uses Theorem 5.2 and Proposition 4.2; the inequality $V^{\mathrm{swF}} \le V^{\mathrm{sw}}$ is not used. □

The four numbers at $T = 3$ say what the switching cost does. Net of the opening penalty, the full-information learner pays $\frac{9}{4} - 1 - \frac{3}{4} = \frac{1}{2}$ for the switching cost and the bandit learner pays $\frac{7}{3} - 1 - \frac{3}{4} = \frac{7}{12}$; since the plain game has no gap at $T = 3$, the entire separation $\frac{1}{12}$ is the extra price the bandit learner pays for switching. This is the mechanism of [9] – a bandit learner must switch to explore, and each switch is charged – in its smallest instance.

**Proposition 5.5** (The harmonic fit)**.** *With $H_T = \sum_{j=1}^{T} \frac{1}{j}$, $H_T + \frac{1}{2} = \frac{3}{2}, 2, \frac{7}{3}, \frac{31}{12}, \frac{167}{60}$ for $T = 1, \ldots, 5$; that is, $V^{\mathrm{sw}}(2,T) = H_T + \frac{1}{2}$ at each of the five horizons of Theorem 5.2.*

This is a fit to five points and is recorded as such; no formula for general $T$ is stated anywhere in this paper. It cannot continue for all $T$: Theorem 1 of [9] gives $\widetilde{\Omega}(k^{1/3}T^{2/3})$ for $k \le T$, and their Section 5.1 ("Binary losses") says in prose that the construction adapts to losses in $\{0,1\}$, this paper's loss model, so the true $V^{\mathrm{sw}}(2,T)$ outgrows $\log T$ (we cite that paragraph and do not use it). No finite-$T$ comparison with their bounds is well posed at these horizons: Theorem 1 carries no constant, and their Theorem 2, the one statement with an explicit constant, assumes $T \ge \max\{k, 6\}$, which excludes every horizon of Theorem 5.2. What the five values show is the exact price of the switching cost at each horizon, the onset of the feedback gap at $T = 3$, and increments $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$ at these horizons; what they cannot show is any onset of the $T^{2/3}$ regime or any constant in front of it. The next theorem, obtained with a different instrument, shows where the fit stops being an equality.

**Theorem 5.6** (The harmonic fit fails at $T = 7$)**.** *For every linearly ordered field $\mathbb{K}$,*

$$\forall v \in \mathcal{W}_6^{\mathrm{sw}}(\mathbb{K}) :\ v \geq \tfrac{59}{20} \qquad \textit{and} \qquad \forall v \in \mathcal{W}_7^{\mathrm{sw}}(\mathbb{K}) :\ v \geq \tfrac{373}{120}.$$

*Moreover $H_6 + \frac{1}{2} = \frac{59}{20}$, while $H_7 + \frac{1}{2} = \frac{433}{140} < \frac{373}{120}$. Hence every horizon-7 behaviour strategy has worst-case switching-cost regret strictly above $H_7 + \frac{1}{2}$: the pattern $V^{\mathrm{sw}}(2, T) = H_T + \frac{1}{2}$, exact for $T \leq 5$, is false at $T = 7$.*

*Certificate.* Every optimal adversary found at $T \leq 5$ has one shape, *delayed commitment*: choose a commitment time $j$, play $(1, 1)$ for $j$ rounds – both arms lose, so the learner pays without learning – and then commit to a single losing arm, $(1, 0)$ or $(0, 1)$, for the remaining $T - j$ rounds. The horizon-7 certificate is the twelve-point distribution on the matrices $((1, 1)^j, (1, 0)^{7-j})$ and $((1, 1)^j, (0, 1)^{7-j})$, $j = 0, \ldots, 5$, with weight

$$\tfrac{1}{6},\ \tfrac{1}{36},\ \tfrac{7}{180},\ \tfrac{7}{120},\ \tfrac{7}{72},\ \tfrac{1}{9}$$

on each of the two matrices with commitment time $j = 0, \ldots, 5$; its value $\mathrm{val}^{\mathrm{sw}}$ is exactly $\frac{373}{120}$, and Lemma 3.2 bounds every behaviour strategy. The horizon-6 certificate is the ten-point member of the same family with weights $\frac{1}{6}, \frac{1}{30}, \frac{1}{20}, \frac{1}{12}, \frac{1}{6}$ per arm for $j = 0, \ldots, 4$, of value exactly $\frac{59}{20}$. Both are one-sided: no learner certificate was computed at $T = 6$ or $T = 7$, whose sequence-form programs have 2730 and 10922 variables, so $V^{\mathrm{sw}}(2, 6)$ and $V^{\mathrm{sw}}(2, 7)$ are bounded below and not pinned. □

*Remark* 5.7 (The closed-form weights; computed but not formalised). The horizon-6 weights above are the instance $T = 6$ of the closed form $w_0 = \frac{1}{T}$, $w_j = 1/((T - j)(T - j + 1))$ for $1 \leq j \leq T - 2$, one copy of $w_j$ on each of the two arms (the $2(T - 1)$ weights sum to 1 because $\sum_{j \geq 1} w_j = \sum_{m=2}^{T-1} \frac{1}{m(m+1)} = \frac{1}{2} - \frac{1}{T}$), and the $T = 5$ adversary of Theorem 5.2 is the instance $T = 5$. Evaluated in exact rational arithmetic, this one closed-form mixture guarantees exactly $H_T + \frac{1}{2}$ at every horizon $T = 2, \ldots, 10$, so $H_T + \frac{1}{2}$ is a lower bound with a reason at least up to $T = 10$; that computation is outside the formal development and no statement for general $T$ is made. Maximising the exact guarantee over all weightings of the delayed-commitment family (a small concave problem, solved by cutting planes each of which is an exact backward-induction best response) gives exactly $H_T + \frac{1}{2}$ for $T = 2, \ldots, 6$, so the family refutes nothing at $T = 6$; it gives $\frac{373}{120}$ at $T = 7$, and at $T = 8$ it gives $\frac{2731}{840}$, which exceeds $H_8 + \frac{1}{2} = \frac{901}{280}$ (computed, not formalised). Some adversary outside this family could in principle refute the fit at $T = 6$; we make no claim either way.

*Remark* 5.8 (The two other readings of the opening convention). With the stated convention of [9] ($X_0$ outside the action set) the values are $H_T + \frac{1}{2}$ for $T \leq 5$ (Theorem 5.2). With no charge on round 1 they are $V^{\mathrm{sw}}(2, T) - 1 = \frac{1}{2}, 1, \frac{4}{3}, \frac{19}{12}$ for $T \leq 4$, i.e. $H_T - \frac{1}{2}$, by Lemma 5.1. With $X_0$ a distinguished arm of the action set – the reading of the sentence quoted after Lemma 5.1 – the same exact two-sided computation, carried out outside the formal development and not formalised, gives $1, \frac{3}{2}, \frac{11}{6}, \frac{25}{12}$ for $T \leq 4$, i.e. $H_T$. That the three readings differ by exactly $\frac{1}{2}$ at every horizon computed is an observed regularity, not a theorem.

*Remark* 5.9 (Controls). The formal development also records that $\frac{33}{40} + 1$ and $\frac{8}{3}$ are not least elements of $\mathcal{W}_4^{\mathrm{sw}}(\mathbb{R})$ (the switching cost costs strictly more than its mandatory opening switch), that $\frac{31}{12} + \frac{1}{4}$ and 3 are not least elements of $\mathcal{W}_5^{\mathrm{sw}}(\mathbb{R})$, that $\frac{7}{3}$ and 2 are not least elements of $\mathcal{W}_3^{\mathrm{swF}}(\mathbb{R})$, that the constant strategies have worst cases 5 (bandit, $T = 4$) and 4 (full information, $T = 3$), that the learner certificates at $T = 4, 5$ and the full-information one at $T = 3$ are tight, and that every adversary certificate is a genuine multi-point distribution with strictly positive weights (supports 6, 8, 4, 10, 12).

## 6. Related work and what we searched

The rate literature for the plain game is [2, 3, 7, 13]; Lattimore and Szepesvári write that "Finding a minimax policy is generally too computationally expensive to be practical. For this reason, we almost always settle for a policy that is nearly minimax optimal" ([13], Chapter 13), and none of these sources prints a value at any finite horizon. The switching-cost game is that of [9]; the words "exact" and "finite horizon" do not occur in that paper, every minimax statement in it is a rate, and so is its own survey of the lineage, on page 3: with full feedback "the Follow the Lazy Leader algorithm (Kalai and Vempala, 2005) and the Shrinking Dartboard algorithm (Geulen et al., 2010) both guarantee a matching upper bound of $O(\sqrt{T \log k})$", while "Arora et al. (2012) presented a simple algorithm with a guaranteed regret of $O(k^{1/3}T^{2/3})$, but a matching lower bound was not known". We did not read Kalai–Vempala or Geulen et al.; those two upper bounds

enter here only as [9] reports them, and the third is [4]. The nearest printed object to our switching-cost values is the zero-sum repeated game with switching costs of Tsodikovich, Venel and Zseleva [14] ("we show that the value of the game exists in stationary strategies, depending solely on the previous action of the minimizing player, not the entire history. We provide important properties of the value and the optimal strategies as a function of the ratio between the switching costs and the stage payoffs"); there the payoff matrix is known to both players, the game has no horizon, and there is neither a comparator "best fixed arm in hindsight" nor bandit feedback, so it is adjacent prior art rather than an overlap. Altschuler and Talwar [5] bound online learning with limited switching by rates, not values.

We searched for a printed value of either game as follows, reading every negative as "not found" rather than "does not exist". A local full-text corpus of arXiv sources (the computer-science, economics and machine-learning slice, about 14 GB, plus the mathematics slice) returned zero files for fixed strings such as "exact minimax regret", "exact minimax value", "minimax value of the bandit", "finite horizon minimax regret", "bandit feedback does not hurt", "gap between bandit and full information", "minimax regret with switching" and "value of the game with switching"; the 389 papers containing "switching cost" were intersected with "minimax value", "exact minimax", "compute the minimax" and "exactly computed", and every hit was read and found to be a false positive; the strings "33/40" and "31/12" never co-occur with a bandit, regret or minimax context. The arXiv search API returned zero results for the corresponding conjunctions and exactly one for "minimax value" with "switching cost", namely [14]. OpenAlex knows one distinct citing work for the COLT 2009 note [1] (two records of the same drifting-games paper), which is full-information and computes no bandit value. For [9], OpenAlex lists 64 distinct citing works across its two records and none of their titles suggests a finite-horizon or exact computation; that sweep is *partial* – the paper has several hundred citations and the 64 are those OpenAlex indexes – and we state it as a strong negative, not an exhaustive one. The OEIS contains neither $\frac{1}{2}, \frac{1}{2}, \frac{3}{4}, \frac{33}{40}$ in any of five encodings nor the numerators $3, 2, 7, 31$; its entries for the numerators of Cover's column (A100071) and for the harmonic numbers (A001008, A002805) mention neither regret nor bandits. A value this small could still sit unlabelled in lecture notes or a solutions manual that no index reaches.

The three- and four-expert full-information values at small horizons are treated in a companion preprint of this project [15]; the two-expert column is not duplicated there beyond its role as a control.

## 7. Verification

Every theorem, proposition and lemma above is a statement checked in Lean 4 with Mathlib. The two lemmas of Section 3 and the transfer of Proposition 3.3 are proved as ordinary structural inductions and carry only the standard axioms (`propext`, `Classical.choice`, `Quot.sound`); Lemma 5.1 carries `propext` and `Quot.sound` only. The value theorems additionally carry the axioms generated by `native_decide` for their own finite checks (validity of the certificate tree, its regret against every one of the $4^T$ matrices, and the distribution and backward-induction value of the adversary mixture); no other axiom and no `sorry` occurs. Appendix A reproduces the formal statements verbatim. The formal definitions match Section 2, with arms named `false` and `true`, and with children indexed by $2a + b$ under bandit feedback, by $2\ell_t(0) + \ell_t(1)$ under full information, and by $4a+2\ell_t(0)+\ell_t(1)$ in the 8-ary trees. Repository reference to be added at submission.

## 8. Open questions

The exact value of $V^{\mathrm{B}}(2, 5)$ is the obvious gap in Table 1; the float-warm-start-then-exact recipe that solved the switching-cost game at $T = 5$ in about 950 seconds applies unchanged to the plain game, whose program is the same with the switching term deleted. The value $V^{\mathrm{swF}}(2, 4)$ needs an 8-ary treeplex with 585 information sets and 1170 sequence variables. The learner side at $T = 6, 7$ in the switching-cost game is far more expensive (1365 and 5461 information sets, 2730 and 10922 sequence variables), so whether $V^{\mathrm{sw}}(2, 6) = H_6 + \frac{1}{2}$ is open. Three arms at $T \le 3$ (a round is a vector in $\{0, 1\}^3$, so $8^3 = 512$ loss matrices and 43 bandit information sets) would give a first data point on how the onset of the feedback gap depends on $K$. The delayed-commitment family of Theorem 5.6 suggests a question we do not answer: whether the optimal adversary of the switching-cost game always has that shape, and if so what its optimal weights are for general $T$.

## Appendix A. The formal statements

The Lean statements corresponding to the results above, verbatim (statements only, proofs and attributes omitted), grouped by module. `Strat K` is $\mathcal{S}(\mathbb{K})$ with children $c_0, \dots, c_3$ at index $2a+b$; `Strat8 K` is $\mathcal{S}^8(\mathbb{K})$; `Ok n S` says $S$ is a complete tree of depth $n$ with probabilities in $[0,1]$; `allLoss n` is $\mathcal{L}_n$; `bestFixed` is m; `breg`, `freg`, `sreg`, `fsreg` are $R^{\mathrm{B}}, R^{\mathrm{F}}, R^{\mathrm{sw}}, R^{\mathrm{swF}}$; `WL K` is the type of mixtures; `brB`, `brF`, `srB`, `srF8` are $\beta^{\mathrm{B}}, \beta^{\mathrm{F}}, \beta^{\mathrm{sw}}, \beta^{\mathrm{swF}}$; `bVal`, `fVal`, `sVal`, `fsVal` are the four val functions; `banditReach`, `fullReach`, `switchReach`, `fswitchReach` are the four sets $\mathcal{W}$; `wbcost`, `wfcost`, `wscost`, `wfscost` are the weighted costs $\sum_i w_i C_{\ell^i}(S)$; `scost` is $C^{\mathrm{sw}}$ with the previous arm first; `harmonic` is Mathlib's $H_T$; `Data.adv*` and `cert*` are the certificate data.

`Lower.lean`

```
theorem brB_le : ∀ (n : ℕ) (L : WL K), (∀ x ∈ L, 0 ≤ x.1) → (∀ x ∈ L, x.2.length = n) →
    ∀ S : Strat K, Ok n S → brB n L ≤ wbcost L S
theorem brF_le : ∀ (n : ℕ) (L : WL K), (∀ x ∈ L, 0 ≤ x.1) → (∀ x ∈ L, x.2.length = n) →
    ∀ S : Strat K, Ok n S → brF n L ≤ wfcost L S
```

`Cert.lean`

```
theorem bandit_lower (n : ℕ) (L : WL K)
    (hpos : ∀ x ∈ L, 0 ≤ x.1) (hlen : ∀ x ∈ L, x.2.length = n)
    (hsum : (L.map Prod.fst).sum = 1)
    (S : Strat K) (hS : Ok n S) (w : K) (hw : ∀ x ∈ L, breg x.2 S ≤ w) :
    bVal n L ≤ w
theorem full_lower (n : ℕ) (L : WL K)
    (hpos : ∀ x ∈ L, 0 ≤ x.1) (hlen : ∀ x ∈ L, x.2.length = n)
    (hsum : (L.map Prod.fst).sum = 1)
    (S : Strat K) (hS : Ok n S) (w : K) (hw : ∀ x ∈ L, freg x.2 S ≤ w) :
    fVal n L ≤ w
theorem bandit_isLeast (n : ℕ) (v : ℚ) (S : Strat ℚ) (hS : Ok n S)
    (hup : ∀ e ∈ allLoss n, breg e S ≤ v)
    (M : WL ℚ) (hpos : ∀ x ∈ M, 0 ≤ x.1) (hlen : ∀ x ∈ M, x.2.length = n)
    (hsum : (M.map Prod.fst).sum = 1) (hmem : ∀ x ∈ M, x.2 ∈ allLoss n)
    (hval : bVal n M = v) :
    IsLeast (banditReach K n) ((v : ℚ) : K)
theorem full_isLeast (n : ℕ) (v : ℚ) (S : Strat ℚ) (hS : Ok n S)
    (hup : ∀ e ∈ allLoss n, freg e S ≤ v)
    (M : WL ℚ) (hpos : ∀ x ∈ M, 0 ≤ x.1) (hlen : ∀ x ∈ M, x.2.length = n)
    (hsum : (M.map Prod.fst).sum = 1) (hmem : ∀ x ∈ M, x.2 ∈ allLoss n)
    (hval : fVal n M = v) :
    IsLeast (fullReach K n) ((v : ℚ) : K)
```

`Value.lean`

```
theorem value_bandit_1 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (banditReach K 1) (1/2)
theorem value_bandit_2 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (banditReach K 2) (1/2)
theorem value_bandit_3 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (banditReach K 3) (3/4)
theorem value_bandit_4 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (banditReach K 4) (33/40)
theorem value_full_1 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fullReach K 1) (1/2)
theorem value_full_2 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fullReach K 2) (1/2)
theorem value_full_3 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fullReach K 3) (3/4)
theorem value_full_4 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fullReach K 4) (3/4)
theorem separation_four :
    IsLeast (banditReach ℝ 4) (33 / 40) ∧ IsLeast (fullReach ℝ 4) (3 / 4) ∧
      (3 : ℝ) / 4 < 33 / 40 ∧ (33 : ℝ) / 40 - 3 / 4 = 3 / 40
theorem bandit_free_below_four :
    IsLeast (banditReach ℝ 1) (1 / 2) ∧ IsLeast (fullReach ℝ 1) (1 / 2) ∧
      IsLeast (banditReach ℝ 2) (1 / 2) ∧ IsLeast (fullReach ℝ 2) (1 / 2) ∧
        IsLeast (banditReach ℝ 3) (3 / 4) ∧ IsLeast (fullReach ℝ 3) (3 / 4)
theorem value_full_5 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fullReach K 5) (15 / 16)
theorem bandit_five_bracket :
    (53 / 40 : ℝ) ∈ banditReach ℝ 5 ∧ ∀ v ∈ banditReach ℝ 5, (15 / 16 : ℝ) ≤ v
```

`Switch.lean`

```
theorem srB_le : ∀ (n : ℕ) (prev : Option Bool) (L : WL K), (∀ x ∈ L, 0 ≤ x.1) →
    (∀ x ∈ L, x.2.length = n) → ∀ S : Strat K, Ok n S → srB prev n L ≤ wscost prev L S
theorem switch_lower (n : ℕ) (L : WL K)
    (hpos : ∀ x ∈ L, 0 ≤ x.1) (hlen : ∀ x ∈ L, x.2.length = n)
    (hsum : (L.map Prod.fst).sum = 1)
    (S : Strat K) (hS : Ok n S) (w : K) (hw : ∀ x ∈ L, sreg x.2 S ≤ w) :
    sVal n L ≤ w
theorem switch_isLeast (n : ℕ) (v : ℚ) (S : Strat ℚ) (hS : Ok n S)
    (hup : ∀ e ∈ allLoss n, sreg e S ≤ v)
    (M : WL ℚ) (hpos : ∀ x ∈ M, 0 ≤ x.1) (hlen : ∀ x ∈ M, x.2.length = n)
    (hsum : (M.map Prod.fst).sum = 1) (hmem : ∀ x ∈ M, x.2 ∈ allLoss n)
```

```
    (hval : sVal n M = v) :
    IsLeast (switchReach K n) ((v : ℚ) : K)
```

SwitchValue.lean

```
theorem scost_none_split {K : Type*} [CommRing K] (l₀ l₁ : Bool) (r : LossSeq) (p : K)
    (c₀ c₁ c₂ c₃ : Strat K) :
    scost none ((l₀, l₁) :: r) (.node p c₀ c₁ c₂ c₃)
      = 1 + (p * (bit l₀ + scost (some false) r (if l₀ then c₁ else c₀))
             + (1 - p) * (bit l₁ + scost (some true) r (if l₁ then c₃ else c₂)))
theorem value_switch_1 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (switchReach K 1) (3/2)
theorem value_switch_2 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (switchReach K 2) (2)
theorem value_switch_3 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (switchReach K 3) (7/3)
theorem value_switch_4 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (switchReach K 4) (31/12)
theorem switch_price_table :
    (IsLeast (banditReach ℝ 1) (1/2) ∧ IsLeast (switchReach ℝ 1) (3/2)
        ∧ (3/2 : ℝ) - 1/2 = 1) ∧
    (IsLeast (banditReach ℝ 2) (1/2) ∧ IsLeast (switchReach ℝ 2) (2)
        ∧ (2 : ℝ) - 1/2 = 3/2) ∧
    (IsLeast (banditReach ℝ 3) (3/4) ∧ IsLeast (switchReach ℝ 3) (7/3)
        ∧ (7/3 : ℝ) - 3/4 = 19/12) ∧
    (IsLeast (banditReach ℝ 4) (33/40) ∧ IsLeast (switchReach ℝ 4) (31/12)
        ∧ (31/12 : ℝ) - 33/40 = 211/120)
theorem switch_price_net :
    (3/2 - 1 : ℝ) - 1/2 = 0 ∧ (2 - 1 : ℝ) - 1/2 = 1/2 ∧
    (7/3 - 1 : ℝ) - 3/4 = 7/12 ∧ (31/12 - 1 : ℝ) - 33/40 = 91/120
theorem harmonic_fit :
    (harmonic 1 : ℚ) + 1/2 = 3/2 ∧ (harmonic 2 : ℚ) + 1/2 = 2 ∧
    (harmonic 3 : ℚ) + 1/2 = 7/3 ∧ (harmonic 4 : ℚ) + 1/2 = 31/12
theorem constS4_bad : ∃ e ∈ allLoss 4, sreg e (constS 4) = (5 : ℚ)
theorem constS4_not_optimal : ¬ ((5 : ℝ) ≤ 31/12)
theorem not_switch_four_plain_plus_one : ¬ IsLeast (switchReach ℝ 4) (33/40 + 1)
theorem not_switch_four_eight_thirds : ¬ IsLeast (switchReach ℝ 4) (8/3)
theorem certS4_tight : ∃ e ∈ allLoss 4, sreg e certS4 = (31/12 : ℚ)
theorem advS4_support : Data.advS4.length = 6 ∧ ∀ x ∈ Data.advS4, 0 < x.1
```

SwitchFull.lean

```
theorem srF8_le : ∀ (n : ℕ) (prev : Option Bool) (L : WL K), (∀ x ∈ L, 0 ≤ x.1) →
    (∀ x ∈ L, x.2.length = n) → ∀ S : Strat8 K, Ok8 n S → srF8 prev n L ≤ wfscost prev L S
theorem fswitch_lower (n : ℕ) (L : WL K)
    (hpos : ∀ x ∈ L, 0 ≤ x.1) (hlen : ∀ x ∈ L, x.2.length = n)
    (hsum : (L.map Prod.fst).sum = 1)
    (S : Strat8 K) (hS : Ok8 n S) (w : K) (hw : ∀ x ∈ L, fsreg x.2 S ≤ w) :
    fsVal n L ≤ w
theorem fswitch_isLeast (n : ℕ) (v : ℚ) (S : Strat8 ℚ) (hS : Ok8 n S)
    (hup : ∀ e ∈ allLoss n, fsreg e S ≤ v)
    (M : WL ℚ) (hpos : ∀ x ∈ M, 0 ≤ x.1) (hlen : ∀ x ∈ M, x.2.length = n)
    (hsum : (M.map Prod.fst).sum = 1) (hmem : ∀ x ∈ M, x.2 ∈ allLoss n)
    (hval : fsVal n M = v) :
    IsLeast (fswitchReach K n) ((v : ℚ) : K)
theorem value_fswitch_1 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fswitchReach K 1) (3/2)
theorem value_fswitch_2 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fswitchReach K 2) (2)
theorem value_fswitch_3 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (fswitchReach K 3) (9/4)
theorem switch_separation_three :
    IsLeast (switchReach ℝ 1) (3/2) ∧ IsLeast (fswitchReach ℝ 1) (3/2) ∧
    IsLeast (switchReach ℝ 2) (2) ∧ IsLeast (fswitchReach ℝ 2) (2) ∧
    IsLeast (switchReach ℝ 3) (7/3) ∧ IsLeast (fswitchReach ℝ 3) (9/4) ∧
    (9/4 : ℝ) < 7/3 ∧ (7/3 : ℝ) - 9/4 = 1/12
theorem onset_moves_earlier :
    IsLeast (banditReach ℝ 3) (3/4) ∧ IsLeast (fullReach ℝ 3) (3/4) ∧
    IsLeast (switchReach ℝ 3) (7/3) ∧ IsLeast (fswitchReach ℝ 3) (9/4) ∧ (9/4 : ℝ) < 7/3
theorem constFS3_bad : ∃ e ∈ allLoss 3, fsreg e (constFS 3) = (4 : ℚ)
theorem constFS3_not_optimal : ¬ ((4 : ℝ) ≤ 9/4)
theorem not_fswitch_three_eq_bandit : ¬ IsLeast (fswitchReach ℝ 3) (7/3)
theorem not_fswitch_three_two : ¬ IsLeast (fswitchReach ℝ 3) (2)
theorem certFS3_tight : ∃ e ∈ allLoss 3, fsreg e certFS3 = (9/4 : ℚ)
theorem advFS3_support : Data.advFS3.length = 4 ∧ ∀ x ∈ Data.advFS3, 0 < x.1
```

SwitchFive.lean

```
theorem value_switch_5 {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    IsLeast (switchReach K 5) (167/60)
theorem harmonic_fit_five :
    (harmonic 1 : ℚ) + 1/2 = 3/2 ∧ (harmonic 2 : ℚ) + 1/2 = 2 ∧
    (harmonic 3 : ℚ) + 1/2 = 7/3 ∧ (harmonic 4 : ℚ) + 1/2 = 31/12 ∧
    (harmonic 5 : ℚ) + 1/2 = 167/60
theorem switch_price_five :
    IsLeast (switchReach ℝ 5) (167/60) ∧ (53/40 : ℝ) ∈ banditReach ℝ 5 ∧
    (∀ v ∈ banditReach ℝ 5, (15/16 : ℝ) ≤ v) ∧
```

```
    (167/60 : ℝ) - 53/40 = 35/24 ∧ (167/60 : ℝ) - 15/16 = 443/240
theorem certS5_tight : ∃ e ∈ allLoss 5, sreg e certS5 = (167/60 : ℚ)
theorem advS5_support : Data.advS5.length = 8 ∧ ∀ x ∈ Data.advS5, 0 < x.1
theorem not_switch_five_quarter_step : ¬ IsLeast (switchReach ℝ 5) (31/12 + 1/4)
theorem not_switch_five_three : ¬ IsLeast (switchReach ℝ 5) (3)
```

`SwitchLower.lean`

```
theorem switch_six_lower {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    ∀ v ∈ switchReach K 6, (59/20 : K) ≤ v
theorem switch_seven_lower {K : Type*} [Field K] [LinearOrder K] [IsStrictOrderedRing K] :
    ∀ v ∈ switchReach K 7, (373/120 : K) ≤ v
theorem harmonic_fit_fails_at_seven :
    (harmonic 7 : ℚ) + 1/2 = 433/140 ∧ (433/140 : ℝ) < 373/120 ∧
    ∀ v ∈ switchReach ℝ 7, (373/120 : ℝ) ≤ v
theorem harmonic_fit_not_refuted_at_six : (harmonic 6 : ℚ) + 1/2 = 59/20
theorem advS67_support :
    Data.advS6.length = 10 ∧ Data.advS7.length = 12 ∧
    (∀ x ∈ Data.advS6, 0 < x.1) ∧ (∀ x ∈ Data.advS7, 0 < x.1)
theorem switch_lower_monotone : (167/60 : ℝ) < 59/20 ∧ (59/20 : ℝ) < 373/120
```

## References


[1] Jacob Abernethy and Manfred K. Warmuth, *Minimax games with bandits*, open problem, Proceedings of the 22nd Annual Conference on Learning Theory (COLT 2009), open problems session. Two pages, no numbered sections; no page numbers and no DOI are recorded for it in OpenAlex (`W1536957971`) or in Warmuth's own publication list, so none is given here. Read from `https://www.learningtheory.org/colt2009/papers/043.pdf`; both quotations in Section 1 are on page 2, left column.

[2] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund and Robert E. Schapire, *The nonstochastic multiarmed bandit problem*, SIAM Journal on Computing **32** (2002), no. 1, 48–77; DOI `10.1137/S0097539701398375`. (Crossref gives 2002; the reference list of [3] prints 2003.)

[3] Jean-Yves Audibert and Sébastien Bubeck, *Minimax policies for adversarial and stochastic bandits*, Proceedings of the 22nd Annual Conference on Learning Theory (COLT 2009). Read from `https://www.learningtheory.org/colt2009/papers/022.pdf` (10 pages). In their notation ($n$ rounds, $K$ arms), Section 1 uses their Section 1 display $\inf\sup R_n \geq \frac{1}{20}\sqrt{nK}$, "where the supremum is taken over all $K$-tuple of probability distributions on $[0,1]$" (recalled there from Auer et al., whose stochastic lower bound holds "thus also in the adversarial case"), and $\sup R_n \leq 100\sqrt{nK}$, read off their Theorem 4 at $q=3$, whose supremum is "over all adversarial environments with rewards in $[0,1]$". Their Remark 3, $\sup R_n \leq 5.7\sqrt{nK}$, sharpens their Theorem 5, whose supremum is again over $K$-tuples of distributions on $[0,1]$, so it is not an adversarial upper bound.

[4] Raman Arora, Ofer Dekel and Ambuj Tewari, *Online bandit learning against an adaptive adversary: from regret to policy regret*, Proceedings of the 29th International Conference on Machine Learning (ICML 2012); arXiv:1206.6400.

[5] Jason M. Altschuler and Kunal Talwar, *Online learning over a finite action set with limited switching*, Mathematics of Operations Research **46** (2021), no. 1, 179–203; DOI `10.1287/moor.2020.1052`; arXiv:1803.01548.

[6] Siddhartha Banerjee, Ramiro N. Deo-Campo Vuong and Robert Kleinberg, *Water-filling is universally minimax optimal*, arXiv:2603.26893; the quotation is from its related-work paragraph "Exact Minimax-Optimal Online Algorithms".

[7] Nicolò Cesa-Bianchi and Gábor Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, 2006.

[8] Thomas M. Cover, *Behavior of sequential predictors of binary sequences*, in: Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Prague, 31 August – 11 September 1965), Publishing House of the Czechoslovak Academy of Sciences, Prague, 1967, pp. 263–272. Read as page images from the scanned offprint at `https://isl.stanford.edu/~cover/papers/paper3.pdf` (11 pages, no text layer; its cover spells the title "Behaviour"); equation numbers are the paper's own.

[9] Ofer Dekel, Jian Ding, Tomer Koren and Yuval Peres, *Bandits with switching costs: $T^{2/3}$ regret*, in: Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing (STOC 2014), ACM, 2014, pp. 459–467; DOI `10.1145/2591796.2591868`; arXiv:1310.2997. Equation, theorem and page numbers here are those of the arXiv version v2 (19 November 2013), whose e-print source we compiled to a PDF of 18 pages, the page count of the posted one; the STOC version was not consulted.

[10] Nadejda Drenska and Robert V. Kohn, *Prediction with expert advice: a PDE perspective*, Journal of Nonlinear Science **30** (2020), no. 1, 137–173; DOI `10.1007/s00332-019-09570-3`; arXiv:1904.11401.

[11] J. Fabius and W. R. van Zwet, *Some remarks on the two-armed bandit*, The Annals of Mathematical Statistics **41** (1970), no. 6, 1906–1916; DOI `10.1214/aoms/1177696692`.

[12] Nick Gravin, Yuval Peres and Balasubramanian Sivan, *Towards optimal algorithms for prediction with expert advice*, in: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2016), SIAM, 2016, pp. 528–547; DOI `10.1137/1.9781611974331.ch39`; arXiv:1409.3040v5 (11 July 2016, the latest), whose e-print supplies the quotations; the SODA version was not consulted.

[13] Tor Lattimore and Csaba Szepesvári, *Bandit Algorithms*, Cambridge University Press, 2020. Chapter and exercise numbers are those of the online edition (`https://tor-lattimore.com/downloads/book/book.pdf`, 597 pages, 2024) read for this paper; they may differ in the printed one.

[14] Yevgeny Tsodikovich, Xavier Venel and Anna Zseleva, *Repeated games with switching costs: stationary vs history-independent strategies*, arXiv:2103.00045. The quotation in Section 6 is from the abstract of the compiled v3 e-print, which differs there from the abstract on the arXiv listing page.

[15] Korea Superintelligence Labs, *The exact finite-horizon minimax regret of prediction with three and four experts, and the exact deficit of the comb adversary*, preprint of this project, 2026.