ADVERSARIAL NONLINEAR SEMI-BANDITS: HIDDEN INTERACTIONS AND AMPLITUDE-SENSITIVE REGRET

Anonymous authors
Paper under double-blind review

ABSTRACT

Semi-bandit feedback reveals every selected base-arm outcome, yet nonlinear rewards can hide information in interactions that are observed only jointly. For every order $s \le K$, we construct an s-way dependence whose proper marginals are unchanged and prove regret $\Omega(\Gamma_s(g;K)\min\{T, \sqrt{(\binom{N}{s}/\binom{K}{s})T}\})$. The finite-difference certificate Γ_s measures how strongly the known link values the hidden interaction. Selected coordinates also make every interaction supported inside the chosen action observable. A Möbius decomposition turns this fact into an amplitude-sensitive upper bound $\widetilde{O}(\sum_s b_s(g)\sqrt{\binom{K}{s}\binom{N}{s}T})$, where $b_s(g)$ is the range of an order-s observable coefficient. For every fixed action size K, the two bounds characterize minimax regret up to logarithmic factors, uniformly over normalized nonnegative polynomial mixtures and a reflected family of monotone concave links. The characterization includes links with a weak high-order component: their rate has a linear baseline and an amplitude-weighted high-order term. Finally, at fixed reward range, we prove that semi-bandit minimax regret is asymptotically smaller than under Bernoulli scalar feedback over a continuous parameter regime. Exact aggregate feedback additionally yields a complementary polynomial-approximation envelope.

1 INTRODUCTION

Semi-bandit feedback breaks the combinatorial barrier for linear rewards because one observed coordinate can be reused across every action containing it. For nonlinear rewards, reuse has a boundary: the value of a coordinate may depend on which other coordinates accompany it. This paper asks where that boundary lies when the link g is known and every selected primitive outcome is observed.

The need for amplitude-sensitive guarantees is already visible in the range-one family $g_\epsilon(z) = (1-\epsilon)z/K + \epsilon(z/K)^d$, where $2 \le d \le K$ and $0 \le \epsilon \le 1/2$. Every $\epsilon > 0$ gives a degree-d link, even though its nonlinear component can be arbitrarily weak. A rate proved separately for each fixed link may absorb this dependence into its constants and therefore does not describe regimes in which ϵ varies with N or T. Which interactions remain statistically costly as their amplitudes vanish? We seek guarantees uniform in the interaction amplitudes, so that the price of weak nonlinear structure remains explicit.

Two earlier results frame the question. Simchowitz et al. (2016) used high-order dependence to make a specific best-of-K objective hard under several feedback models, including semi-bandit feedback, and discussed both regret consequences and the small gaps forced by marginal hiding. Han et al. (2021) showed that, for general known links under noisy scalar feedback, polynomial degree controls the adversarial regret exponent. What remains unresolved is a link-dependent account of what full selected-coordinate observations cannot reveal, and what they can exploit.

The answer is two-sided. Coordinate observations do not erase interaction order; they change the unit at which information can be reused, from individual arms to subsets that have been observed jointly. Interaction order therefore controls how rarely decisive evidence is exposed, while interaction amplitude controls how much the learner should pay to resolve it. The paper develops matching lower and upper descriptions of these two effects and then uses them to separate coordinate feedback from noisy scalar feedback.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Hidden interactions. For every $s \le K$, we plant a hidden s-set whose proper marginals have exactly the null law. A finite difference converts that invisible dependence into a reward gap, giving, for $N \ge 2K$,

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \ge \frac{1}{16}\Gamma_s(g; K) \min\left\{T, \sqrt{\frac{\binom{N}{s}}{\binom{K}{s}}T}\right\}.$$

Here $\Gamma_s(g; K)$ is an explicit finite-difference amplitude, while $\binom{N}{s}/\binom{K}{s}$ is the inverse fraction of candidate interactions exposed by one action. The result turns marginal hiding into an all-order, link-dependent factorization: reward amplitude times joint-coverage cost. For $s < K$, a single action covers $\binom{K}{s}$ overlapping candidate s-sets; the coverage identity accounts for these simultaneous observations.

Observable interactions. Can coordinate feedback pay for an interaction in proportion to its strength? A Möbius decomposition reconstructs the reward from subset coefficients, and every coefficient supported inside the selected action is observable. If $b_s(g)$ is the range of an order-s coefficient, we prove

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \le \sqrt{2T \log \binom{N}{K}} \sum_{s=1}^{K} b_s(g) \sqrt{\binom{K}{s}\binom{N}{s}},$$

intersected with the trivial range bound. Exponential weighting on these subset coordinates preserves the width of each order. Its second-moment analysis accounts for overlapping observation events through an independent auxiliary action; no independence among the observed subsets is assumed.

Matching laws and feedback separation. The lower and upper descriptions match on the normalized polynomial mixtures $g^+_\beta(z) = \sum_{j=1}^{K} \beta_j (z/K)^j$, where $\beta_j \ge 0$ and $\sum_j \beta_j = 1$. For every fixed K, uniformly over the coefficient simplex,

$$\mathfrak{R}_{\mathrm{SB}}(g^+_\beta, N, K, T) = \widetilde{\Theta}_K\left(\sum_{j=1}^{K} \beta_j \min\left\{T, \sqrt{\frac{\binom{N}{j}}{\binom{K}{j}}T}\right\}\right).$$

The same law holds for a reflected family of monotone concave links. For the two-term example $(1-\epsilon)z/K + \epsilon(z/K)^d$, the minimax rate consists of a linear baseline and an amplitude-weighted order-d term.

Positivity links the finite-difference lower certificates to the observable widths. For fixed K, this comparison is uniform even when coefficients vanish with N or T, so it captures regimes hidden by degree-only rates.

The same amplitude sensitivity yields a strict feedback separation. For $K = 2$, a range-one quadratic link satisfies over a continuous family of (ϵ, T) scales

$$\frac{\mathfrak{R}_{\mathrm{SB}}}{\mathfrak{R}_{\mathrm{Ber}}} \longrightarrow 0,$$

where $\mathfrak{R}_{\mathrm{Ber}}$ denotes Bernoulli scalar feedback with the same conditional reward mean. The reward range remains one, and the separation holds even when the nonlinear component controls the semi-bandit rate. Section 5 identifies the corresponding feedback-dependent information attenuation.

Contribution and positioning. The main contribution is a coefficient-uniform characterization of how interaction amplitudes affect regret on structured link families, together with a quantitative comparison of feedback models. For example, at $K = 2$, $\epsilon = N^{-1/4}$, and $T = \lceil N^{5/2} \rceil$, the range-one quadratic family has semi-bandit regret $\widetilde{\Theta}(N^2)$ and Bernoulli-scalar regret $\Theta(N^{9/4})$. The quadratic component already dominates the semi-bandit rate in this regime. Thus coordinate observations yield a polynomial improvement while the nonlinear component remains statistically consequential. Table 1 compares the closest results, and Appendix F gives the broader context.

Classical combinatorial semi-bandit work supplies the linear estimation tools (Cesa-Bianchi & Lugosi, 2012; Audibert et al., 2014; Neu, 2015; Bubeck et al., 2012); stochastic nonlinear CMAB instead relies on stationary or structured reward laws (Chen et al., 2013; Kveton et al., 2015; Chen et al., 2016b;a).

Table 1: Closest results and the distinctions addressed here.

Closest work	Result and distinction
Simchowitz et al. (2016)	Dependence hiding for a max objective, including semi-bandit identification lower bounds and regret consequences. **Our work:** general known links and arbitrary interaction orders, explicit finite-difference amplitudes, general observable-interaction upper bounds, and matching laws on structured link families.
Han et al. (2021)	Degree-dependent adversarial regret under noisy scalar feedback. **Our work:** hardness under selected-coordinate feedback and a fixed-range minimax separation from Bernoulli scalar feedback.

2 Problem Setup and Background

2.1 Adversarial nonlinear semi-bandits

Let $[N] = \{1, \dots, N\}$ denote the set of base arms, and let $\mathcal{S}_K = \{S \subseteq [N] : |S| = K\}$ be the family of feasible actions. Before play begins, an oblivious adversary fixes a sequence $v_{1:T} \in ([0,1]^N)^T$. At round t, the learner faces $v_t = (v_{t,1}, \dots, v_{t,N}) \in [0,1]^N$, which is not revealed before the action is chosen. The learner then selects an action $S_t \in \mathcal{S}_K$ and receives reward

$$r_t(S_t) = g\left(\sum_{i \in S_t} v_{t,i}\right),$$

where the link $g : [0, K] \to [0, 1]$ is known. Under *semi-bandit feedback*, the learner observes the full selected vector $v_{t,S_t} = (v_{t,i})_{i \in S_t}$. In particular, it also knows the selected aggregate $z_t = \sum_{i \in S_t} v_{t,i}$ and hence the realized reward $g(z_t)$ exactly.

For a possibly randomized policy π, we measure regret against the best fixed K-subset in hindsight:

$$R_T(\pi; v_{1:T}) = \mathbb{E}_\pi \left[\max_{S \in \mathcal{S}_K} \sum_{t=1}^{T} g\left(\sum_{i \in S} v_{t,i}\right) - \sum_{t=1}^{T} g\left(\sum_{i \in S_t} v_{t,i}\right) \right].$$

The expectation is over the learner's internal randomization. The corresponding minimax value is

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) = \inf_{\pi} \sup_{v_{1:T} \in ([0,1]^N)^T} R_T(\pi; v_{1:T}).$$

Our lower bounds use i.i.d. random environments as a stochastic subclass of this adversarial model; the resulting Bayes lower bounds therefore also lower bound the adversarial minimax regret.

We distinguish three feedback levels. Selected-coordinate feedback reveals v_{t,S_t}; exact-aggregate feedback reveals only $z_t = \sum_{i \in S_t} v_{t,i}$; noisy scalar feedback reveals a bounded random variable whose conditional mean is $g(z_t)$. Selected coordinates determine the exact aggregate, and the exact aggregate determines the exact reward, but the reverse implications need not hold. Our interaction upper bound uses the first level, whereas the surrogate reduction requires only the second. In the Bernoulli model, Y_t is drawn afresh conditional on the past and the current action, with mean $g(z_t)$. We write $\mathfrak{R}_{\mathrm{Ber}}$ for the same minimax comparator as above, with expectation over both policy and observation randomness.

2.2 Scalar-feedback benchmark

For comparison, Han et al. (2021) study a closely related adversarial model in which the learner observes a bounded scalar Y_t with conditional mean $g(\sum_{i \in S_t} v_{t,i})$, but not the selected coordinates. For fixed K, their benchmark is

$$\widetilde{\Theta}_{g,d,K}\left(\min\{T, \sqrt{N^d T}\}\right) \quad \text{for degree } 1 \le d < K,$$

$$\Theta_{g,K}\left(\min\{T, \sqrt{N^K T}\}\right) \quad \text{for genuinely } K\text{-th-order links.}$$

We compare semi-bandit feedback against these N, T rates for fixed K; no claim of optimal dependence on K is made unless stated explicitly.

2.3 FINITE-DIFFERENCE NOTATION

For an integer $s \geq 1$ and step size $h \geq 0$, define the s-th forward finite difference of g by $\Delta_h^s g(x) = \sum_{j=0}^{s} (-1)^{s-j} \binom{s}{j} g(x + jh)$, whenever all evaluation points lie in $[0, K]$. We also write

$$q_s := \frac{\binom{K}{s}}{\binom{N}{s}}, \qquad \psi_s(T) := \min\left\{T, \sqrt{T/q_s}\right\}.$$

Here q_s is the fraction of all s-subsets contained in any fixed K-subset, while $\psi_s(T)$ is the corresponding finite-horizon coverage scale.

3 HIDDEN INTERACTIONS SURVIVE SEMI-BANDIT FEEDBACK

Semi-bandit feedback reveals every selected coordinate, but it only reveals interactions among coordinates that are selected jointly. We exploit this distinction by constructing environments whose lower-order marginals are identical while a single hidden higher-order interaction changes the reward. The learner can therefore observe many coordinates without learning anything about the interaction that matters.

3.1 HIDING AN INTERACTION FROM EVERY PROPER MARGINAL

Fix an interaction order $s \in [K]$ and a hidden set $A \subseteq [N]$ with $|A| = s$. Let $X \in \{-1, +1\}^N$. Under the null environment, X is uniform on the Boolean cube: $P_0(x) = 2^{-N}$. Under the alternative indexed by A, define

$$P_A(x) = 2^{-N}\big(1 + \eta\sigma\chi_A(x)\big), \qquad \chi_A(x) := \prod_{i \in A} x_i,$$

where $\eta \in (0, 1/2]$ and $\sigma \in \{-1, +1\}$. Since $\chi_A(x) \in \{-1, +1\}$ and has zero mean under P_0, P_A is a valid probability distribution.

The base-arm values are obtained through the affine embedding $v_i = a + h(1 + X_i)/2$, where $a, h \geq 0$ and $a + h \leq 1$.

Exact marginal hiding. For every $S \subseteq [N]$ such that $A \not\subseteq S$, the marginal law of X_S under P_A is exactly the same as under P_0: choosing any $i \in A \setminus S$ and marginalizing over X_i kills the parity term because $\sum_{x_i \in \{-1,+1\}} x_i = 0$.

Thus, before jointly selecting A, the complete observation has exactly the null law: there is no lower-order signal to accumulate.

3.2 FROM A HIDDEN INTERACTION TO A REWARD GAP

For $1 \leq s \leq K$, define the order-s parity-family certificate by

$$\Gamma_s(g; K) := \sup_{\substack{a,h \geq 0 \\ a+h \leq 1}} 2^{-s} \left| \mathbb{E}_{J \sim \mathrm{Bin}(K-s, 1/2)} \left[\Delta_h^s g(Ka + hJ)\right] \right|. \tag{1}$$

This is the largest reward-gap coefficient per unit perturbation strength within the chosen parity family.

Indeed, fix a, h and suppose $A \subseteq S$. Under P_A, the expected reward of S exceeds its null value by $\eta\sigma$ times the corresponding finite-difference term. Choosing σ to match its sign yields a gap $\eta\gamma_s(a, h)$, where $\gamma_s(a, h)$ denotes the quantity inside the supremum defining $\Gamma_s(g; K)$. If $A \not\subseteq S$, exact marginal hiding implies that the expected reward is exactly the null value. Thus actions containing A are optimal and all others incur this gap.

3.3 THE COST OF DISCOVERING A HIDDEN INTERACTION

Let $N_A(T) := \sum_{t=1}^{T} \mathbf{1}\{A \subseteq S_t\}$ be the number of rounds on which the learner jointly observes the hidden interaction. By exact marginal hiding, the KL divergence between the null and alternative

histories can increase only on these rounds. For $h > 0$, the one-round informative KL divergence is $-\frac{1}{2}\log(1-\eta^2) \le \eta^2$, and hence

$$D_{\mathrm{KL}}\big(\mathbb{P}_0^T \,\big\|\, \mathbb{P}_A^T\big) \le \eta^2\, \mathbb{E}_0[N_A(T)].$$

The case $h = 0$ has zero reward gap and is trivial. The second ingredient is purely combinatorial. Every selected K-subset contains exactly $\binom{K}{s}$ candidate hidden s-subsets. Therefore, pathwise,

$$\sum_{\substack{A\subseteq[N]\\|A|=s}} N_A(T) = \binom{K}{s} T.$$

Averaging over the $\binom{N}{s}$ possible hidden interactions, transferring coverage from the null to the alternatives via Pinsker's inequality, and optimizing the perturbation size η gives the main lower bound.

Theorem 1 (Semi-bandit interaction lower bound)**.** *Assume $N \ge 2K$. For every $s \in [K]$,*

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \ge \frac{1}{16}\,\Gamma_s(g; K)\,\psi_s(T). \tag{2}$$

The two factors separate reward amplitude from coverage: among $\binom{N}{s}$ candidate interactions, one action exposes only $\binom{K}{s}$, so $\psi_s(T) = \min\{T, \sqrt{((\binom{N}{s}/\binom{K}{s})T}\}$.

The assumption $N \ge 2K$ is only used to present the bound in this clean form. Appendix A gives the all-regime statement in terms of $q_s = \binom{K}{s}/\binom{N}{s}$, with factors of $1 - q_s$ that vanish when $K = N$.

Standard link classes follow by evaluating Γ_s.

3.4 Degree consequences

Polynomial links. Let $g(x) = a_d x^d + \cdots + a_0$ be a polynomial of degree $1 \le d < K$ with $a_d \ne 0$. Its d-th forward difference is independent of the evaluation point: $\Delta_h^d g(x) = a_d d! h^d$, so $\Gamma_d(g; K) = |a_d| d!/2^d$. Combining Theorem 1 with the scalar-feedback upper bound that semi-bandit feedback can simulate gives, for fixed K, d and $N \ge 2K$,

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) = \widetilde{\Theta}_{g,K,d}\Big(\min\{T, \sqrt{N^d T}\}\Big).$$

Genuinely higher-order links. Assume $g \in C^K([0, K])$ and that g is not a polynomial of degree at most $K - 1$. Then $g^{(K)}$ is not identically zero, so there exist an interior point x and a sufficiently small $h > 0$ for which $\Delta_h^K g(x) \ne 0$. Hence $\Gamma_K(g; K) > 0$; treating the $\binom{N}{K}$ actions as separate arms matches the resulting lower bound. Thus, for fixed K and $N \ge 2K$,

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) = \Theta_{g,K}\Big(\min\{T, \sqrt{N^K T}\}\Big).$$

These statements concern fixed K and the N, T exponent; they do not optimize the dependence on K. Appendix A.9 gives the certificates and complete proofs; Appendix A gives the all-regime lower bound, including the degenerate case $K = N$.

Degree identifies the asymptotic exponent. Retaining the size of each interaction is necessary to distinguish a dominant nonlinear component from one with vanishing weight.

4 Observable Interactions and Matching Rates

The lower bound charges the learner for discovering a hidden interaction. We now show that selected-coordinate feedback also lets the learner exploit weak interactions. The key is a decomposition that is observable on every chosen action.

4.1 OBSERVABLE SUBSET COEFFICIENTS

For a known bounded link $f : [0, K] \to \mathbb{R}$ and $U \subseteq [N]$ with $1 \le |U| \le K$, define

$$c_{f,U}(v_U) := \sum_{V \subseteq U} (-1)^{|U|-|V|} f\left(\sum_{i \in V} v_i\right).$$

For every feasible S, Boolean-lattice inversion gives, pathwise,

$$f\left(\sum_{i \in S} v_i\right) = f(0) + \sum_{\varnothing \neq U \subseteq S} c_{f,U}(v_U). \tag{3}$$

Whenever $U \subseteq S_t$, semi-bandit feedback reveals $v_{t,U}$ and hence $c_{f,U}(v_{t,U})$. For $s \in [K]$, let

$$b_s(f) := \sup_{u \in [0,1]^s} c_{f,[s]}(u) - \inf_{u \in [0,1]^s} c_{f,[s]}(u), \qquad \mathcal{B}_K(f; N) := \sum_{s=1}^{K} b_s(f) \sqrt{\binom{K}{s}\binom{N}{s}}.$$

The width b_s controls estimation of the observable order-s coefficients. We use the same regret definition for bounded real-valued links f.

Theorem 2 (Amplitude-sensitive semi-bandit upper bound)**.** *Let $M = \binom{N}{K} > 1$. For every known bounded link f, there exists an information-theoretic semi-bandit policy satisfying*

$$\mathfrak{R}_{\mathrm{SB}}(f, N, K, T) \le \min\left\{T \operatorname{osc}(f), \sqrt{2T \log M}\, \mathcal{B}_K(f; N)\right\}. \tag{4}$$

The proof lifts every nonempty subset to a semi-bandit coordinate and uses importance-weighted exponential weights. Let $S_t \sim p_t$, set $q_t(U) = \Pr_t(U \subseteq S_t)$, and draw an auxiliary $S' \sim p_t$ conditionally independently of S_t. For the shifted nonnegative coefficient losses $0 \le \lambda_{t,U} \le b_s(f)$, Cauchy–Schwarz gives, without assuming independence among overlapping observation events,

$$\mathbb{E}_t[H_{t,s}^2] \le \binom{K}{s} \sum_{|U|=s} \lambda_{t,U}^2 \le \binom{K}{s}\binom{N}{s} b_s(f)^2, \quad H_{t,s} := \sum_{\substack{U \subseteq S' \\ |U|=s}} \frac{\lambda_{t,U} \mathbf{1}\{U \subseteq S_t\}}{q_t(U)}. \tag{5}$$

The two inclusion probabilities cancel the squared importance weight. Minkowski across orders yields the complexity in equation 4; the remaining potential argument is standard. The policy may be computationally inefficient; details are in Appendix B.

The nonnegative shift is essential to the potential argument. Write $z_t(S) = \sum_{i \in S} v_{t,i}$, $u_s = \sup c_{f,[s]}$, and $\lambda_{t,U} = u_{|U|} - c_{f,U}(v_{t,U})$. Then

$$\sum_{\varnothing \neq U \subseteq S} \lambda_{t,U} = \sum_{s=1}^{K} \binom{K}{s} u_s + f(0) - f(z_t(S)).$$

Thus the shift leaves regret unchanged. Since the estimated losses are nonnegative, $e^{-x} \le 1 - x + x^2/2$ applies for every $x \ge 0$; no additional exploration is needed to bound their magnitudes.

Appendix B also combines the theorem with a surrogate comparison, allowing the learner to trade observable interaction width against uniform approximation error. In the square-root regime and for fixed K, each order-s term in the upper bound carries the same joint-coverage scale $\sqrt{T/q_s}$ as the lower certificate, but with the observable width b_s in place of Γ_s.

4.2 A MATCHING AMPLITUDE LAW

Consider the normalized polynomial mixtures and their endpoint reflections

$$g_\beta^+(z) := \sum_{j=1}^{K} \beta_j (z/K)^j, \qquad g_\beta^-(z) := 1 - \sum_{j=1}^{K} \beta_j (1 - z/K)^j, \qquad \beta_j \ge 0, \qquad \sum_{j=1}^{K} \beta_j = 1.$$

Both families map $[0, K]$ into $[0, 1]$ and have range one; g_β^+ is convex and g_β^- is concave. Define $F_\beta := \sum_{j=1}^{K} \beta_j \psi_j(T)$ and retain the stronger order-resolved certificate $L_\beta := \frac{1}{16} \max_s \{\beta_s s! \psi_s(T)/(2^s K^s)\}$.

Theorem 3 (Uniform amplitude law for polynomial mixtures). *Let $M = \binom{N}{K}$ and assume $N \geq 2K$. For either $g \in \{g_\beta^+, g_\beta^-\}$, uniformly over the coefficient simplex,*

$$c_K^\sharp F_\beta \leq L_\beta \leq \mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \leq 2\max\{1, \sqrt{\log M}\} F_\beta, \tag{6}$$

where $c_K^\sharp = (16 \sum_{s=1}^K 2^s K^s / s!)^{-1}$. Thus, for every fixed K, $\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) = \widetilde{\Theta}_K(F_\beta)$ uniformly over the full coefficient simplex.

The theorem closes a genuine max–sum gap. A hidden-interaction alternative isolates one order, so the lower-bound geometry is a maximum over s; the observable-interaction algorithm estimates every order simultaneously, so its cost is a sum. The occupancy weights below express each monomial as a distribution over subset orders, while positivity connects the same coefficient to a finite-difference certificate. The factor $c_K^\sharp$ is the uniform price of passing from the maximum to the sum. Crucially, this comparison is coefficientwise: β may depend on N or T and approach the boundary of the simplex. The theorem therefore captures vanishing nonlinear components that fixed-link degree statements leave inside their link-dependent constants.

For the positive family, let $\omega_{j,s} := \binom{K}{s} s! \left\{ {j \atop s} \right\} / K^j$. These are the probabilities that j independent uniform draws from $[K]$ visit exactly s distinct coordinates, so $\sum_{s=1}^j \omega_{j,s} = 1$. The exact monomial widths therefore give

$$\mathcal{B}_K(g_\beta^+; N) = \sum_{j=1}^K \beta_j \sum_{s=1}^j \frac{\omega_{j,s}}{\sqrt{q_s}} \leq \sum_{j=1}^K \frac{\beta_j}{\sqrt{q_j}}.$$

Truncating monomials whose estimation cost exceeds their residual cost gives the finite-horizon upper law. Positivity also gives $\Gamma_s(g_\beta^+; K) \geq \beta_s s!/(2^s K^s)$, hence L_β. Writing $a_s = s!/(2^s K^s)$, the inequality $F_\beta \leq 16 L_\beta \sum_s a_s^{-1}$ yields $c_K^\sharp$. The constant remains exponential in K.

For the reflected family, set $w_i = 1 - v_i$ and $f_\beta = 1 - g_\beta^+$. Then $g_\beta^-(\sum_{i \in S} v_i) = f_\beta(\sum_{i \in S} w_i)$. This observation-preserving transformation reduces the problem to f_β, whose nonempty Möbius coefficients are the negatives of those of g_β^+. The same upper and lower arguments apply. Appendix B gives the full proof.

Corollary 4 (A weak high-order component). *Assume $N \geq 2K$ and fix $2 \leq d \leq K$ and*

$$g_\epsilon(z) = (1 - \epsilon) z/K + \epsilon (z/K)^d, \qquad 0 \leq \epsilon \leq 1/2.$$

Uniformly in ϵ, N, T under these assumptions,

$$\mathfrak{R}_{\mathrm{SB}}(g_\epsilon, N, K, T) = \widetilde{\Theta}_{K,d}(\psi_1(T) + \epsilon \psi_d(T)). \tag{7}$$

A linear surrogate $p(z) = (1-\epsilon)z/K$ has uniform error ϵ, so the surrogate comparison pays a residual cost $2\epsilon T$. By contrast, once $T \geq q_d^{-1}$, Corollary 4 charges the order-d term at the statistical scale $\widetilde{O}_{K,d}(\epsilon \sqrt{T/q_d})$ within the matched total rate. The gain comes from estimating the weak nonlinear component through observable subset coefficients rather than treating it entirely as approximation bias.

When does the weak interaction matter? Write $r_s = q_s^{-1} = \binom{N}{s} / \binom{K}{s}$. Since $r_1 \leq r_d$, the relative size of the two terms in equation 7 is exactly

$$\frac{\epsilon \psi_d(T)}{\psi_1(T)} = \begin{cases} \epsilon, & T \leq r_1, \\ \epsilon \sqrt{T/r_1}, & r_1 < T \leq r_d, \\ \epsilon \sqrt{r_d/r_1}, & T > r_d. \end{cases} \tag{8}$$

For $\epsilon > 0$, the high-order term reaches the baseline only if $\epsilon \geq \epsilon_c := \sqrt{r_1/r_d}$, at $T_\times = r_1/\epsilon^2$. At $\epsilon = \epsilon_c$, the terms remain equal for $T \geq r_d$; strict dominance requires $\epsilon > \epsilon_c$. For $\epsilon = 0$, the high-order term is absent. These are exact crossovers of the normalized rate expression, not sharp transitions in the minimax value.

5 A Fixed-Range Feedback Separation

We now use the amplitude-sensitive upper bound to separate selected-coordinate feedback from Bernoulli scalar feedback while keeping the reward range fixed.

Theorem 5 (Fixed-range feedback separation)**.** *Let $N \geq 4$, $K = 2$, and $g_\epsilon(z) = (1-\epsilon)z/2 + \epsilon(z/2)^2$ for $0 < \epsilon \leq 1/2$. Writing $M = \binom{N}{2}$, the semi-bandit and Bernoulli-scalar minimax values satisfy*

$$\mathfrak{R}_{\mathrm{SB}}(g_\epsilon, N, 2, T) \leq \sqrt{2T \log M}\left[\left(\frac{1}{2} - \frac{\epsilon}{4}\right)\sqrt{2N} + \frac{\epsilon}{2}\sqrt{M}\right], \tag{9}$$

$$\mathfrak{R}_{\mathrm{Ber}}(g_\epsilon, N, 2, T) \geq c \min\{\epsilon T, \sqrt{MT}\} \tag{10}$$

for a universal $c > 0$. For any fixed $\alpha > 0$, let $\epsilon_N = N^{-\alpha}$ and $T_N = \lceil N^{2+2\alpha} \rceil$. For all sufficiently large N, the two minimax values satisfy

$$\begin{aligned} \mathfrak{R}_{\mathrm{SB}}(g_{\epsilon_N}, N, 2, T_N) &= \widetilde{\Theta}\left(N^2 + N^{3/2+\alpha}\right), \\ \mathfrak{R}_{\mathrm{Ber}}(g_{\epsilon_N}, N, 2, T_N) &= \Theta\left(N^{2+\alpha}\right). \end{aligned} \tag{11}$$

Consequently,

$$\frac{\mathfrak{R}_{\mathrm{SB}}(g_{\epsilon_N}, N, 2, T_N)}{\mathfrak{R}_{\mathrm{Ber}}(g_{\epsilon_N}, N, 2, T_N)} \lesssim \sqrt{\log N}\left(N^{-1/2} + N^{-\alpha}\right) \longrightarrow 0. \tag{12}$$

The link always satisfies $g_\epsilon(0) = 0$ and $g_\epsilon(2) = 1$, so the separation is not obtained by shrinking the reward range. The mechanism is instead a feedback-dependent attenuation: the parity perturbation is visible in the coordinates at scale η, whereas the Bernoulli mean changes only at scale $\epsilon\eta$. On an informative covering round, the two history channels therefore accumulate information at the respective scales

$$D_{\mathrm{KL}}^{\mathrm{coord}} = \Theta(\eta^2), \qquad D_{\mathrm{KL}}^{\mathrm{Ber}} = \Theta(\epsilon^2\eta^2).$$

For fixed testing accuracy on the covered interaction, the Bernoulli channel therefore requires an order $1/\epsilon^2$ more informative observations. No separation from a noiseless exact aggregate is claimed. Appendix B gives the proof.

When $0 < \alpha < 1/2$, the order-two term dominates F_β, and equation 11 simplifies to $\mathfrak{R}_{\mathrm{SB}} = \widetilde{\Theta}(N^2)$ while $\mathfrak{R}_{\mathrm{Ber}} = \Theta(N^{2+\alpha})$. Thus the polynomial separation occurs even in a regime where the nonlinear component, rather than the linear baseline, controls the semi-bandit rate. Figure 1 illustrates the normalized rate expression and separation certificate.

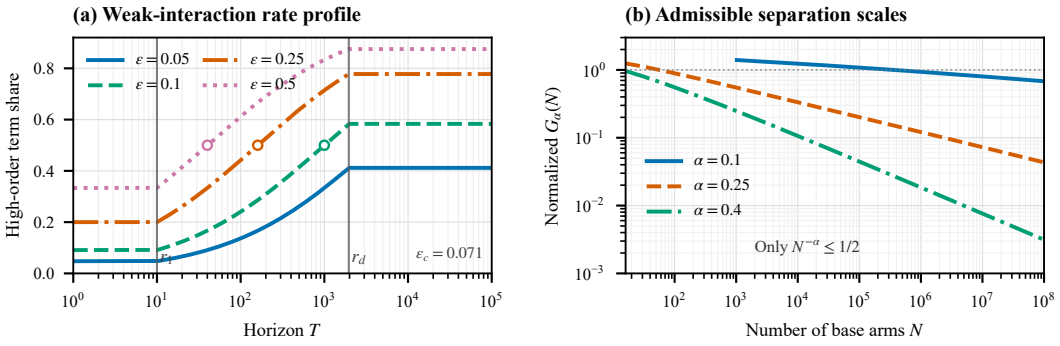

Figure 1: Expressions underlying the finite-horizon results. **Left:** $\epsilon\psi_d(T)/(\psi_1(T) + \epsilon\psi_d(T))$ in Corollary 4, for $N = 50$, $K = 5$, $d = 3$. Open circles mark equality at $T_\times = r_1/\epsilon^2$; below $\epsilon_c = \sqrt{r_1/r_d}$ the high-order term stays smaller. **Right:** $G_\alpha(N) = \sqrt{\log N}(N^{-1/2} + N^{-\alpha})$ from Theorem 5, restricted to $N^{-\alpha} \leq 1/2$. The left share belongs to the normalized rate expression; the right certificate suppresses a universal multiplicative constant, so $G_\alpha = 1$ is not a certified threshold for the regret ratio.

6 A Complementary Exact-Aggregate Consequence

The coordinate-level results above are distinct from a simpler reduction that needs only the exact selected aggregate $z_t = \sum_{i \in S_t} v_{t,i}$. Given a surrogate link p, the learner can feed $p(z_t)$ to an algorithm for p; if $\epsilon = \|g - p\|_\infty$, the pathwise comparison gives $R_T(g) \le R_T(p) + 2\epsilon T$. For $0 \le d < K$, write $E_d(g; K) := \inf_{\deg(p) \le d} \|g - p\|_{\infty,[0,K]}$.

Proposition 6 (Polynomial-approximation envelope)**.** *Let $M = \binom{N}{K}$ and $E_d = E_d(g; K)$. There are universal constants $C, C_0 \ge 1$ such that*

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \le \min \left\{ T,\ 2TE_0,\ \min_{1 \le d < K} \left[C \sqrt{\binom{N}{d} T \log M} + 2TE_d \right],\ C_0 \sqrt{(M-1)T} \right\}. \tag{13}$$

The alternatives use a constant surrogate, resolve the link to degree d, or treat each action separately. Since g is known, the minimizing degree can be selected before play for the chosen horizon. Appendix C proves the pathwise reduction, the exact $\binom{N}{d}$-dimensional subset lift, and the full range-sensitive statement. The one-sided interaction–approximation and hidden–observable bridges are recorded there and in Appendix B. The quantities Γ_s, b_s, and E_d therefore serve different statistical roles. The first is a lower certificate for a marginally hidden parity perturbation, the second is the width of an interaction exposed by selected-coordinate feedback, and the third is the bias incurred when exact aggregate feedback ignores structure above degree d. The one-sided bridges ensure consistency among these descriptions, not their equivalence. Positivity closes the lower–upper gap for the polynomial mixtures studied above; for general links, no minimax "effective order" is claimed.

MNL as a link-level example. The multinomial-logit link $g(x) = x/(1+x)$ is central to dynamic assortment learning (Caro & Gallien, 2007; Rusmevichientong et al., 2010; Sauré & Zeevi, 2013; Agrawal et al., 2019; Chen et al., 2021b). It is non-polynomial but accurately approximable at low degree. For every $s \in [K]$, the reciprocal finite-difference identity and a Chebyshev construction give

$$\Gamma_s(g; K) \ge \frac{s!}{2^{2s+1} K^s}, \qquad E_d(g; K) \le 2e^{-(d+1)\,\mathrm{arcosh}(1+2/K)}.$$

Theorem 1 and Proposition 6 therefore yield competing all-order lower certificates and a low-degree upper envelope. They illustrate the distinction between exact interaction order and finite-horizon approximability, but do not characterize the finite-horizon minimax value. Quantitatively, $\Gamma_1(g; K) \ge 1/(8K)$ whereas $\Gamma_K(g; K) = 1/[2^K(K+1)]$: the highest-order certificate fixes the eventual N exponent for fixed K, yet can be exponentially smaller than a low-order certificate that is more informative at accessible horizons. Appendix D gives the exact calculations, the full envelopes, and the degree-selection analysis.

7 Scope and Limitations

Our minimax equivalences characterize the dependence on N and T for fixed K in the nondegenerate regime $N \ge 2K$; the all-regime lower bound is given in Appendix A. We do not optimize the dependence on K: the uniform matching constant decays exponentially, and cancellations in broader signed link classes remain open. All upper bounds are information-theoretic and may be computationally inefficient. The strict feedback separation is from noisy scalar observations, not from a noiseless aggregate. Unknown links, noisy primitive observations, and efficient algorithms are left for future work.

Taken together, the results identify jointly observed subsets—rather than individual coordinates—as the reusable statistical units of nonlinear semi-bandit feedback. This viewpoint explains both the hidden-interaction barrier and the amplitude-sensitive gains established above.

AI USE STATEMENT

The authors used LLM tools after preparing the initial manuscript. The tools served as interactive editorial and review aids: they assisted with the organization and wording of the paper, searches for potentially relevant literature, and refinements to the presentation and derivation of selected results. They were also used to stress-test the manuscript by flagging potential ambiguities, missing assumptions, overstatements, and edge cases, and to assist with LaTeX editing and with writing and revising code for the theorem-derived numerical illustrations.

The research problem, model, core constructions, and initial mathematical arguments were supplied by the authors. All tool outputs were treated as unverified suggestions. The authors manually rederived and verified the mathematical claims and proofs, checked the cited sources, inspected the illustration code and its outputs, and made all final decisions about the paper. No generative AI was used to generate datasets or empirical results. The authors take full responsibility for the final content.

REFERENCES

Shipra Agrawal, Vashist Avadhanula, Vineet Goyal, and Assaf Zeevi. MNL-bandit: A dynamic learning approach to assortment selection. *Operations Research*, 67(5):1453–1485, 2019. doi: 10.1287/opre.2018.1832.

Noga Alon, Nicolò Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Proceedings of the 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pp. 23–35. PMLR, 2015. URL https://proceedings.mlr.press/v40/Alon15.html.

Jean-Yves Audibert and Sébastien Bubeck. Minimax policies for adversarial and stochastic bandits. In *Proceedings of the 22nd Annual Conference on Learning Theory*, pp. 217–226, 2009.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2014. doi: 10.1287/moor.2013.0598.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pp. 322–331, 1995. doi: 10.1109/SFCS.1995.492488.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002. doi: 10.1137/S0097539701398375.

Gábor Bartók, Dean P. Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring—classification, regret bounds, and algorithms. *Mathematics of Operations Research*, 39(4):967–997, 2014. doi: 10.1287/moor.2014.0663.

Christoph Bauschmann and Setareh Maghsudi. Graph structured combinatorial semi-bandit with nonlinear reward associations through separable signals. *arXiv preprint arXiv:2606.14650*, 2026. URL https://arxiv.org/abs/2606.14650.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012. doi: 10.1561/2200000024.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Sham M. Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pp. 41.1–41.14. PMLR, 2012. URL https://proceedings.mlr.press/v23/bubeck12a.html.

Felipe Caro and Jérémie Gallien. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*, 53(2):276–292, 2007. doi: 10.1287/mnsc.1060.0613.

Nicolò Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012. doi: 10.1016/j.jcss.2012.01.001.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 151–159. PMLR, 2013. URL https://proceedings.mlr.press/v28/chen13a.html.

Wei Chen, Wei Hu, Fu Li, Jian Li, Yu Liu, and Pinyan Lu. Combinatorial multi-armed bandit with general reward functions. In *Advances in Neural Information Processing Systems*, volume 29, pp. 1651–1659, 2016a.

Wei Chen, Yajun Wang, Yang Yuan, and Qinshi Wang. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *Journal of Machine Learning Research*, 17(50):1–33, 2016b. URL https://jmlr.org/papers/v17/14-298.html.

Wei Chen, Liwei Wang, Haoyu Zhao, and Kai Zheng. Combinatorial semi-bandit in the non-stationary environment. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 865–875. PMLR, 2021a. URL https://proceedings.mlr.press/v161/chen21a.html.

Xi Chen and Yining Wang. A note on a tight lower bound for capacitated MNL-bandit assortment selection models. *Operations Research Letters*, 46(5):534–537, 2018. doi: 10.1016/j.orl.2018.08.006.

Xi Chen, Yining Wang, and Yuan Zhou. Optimal policy for dynamic assortment planning under multinomial logit models. *Mathematics of Operations Research*, 46(4):1639–1657, 2021b. doi: 10.1287/moor.2021.1133.

Richard Combes, Mohammad Sadegh Talebi, Alexandre Proutière, and Marc Lelarge. Combinatorial bandits revisited. In *Advances in Neural Information Processing Systems*, volume 28, pp. 2116–2124, 2015.

Yi Gai, Bhaskar Krishnamachari, and Ramesh Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. *IEEE/ACM Transactions on Networking*, 20(5):1466–1478, 2012. doi: 10.1109/TNET.2011.2181864.

Yanjun Han, Yining Wang, and Xi Chen. Adversarial combinatorial bandits with general non-linear reward functions. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 4030–4039. PMLR, 2021. URL https://proceedings.mlr.press/v139/han21b.html.

Nicholas J. A. Harvey, Christopher Liaw, and Tasuku Soma. Improved algorithms for online submodular maximization via first-order regret bounds. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Yuko Kuroki, Liyuan Xu, Atsushi Miyauchi, Junya Honda, and Masashi Sugiyama. Polynomial-time algorithms for multiple-arm identification with full-bandit feedback. *Neural Computation*, 32(9):1733–1773, 2020. doi: 10.1162/neco_a_01299.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvári. Tight regret bounds for stochastic combinatorial semi-bandits. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 535–543. PMLR, 2015. URL https://proceedings.mlr.press/v38/kveton15.html.

Xutong Liu, Jinhang Zuo, Siwei Wang, Carlee Joe-Wong, John C. S. Lui, and Wei Chen. Batch-size independent regret bounds for combinatorial semi-bandits with probabilistically triggered arms or independent arms. In *Advances in Neural Information Processing Systems*, volume 35, pp. 14904–14916, 2022.

Nadav Merlis and Shie Mannor. Batch-size independent regret bounds for the combinatorial multi-armed bandit problem. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 2465–2489. PMLR, 2019. URL https://proceedings.mlr.press/v99/merlis19a.html.

Nadav Merlis and Shie Mannor. Tight lower bounds for combinatorial multi-armed bandits. In *Proceedings of the Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 2830–2857. PMLR, 2020. URL https://proceedings.mlr.press/v125/merlis20a.html.

Gergely Neu. First-order regret bounds for combinatorial semi-bandits. In *Proceedings of the 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pp. 1360–1375. PMLR, 2015. URL https://proceedings.mlr.press/v40/Neu15.html.

Stephen U. Pasteris, Alberto Rumi, Fabio Vitale, and Nicolò Cesa-Bianchi. Sum-max submodular bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2323–2331. PMLR, 2024. URL https://proceedings.mlr.press/v238/pasteris24a.html.

Idan Rejwan and Yishay Mansour. Top-k combinatorial bandits with full-bandit feedback. In *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, volume 117 of *Proceedings of Machine Learning Research*, pp. 752–776. PMLR, 2020. URL https://proceedings.mlr.press/v117/rejwan20a.html.

Paat Rusmevichientong, Zuo-Jun Max Shen, and David B. Shmoys. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research*, 58(6):1666–1680, 2010. doi: 10.1287/opre.1100.0866.

Denis Sauré and Assaf Zeevi. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, 15(3):387–404, 2013. doi: 10.1287/msom.2013.0429.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. Best-of-k bandits. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 1440–1489. PMLR, 2016. URL https://proceedings.mlr.press/v49/simchowitz16.html.

Matthew J. Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Information Processing Systems*, volume 21, pp. 1577–1584, 2008.

Qinshi Wang and Wei Chen. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Siwei Wang and Wei Chen. Thompson sampling for combinatorial semi-bandits. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5114–5122. PMLR, 2018. URL https://proceedings.mlr.press/v80/wang18a.html.

Yuxiao Wen. Adversarial combinatorial semi-bandits with graph feedback. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 66419–66436. PMLR, 2025. URL https://proceedings.mlr.press/v267/wen25a.html.

A Interaction Lower Bound and Minimax Consequences

This appendix proves the interaction lower bound stated in Theorem 1. We first establish an all-regime result that keeps explicit track of the coverage fraction

$$q_s := \frac{\binom{K}{s}}{\binom{N}{s}}.$$

The factor $1 - q_s$ is necessary when K is close to N: if $K = N$, the learner has only one feasible action and the minimax regret is identically zero. The cleaner statement in the main text follows because $q_s \le 1/2$ whenever $N \ge 2K$.

A.1 An all-regime statement

Theorem 7 (All-regime interaction lower bound)**.** *Let $N \ge K \ge s \ge 1$ and $T \ge 1$. Then*

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \ge \frac{1}{4}\,\Gamma_s(g; K)\min\left\{(1-q_s)T, (1-q_s)^2\sqrt{\frac{T}{q_s}}\right\}.$$

Consequently, if $N \ge 2K$, then

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \ge \frac{1}{16}\,\Gamma_s(g; K)\min\left\{T, \sqrt{\frac{\binom{N}{s}}{\binom{K}{s}}\,T}\right\}.$$

The numerical constants are not optimized. The proof proceeds through a family of i.i.d. environments indexed by a hidden s-subset. Averaging over these random environments produces a distribution over deterministic oblivious sequences, so the resulting Bayes bound lower bounds the minimax value defined in Section 2.

A.2 The parity-perturbed family

If $N = K$, the learner has one action and the claimed bound is zero. We therefore assume $N > K$ below, so $0 < q_s < 1$. Fix $a, h \ge 0$ with $a + h \le 1$, and define

$$G_s(a, h) := 2^{-s}\mathbb{E}_{J\sim\mathrm{Bin}(K-s,1/2)}\left[\Delta_h^s g(Ka + hJ)\right].$$

We first prove a lower bound with $|G_s(a, h)|$ in place of $\Gamma_s(g; K)$ and then take a supremum over (a, h). If $h = 0$, then $G_s(a, 0) = 0$ and the desired bound for this pair is trivial. Hence the nontrivial argument below assumes $h > 0$; in that case the two-point affine embedding is injective and preserves the likelihood-ratio calculation.

Let P_0 be the uniform distribution on $\{-1, +1\}^N$. For each $A \subseteq [N]$ with $|A| = s$, let

$$\chi_A(x) := \prod_{i\in A} x_i$$

and define

$$P_A(x) = 2^{-N}\big(1 + \eta\sigma\chi_A(x)\big), \qquad x \in \{-1, +1\}^N, \tag{14}$$

where $0 < \eta \le 1/2$ and $\sigma \in \{-1, +1\}$ is chosen so that $\sigma G_s(a, h) = |G_s(a, h)|$. If $G_s(a, h) = 0$, the desired lower bound for this pair (a, h) is trivial and the choice of σ is immaterial.

Because χ_A is balanced under P_0 and takes values in $\{-1, +1\}$, the right-hand side of equation 14 is nonnegative and sums to one. Thus P_A is a probability distribution.

At every round t, the environment draws X_t independently from P_0 under the null and from P_A under alternative A. It then assigns the base-arm values

$$v_{t,i} = a + \frac{h}{2}(1 + X_{t,i}), \qquad i \in [N]. \tag{15}$$

The constraint $a + h \le 1$ ensures that $v_{t,i} \in [0, 1]$.

A.3 EXACT MARGINAL INVARIANCE

We begin with the property that makes the construction compatible with semi-bandit feedback.

Lemma 8 (Exact marginal invariance)**.** *For every $S \subseteq [N]$ such that $A \nsubseteq S$, the marginal law of X_S under P_A is identical to its marginal law under P_0.*

Proof. Fix S with $A \nsubseteq S$ and choose $i \in A \setminus S$. For every $x_S \in \{-1,+1\}^{|S|}$, marginalizing equation 14 over the unobserved coordinates includes the sum

$$\sum_{x_i \in \{-1,+1\}} x_i = 0.$$

Consequently, the perturbation term has zero marginal mass at every value of x_S. The remaining uniform term gives $P_A(X_S = x_S) = 2^{-|S|} = P_0(X_S = x_S)$. □

In particular, if a selected action does not contain all of A, then the entire semi-bandit observation has the null distribution. This statement is distributional, not merely a matching of finitely many moments.

A.4 THE FINITE-DIFFERENCE REWARD IDENTITY

Under the null, all K-subsets have the same expected reward. Denote this common value by

$$\mu_0(a,h) := \mathbb{E}_0[g(Ka + hB)], \qquad B \sim \mathrm{Bin}(K, 1/2).$$

The next lemma identifies the exact reward shift under an alternative.

Lemma 9 (Reward separation)**.** *For every $S \in \mathcal{S}_K$,*

$$\mathbb{E}_A\left[g\left(\sum_{i\in S} v_i\right)\right] = \begin{cases} \mu_0(a,h) + \eta|G_s(a,h)|, & A \subseteq S, \\ \mu_0(a,h), & A \nsubseteq S. \end{cases}$$

Proof. If $A \nsubseteq S$, the conclusion follows immediately from Lemma 8 and the fact that the reward is a measurable function of X_S.

Now suppose that $A \subseteq S$. Since $P_A/P_0 = 1 + \eta\sigma\chi_A$,

$$\begin{aligned} &\mathbb{E}_A\left[g\left(\sum_{i\in S} v_i\right)\right] - \mathbb{E}_0\left[g\left(\sum_{i\in S} v_i\right)\right] \\ &\quad = \eta\sigma\,\mathbb{E}_0\left[\chi_A(X) g\left(Ka + h\sum_{i\in S}\frac{1+X_i}{2}\right)\right]. \end{aligned}$$

Let

$$J := \sum_{i\in S\setminus A} \frac{1+X_i}{2}.$$

Under P_0, $J \sim \mathrm{Bin}(K-s, 1/2)$ and is independent of X_A. Conditional on J, summing over X_A gives

$$\begin{aligned} &\mathbb{E}_0\left[\chi_A(X) g\left(Ka + hJ + h\sum_{i\in A}\frac{1+X_i}{2}\right)\,\middle|\, J\right] \\ &\quad = 2^{-s}\sum_{\ell=0}^{s}(-1)^{s-\ell}\binom{s}{\ell} g(Ka + hJ + h\ell) \\ &\quad = 2^{-s}\Delta_h^s g(Ka + hJ). \end{aligned}$$

Averaging over J shows that the reward shift is $\eta\sigma G_s(a,h) = \eta|G_s(a,h)|$. □

Thus every action containing A is optimal under alternative A, and every action not containing A has the same one-round expected gap $\eta|G_s(a,h)|$.

A.5 Adaptive information accumulation

Fix an arbitrary, possibly randomized, learner. Let H_T denote its complete interaction history through round T, including its selected actions and semi-bandit observations. Rewards need not be added separately because they are deterministic functions of the selected coordinates and the known link. Write $\mathbb{P}_0^T$ and $\mathbb{P}_A^T$ for the laws of H_T under the null and alternative A, respectively, and let $\mathbb{E}_0$ and $\mathbb{E}_A$ denote the corresponding expectations.

Define the number of rounds on which the learner covers A by

$$N_A(T) := \sum_{t=1}^{T} \mathbf{1}\{A \subseteq S_t\}.$$

Lemma 10 (Adaptive KL identity)**.** *For every $A \subseteq [N]$ with $|A| = s$,*

$$D_{\mathrm{KL}}\big(\mathbb{P}_0^T \,\big\|\, \mathbb{P}_A^T\big) = \kappa(\eta)\, \mathbb{E}_0[N_A(T)], \qquad \kappa(\eta) := -\frac{1}{2}\log(1-\eta^2).$$

In particular, for $0 < \eta \le 1/2$,

$$D_{\mathrm{KL}}\big(\mathbb{P}_0^T \,\big\|\, \mathbb{P}_A^T\big) \le \eta^2 \mathbb{E}_0[N_A(T)].$$

Proof. Apply the chain rule for relative entropy to the successive conditional laws of the action and observation at each round. Conditional on a realized past, the learner uses the same action kernel under the null and alternative, so the action itself contributes no conditional KL divergence.

If $A \nsubseteq S_t$, Lemma 8 shows that the conditional observation laws coincide, and the conditional KL contribution is zero. If $A \subseteq S_t$, the likelihood ratio of the selected observation under P_A relative to P_0 is $1 + \eta\sigma\chi_A$. Under P_0, χ_A is equally likely to be -1 or $+1$, and hence

$$\begin{aligned} D_{\mathrm{KL}}(P_{0,S_t} \| P_{A,S_t}) &= -\frac{1}{2}\log(1+\eta) - \frac{1}{2}\log(1-\eta) \\ &= -\frac{1}{2}\log(1-\eta^2) = \kappa(\eta). \end{aligned}$$

Summing the conditional contributions and taking expectation under the null proves the identity. Finally, for $\eta \le 1/2$, $-\frac{1}{2}\log(1-\eta^2) \le \eta^2$. □

The identity makes the information constraint exact: observing any proper subset of A contributes zero information about the alternative.

A.6 Coverage and change of measure

Let

$$\mathcal{A}_s := \{A \subseteq [N] : |A| = s\}, \qquad M_s := |\mathcal{A}_s| = \binom{N}{s}.$$

For every realized action sequence, independently of the learner and the environment,

$$\begin{aligned} \sum_{A \in \mathcal{A}_s} N_A(T) &= \sum_{t=1}^{T} \sum_{A \in \mathcal{A}_s} \mathbf{1}\{A \subseteq S_t\} \\ &= \binom{K}{s} T. \end{aligned} \tag{16}$$

Consequently,

$$\frac{1}{M_s} \sum_{A \in \mathcal{A}_s} \mathbb{E}_0[N_A(T)] = q_s T. \tag{17}$$

For any random variable $Z \in [0, T]$, total variation and Pinsker's inequality give

$$\mathbb{E}_A[Z] \le \mathbb{E}_0[Z] + T\ \mathrm{TV}(\mathbb{P}_A^T, \mathbb{P}_0^T) \le \mathbb{E}_0[Z] + T\sqrt{\frac{1}{2} D_{\mathrm{KL}}(\mathbb{P}_0^T \| \mathbb{P}_A^T)}.$$

Applying this inequality to $Z = N_A(T)$ and using Lemma 10,

$$\mathbb{E}_A[N_A(T)] \le \mathbb{E}_0[N_A(T)] + \eta T \sqrt{\frac{\mathbb{E}_0[N_A(T)]}{2}}.$$

Averaging over A, using equation 17, and applying Jensen's inequality to the square root yields

$$\frac{1}{M_s} \sum_{A \in \mathcal{A}_s} \mathbb{E}_A[N_A(T)] \le q_s T + \eta T \sqrt{\frac{q_s T}{2}}. \tag{18}$$

A.7 Regret calculation and optimization

For each alternative A, fix any comparator $S^A \in \mathcal{S}_K$ containing A. Such a comparator exists because $s \le K$. Lemma 9 and independence across rounds imply

$$\begin{aligned}\mathbb{E}_A[R_T] &\ge \mathbb{E}_A\left[\sum_{t=1}^{T} g\left(\sum_{i \in S^A} v_{t,i}\right) - \sum_{t=1}^{T} g\left(\sum_{i \in S_t} v_{t,i}\right)\right] \\ &= \eta |G_s(a,h)| \left(T - \mathbb{E}_A[N_A(T)]\right).\end{aligned}$$

Average this inequality over A and use equation 18 to obtain

$$\frac{1}{M_s} \sum_{A \in \mathcal{A}_s} \mathbb{E}_A[R_T] \ge \eta |G_s(a,h)| T \left(1 - q_s - \eta \sqrt{\frac{q_s T}{2}}\right). \tag{19}$$

Choose

$$\eta = \min\left\{\frac{1}{2}, \frac{1 - q_s}{\sqrt{2 q_s T}}\right\}. \tag{20}$$

Then

$$\eta \sqrt{\frac{q_s T}{2}} \le \frac{1 - q_s}{2},$$

so equation 19 is at least

$$\frac{1}{2} \eta |G_s(a,h)| (1 - q_s) T.$$

If the first term in equation 20 is active, this lower bound is $|G_s(a,h)|(1-q_s)T/4$. If the second term is active, it is

$$\frac{|G_s(a,h)|}{2\sqrt{2}} (1 - q_s)^2 \sqrt{\frac{T}{q_s}}.$$

In either case,

$$\frac{1}{M_s} \sum_{A \in \mathcal{A}_s} \mathbb{E}_A[R_T] \ge \frac{1}{4} |G_s(a,h)| \min\left\{(1 - q_s)T, (1 - q_s)^2 \sqrt{\frac{T}{q_s}}\right\}. \tag{21}$$

It remains to connect this Bayes calculation to the adversarial minimax value. For the fixed learner under consideration, the average in equation 21 implies that at least one alternative A attains the displayed lower bound. Moreover, the expected regret under that i.i.d. alternative is an average of the learner's regret over deterministic sequences $v_{1:T} \in ([0,1]^N)^T$. It therefore cannot exceed the supremum of the learner's regret over deterministic sequences. Since the learner was arbitrary, taking the infimum over policies proves the all-regime lower bound with $|G_s(a,h)|$.

Finally, take the supremum over all admissible (a,h). The supremum in the definition of $\Gamma_s(g;K)$ need not be attained; applying equation 21 to an approximating sequence of pairs and passing to the limit gives the first claim of Theorem 7.

If $N \ge 2K$, then

$$q_s = \prod_{j=0}^{s-1} \frac{K - j}{N - j} \le \left(\frac{K}{N}\right)^s \le \frac{1}{2}.$$

Hence $1 - q_s \geq 1/2$ and $(1 - q_s)^2 \geq 1/4$. Substituting these inequalities into the all-regime result gives

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \geq \frac{1}{16}\Gamma_s(g; K) \min\left\{T, \sqrt{\frac{T}{q_s}}\right\},$$

which is the second claim and proves Theorem 1.

A.8 What the proof uses

The argument depends on three exact identities rather than an asymptotic approximation. First, every marginal omitting a member of A is identical under the null and alternative. Second, the reward separation is precisely a finite difference of the known link. Third, every action covers exactly $\binom{K}{s}$ candidate hidden interactions. Adaptivity affects the random allocation of coverage across the alternatives, but it cannot change any of these three identities.

A.9 Finite-difference certificates for link classes

We next derive the degree consequences stated in Section 3.4. The lower bounds follow from explicit finite-difference certificates, while the upper bounds follow by simulating scalar bandit feedback and invoking the corresponding bandit-feedback algorithms.

We will use the following standard integral representation. If $f \in C^s([0, K])$ and $x, x + sh \in [0, K]$, then

$$\Delta_h^s f(x) = \int_{[0,h]^s} f^{(s)}(x + u_1 + \cdots + u_s)\, du_1 \cdots du_s. \tag{22}$$

For completeness, the identity follows by applying the fundamental theorem of calculus to $\Delta_h f(x) = \int_0^h f'(x + u)\, du$ and iterating s times.

A.9.1 Polynomial links

Let

$$g(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_0, \qquad 1 \leq d < K, \qquad a_d \neq 0.$$

The d-th finite difference annihilates every monomial of degree below d and satisfies

$$\Delta_h^d x^d = d! h^d.$$

Consequently,

$$\Delta_h^d g(x) = a_d d! h^d \qquad \text{for every } x, h. \tag{23}$$

The binomial expectation in the definition of $\Gamma_d(g; K)$ therefore has no dependence on J or on a. Since every admissible pair satisfies $0 \leq h \leq 1$, and $(a, h) = (0, 1)$ is admissible, we obtain the exact identity

Lemma 11 (Polynomial interaction certificate)**.** *If g is a polynomial of degree $d \in \{1, \ldots, K - 1\}$ with leading coefficient $a_d \neq 0$, then*

$$\Gamma_d(g; K) = \frac{|a_d| d!}{2^d}.$$

Combining Lemma 11 with Theorem 7 gives, for $N \geq 2K$,

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \geq \frac{|a_d| d!}{2^{d+4}} \min\left\{T, \sqrt{\frac{\binom{N}{d}}{\binom{K}{d}} T}\right\}. \tag{24}$$

The restriction $d \geq 1$ is substantive. If g is constant, all actions have the same reward on every round and the minimax regret is zero.

A.9.2 GENUINELY K-TH-ORDER LINKS

At the highest interaction order, the binomial variable in the definition of Γ_K is identically zero, so

$$\Gamma_K(g;K) = 2^{-K} \sup_{\substack{a,h\geq 0\\ a+h\leq 1}} \left|\Delta_h^K g(Ka)\right|. \tag{25}$$

Lemma 12 (Nonvanishing highest-order certificate)**.** *Suppose $g \in C^K([0,K])$. Then*

$$\Gamma_K(g;K) = 0 \quad \Longleftrightarrow \quad g \text{ is a polynomial of degree at most } K-1.$$

Proof. If g is a polynomial of degree at most $K-1$, then $\Delta_h^K g \equiv 0$, and hence $\Gamma_K(g;K) = 0$.

Conversely, suppose that g is not a polynomial of degree at most $K-1$. Then $g^{(K)}$ is not identically zero. By continuity, there is a nonempty interval $I \subseteq [0,K]$ on which $g^{(K)}$ has a fixed nonzero sign. We may choose $x \in [0,K)$ and $h > 0$ such that

$$[x, x+Kh] \subseteq I.$$

Writing $a = x/K$, this choice satisfies $a + h \leq 1$. The integral identity equation 22 gives

$$\Delta_h^K g(Ka) = \int_{[0,h]^K} g^{(K)}(Ka + u_1 + \cdots + u_K)\, du_1 \cdots du_K,$$

which is nonzero because the integrand has a fixed nonzero sign. Therefore $\Gamma_K(g;K) > 0$. □

Applying Theorem 7 at $s = K$ yields, for $N \geq 2K$,

$$\mathfrak{R}_{\mathrm{SB}}(g,N,K,T) \geq \frac{\Gamma_K(g;K)}{16} \min\left\{T, \sqrt{\binom{N}{K} T}\right\}. \tag{26}$$

This conclusion covers both non-polynomial links and polynomials of degree at least K, provided that $g \in C^K([0,K])$ and is not a polynomial of degree at most $K-1$.

A.10 TRANSFERRING SCALAR-FEEDBACK UPPER BOUNDS

The additional observations in the semi-bandit model cannot make the problem harder. The following simulation makes that monotonicity precise in the form needed here.

Lemma 13 (Feedback simulation)**.** *Suppose an algorithm for the scalar-feedback model has regret at most $B(g,N,K,T)$ when, after selecting S_t, it receives a Bernoulli observation with mean $g(\sum_{i\in S_t} v_{t,i})$. Then there is a semi-bandit algorithm with the same regret guarantee.*

Proof. After selecting S_t, a semi-bandit learner observes v_{t,S_t} and hence computes

$$z_t = \sum_{i\in S_t} v_{t,i} \qquad \text{and} \qquad g(z_t).$$

It may therefore draw, using its own randomness, $Y_t \sim \mathrm{Bernoulli}(g(z_t))$ and provide Y_t to the scalar-feedback algorithm. Conditional on the action and reward vector, the simulated observation has exactly the scalar-feedback law. The resulting action process and expected regret are consequently identical to those of the scalar-feedback algorithm. □

For a degree-d polynomial link, the polynomial-link algorithm of Han et al. (2021) gives

$$\mathfrak{R}_{\mathrm{B}}(g,N,K,T) \leq C_{g,d,K} \min\left\{T, \sqrt{N^d T \log N}\right\}. \tag{27}$$

Their reduction writes the degree-d reward as an adversarial linear-bandit loss in an N^d-dimensional tensor space and applies an EXP2-type algorithm with John's exploration. Lemma 13 transfers equation 27 directly to the semi-bandit model.

For an arbitrary bounded link, one may instead treat every member of $\mathcal{S}_K$ as a separate arm. A minimax adversarial multi-armed bandit algorithm, such as INF, gives

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \le C \min\left\{T, \sqrt{\binom{N}{K} T}\right\} \tag{28}$$

for a universal constant C. This bound requires only the realized scalar reward, which the semi-bandit learner computes from the observed selected coordinates.

A.11 Proofs of the degree consequences

We can now combine the explicit lower bounds with the preceding upper bounds.

Proof of the polynomial-link consequence. Let g be a polynomial of degree d with $1 \le d < K$ and leading coefficient $a_d \neq 0$. For fixed K, d and $N \ge 2K$,

$$\frac{\binom{N}{d}}{\binom{K}{d}} = \Theta_{K,d}(N^d). \tag{29}$$

Indeed,

$$\frac{\binom{N}{d}}{\binom{K}{d}} = \prod_{j=0}^{d-1} \frac{N-j}{K-j},$$

and every denominator depends only on K, d, while every numerator is between $N/2$ and N when $N \ge 2K$. The explicit lower bound equation 24 is therefore

$$\Omega_{g,K,d}\Big(\min\{T, \sqrt{N^d T}\}\Big).$$

The transferred scalar-feedback upper bound equation 27 is

$$\widetilde{O}_{g,K,d}\Big(\min\{T, \sqrt{N^d T}\}\Big).$$

Combining the two proves

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) = \widetilde{\Theta}_{g,K,d}\Big(\min\{T, \sqrt{N^d T}\}\Big)$$

throughout the fixed-(K, d) regime with $N \ge 2K$. □

Proof of the highest-order consequence. Suppose $g \in C^K([0, K])$ is not a polynomial of degree at most $K - 1$. Lemma 12 shows that $\Gamma_K(g; K) > 0$. Moreover, for fixed K and $N \ge 2K$,

$$\binom{N}{K} = \Theta_K(N^K).$$

The lower bound equation 26 and the action-as-arm upper bound equation 28 therefore give

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) = \Theta_{g,K}\Big(\min\{T, \sqrt{N^K T}\}\Big)$$

throughout the fixed-K regime with $N \ge 2K$. In particular, the equality holds without an additional logarithmic loss in this highest-order regime. □

Scope of the asymptotic statements. The restriction $N \ge 2K$ in the two degree consequences isolates the nondegenerate combinatorial regime and permits the compact powers N^d and N^K. When K is close to N, the exact lower bound is instead Theorem 7, whose factors of $1 - q_s$ correctly reflect the collapse of the action family. In particular, when $N = K$ there is only one feasible action and the regret is zero. We therefore do not claim a uniform $\Theta(\sqrt{N^d T})$ or $\Theta(\sqrt{N^K T})$ statement over all $N \ge K$.

B Proofs for Observable Interactions and Amplitude-Sensitive Regret

This appendix proves the results of Sections 4 and 5. All policies are information-theoretic. The primitive sequence $v_{1:T} \in ([0,1]^N)^T$ is fixed before learning begins, and regret is measured against the best fixed K-subset in hindsight.

B.1 Observable subset coefficients

For $U \subseteq [N]$ with $1 \le |U| \le K$, recall

$$c_{f,U}(v_U) = \sum_{V \subseteq U} (-1)^{|U|-|V|} f\left(\sum_{i \in V} v_i\right), \qquad c_{f,\varnothing} = f(0).$$

Lemma 14 (Boolean-lattice reconstruction)**.** *For every $S \subseteq [N]$ with $|S| \le K$ and $v \in [0,1]^N$,*

$$f\left(\sum_{i \in S} v_i\right) = \sum_{U \subseteq S} c_{f,U}(v_U).$$

Proof. Expanding the right-hand side, the coefficient multiplying $f(\sum_{i \in V} v_i)$ for a fixed $V \subseteq S$ is

$$\sum_{U : V \subseteq U \subseteq S} (-1)^{|U|-|V|} = (1-1)^{|S|-|V|}.$$

It vanishes unless $V = S$, in which case it equals one. □

If f is a polynomial of degree at most d, then $c_{f,U} \equiv 0$ whenever $|U| > d$. This follows either from the multinomial expansion or from the fact that an iterated finite difference of order greater than the polynomial degree vanishes.

B.2 Proof of the amplitude-sensitive upper bound

For each $s \in [K]$, choose known bounds

$$\ell_s \le c_{f,U}(u) \le u_s, \qquad u \in [0,1]^U, \qquad |U| = s,$$

and write $b_s = u_s - \ell_s$. The exact infimum and supremum defining $b_s(f)$ may be used; any certified wider interval also gives a valid bound. Let

$$m_s = \binom{K}{s}, \qquad D_s = \binom{N}{s}, \qquad \mathcal{B} = \sum_{s=1}^{K} b_s \sqrt{m_s D_s}.$$

Define the nonnegative coordinate losses

$$\lambda_{t,U} := u_{|U|} - c_{f,U}(v_{t,U}) \in [0, b_{|U|}]. \tag{30}$$

The induced action loss is

$$\begin{aligned} L_t(S) &:= \sum_{\varnothing \ne U \subseteq S} \lambda_{t,U} \\ &= \sum_{s=1}^{K} m_s u_s + f(0) - f\left(\sum_{i \in S} v_{t,i}\right). \end{aligned} \tag{31}$$

The first two terms on the last line are independent of S. Hence regret for the losses L_t is exactly regret for the original rewards.

Initialize a positive weight $w_1(S) = 1$ for each $S \in \mathcal{S}_K$. At round t, let

$$\rho_t(S) := \frac{w_t(S)}{\sum_{A \in \mathcal{S}_K} w_t(A)},$$

draw $S_t \sim \rho_t$, and define

$$q_t(U) := \Pr_{S\sim\rho_t}(U \subseteq S).$$

All weights remain strictly positive, so $q_t(U) > 0$ for every U with $|U| \le K$. Semi-bandit feedback reveals $\lambda_{t,U}$ for all $U \subseteq S_t$. Form the estimates

$$\widehat{\lambda}_{t,U} := \frac{\lambda_{t,U}\mathbf{1}\{U \subseteq S_t\}}{q_t(U)}, \qquad \widehat{L}_t(S) := \sum_{\varnothing\neq U\subseteq S} \widehat{\lambda}_{t,U}. \tag{32}$$

Conditionally on the past,

$$\mathbb{E}_t[\widehat{\lambda}_{t,U}] = \lambda_{t,U}, \qquad \mathbb{E}_t[\widehat{L}_t(S)] = L_t(S).$$

Update

$$w_{t+1}(S) = w_t(S)\exp(-\eta\widehat{L}_t(S)).$$

The central estimate is the following.

Lemma 15 (Order-sensitive conditional second moment)**.** *For every round t,*

$$\mathbb{E}_t\left[\sum_{S\in\mathcal{S}_K} \rho_t(S)\widehat{L}_t(S)^2\right] \le \mathcal{B}^2.$$

Proof. Let $S' \sim \rho_t$ be conditionally independent of S_t and define

$$H_s(S', S_t) := \sum_{\substack{U\subseteq S'\\ |U|=s}} \frac{\lambda_{t,U}\mathbf{1}\{U \subseteq S_t\}}{q_t(U)}.$$

Each S' contains exactly m_s subsets of size s. Cauchy–Schwarz gives

$$\begin{aligned}\mathbb{E}_t[H_s^2] &\le m_s \sum_{|U|=s} \frac{\lambda_{t,U}^2}{q_t(U)^2}\Pr(U \subseteq S')\Pr(U \subseteq S_t)\\ &= m_s \sum_{|U|=s} \lambda_{t,U}^2 \le m_s D_s b_s^2.\end{aligned}$$

The equality uses the conditional independence of S' and S_t; it does not assume independence among the overlapping subset-observation events. Since $\widehat{L}_t(S') = \sum_s H_s(S', S_t)$, Minkowski's inequality in L_2 yields

$$\left(\mathbb{E}_t[\widehat{L}_t(S')^2]\right)^{1/2} \le \sum_{s=1}^{K}\left(\mathbb{E}_t[H_s^2]\right)^{1/2} \le \sum_{s=1}^{K} b_s\sqrt{m_s D_s}.$$

Squaring proves the claim. □

We now complete the regret proof. Since every estimated loss is nonnegative, $e^{-x} \le 1 - x + x^2/2$ applies with $x = \eta\widehat{L}_t(S)$ without requiring an upper bound on x. The standard potential calculation gives, for every fixed $S^* \in \mathcal{S}_K$,

$$\sum_{t=1}^{T}\sum_{S} \rho_t(S)\widehat{L}_t(S) - \sum_{t=1}^{T}\widehat{L}_t(S^*) \le \frac{\log M}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{S}\rho_t(S)\widehat{L}_t(S)^2.$$

Because the primitive sequence is oblivious, a best comparator S^* is fixed and deterministic. Taking expectations, using conditional unbiasedness and Lemma 15, gives

$$\mathbb{E}[\mathrm{Reg}_T(f)] \le \frac{\log M}{\eta} + \frac{\eta T\mathcal{B}^2}{2}.$$

If $\mathcal{B} > 0$, choose $\eta = \sqrt{2\log M/(T\mathcal{B}^2)}$. This gives

$$\mathbb{E}[\mathrm{Reg}_T(f)] \le \sqrt{2T\log M}\,\mathcal{B}.$$

If $\mathcal{B} = 0$, every nonempty coefficient is identically zero: each such coefficient vanishes when one of its arguments is zero, so a zero-width coefficient must vanish everywhere. Lemma 14 then shows that every action has reward $f(0)$ and regret is zero. Finally, the pathwise bound $T\,\mathrm{osc}(f)$ proves Theorem 2.

For an unknown horizon, the same rate up to a universal constant follows from a standard doubling schedule.

B.3 An amplitude-aware surrogate envelope

Let p be any known surrogate. Running the preceding policy on the coefficients of p is implementable because semi-bandit feedback reveals all $v_{t,U}$ with $U \subseteq S_t$. The pathwise inequality

$$g(z_t(S)) - g(z_t(S_t)) \le p(z_t(S)) - p(z_t(S_t)) + 2\|g - p\|_\infty$$

gives

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \le \sqrt{2T \log M}\, \mathcal{B}_K(p; N) + 2T\|g - p\|_\infty. \tag{33}$$

Taking the infimum over bounded surrogates p gives the corresponding amplitude-aware approximation envelope.

B.4 Proof of the hidden–observable bridge

Proposition 16 (Hidden–observable interaction bridge)**.** *For every $s \in [K]$,*

$$\Gamma_s(g; K) \le \frac{1}{2} \sum_{r=s}^{K} \binom{K-s}{r-s} b_r(g).$$

Proof. Fix $s \in [K]$, an s-subset $A \subseteq [K]$, and admissible (a, h). Let $B_i = (1 + X_i)/2$ under the uniform Boolean law and $v_i = a + hB_i$. The finite-difference quantity in the definition of Γ_s is the Boolean Fourier coefficient

$$G_s(a, h) = \mathbb{E}_0\left[\chi_A(X) g\left(\sum_{i=1}^{K} v_i\right)\right].$$

Apply Lemma 14 to the reward. If $A \not\subseteq U$, a coordinate in $A \setminus U$ appears in χ_A but not in $c_{g,U}(v_U)$, so independence and centering make that term vanish. Hence

$$G_s(a, h) = \sum_{U \supseteq A} \mathbb{E}_0[\chi_A(X) c_{g,U}(v_U)].$$

For fixed U and conditional on $B_{U \setminus A}$, the sign χ_A is balanced. If a random variable Z lies in an interval of width b, then $|\mathbb{E}[\chi_A Z]| \le b/2$. Therefore a term with $|U| = r$ has absolute value at most $b_r(g)/2$. There are $\binom{K-s}{r-s}$ supersets of A of size r, so the triangle inequality gives

$$|G_s(a, h)| \le \frac{1}{2} \sum_{r=s}^{K} \binom{K-s}{r-s} b_r(g).$$

Taking the supremum over (a, h) proves Proposition 16. □

B.5 Positive polynomials

Let $f_j(z) = (z/K)^j$. We first compute its subset-coefficient range.

Lemma 17 (Monomial subset coefficients)**.** *For $1 \le s \le j$, the s-subset coefficient of f_j is a polynomial in its s arguments with nonnegative coefficients. Its exact range on $[0, 1]^s$ is*

$$0 \le c_{f_j,[s]}(u) \le \frac{s! \left\{ {j \atop s} \right\}}{K^j}.$$

For $s > j$, the coefficient vanishes identically.

Proof. Expand $(\sum_{i\in V} u_i)^j$ as a sum over maps $\varphi : [j] \to V$. Inclusion–exclusion over $V \subseteq [s]$ cancels every map whose image omits at least one element of $[s]$. Hence

$$c_{f_j,[s]}(u) = \frac{1}{K^j} \sum_{\substack{\varphi:[j]\to[s] \\ \mathrm{im}(\varphi)=[s]}} \prod_{r=1}^{j} u_{\varphi(r)}.$$

All coefficients are nonnegative. The minimum is attained when any coordinate is zero, and the maximum is attained at $u = (1, \dots, 1)$. The number of surjections from a j-element set onto an s-element set is $s!\left\{{j \atop s}\right\}$. □

For $g_\beta^+ = \sum_j \beta_j f_j$ with $\beta_j \geq 0$, the exact width at order s is therefore

$$b_s(g_\beta^+) = \sum_{j=s}^{K} \beta_j \frac{s!\left\{{j \atop s}\right\}}{K^j}.$$

Define

$$V_j := \sum_{s=1}^{j} \frac{s!\left\{{j \atop s}\right\}}{K^j} \sqrt{m_s D_s}.$$

Counting all functions from $[j]$ to $[K]$ by image size gives

$$\sum_{s=1}^{j} \binom{K}{s} s!\left\{{j \atop s}\right\} = K^j.$$

Thus, with

$$\omega_{j,s} := \frac{\binom{K}{s} s!\left\{{j \atop s}\right\}}{K^j},$$

we have $\omega_{j,s} \geq 0$, $\sum_s \omega_{j,s} = 1$, and

$$V_j = \sum_{s=1}^{j} \omega_{j,s} \sqrt{r_s}, \qquad r_s = \frac{D_s}{m_s}.$$

Since

$$\frac{r_{s+1}}{r_s} = \frac{N-s}{K-s} \geq 1,$$

we obtain $V_j \leq \sqrt{r_j}$. Theorem 2 therefore implies

$$\mathfrak{R}_{\mathrm{SB}}(g_\beta^+, N, K, T) \leq \sqrt{2T \log M} \sum_{j=1}^{K} \beta_j \sqrt{r_j}. \tag{34}$$

For short horizons, choose any $I \subseteq [K]$ and use the surrogate

$$p_I(z) = \sum_{j\in I} \beta_j (z/K)^j.$$

Because $(z/K)^j \in [0,1]$ on $[0, K]$,

$$\|g_\beta^+ - p_I\|_\infty \leq \sum_{j\notin I} \beta_j.$$

The amplitude-aware envelope and $V_j \leq \sqrt{r_j}$ give

$$\mathfrak{R}_{\mathrm{SB}}(g_\beta^+, N, K, T) \leq \sqrt{2T \log M} \sum_{j\in I} \beta_j \sqrt{r_j} + 2T \sum_{j\notin I} \beta_j.$$

Choosing membership in I separately for each j yields

$$\mathfrak{R}_{\mathrm{SB}}(g_\beta^+, N, K, T) \leq \sum_{j=1}^{K} \beta_j \min\{2T, \sqrt{2 r_j T \log M}\}. \tag{35}$$

Intersecting with the trivial bound T is harmless.

We now invoke the paper's all-order lower theorem in the form

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \geq \frac{1}{16}\,\Gamma_s(g; K)\min\{T, \sqrt{r_s T}\}, \qquad s \in [K], \tag{36}$$

where $N \geq 2K$.

At $a = 0, h = 1$, every forward difference $\Delta_1^s z^j(x)$ is nonnegative for $x \geq 0$ and $j \geq s$. The contribution of $\beta_s(z/K)^s$ is the constant $\beta_s s!/K^s$. Consequently,

$$\Gamma_s(g_\beta^+; K) \geq \frac{\beta_s s!}{2^s K^s}. \tag{37}$$

Combining equation 36 and equation 37,

$$\mathfrak{R}_{\mathrm{SB}}(g_\beta^+, N, K, T) \geq \frac{1}{16}\max_{s\in[K]}\left\{\frac{\beta_s s!}{2^s K^s}\min\{T, \sqrt{r_s T}\}\right\}.$$

The right-hand side is exactly L_β from the main text.

Set $a_s = s!/(2^s K^s)$ and $x_s = \beta_s \psi_s(T)$. Since $16L_\beta = \max_s a_s x_s$,

$$F_\beta = \sum_{s=1}^{K} x_s \leq 16 L_\beta \sum_{s=1}^{K} a_s^{-1}.$$

Consequently,

$$\mathfrak{R}_{\mathrm{SB}}(g_\beta^+, N, K, T) \geq L_\beta \geq c_K^\sharp F_\beta, \qquad c_K^\sharp := \left(16\sum_{s=1}^{K}\frac{2^s K^s}{s!}\right)^{-1}. \tag{38}$$

The inverse weights satisfy $a_{s-1}^{-1}/a_s^{-1} = s/(2K) \leq 1/2$ for $2 \leq s \leq K$. Thus

$$\frac{K!}{32\,2^K K^K} \leq c_K^\sharp \leq \frac{K!}{16\,2^K K^K}.$$

The constant is sharper than the bound obtained by taking the smallest a_s and dividing by $16K$, but it still decays exponentially with K.

Finally, for $x, y \geq 0$ and $a, b > 0$, $\min\{ax, by\} \leq \max\{a, b\}\min\{x, y\}$. Applying this termwise in equation 35 gives

$$\mathfrak{R}_{\mathrm{SB}}(g_\beta^+, N, K, T) \leq 2\max\{1, \sqrt{\log M}\}\, F_\beta.$$

Together with equation 38, this proves the positive-family part of Theorem 3. The weak-high-order corollary follows by setting $\beta_1 = 1 - \epsilon$, $\beta_d = \epsilon$, and using $1 - \epsilon \in [1/2, 1]$.

B.6 The reflected concave family

We prove the reflected-family part of Theorem 3. Let

$$g_\beta^-(z) = 1 - g_\beta^+(K - z), \qquad f_\beta(z) := 1 - g_\beta^+(z).$$

For any primitive vector $v \in [0, 1]^N$, set $w_i = 1 - v_i$. Since every feasible action has cardinality K,

$$g_\beta^-\left(\sum_{i\in S} v_i\right) = f_\beta\left(\sum_{i\in S} w_i\right).$$

The map $v \mapsto w$ is a bijection on primitive sequences, and either selected vector determines the other. Hence the two semi-bandit minimax problems have the same value. For every nonempty U, the constant in $f_\beta = 1 - g_\beta^+$ vanishes under Möbius inversion, so $c_{f_\beta, U} = -c_{g_\beta^+, U}$ and therefore $b_s(f_\beta) = b_s(g_\beta^+)$. Likewise, $\Delta_h^s f_\beta = -\Delta_h^s g_\beta^+$ for $s \geq 1$, giving $\Gamma_s(f_\beta; K) = \Gamma_s(g_\beta^+; K)$. For the finite-horizon upper bound, use $1 - \sum_{j\in I}\beta_j(z/K)^j$ as the surrogate for f_β. Its coefficient widths and uniform residual bound are the same as for the positive-family truncation. The lower bound uses the unchanged absolute finite differences. Both arguments therefore transfer with the same constants, proving the reflected-family claim. This argument does not require $b_s(g_\beta^-) = b_s(g_\beta^+)$, which need not hold.

B.7 NOISY SCALAR SEPARATION

We prove a slightly more general lower bound before specializing to $K = 2$. Fix $K \geq 2$, $N \geq 2K$, and

$$g_\epsilon(z) = (1-\epsilon)\frac{z}{K} + \epsilon\left(\frac{z}{K}\right)^K, \qquad 0 < \epsilon \leq \frac{1}{2}.$$

In the noisy scalar model, after selecting S_t the learner observes only

$$Y_t \sim \text{Bernoulli}\left(g_\epsilon\left(\sum_{i\in S_t} v_{t,i}\right)\right),$$

using a fresh random seed independent of the past at each round. Equivalently, conditional on the history $\mathcal{H}_{t-1}$, the current action S_t, and the fixed primitive sequence, the observation has the displayed Bernoulli law.

Let P_0 be uniform on $\{0,1\}^N$. For each hidden $A \in \mathcal{S}_K$, define

$$P_A(v) = 2^{-N}\left(1 + \eta\chi_A(v)\right), \qquad \chi_A(v) = \prod_{i\in A}(2v_i - 1), \qquad 0 < \eta \leq \frac{1}{2}.$$

If $S \neq A$, then $A \not\subseteq S$ and the marginal law of v_S is exactly the null marginal. If $S = A$, the expected reward increases by

$$\delta = \eta\epsilon a_K, \qquad a_K := \frac{K!}{2^K K^K}.$$

Indeed, the linear component has zero K-th finite difference and $\Delta_1^K (z/K)^K = K!/K^K$.

Under the null, for $B \sim \text{Bin}(K, 1/2)$ the scalar observation has mean

$$\mu_0 = \frac{1-\epsilon}{2} + \epsilon\mathbb{E}[(B/K)^K].$$

Since $(B/K)^K \leq B/K$, we have $\mu_0 \in [1/4, 1/2]$. Also $\delta \leq 1/4$, so $\mu_0 + \delta \in [1/4, 3/4]$. A mixture of Bernoulli variables is Bernoulli; hence an informative scalar observation has KL divergence

$$d_{\text{KL}}(\text{Ber}(\mu_0) \,\|\, \text{Ber}(\mu_0+\delta)) \leq \frac{\delta^2}{(\mu_0+\delta)(1-\mu_0-\delta)} \leq 6\delta^2. \tag{39}$$

Conversely, Pinsker's inequality for Bernoulli laws gives $d_{\text{KL}}(\text{Ber}(\mu_0)\|\,\text{Ber}(\mu_0+\delta)) \geq 2\delta^2$. For $K = 2$, $\delta = \eta\epsilon/8$, so the informative Bernoulli divergence is $\Theta(\epsilon^2\eta^2)$, compared with $-\frac{1}{2}\log(1-\eta^2) = \Theta(\eta^2)$ for coordinates. Let $N_A(T) = \sum_{t=1}^T \mathbf{1}\{S_t = A\}$ and $M = \binom{N}{K}$ and $q = 1/M$. The adaptive KL chain rule gives

$$D_{\text{KL}}(\mathbb{P}_0^T \| \mathbb{P}_A^T) \leq 6\delta^2\mathbb{E}_0[N_A(T)].$$

Under the null, pathwise $\sum_A N_A(T) = T$, and hence

$$\frac{1}{M}\sum_A \mathbb{E}_0[N_A(T)] = qT.$$

Pinsker's inequality and Cauchy–Schwarz imply

$$\begin{aligned}\frac{1}{M}\sum_A \mathbb{E}_A[N_A(T)] &\leq qT + \frac{T}{M}\sum_A \sqrt{\frac{1}{2}D_{\text{KL}}(\mathbb{P}_0^T\|\mathbb{P}_A^T)} \\ &\leq qT + T\delta\sqrt{3qT}.\end{aligned}$$

Comparing with the fixed action A, the average expected regret is therefore at least

$$\delta T\left(1 - q - \delta\sqrt{3qT}\right).$$

Choose

$$\delta = \min\left\{\frac{a_K\epsilon}{2}, \frac{1}{4\sqrt{3qT}}\right\},$$

which is realized by some $\eta \le 1/2$. Since $q \le 1/2$, the term in parentheses is at least $1/4$. Consequently,

$$\mathfrak{R}_{\mathrm{Ber}}(g_\epsilon, N, K, T) \ge \min\left\{\frac{a_K \epsilon T}{8}, \frac{\sqrt{MT}}{16\sqrt{3}}\right\} \ge c'_K \min\{\epsilon T, \sqrt{MT}\}, \tag{40}$$

where $c'_K = \min\{a_K/8, 1/(16\sqrt{3})\}$. The stochastic environments above form a distribution over deterministic oblivious primitive sequences, so the Bayes lower bound also lower bounds the worst-case expected regret.

For $K = 2$, the observable coefficients are

$$c_{\{i\}}(v_i) = \frac{1-\epsilon}{2} v_i + \frac{\epsilon}{4} v_i^2,$$
$$c_{\{i,j\}}(v_i, v_j) = \frac{\epsilon}{2} v_i v_j.$$

Their exact widths are $b_1 = 1/2 - \epsilon/4$ and $b_2 = \epsilon/2$. Substitution into Theorem 2 proves equation 9; equation 40 proves equation 10.

Fix $\alpha > 0$, take $\epsilon = N^{-\alpha}$, and let $T = \lceil N^{2+2\alpha} \rceil$. For all sufficiently large N, $\epsilon \le 1/2$, and $M = \Theta(N^2)$ gives

$$\mathfrak{R}_{\mathrm{SB}} = O\Big(\sqrt{\log N}\left[N^{3/2+\alpha} + N^2\right]\Big), \qquad \mathfrak{R}_{\mathrm{Ber}} = \Omega(N^{2+\alpha}).$$

This proves equation 11 and completes the proof of Theorem 5. Dividing the two displays proves equation 12.

The matching rates stated in the main text follow as well. Since $T \ge r_2$ for this choice of horizon, Theorem 3 gives

$$\mathfrak{R}_{\mathrm{SB}} = \widetilde{\Theta}\Big(\sqrt{NT} + \epsilon N\sqrt{T}\Big) = \widetilde{\Theta}\Big(N^{3/2+\alpha} + N^2\Big).$$

For the Bernoulli channel, a minimax adversarial M-armed bandit policy (Audibert & Bubeck, 2009) gives $O(\sqrt{MT})$ expected regret against the fixed mean-reward comparator. This bound applies to the bounded random observations by conditional unbiasedness. Combining it with equation 40 yields $\mathfrak{R}_{\mathrm{Ber}} = \Theta(N^{2+\alpha})$. For $0 < \alpha < 1/2$, moreover, $\epsilon\psi_2(T)/\psi_1(T) = N^{-\alpha}\sqrt{N-1} \to \infty$. The choice $\alpha = 1/4$ recovers the concrete scale $\epsilon = N^{-1/4}$, $T = \lceil N^{5/2} \rceil$; it is admissible for $N \ge 16$.

C Surrogate and Approximation Upper Bounds

This appendix proves the surrogate-feedback reduction and the resulting polynomial-approximation envelope. Only the exact selected aggregate is needed; individual coordinates play no role in this reduction. The learner can therefore run an algorithm for a simpler link without receiving that link as part of the environmental feedback.

Throughout this section, for $S \in \mathcal{S}_K$ and round t, write

$$z_t(S) := \sum_{i \in S} v_{t,i} \in [0, K].$$

For a link $f : [0, K] \to \mathbb{R}$, define the regret of an action sequence $S_1, \dots, S_T$ by

$$\mathrm{Reg}_T(f) := \max_{S \in \mathcal{S}_K} \sum_{t=1}^{T} f(z_t(S)) - \sum_{t=1}^{T} f(z_t(S_t)). \tag{41}$$

This notation is pathwise; expectations are introduced only when the action sequence is randomized.

C.1 A pathwise surrogate reduction

Theorem 18 (Surrogate-feedback reduction)**.** *Let $g, p : [0, K] \to \mathbb{R}$ and suppose*

$$\|g - p\|_{\infty,[0,K]} \le \epsilon.$$

Assume that an algorithm for link p, when given the selected scalar rewards $p(z_t(S_t))$, guarantees

$$\mathbb{E}[\mathrm{Reg}_T(p)] \le B_p(T)$$

against every admissible reward-vector sequence. Then there is an exact-aggregate algorithm for g satisfying

$$\mathbb{E}[\mathrm{Reg}_T(g)] \le B_p(T) + 2\epsilon T. \tag{42}$$

Proof. The learner runs the algorithm for p. After selecting S_t, exact-aggregate feedback supplies $z_t(S_t) = \sum_{i\in S_t} v_{t,i}$, and the learner passes $p(z_t(S_t))$ to the internal algorithm. Thus the internal algorithm receives exactly the feedback required by its guarantee.

For every fixed comparator $S \in \mathcal{S}_K$ and every realized action sequence,

$$\sum_{t=1}^{T}\big[g(z_t(S)) - g(z_t(S_t))\big] \le \sum_{t=1}^{T}\big[p(z_t(S)) - p(z_t(S_t))\big] + 2\epsilon T.$$

Taking the maximum over S preserves the inequality, so pathwise

$$\mathrm{Reg}_T(g) \le \mathrm{Reg}_T(p) + 2\epsilon T.$$

Taking expectation proves equation 42. □

The reduction uses uniform approximation twice per round: once for the fixed comparator and once for the learner's action. No stochastic assumption on $v_{1:T}$ is involved.

C.2 Polynomial links as subset-linear bandits

We give an exact $\binom{N}{d}$-dimensional representation, avoiding the repeated coordinates of the usual ordered-tensor lift. For a polynomial p of degree at most d and $U \subseteq [N]$, define its subset coefficient

$$c_{p,U}(v_U) := \sum_{V\subseteq U} (-1)^{|U|-|V|} p\left(\sum_{i\in V} v_i\right).$$

Coefficients with $|U| > d$ vanish.

Lemma 19 (Exact subset-feature linearization)**.** *Let $1 \le d \le K$. For each d-subset $W \subseteq [N]$, define*

$$\theta_{t,W}(p) := \sum_{U\subseteq W} \frac{c_{p,U}(v_{t,U})}{\binom{K-|U|}{d-|U|}}.$$

Then every $S \in \mathcal{S}_K$ satisfies

$$p\left(\sum_{i\in S} v_{t,i}\right) = \sum_{\substack{W\subseteq S\\ |W|=d}} \theta_{t,W}(p). \tag{43}$$

Proof. Summing the definition of $\theta_{t,W}$ over the d-subsets $W \subseteq S$, a fixed $U \subseteq S$ appears exactly $\binom{K-|U|}{d-|U|}$ times. The denominators cancel, leaving $\sum_{U\subseteq S} c_{p,U}(v_{t,U})$, which equals the reward by Boolean-lattice reconstruction. □

Thus the finite feature-action set is

$$\mathcal{X}_d := \Big\{(\mathbf{1}\{W \subseteq S\})_{W\subseteq[N],\,|W|=d} : S \in \mathcal{S}_K\Big\} \subseteq \mathbb{R}^{\binom{N}{d}}.$$

Each action is the incidence vector of its contained d-subsets. For $1 \le d \le K$, their union recovers S, so $|\mathcal{X}_d| = \binom{N}{K}$.

Lemma 20 (Range-sensitive polynomial bound)**.** *Let p be a polynomial of degree at most d, where $1 \le d < K$, and define*

$$\operatorname{osc}(p) := \sup_{x\in[0,K]} p(x) - \inf_{x\in[0,K]} p(x).$$

There is a universal constant $C > 0$ and an exact-aggregate algorithm satisfying

$$\mathbb{E}[\mathrm{Reg}_T(p)] \le C \operatorname{osc}(p) \min\left\{T, \sqrt{\binom{N}{d} T \log \binom{N}{K}}\right\}. \tag{44}$$

Proof. If $\operatorname{osc}(p) = 0$, the regret is identically zero. Otherwise let

$$m_p := \inf_{x\in[0,K]} p(x), \qquad q(x) := \frac{p(x) - m_p}{\operatorname{osc}(p)}.$$

Then $q([0, K]) \subseteq [0, 1]$, q has degree at most d, and

$$\mathrm{Reg}_T(p) = \operatorname{osc}(p)\, \mathrm{Reg}_T(q). \tag{45}$$

By Lemma 19, the rewards generated by q are linear in d-subset incidence vectors in dimension $D = \binom{N}{d}$, and all action rewards lie in $[0, 1]$. Applying the finite-action linear-bandit bound based on exponential weights and John's exploration (Bubeck et al., 2012; Audibert & Bubeck, 2009) to $\mathcal{X}_d$ gives

$$\mathbb{E}[\mathrm{Reg}_T(q)] \le C \min\left\{T, \sqrt{DT \log |\mathcal{X}_d|}\right\}.$$

Since the action set has size $\binom{N}{K}$, multiplying by $\operatorname{osc}(p)$ proves equation 44. The learner implements the scalar-feedback linear-bandit algorithm from the exact aggregate by computing $q(z_t(S_t))$. □

The polynomial coefficients enter the loss vector $\theta_t(p)$, but they do not enter the regret guarantee once the realized rewards on the action set are normalized to $[0, 1]$. This is the precise sense in which the bound has no hidden coefficient dependence.

C.3 The polynomial-approximation envelope

For $0 \le d < K$, recall

$$E_d(g; K) := \inf_{\deg(p)\le d} \|g - p\|_{\infty,[0,K]}.$$

We keep the dependence on K implicit below and write $E_d = E_d(g; K)$.

Theorem 21 (Polynomial-approximation envelope)**.** *Suppose $g : [0, K] \to [0, 1]$, and let $M = \binom{N}{K}$. There are universal constants $C, C_0 > 0$ such that*

$$\begin{aligned}\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \le \min\Bigg\{ & T, 2TE_0, \\ & \min_{1\le d<K}\left[C \min\left\{T, \sqrt{\binom{N}{d} T \log M}\right\} + 2TE_d\right], \\ & C_0 \min\left\{T, \sqrt{(M-1)T}\right\}\Bigg\}.\end{aligned} \tag{46}$$

When $M = 1$, the last term is interpreted as zero.

Proof. The first term is the trivial bound for rewards in $[0, 1]$.

For $d = 0$, choose a constant p whose approximation error is arbitrarily close to E_0. Every action receives the same surrogate reward, so $\mathrm{Reg}_T(p) = 0$. Theorem 18 gives

$$\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \le 2TE_0$$

after taking a limit over an approximating sequence of constants.

Now fix $1 \le d < K$ and $\delta > 0$. Choose a polynomial p_d of degree at most d satisfying

$$\|g - p_d\|_{\infty,[0,K]} \le E_d + \delta.$$

Because $g([0,K]) \subseteq [0,1]$,

$$p_d([0,K]) \subseteq [-E_d - \delta, 1 + E_d + \delta]$$

and hence

$$\operatorname{osc}(p_d) \le 1 + 2(E_d + \delta). \tag{47}$$

Moreover, the constant polynomial $1/2$ shows that $E_d \le E_0 \le 1/2$. Lemma 20, Theorem 18, and equation 47 therefore give

$$\mathfrak{R}_{\mathrm{SB}}(g,N,K,T) \le C\big(1 + 2E_d + 2\delta\big) \min\left\{T, \sqrt{\binom{N}{d} T \log M}\right\} + 2T(E_d + \delta).$$

Letting $\delta \downarrow 0$ and absorbing $1 + 2E_d \le 2$ into the universal constant gives the middle family of terms in equation 46.

Finally, treat the M feasible subsets as unrelated arms and run a minimax adversarial multi-armed bandit algorithm (Audibert & Bubeck, 2009) on the exact rewards $g(z_t(S_t))$. For $M \ge 2$, the standard $O(\min\{T, \sqrt{MT}\})$ guarantee is at most a universal constant times $\min\{T, \sqrt{(M-1)T}\}$. If $M = 1$, there is no decision to make and the regret is zero. This proves the last term and completes the proof. □

The minimizing degree in equation 46 depends on the horizon. The link g is known, so this is not adaptation to an unknown statistical parameter: for a known horizon, the learner may select the degree and surrogate before the interaction begins. The statement is an information-theoretic oracle envelope and does not assert that computing a best uniform approximant or implementing the resulting linear-bandit algorithm is efficient.

Unknown horizon. The known-horizon assumption is inessential at the level of rates. Partition time into geometrically increasing epochs and, at the beginning of each epoch, select the degree and surrogate minimizing the bound for that epoch length. For any fixed degree, the square-root costs form a geometric series dominated by its final term, while the approximation costs sum linearly with the total number of rounds. Comparing with the degree minimizing the final-horizon bound shows that restarting incurs only a universal constant factor.

C.4 The interaction–approximation bridge

We finally prove that every interaction certificate above degree d is an obstruction to degree-d uniform approximation.

Proposition 22 (Interaction–approximation bridge)**.** *For every* $0 \le d < K$*,*

$$E_d(g;K) \ge \max_{s \in \{d+1,\dots,K\}} \Gamma_s(g;K). \tag{48}$$

Proof. Fix $s > d$ and a polynomial p of degree at most d. Since $\Delta_h^s p \equiv 0$,

$$\begin{aligned} 2^{-s} \left|\mathbb{E}_{J\sim\mathrm{Bin}(K-s,1/2)} \left[\Delta_h^s g(Ka + hJ)\right]\right| \\ = 2^{-s} \left|\mathbb{E}_J \left[\Delta_h^s (g-p)(Ka + hJ)\right]\right|. \end{aligned}$$

For every x at which the finite difference is defined,

$$\begin{aligned} |\Delta_h^s (g-p)(x)| &\le \sum_{j=0}^{s} \binom{s}{j} \|g - p\|_{\infty,[0,K]} \\ &= 2^s \|g - p\|_{\infty,[0,K]}. \end{aligned}$$

It follows that

$$\Gamma_s(g;K) \le \|g - p\|_{\infty,[0,K]}.$$

Taking the infimum over all polynomials of degree at most d gives $\Gamma_s(g;K) \le E_d(g;K)$. Taking the maximum over $s > d$ proves equation 48. □

Proposition 22 is intentionally one-sided. A large high-order certificate rules out accurate low-degree approximation, but the restricted family of finite-difference tests defining Γ_s need not capture every obstruction to uniform approximation. No converse inequality is used or claimed.

C.5 Reading the lower and upper bounds together

For every $s > d$, Theorem 7 assigns a statistical price to an s-way interaction of amplitude $\Gamma_s(g; K)$. Theorem 21 permits the learner to ignore all structure above degree d at deterministic cost $2TE_d(g; K)$. Proposition 22 ensures that these two descriptions are consistent: an interaction large enough to certify lower-bound hardness cannot simultaneously disappear inside a more accurate lower-degree approximation.

The two sides nevertheless answer different questions. The lower bound asks what an adversary can force when a particular interaction must be discovered. The upper bound asks whether resolving that interaction is worthwhile at the current horizon. Their one-sided consistency does not identify a minimax "effective order."

D Detailed Calculations for the MNL Link

This appendix specializes the interaction certificate and the approximation envelope to the multinomial-logit link

$$g(x) = \frac{x}{1+x}, \qquad x \in [0, K].$$

We first compute its highest-order interaction strength exactly. We then give an explicit Chebyshev polynomial achieving the approximation bound used in the main text, and finally make the horizon-dependent degree calculation precise.

D.1 Exact finite differences of the reciprocal link

The required calculation follows from a general identity for reciprocal functions.

Lemma 23 (Finite differences of a reciprocal)**.** *For every integer $s \geq 1$, $h \geq 0$, and $x > -1$,*

$$\Delta_h^s \frac{1}{1+x} = \frac{(-1)^s s! h^s}{\prod_{j=0}^{s}(1+x+jh)}. \tag{49}$$

Proof. The claim is immediate for $s = 1$:

$$\frac{1}{1+x+h} - \frac{1}{1+x} = -\frac{h}{(1+x)(1+x+h)}.$$

Suppose the identity holds at order s. Taking one more forward difference and placing the two fractions over a common denominator gives

$$\begin{aligned}\Delta_h^{s+1} \frac{1}{1+x} &= (-1)^s s! h^s \left[\frac{1}{\prod_{j=1}^{s+1}(1+x+jh)} - \frac{1}{\prod_{j=0}^{s}(1+x+jh)}\right] \\ &= \frac{(-1)^{s+1}(s+1)! h^{s+1}}{\prod_{j=0}^{s+1}(1+x+jh)}.\end{aligned}$$

The result follows by induction. □

Since $g(x) = 1 - (1+x)^{-1}$, Lemma 23 implies that, for every $s \geq 1$,

$$\Delta_h^s g(x) = \frac{(-1)^{s+1} s! h^s}{\prod_{j=0}^{s}(1+x+jh)}. \tag{50}$$

Proposition 24 (All-order MNL certificates)**.** *For every $s \in [K]$,*

$$\Gamma_s(g; K) \geq \frac{s!}{2^{2s+1} K^s}.$$

In particular, $\Gamma_1(g; K) \geq 1/(8K)$.

Proof. Take $a = 0$ and $h = 1/K$ in the definition of Γ_s. If $J \sim \mathrm{Bin}(K-s, 1/2)$, then $0 \le J \le K-s$ and hence, for every $r \in \{0, \dots, s\}$,

$$1 + \frac{J+r}{K} \le 2.$$

Equation equation 50 therefore gives

$$\left|\Delta^s_{1/K} g(J/K)\right| = \frac{s! K^{-s}}{\prod_{r=0}^{s}(1 + (J+r)/K)} \ge \frac{s!}{2^{s+1} K^s}.$$

The sign is constant, so taking the binomial expectation and multiplying by the 2^{-s} normalization in Γ_s proves the claim. □

Together with Proposition 22, the same calculation yields the explicit approximation obstruction

$$E_d(g; K) \ge \Gamma_{d+1}(g; K) \ge \frac{(d+1)!}{2^{2d+3} K^{d+1}}, \qquad 0 \le d < K.$$

At order $s = K$, the auxiliary binomial variable in Γ_K is zero almost surely. Therefore

$$\Gamma_K(g; K) = 2^{-K} \sup_{\substack{a,h \ge 0 \\ a+h \le 1}} \frac{K! h^K}{\prod_{j=0}^{K}(1 + Ka + jh)}. \tag{51}$$

For fixed h, the expression in equation 51 is nonincreasing in a, so its maximum occurs at $a = 0$. At $a = 0$, let

$$F_K(h) := \frac{K! h^K}{\prod_{j=0}^{K}(1 + jh)}, \qquad 0 < h \le 1.$$

Its logarithmic derivative is

$$\frac{d}{dh} \log F_K(h) = \frac{K}{h} - \sum_{j=1}^{K} \frac{j}{1 + jh} > 0,$$

because $j/(1 + jh) < 1/h$ for every $j \ge 1$. Hence F_K is strictly increasing and is maximized at $h = 1$. Substitution into equation 51 gives the exact certificate.

Proposition 25 (Exact MNL interaction strength)**.** *For $g(x) = x/(1 + x)$,*

$$\Gamma_K(g; K) = \frac{1}{2^K (K+1)}. \tag{52}$$

Indeed, the maximizing parameters are $a = 0$ and $h = 1$, for which

$$\Delta^K_1 g(0) = \frac{(-1)^{K+1}}{K+1}.$$

Thus the lower bound in the main text is not merely obtained from one convenient choice of embedding: within the parity family defining Γ_K, that choice is optimal.

D.2 An explicit Chebyshev approximant

We next construct a degree-d polynomial approximation to g. Set

$$y = 1 + x, \qquad y \in [1, K+1],$$

so it suffices to approximate $y \mapsto 1/y$. Let T_m denote the Chebyshev polynomial of the first kind and put $m = d + 1$. Define the affine map

$$\tau(y) := \frac{K + 2 - 2y}{K}.$$

This map sends $[1, K+1]$ onto $[1, -1]$ and satisfies

$$\tau(0) = 1 + \frac{2}{K} > 1.$$

Consider the degree-m residual polynomial

$$r_m(y) := \frac{T_m(\tau(y))}{T_m(1+2/K)}. \tag{53}$$

Since $r_m(0) = 1$, the polynomial $1 - r_m(y)$ has a root at $y = 0$. Hence

$$q_d(y) := \frac{1 - r_m(y)}{y} \tag{54}$$

is a polynomial of degree at most $m - 1 = d$. Moreover, for every $y \in [1, K+1]$,

$$\begin{aligned}\left|\frac{1}{y} - q_d(y)\right| &= \frac{|r_m(y)|}{y} \\ &\le \frac{1}{T_m(1+2/K)},\end{aligned}$$

where we used $y \ge 1$ and $|T_m(u)| \le 1$ for $u \in [-1, 1]$.

Finally, define

$$p_d(x) := 1 - q_d(1+x). \tag{55}$$

Then p_d has degree at most d and

$$|g(x) - p_d(x)| = \left|\frac{1}{1+x} - q_d(1+x)\right|.$$

We have proved the following bound.

Proposition 26 (Chebyshev approximation of the MNL link)**.** *For every $d \ge 0$,*

$$E_d(g; K) \le \frac{1}{T_{d+1}(1+2/K)}. \tag{56}$$

To expose the rate in d, define

$$\theta_K := \operatorname{arcosh}\left(1 + \frac{2}{K}\right) = 2\operatorname{arsinh}\left(\frac{1}{\sqrt{K}}\right). \tag{57}$$

Since $T_m(\cosh\theta) = \cosh(m\theta)$ and $\cosh u \ge e^u/2$ for $u \ge 0$,

$$E_d(g; K) \le \frac{1}{\cosh((d+1)\theta_K)} \le 2e^{-(d+1)\theta_K}. \tag{58}$$

The scale of the exponent is explicit. Since $x/\sqrt{2} \le \operatorname{arsinh}(x) \le x$ for $x \in [0, 1]$,

$$\sqrt{\frac{2}{K}} \le \theta_K \le \frac{2}{\sqrt{K}}. \tag{59}$$

Thus polynomial degree of order $\sqrt{K}\log(1/\epsilon)$ is sufficient to approximate the MNL link uniformly to accuracy ϵ on $[0, K]$.

D.3 The resulting finite-horizon envelope

Substituting Proposition 26 into Theorem 21 gives, for $M = \binom{N}{K}$,

$$\begin{aligned}\mathfrak{R}_{\mathrm{SB}}(g, N, K, T) \le \min_{0 \le d < K} \Bigg\{ & C\mathbf{1}\{d \ge 1\} \min\left\{T, \sqrt{\binom{N}{d} T \log M}\right\} \\ & + \frac{2T}{T_{d+1}(1+2/K)}\Bigg\},\end{aligned} \tag{60}$$

up to also taking the minimum with the trivial and action-as-arm terms in equation 46. Using equation 58 yields the simpler display

$$\mathfrak{R}_{\mathrm{SB}}(g,N,K,T) \le \min_{0\le d<K}\left\{C\mathbf{1}\{d\ge 1\}\min\left\{T,\sqrt{\binom{N}{d}T\log M}\right\} + 4Te^{-(d+1)\theta_K}\right\}. \tag{61}$$

For $d=0$, the statistical term vanishes because a constant surrogate requires no learning. At the other extreme, if every degree $d<K$ gives a larger bound than the action-as-arm term, the learner may resolve the link exactly at cost $O(\sqrt{MT})$.

The approximation and interaction calculations can be read simultaneously. Propositions 25 and 26, together with the bridge in Proposition 22, give

$$\frac{1}{2^K(K+1)} = \Gamma_K(g;K) \le E_{K-1}(g;K) \le \frac{1}{T_K(1+2/K)}. \tag{62}$$

The first term is strictly positive, which forces full K-th-order hardness asymptotically. The last term can nevertheless be small, so a lower-degree surrogate can have the smallest certified upper cost at finite horizons. The inequalities are not expected to match: Proposition 22 is a one-sided consistency relation, not an equivalence between interaction strength and best polynomial approximation.

D.4 Degree selected by the upper envelope

We now formalize the horizon dependence suggested by equation 60. Ignore the outer trivial and action-as-arm caps momentarily and define the certified surrogate costs

$$U_d(T) := a_d\sqrt{T} + b_dT, \qquad 0\le d<K, \tag{63}$$

where

$$a_0 := 0, \qquad a_d := C\sqrt{\binom{N}{d}\log M} \quad (d\ge 1), \qquad b_d := \frac{2}{T_{d+1}(1+2/K)}.$$

Define a bound-selected effective degree by choosing the largest minimizer,

$$d_{\mathrm{eff}}^{\mathrm{cert}}(T) := \max \arg\min_{0\le d<K} U_d(T). \tag{64}$$

This is a property of the proved upper envelope, not a claim that the minimax regret itself has a uniquely defined degree.

Lemma 27 (Monotonicity of the certified degree)**.** *Assume $N\ge 2K$. The function $T\mapsto d_{\mathrm{eff}}^{\mathrm{cert}}(T)$ is nondecreasing.*

Proof. Let $x=\sqrt{T}$. Dividing equation 63 by $\sqrt{T}$ shows that minimizing $U_d(T)$ is equivalent to minimizing the line

$$L_d(x) := a_d + b_dx.$$

The intercepts a_d are nondecreasing in d, while the slopes b_d are strictly decreasing because $T_{d+1}(1+2/K)=\cosh((d+1)\theta_K)$ is strictly increasing. For any $d_1>d_0$, the difference

$$L_{d_1}(x)-L_{d_0}(x) = (a_{d_1}-a_{d_0})+(b_{d_1}-b_{d_0})x$$

is nonincreasing in x. Once the higher-degree line is no larger than the lower-degree line, the lower-degree line cannot become strictly preferable at a larger x. Choosing the largest minimizer resolves ties consistently and proves the claim. □

Thus the statement that the certified resolution becomes finer with the horizon is literal for the surrogate envelope; it is not inferred merely from a plot.

For a closed-form heuristic, replace the exact Chebyshev term by its exponential upper bound and suppress multiplicative constants and the factor $\log M$. Balancing the two terms gives

$$N^{d/2}T^{1/2} \asymp Te^{-(d+1)\theta_K}.$$

Solving for a continuous d yields

$$d_{\mathrm{bal}}(T) = \frac{\log T - 2\theta_K}{\log N + 2\theta_K}. \tag{65}$$

Dropping the additive shift caused by $d+1$ gives the simpler proxy

$$d_{\mathrm{eff}}(T) \approx \frac{\log T}{\log N + 2\theta_K}. \tag{66}$$

Both expressions must be rounded to an integer and truncated to $\{0, \dots, K-1\}$. One may assign the value K when the action-as-arm term is selected, corresponding to exact rather than surrogate resolution.

The transition scale between adjacent proxy degrees can also be made explicit. Let

$$\widetilde{U}_d(T) := AN^{d/2}\sqrt{T} + BTe^{-(d+1)\theta_K},$$

where $A, B > 0$ absorb the suppressed constants and logarithmic factor. Direct comparison shows that degree $d+1$ improves on degree d precisely when

$$\sqrt{T} \geq \frac{A(\sqrt{N}-1)}{B(1-e^{-\theta_K})} N^{d/2} e^{(d+1)\theta_K}. \tag{67}$$

Up to a prefactor independent of d, the corresponding horizon thresholds therefore grow as

$$T_d \asymp N^d e^{2(d+1)\theta_K}, \qquad \frac{T_{d+1}}{T_d} \asymp Ne^{2\theta_K}. \tag{68}$$

These thresholds explain the staircase behavior expected from a theoretical plot of $d_{\mathrm{eff}}(T)$. They are consequences of the displayed upper certificate, not empirical phase transitions and not a separate minimax characterization.

D.5 Conclusion for the MNL link

The MNL link exhibits both sides of the theory quantitatively. Its highest interaction certificate is exactly $2^{-K}/(K+1)$, so exact K-th-order hardness never disappears. At the same time, its degree-d approximation error decays at least as $2e^{-(d+1)\theta_K}$, with $\theta_K = \Theta(K^{-1/2})$. The former determines the eventual worst-case exponent; the latter determines how long a lower-degree surrogate can remain preferable. Semi-bandit feedback does not alter the first fact; exact aggregate feedback is already sufficient to make the second operational.

E Numerical Illustrations of the Theoretical Bounds

This section expands Figure 1 over continuous parameter ranges. The plots deterministically evaluate expressions from the theorems; no regret trajectory is sampled and no minimax value is estimated. The two accompanying Python scripts regenerate the PDF figures and PNG previews without random seeds or external data.

For the weak-interaction family, we plot the normalized order-d contribution

$$S_{d,\epsilon}(T) := \frac{\epsilon\psi_d(T)}{\psi_1(T) + \epsilon\psi_d(T)}.$$

The contour $S_{d,\epsilon}(T) = 1/2$ marks equality of the two terms in the normalized rate expression. For $\epsilon_c \leq \epsilon \leq 1/2$, its intermediate branch is $T = r_1/\epsilon^2$. At $(T, \epsilon) = (r_d, \epsilon_c)$ it continues along the horizontal ray $\epsilon = \epsilon_c$, $T \geq r_d$. Below $\epsilon_c = \sqrt{r_1/r_d}$, the high-order term remains smaller for every horizon. These algebraic boundaries do not locate an exact transition in the minimax value, whose characterization allows K-dependent constants and logarithmic factors.

For the feedback separation, we evaluate the normalized certificate

$$G_\alpha(N) := \sqrt{\log N}\left(N^{-1/2} + N^{-\alpha}\right),$$

which upper bounds the regret ratio up to a universal multiplicative constant along the parameter choice in Theorem 5. We restrict both separation panels to $N \geq 4$ and $\alpha \log N \geq \log 2$, ensuring $\epsilon = N^{-\alpha} \leq 1/2$. Figure 2 masks the remaining region and marks its boundary $\alpha = \log 2/\log N$. The level sets describe the normalized expression: without the suppressed constant, $G_\alpha = 1$ is not a certified boundary for a regret ratio below one. Every fixed $\alpha > 0$ has $G_\alpha(N) \to 0$ as $N \to \infty$. The line plots use integer values of N; the heat maps interpolate the analytic expressions over a continuous plotting grid.

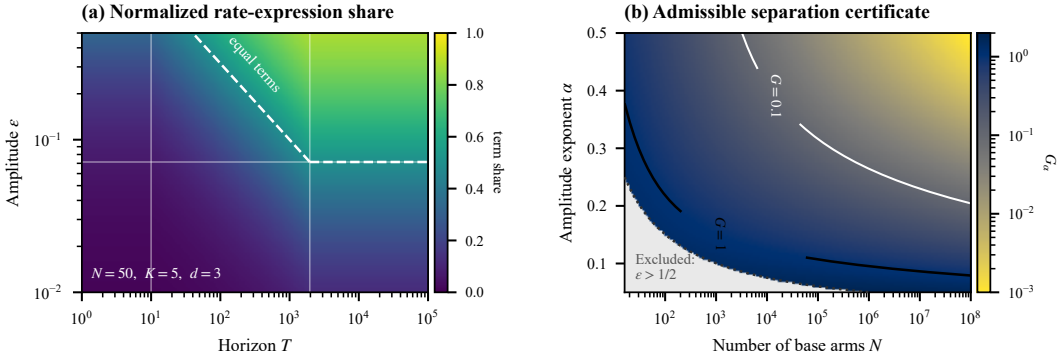

Figure 2: Continuous plots of the theorem expressions. **Left:** the normalized order-d term share for $N = 50$, $K = 5$, $d = 3$. The dashed equal-term boundary follows $T = r_1/\epsilon^2$ to (r_d, ϵ_c) and continues horizontally for $T \geq r_d$. Faint guides mark r_1, r_d, and ϵ_c. **Right:** $G_\alpha(N)$ on the admissible domain $N^{-\alpha} \leq 1/2$; gray points are outside the theorem assumptions. The dotted boundary is $\alpha = \log 2/\log N$. Labeled contours are levels of G_α, with its universal multiplicative constant suppressed.

F Detailed Discussion of Related Work

This appendix expands the positioning in the introduction. We organize the literature by the distinctions that matter for our results: whether the environment is adversarial or stochastic, whether feedback reveals an aggregate or the selected coordinates, whether nonlinearity is arbitrary or structurally restricted, and whether regret is exact or measured relative to an approximation benchmark. Table 2 summarizes these axes before we discuss them in detail.

F.1 Adversarial combinatorial bandits and the closest nonlinear model

Adversarial multi-armed bandits provide the basic change-of-measure and finite-action benchmarks used throughout online learning (Auer et al., 1995; 2002; Audibert & Bubeck, 2009; Bubeck & Cesa-Bianchi, 2012). In combinatorial online optimization, actions are incidence vectors and the canonical reward or loss is linear in the selected coordinates. The full-bandit and semi-bandit versions of this problem have been studied from minimax, computational, and first-order perspectives (Cesa-Bianchi & Lugosi, 2012; Audibert et al., 2014; Bubeck et al., 2012; Neu, 2015). Graph feedback provides an intermediate and more general observation pattern for linear combinatorial losses (Wen, 2025). These results demonstrate the usual value of semi-bandit observations: an observed coordinate can be reused to estimate the linear reward of every action containing that coordinate. Our construction shows that this reuse principle can fail for a nonlinear function of a jointly distributed outcome vector. Revealing all coordinates in the chosen action still leaves an s-way interaction completely unobserved until its entire support is covered.

Han et al. (2021) are the closest point of comparison. They study the same action class and known rewards of the form $g(\sum_{i \in S} v_{t,i})$ in an adversarial environment, but the learner sees only scalar

Table 2: Representative neighboring formulations. "Exact" refers to regret against the best feasible action for the stated reward, rather than to an approximation ratio caused by an offline optimization barrier.

Line of work	Environment	Feedback	Principal distinction from this paper
Adversarial nonlinear combinatorial bandits	Adversarial	Scalar reward	Same known-link model; establishes the scalar-feedback benchmark.
Linear combinatorial semi-bandits	Adversarial or stochastic	Selected coordinates	Additivity makes coordinate observations directly reusable across actions.
Stochastic nonlinear CMAB	Stationary or structured nonstationary	Selected or triggered outcomes	Tractability rests on distributional and reward regularity assumptions.
Best-of-K with dependent arms	Primarily stochastic; adversarial consequences	Bandit, marked, and selected coordinates	Establishes dependence hiding, including a semi-bandit lower bound, regret consequences, and exponentially small gaps for a max objective.
Online submodular maximization	Often adversarial	Bandit or richer feedback	Commonly targets multiplicative approximate regret for restricted function classes.
This paper	Adversarial	Selected coordinates	General known links; all-order finite-difference certificates, coverage geometry, amplitude-sensitive matching laws, and a noisy-feedback separation.

bandit feedback. They identify the polynomial-degree transition: a degree-$d < K$ link has minimax rate $\widetilde{\Theta}(\sqrt{N^d T})$, whereas a link with genuine K-th-order structure has rate $\Theta(\sqrt{N^K T})$ for fixed K. Their work leaves open whether observing the selected primitive outcomes changes this conclusion. Our lower bound answers that feedback question negatively at the level of the N, T exponent. It also localizes the lower-bound amplitude through $\Gamma_s(g; K)$. On the positive side, selected coordinates expose observable Möbius interactions and yield an amplitude-sensitive bound that matches the lower theory on normalized positive polynomials and their reflected concave family. The fixed-range separation from Bernoulli scalar feedback makes this coordinate advantage quantitative. A separate surrogate result uses only the exact aggregate; distinguishing these information levels is essential.

F.2 Stochastic combinatorial semi-bandits and nonlinear rewards

For stochastic linear rewards, individual observations make it possible to learn shared base-arm means rather than treating every feasible subset as an independent arm. This principle underlies early network and general CMAB formulations (Gai et al., 2012; Chen et al., 2013) and sharp analyses of stochastic combinatorial semi-bandits (Kveton et al., 2015; Combes et al., 2015; Merlis & Mannor, 2019; 2020). The framework has been extended to probabilistically triggered arms and to Thompson sampling (Chen et al., 2016b; Wang & Chen, 2017; 2018; Liu et al., 2022). These works primarily exploit stable arm-level distributions, independence or smoothness conditions, and oracle access to the corresponding offline optimization problem.

Nonlinearity itself is therefore not absent from the stochastic CMAB literature. In particular, Chen et al. (2016a) allow an expected reward to depend on the full distributions of the selected random variables, covering objectives such as maximum and expected utility. Their distributional confidence method yields stochastic regret guarantees under monotonicity and continuity conditions and, when exact offline optimization is unavailable, approximation regret. This is complementary to our question. We allow the joint outcome law to encode adversarially hard dependence and measure exact regret against the best fixed subset; the link itself is known and has the particular aggregate form $g(\sum_i v_i)$. Neither formulation contains the other.

Structured nonstationarity has also been considered for combinatorial semi-bandits (Chen et al., 2021a). Such variation-budget or change-point structure differs from the unrestricted adversarial sequence used here. A concurrent graph-structured model studies nonlinear associations generated by

separable analytic signals and independent stochastic base-arm distributions (Bauschmann & Maghsudi, 2026). Its goal is to learn the reward graph and distributional moments under strong feedback; ours instead asks what remains information-theoretically hidden even when the aggregate link is already known. The two uses of polynomial approximation are accordingly different: estimating a structured stochastic reward model versus choosing a surrogate for adversarial exact regret.

F.3 Higher-order dependence and information hiding

The use of dependence to conceal subset quality predates our construction. Most directly, Simchowitz et al. (2016) study stochastic best-of-K identification for the maximum of a jointly distributed reward vector. Their Theorem 2.1 includes a semi-bandit lower bound; they also derive stochastic and adversarial regret consequences and show that matching all $(K-1)$-order statistics forces an exponentially small gap in their binary max objective. Thus neither dependence hiding nor its regret relevance is new here.

Our lower-bound contribution is the general known-link formulation. For every order $s \le K$, the finite difference of g gives the explicit reward amplitude $\Gamma_s(g;K)$, while $\binom{N}{s}/\binom{K}{s}$ records the exact joint-coverage cost. The substantive advance of the full theory is that this lower description can then be compared with observable Möbius widths, yielding coefficient-uniform matching laws on structured link families.

Other subset-learning results study full-bandit feedback, where even an additive subset reward is observed only through its aggregate. Examples include top-k regret minimization and multiple-arm identification (Rejwan & Mansour, 2020; Kuroki et al., 2020). Their information loss is caused by aggregation of otherwise additive arm outcomes. Our learner sees each selected primitive outcome; the obstruction survives because the relevant signal is a joint distributional interaction rather than an unobserved sum. General partial-monitoring and feedback-graph theories classify how an action's signal reveals other losses (Bartók et al., 2014; Alon et al., 2015). They provide useful high-level context, but applying their generic classifications directly would not expose the link-dependent finite-difference amplitude or the combinatorial joint-coverage identity used here.

F.4 Approximate objectives and surrogate models

Online submodular maximization is another major source of nonlinear set objectives. Because exact offline maximization can itself be intractable, this literature often evaluates an online learner against a multiplicative fraction of the offline optimum (Streeter & Golovin, 2008; Harvey et al., 2020; Pasteris et al., 2024). Sum-max functions, for example, include best-of-K-type objectives and admit efficient $(1-1/e)$-regret guarantees under bandit feedback (Pasteris et al., 2024). Likewise, stochastic general-reward CMAB may use approximation regret when its offline problem is solved only approximately (Chen et al., 2016a).

Our approximation term has a different origin and interpretation. The comparator remains the best action under the true known link g; there is no multiplicative relaxation of the benchmark. Instead, semi-bandit feedback reveals the selected aggregate (although the individual coordinates are not needed for this step), so the learner may run an exact-regret algorithm for a surrogate p and pay the pathwise additive discrepancy $2T\|g-p\|_\infty$. Optimizing over polynomial degree gives a statistical bias–complexity envelope. The bridge $E_d(g;K) \ge \max_{s>d} \Gamma_s(g;K)$ further ties this algorithmic approximation choice to the same interaction that appears in the lower bound. Selected coordinates support the stronger observable-interaction envelope and the matching positive-polynomial law. We claim these quantitative connections, not polynomial or surrogate approximation as a general methodology.

F.5 MNL bandits and assortment learning

The multinomial-logit model has generated a substantial literature on dynamic assortment selection. Early work studies demand learning and capacity constraints (Caro & Gallien, 2007; Rusmevichientong et al., 2010; Sauré & Zeevi, 2013); later MNL-bandit results establish regret upper and lower bounds for learning unknown attraction parameters (Agrawal et al., 2019; Chen & Wang, 2018; Chen et al., 2021b). These are stochastic choice models: feedback is typically a customer's realized choice,

and the learner's uncertainty concerns persistent item parameters. Han et al. (2021) instead use nonlinear combinatorial bandits to derive adversarial assortment hardness under scalar feedback.

Our MNL section should be read as a link-level worked example rather than as a solution to the full stochastic assortment problem. With unit revenues and a normalized outside option, the aggregate purchase-probability link is $g(x) = x/(1+x)$. This scalar function lets us compute both the exact highest finite difference and an explicit Chebyshev approximation rate. It therefore demonstrates, in a familiar nonlinear objective, how exact K-th-order hardness can coexist with a substantially smaller degree selected by a finite-horizon upper envelope. This is not a minimax identification of an effective order. The example does not estimate unknown choice parameters, impose assortment-specific revenue structure, or compete with the stochastic MNL algorithms cited above.