
Structure, Not Belief: Correlated Thompson Sampling from LLM-Derived Covariance in Combinatorial Semi-Bandits

Vikram Kakaria
Princeton University

Anish Kataria
Princeton University

Anany Kotawala
Princeton University

Abstract

Combinatorial Thompson sampling (CTS) draws independent posterior samples for every arm, so its exploration dynamics ignore any relation among arms. We study a minimal change to those dynamics: an LLM is queried *once* for a partition of the arms, the partition becomes a positive-definite correlation matrix Σ through an RBF kernel on cluster ranks, and the per-round posterior sample is drawn with covariance Σ while the Beta posteriors are updated from real rewards only, so the LLM shapes *how* the sampler moves, not *what* it believes. We give a self-contained Bayesian regret bound for the idealized Gaussian sampler whose information gain splits into a $K \log T$ term from the K -cluster structure and a ridge term that grows to $d \log T$: the $\sqrt{d/K}$ improvement over independent sampling is a finite-horizon transient, exact only as the within-cluster correlation tends to one. The correlated sampler reduces regret by 19% over CTS on 16 synthetic Bernoulli families at $T=2,500$ (6–7% at $T=25,000$ with data-adaptive kernels) and by 41% on the Microsoft MIND-small news benchmark ($d=200$ real articles), while pseudo-observation warm starts give nothing. An LLM-free ablation with a simulated oracle of controlled quality shows that on unstructured instances the gain is a property of the kernel *shape* (a random partition, or a plain tempering of the sampling noise, reproduces it), while belief injection at matched oracle quality never helps.

1 Introduction

Thompson sampling is a stochastic dynamical system: the learner’s state is a posterior, the control is a random draw from it, and the reward feedback closes the loop; this paper is about how the covariance of that draw changes exploration, how the effect decays with the horizon, and how an oracle in the loop couples to the learner’s state. In the combinatorial semi-bandit [5, 9], where m of d arms are played each round and each played arm returns its own Bernoulli reward, Combinatorial Thompson Sampling (CTS) [13] draws the d coordinates of that control *independently*, so its exploration dynamics carry no information about which arms are alike.

Large language models are an obvious source of such information, and most LLM-augmented bandits inject it as a *belief*: a prior over means [2, 3], pseudo-observations, or a recommendation mixed into the action distribution [4, 8]. Belief injection has a feedback-loop failure mode that we observed directly (Section G): the LLM’s answer quality depends on the estimates it is shown, which depend on the learner’s own trajectory, and a wrong belief biases the posterior at a cost of real observations. We instead use the LLM for *structure*: one cluster query, converted to a correlation matrix Σ , changes only the covariance of the posterior draw, so posterior means and variances stay those of CTS (Proposition 1) and an arbitrarily bad Σ can distort the sampler’s motion but not corrupt its state.

Contributions. (i) A statement of what Σ represents (Assumption 2) and a self-contained Bayesian regret bound for GP-CTS(Σ) (Theorem 4) whose $\sqrt{d/K}$ advantage is a transient, exact only as $\rho_{\max} \rightarrow 1$. (ii) CORRCTS-FULL as a *fixed-correlation, calibrated-marginal* approximation to GP-CTS(Σ), partition-free accounts of why a fixed correlation changes top- m selection (Lemmas 5 and 6; in the equicorrelated case it is exactly a tempering), and a total-variation guarantee for the credibility gate (Proposition 7). (iii) LLM experiments on synthetic families and real MIND click data, plus an LLM-free ablation in which every injection mode receives the *same* simulated partition of controlled quality, separating kernel shape from partition information and structure from belief.

2 Setting and algorithm

Combinatorial semi-bandit. There are d arms with unknown means $\mu \in [0, 1]^d$. At round t the learner plays $S_t \subseteq [d]$ with $|S_t| = m$ and observes $X_{i,t} \sim \text{Bernoulli}(\mu_i)$ for each $i \in S_t$. With $S^* = \arg \max_{|S|=m} \sum_{i \in S} \mu_i$, the regret is $R(T) = \sum_{t \leq T} (\sum_{i \in S^*} \mu_i - \sum_{i \in S_t} \mu_i)$. CTS [13] keeps $\text{Beta}(\alpha_i, \beta_i)$ posteriors, draws θ_i independently, plays the top- m of θ , and updates (α_i, β_i) for $i \in S_t$. At a warmup time T_w the learner may send the empirical means $\hat{\mu}$ (and, for MIND, article titles) to an oracle that returns a partition $\mathcal{C} = \{C_1, \dots, C_K\}$ of $[d]$ of unknown quality, which may depend on $\hat{\mu}$.

From partition to covariance. Clusters are ranked by their mean $\hat{\mu}$ at T_w (rank 0 is best), each arm receives the normalized rank $f_i = r_{k(i)}/(K-1) \in [0, 1]$ of its cluster, and

$$\Sigma_{ij} = \rho_{\max} \exp(-(f_i - f_j)^2 / (2\ell^2)) \quad (i \neq j), \quad \Sigma_{ii} = 1. \quad (1)$$

Writing G for the RBF Gram matrix on the features f , $\Sigma = \rho_{\max} G + (1 - \rho_{\max}) I_d \succeq (1 - \rho_{\max}) I_d$ is positive definite for any partition, and $\text{rank}(G) \leq K$ because f takes at most K distinct values; same-cluster arms have correlation $\rho_{\max}$ and adjacent clusters keep positive correlation.

CORRCTS-FULL (pseudocode in Section B). Rounds $t \leq T_w$ are plain CTS. At T_w the partition is turned into Σ and its Cholesky factor L . Afterwards each round takes the Beta posterior mean and standard deviation $(\bar{\mu}_i, \sigma_i)$ of every arm, draws $\epsilon \sim \mathcal{N}(0, I_d)$ and plays the top- m of $\theta = \bar{\mu} + \sigma \odot (L\epsilon)$; the marginal of θ_i is the moment-matched Gaussian $\mathcal{N}(\bar{\mu}_i, \sigma_i^2)$, and only the joint law changes. ROBUSTCORRCTS adds a credibility weight $w \in [0, 1]$ ($w = 1$ initially), updated every Δ_{check} rounds as $w \leftarrow 0.7w + 0.3 \text{overlap}_t$ (overlap_t : fraction of the empirical top- m in the oracle’s best cluster), and plays the top- m of $w\theta^{\text{corr}} + (1-w)\theta^{\text{ind}}$ (independent draws). CORRCTS-PRUNE and CORRCTS-KM refresh Σ from data at long horizons without new queries (Section D).

Proposition 1 (Posterior integrity). *For every reward sequence, the Beta parameters (α_i, β_i) of every arm coincide under CORRCTS-FULL, ROBUSTCORRCTS and CTS; Σ enters only the sampling step, and $\Sigma = I_d$ gives CTS with moment-matched Gaussian marginals (“CTS-Gaussian”).*

3 What a fixed correlation does to the sampling dynamics

3.1 What Σ stands for

Assumption 2 (Structural prior). $\mu \sim \mathcal{N}(\mu_0 \mathbf{1}, \tau^2 \Sigma)$ and, for $i \in S_t$, the observation is $y_{i,t} = \mu_i + \eta_{i,t}$ with $\eta_{i,t} \sim \mathcal{N}(0, \sigma^2)$ independent.

Thus Σ is the prior correlation of the means and the oracle supplies an *estimate* of it; when the estimate is wrong the bound below does not apply, the situation Lemmas 5 and 6, the credibility gate and Section 4.2 address. GP-CTS(Σ) denotes Thompson sampling under this model: $\theta_t \sim \mathcal{N}(\mu_t, \Sigma_t)$ from the exact Gaussian posterior, then top- m .

3.2 Bayesian regret of GP-CTS(Σ)

Let $D_T = \text{diag}(n_1, \dots, n_d)$ count pulls after T rounds ($\sum_i n_i = mT$) and let $\gamma_T = \frac{1}{2} \log \det(I_d + \sigma^{-2} \tau^2 \Sigma^{1/2} D_T \Sigma^{1/2})$ be the information gain of the play sequence.

Lemma 3 (Two-term information gain). *For $\Sigma = \rho G + (1 - \rho)I_d$ with $G \succeq 0$, $G_{ii} = 1$, $\text{rank}(G) \leq K$, and $N = mT$,*

$$\gamma_T \leq \frac{K}{2} \log\left(1 + \frac{\rho\tau^2 N}{K\sigma^2}\right) + \frac{1}{2} \min\left\{\frac{(1 - \rho)\tau^2 N}{\sigma^2}, d \log\left(1 + \frac{(1 - \rho)\tau^2 N}{d\sigma^2}\right)\right\}.$$

For any correlation matrix (e.g. $\Sigma = I_d$), $\gamma_T \leq \frac{d}{2} \log(1 + \tau^2 N/(d\sigma^2))$; conversely, if every arm has been pulled at least n times, $\gamma_T \geq \frac{d}{2} \log(1 + (1 - \rho)\tau^2 n/\sigma^2)$, so the ridge term is not an artifact of the upper bound.

Theorem 4 (Bayesian regret of GP-CTS(Σ)). *Under Assumption 2, with $c = m\tau^2/\sigma^2$, $C_c = c/\log(1 + c)$ and $\beta_T = 2 \log\left(\binom{d}{m} T^2\right) \leq 2m \log d + 4 \log T$,*

$$\mathbb{E}[R(T)] \leq \sqrt{2 C_c m \sigma^2 \beta_T T \gamma_T} + \frac{m\tau}{T},$$

with γ_T bounded as in Lemma 3; the proof (Section A) follows Russo and Van Roy [11] with an elliptical-potential step.

Reading the bound. With $\beta_T = O(m \log(dT))$ and the lemma,

$$\mathbb{E}[R(T)] = O\left(\sqrt{C_c} m \sqrt{T \log(dT)} [K \log T + \min\{(1 - \rho)\tau^2 mT/\sigma^2, d \log T\}]\right).$$

Because Σ has full rank d whenever $\rho < 1$, the bound never depends on K alone: the ridge term vanishes only as $\rho \rightarrow 1$, where the rate becomes $\tilde{O}(m\sqrt{KT})$, a $\sqrt{d/K}$ improvement over the $\tilde{O}(m\sqrt{dT})$ that the same argument gives for $\Sigma = I_d$; for fixed $\rho < 1$ the two rates coincide once $(1 - \rho)\tau^2 mT/\sigma^2 \gtrsim K \log T$. At the experimental values ($d=50, m=5, \rho=0.7, \tau^2=1/12, \sigma^2=1/4$) the crossover is below 150 rounds (Proposition 9), so a proper Bayesian sampler should gain little at $T = 2,500$, as Section 4.2 confirms.

3.3 CORRECTS-FULL versus GP-CTS(Σ)

CORRECTS-FULL is not GP-CTS(Σ): its marginals are the exact Beta moments, but its correlation stays fixed at Σ whereas the Gaussian posterior correlation shrinks as data accumulate; a fixed positive correlation changes top- m selection in two specific ways.

Lemma 5 (Marginals kept, joint exceedance raised). *Let $\theta \sim \mathcal{N}(\bar{\mu}, \text{diag}(\sigma)\Sigma \text{diag}(\sigma))$ for a correlation matrix Σ with $\Sigma_{ij} \geq 0$ for all i, j (every kernel of Equation (1), for any partition). Then each $\theta_i \sim \mathcal{N}(\bar{\mu}_i, \sigma_i^2)$ exactly as under $\Sigma = I_d$, and for every $A \subseteq [d]$ and threshold vector c , $\mathbb{P}_\Sigma(\theta_i > c_i \forall i \in A) \geq \prod_{i \in A} \mathbb{P}(\theta_i > c_i)$, with the left side nondecreasing in each off-diagonal entry of Σ .*

This is Slepian’s inequality [12] and holds for any partition: per-arm concentration is untouched, while the joint optimism event, which independent sampling grants m near-optimal arms only with probability decaying like q^m , can only become more likely; correlation also shrinks the noise in the pairwise comparisons that top- m selection is made of, and in the equicorrelated case this is *exactly* a tempering.

Lemma 6 (Comparison noise and tempering). *Let $\theta \sim \mathcal{N}(\bar{\mu}, \text{diag}(\sigma)\Sigma \text{diag}(\sigma))$. (i) $\text{Var}(\theta_i - \theta_j) = \sigma_i^2 + \sigma_j^2 - 2\Sigma_{ij}\sigma_i\sigma_j$, which for $\sigma_i = \sigma_j$ equals the independent-sampling value times $1 - \Sigma_{ij}$. (ii) If $\Sigma = \rho \mathbf{1}\mathbf{1}^\top + (1 - \rho)I_d$ and $\sigma_i \equiv \sigma$, then top- $m(\theta)$ has exactly the law of top- $m(\bar{\mu} + \sqrt{1 - \rho}\sigma\epsilon)$ with $\epsilon \sim \mathcal{N}(0, I_d)$, i.e. of independent sampling tempered by $c = \sqrt{1 - \rho}$.*

With $\rho_{\max} = 0.7$, $\sqrt{1 - \rho_{\max}} \approx 0.55$, and Section 4.2 finds the best tempering factor near $c = 0.6$; CORRECTS-FULL, whose σ_i differ across arms, is a heteroscedastic, kernel-shaped relaxation of (ii).

3.4 The credibility gate

Proposition 7 (Gate convergence in total variation). *Given a common posterior, the per-round sampling distributions of ROBUSTCORRECTS (correlation $\Sigma_w = w^2\Sigma + (1 - w)^2I_d$ at credibility w) and CTS-Gaussian satisfy $d_{\text{TV}} \leq \frac{3}{2} w_t (\rho_{\max}d + 2\sqrt{d})$. If the overlap statistic is 0 at every check after T_w , then $\sum_{t > T_w} w_t \leq \frac{10}{3} \Delta_{\text{check}}$ and $\sum_{t > T_w} d_{\text{TV}} \leq 5(\rho_{\max}d + 2\sqrt{d})\Delta_{\text{check}}$ for every T .*

Table 1: LLM experiments E1 and E3: paired regret reduction vs CTS (CTS mean regret 330.7 on E1, 1573.7 on E3), paired wins/losses, and Holm-corrected p (sign test for E1, Wilcoxon for E3).

Exp.	Algorithm	Regret	vs CTS	W/L	p
E1 ($n=320$)	CORRCTS-FULL	268.2	+18.9%	262/58	$<10^{-4}$
	Rec.-mixing (belief)	279.5	+15.5%	143/97	0.0036
	ROBUSTCORRCTS	284.2	+14.1%	237/83	$<10^{-4}$
	RandomCorr (block, random partition)	307.2	+7.1%	227/93	$<10^{-4}$
E3 ($n=135$)	CORRCTS-FULL	927.5	+41.1%	134/1	$<10^{-6}$
	CORRCTS-KM	928.6	+41.0%	134/1	$<10^{-6}$
	CORRCTS-PRUNE	943.7	+40.0%	133/2	$<10^{-6}$
	Warm-start CTS (belief)	1591.4	-1.1%	53/82	0.014

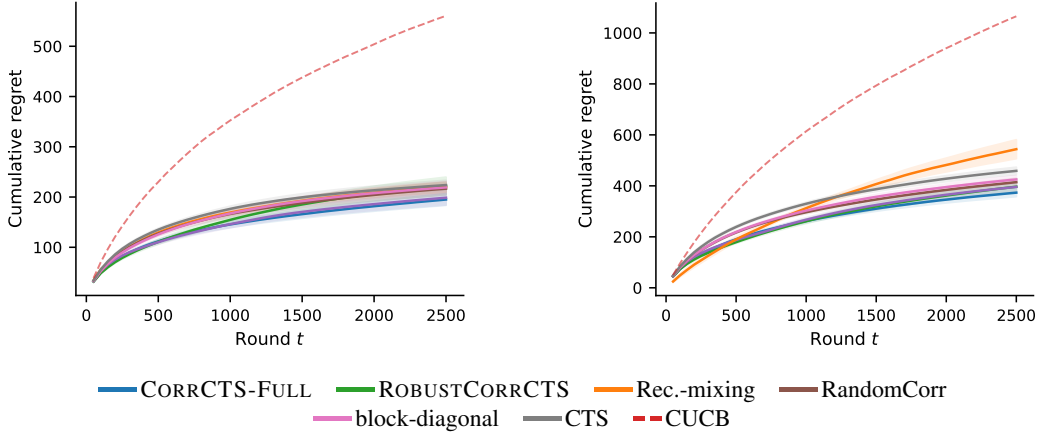


Figure 1: E1 cumulative regret on the uniform-small family at $d=25$ (left) and $d=50$ (right), $m=5$ (20 seeds; bands: one standard error; purple: a gated block-diagonal pilot variant). The correlated samplers separate from CTS just after the query at $T_w=30$ and the gap widens with d ; recommendation mixing competes only at $d=25$.

We claim no regret inequality (that would need a coupling of trajectories sharing a posterior only until their first disagreement): the guarantee is that under sustained disagreement the gate returns to CTS-Gaussian within $O(\Delta_{\text{check}})$ rounds. Corraling with CTS [1] is the principled alternative (Section C).

4 Experiments

4.1 LLM experiments

All comparisons are paired (same instance, same seed) with Holm-Bonferroni correction [7]; $\rho_{\max} = 0.7$, $\ell = 0.5$, $K = 8$, $T_w = 30$ were chosen on a disjoint validation grid (Section E). Table 1 summarizes three experiments with an LLM partition (GPT-5.4 for E1-E2, Claude Sonnet 4.6 for E3; breakdowns in Section G). **E1** (synthetic Bernoulli, $T=2,500$): $d \in \{25, 50\}$, $m \in \{3, 5\}$, four gap families, 20 seeds ($n=320$); gains grow with d (12–15% at $d=25$, 19–25% at $d=50$) and recommendation mixing collapses at $d=50, m=5$ (Figure 1). **E2** ($T=25,000$; the 16 E1 families plus 16 held-out ones, 15 seeds, $n=480$; Section G): CTS 615.9, static CORRCTS-FULL 587.9 (+4.5%), CORRCTS-PRUNE 578.5 (+6.1%), CORRCTS-KM 573.8 (+6.8%), all $p < 10^{-4}$; the static advantage decays and the adaptive variants overtake it, as Theorem 4 predicts. **E3** (MIND-small [14], $d=200$ articles with empirical CTRs, $m=5$, $T=10,000$, three user profiles $\times$ three article samples, 15 seeds, $n=135$): the reduction is stable across profiles (40–42%) while warm start is flooded by real data. Under adversarial clusters (true top- m grouped with the m worst arms, ranked first) CORRCTS-FULL loses 16.5% to CTS on E1 while ROBUSTCORRCTS loses 1.8%.

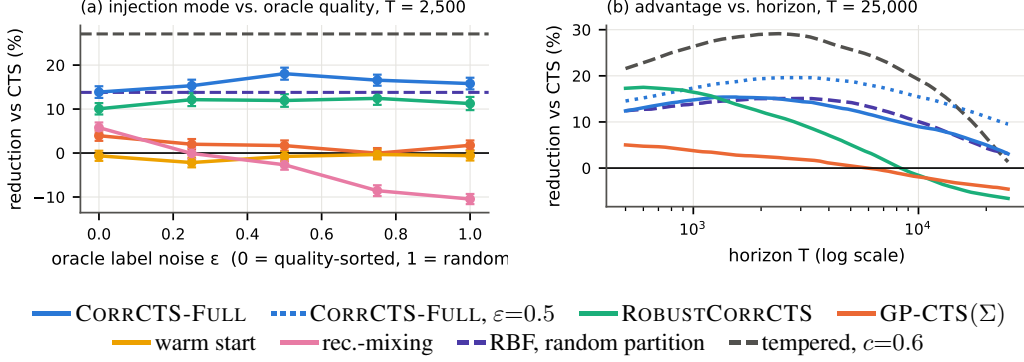


Figure 2: LLM-free ablation. (a) Paired regret reduction versus CTS against oracle label noise ε (dashed: partition-free references; error bars: one standard error over 320 paired trials). (b) The same against the horizon (16 configurations, 10 seeds).

4.2 Controlled ablation with a simulated oracle

The RandomCorr row of Table 1 changes the partition *and* the kernel, so it cannot attribute the gain to partition quality; we therefore re-implemented the E1 family and all LLM-free samplers (Table 3) and replaced the LLM by a *simulated oracle*: the quality-sorted partition into $K = 8$ groups with each label replaced by a random one with probability ε . All modes receive the same partition; the belief baselines inject its best-ranked cluster’s m best- $\hat{\mu}$ arms as 5 pseudo-successes (warm start) or by recommendation mixing (Figure 2; Section F).

Kernel shape versus partition information (Table 4). A *random* partition in *random* order through the RBF kernel already gives +13.8% over CTS, the quality-sorted partition +13.9%, and an adversarial partition that groups the true top- m with the m worst arms (keeping the $\hat{\mu}$ rank order) +14.2%: on these unstructured instances the gain is a property of the kernel shape, not of the partition.

Structure versus belief at matched oracle quality (Figure 2a, Table 6). Across $\varepsilon \in [0, 1]$ the structure samplers stay between +14% and +18%, warm start never differs from CTS, recommendation mixing falls from +5.8% to −10.5% as ε goes from 0 to 1, and the exact Bayesian GP-CTS(Σ) gains at most +3.9% (Theorem 4).

Tempering and the horizon (Figure 2b, Tables 4 and 5). The tempering reading of Lemma 6 is confirmed by the simplest control: independent CTS-Gaussian with posterior standard deviation scaled by $c = 0.6$ reduces regret by +27.1% at $T = 2,500$, more than any kernel. All are transient: at $T = 25,000$ the RBF samplers retain +3%, the tempered sampler +1%, and GP-CTS(Σ) and the gated sampler fall slightly below CTS.

5 Discussion and limitations

Moving LLM information from the posterior to the sampling covariance is safe by construction (Proposition 1) and beneficial at moderate horizons; belief injection at matched oracle quality is not. A *fixed* correlation changes the exploration dynamics (Lemmas 5 and 6) whether or not the partition is right, so a tempered sampler is the LLM-free competitor to beat; partition *information* buys a proper Bayesian sampler only a transient (Theorem 4) that matters where latent structure exists (E3). *Limitations:* Theorem 4 concerns GP-CTS(Σ) rather than CORRCTS-FULL; misspecification is covered only by Lemmas 5 and 6 and Proposition 7; E3 uses one dataset and one LLM.

References

- [1] Alekh Agarwal, Haipeng Luo, Behnam Neyshabur, and Robert E. Schapire. Corraling a band of bandit algorithms. In *Proceedings of the 30th Annual Conference on Learning Theory (COLT)*, 2017.

- [2] Parand A. Alamdari, Yanshuai Cao, and Kevin H. Wilson. Jump starting bandits with LLM-generated prior knowledge. *arXiv preprint arXiv:2406.19317*, 2024.
- [3] Adam Bayley, Kevin H. Wilson, Yanshuai Cao, Raquel Aoki, and Xiaodan Zhu. Robustness of LLM-initialized bandits for recommendation under noisy priors. In *Workshop on Generative AI for Recommender Systems and Personalization (GenAIRecP)*, 2025.
- [4] Djallel Bouneffouf and Raphael Feraud Ferreira. Survey: Multi-armed bandits meet large language models. *arXiv preprint arXiv:2505.13355*, 2025.
- [5] 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 (ICML)*, 2013.
- [6] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional Gaussians with the same mean. *arXiv preprint arXiv:1810.08693*, 2018.
- [7] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.
- [8] Akshay Krishnamurthy, Keegan Harris, Dylan J. Foster, and Aleksandrs Slivkins. Can large language models explore in-context? In *Advances in Neural Information Processing Systems 37 (NeurIPS)*, 2024.
- [9] Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvári. Tight regret bounds for stochastic combinatorial semi-bandits. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015.
- [10] Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.
- [11] Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.
- [12] David Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41(2):463–501, 1962.
- [13] Siwei Wang and Wei Chen. Thompson sampling for combinatorial semi-bandits. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.
- [14] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Dongsheng Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. MIND: A large-scale dataset for news recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

A Proofs

A.1 Proof of Lemma 3

Write $\Sigma_1 = \tau^2 \Sigma$ and $M = \sigma^{-2} \Sigma_1^{1/2} D_T \Sigma_1^{1/2}$; then $\gamma_T = \frac{1}{2} \log \det(I + M)$ and, since $\det(I + AB) = \det(I + BA)$, also $\gamma_T = \frac{1}{2} \log \det(I + \sigma^{-2} D_T^{1/2} \Sigma_1 D_T^{1/2})$. Decompose $\sigma^{-2} D_T^{1/2} \Sigma_1 D_T^{1/2} = A + B$ with $A = \sigma^{-2} \tau^2 \rho D_T^{1/2} K D_T^{1/2}$ and $B = \sigma^{-2} \tau^2 (1 - \rho) D_T$, both positive semidefinite.

Step 1: $\det(I + A + B) \leq \det(I + A) \det(I + B)$. $\det(I + A + B) = \det(I + A) \det(I + C)$ with $C = (I + A)^{-1/2} B (I + A)^{-1/2}$. Since $(I + A)^{-1/2}$ is a contraction, $\lambda_j(C) \leq \lambda_j(B)$ for every j (Courant–Fischer), so $\det(I + C) \leq \det(I + B)$.

Step 2: AM–GM. For a positive semidefinite matrix M of rank at most r , $\log \det(I + M) = \sum_{j \leq r} \log(1 + \lambda_j) \leq r \log(1 + \text{tr}(M)/r)$ by concavity of $\log$. A has rank at most $\text{rank}(K) \leq K$ and $\text{tr}(A) = \sigma^{-2} \tau^2 \rho \sum_i n_i K_{ii} = \sigma^{-2} \tau^2 \rho N$; B is diagonal with trace $\sigma^{-2} \tau^2 (1 - \rho) N$ and rank at most d . Combining the two steps gives the first claim (the min uses $\log(1 + x) \leq x$ on the second term); the second claim is Step 2 applied to the full matrix, whose trace is $\sigma^{-2} \tau^2 N$ for any unit-diagonal Σ . For the lower bound, $\Sigma \succeq (1 - \rho) I_d$ gives $D_T^{1/2} \Sigma_1 D_T^{1/2} \succeq \tau^2 (1 - \rho) D_T$, and $\log \det$ is monotone in the Loewner order, so $2\gamma_T \geq \sum_i \log(1 + \sigma^{-2} \tau^2 (1 - \rho) n_i) \geq d \log(1 + \sigma^{-2} \tau^2 (1 - \rho) n)$. $\square$

A.2 Proof of Theorem 4

Let $\mathcal{A} = \{S \subseteq [d] : |S| = m\}$, $f(S) = \mathbf{1}_S^\top \mu$, and let (μ_t, Σ_t) be the Gaussian posterior mean and covariance of μ given the history $\mathcal{H}_t$ before round t ($\Sigma_1 = \tau^2 \Sigma$). Define the upper index $U_t(S) = \mathbf{1}_S^\top \mu_t + \sqrt{\beta_T} \|\mathbf{1}_S\|_{\Sigma_t}$, where $\|x\|_{\Sigma_t}^2 = x^\top \Sigma_t x$.

Step 1: decomposition. Under Thompson sampling, S_t and S^* have the same conditional law given $\mathcal{H}_t$, and U_t is $\mathcal{H}_t$ -measurable, so $\mathbb{E}[U_t(S^*) \mid \mathcal{H}_t] = \mathbb{E}[U_t(S_t) \mid \mathcal{H}_t]$ [11, Prop. 1]. Hence

$$\mathbb{E}[R(T)] = \mathbb{E} \sum_{t \leq T} [f(S^*) - U_t(S^*)] + \mathbb{E} \sum_{t \leq T} [U_t(S_t) - f(S_t)].$$

Step 2: the first sum is small. Given $\mathcal{H}_t$, $\mathbf{1}_S^\top (\mu - \mu_t) \sim \mathcal{N}(0, \|\mathbf{1}_S\|_{\Sigma_t}^2)$. For $Z \sim \mathcal{N}(0, s^2)$ and $b > 0$, $\mathbb{E}[(Z - bs)_+] = s \int_b^\infty (z - b) \phi(z) dz \leq s e^{-b^2/2}$. Therefore $\mathbb{E}[(f(S^*) - U_t(S^*))_+ \mid \mathcal{H}_t] \leq \sum_{S \in \mathcal{A}} \mathbb{E}[(\mathbf{1}_S^\top (\mu - \mu_t) - \sqrt{\beta_T} \|\mathbf{1}_S\|_{\Sigma_t})_+ \mid \mathcal{H}_t] \leq |\mathcal{A}| e^{-\beta_T/2} \max_S \|\mathbf{1}_S\|_{\Sigma_t}$. Posterior variances never exceed prior variances, so $\|\mathbf{1}_S\|_{\Sigma_t}^2 \leq m \sum_{i \in S} (\Sigma_t)_{ii} \leq m^2 \tau^2$, and with $\beta_T = 2 \log(|\mathcal{A}| T^2)$ the first sum is at most $T \cdot m\tau/T^2 = m\tau/T$.

Step 3: the second sum. S_t is a function of $\mathcal{H}_t$ and of sampling noise independent of μ , so $\mathbb{E}[\mathbf{1}_{S_t}^\top (\mu_t - \mu) \mid \mathcal{H}_t] = 0$ and the second sum equals $\sqrt{\beta_T} \mathbb{E} \sum_t \|\mathbf{1}_{S_t}\|_{\Sigma_t} \leq \sqrt{\beta_T} \mathbb{E} \sqrt{T \sum_t \|\mathbf{1}_{S_t}\|_{\Sigma_t}^2}$ by Cauchy–Schwarz.

Step 4: elliptical potential. Let $x_t = \sigma^{-2} \sum_{i \in S_t} (\Sigma_t)_{ii} \leq m\tau^2/\sigma^2 = c$. Because $x \mapsto \log(1 + x)/x$ is decreasing, $x_t \leq C_c \log(1 + x_t)$ with $C_c = c/\log(1 + c)$. The precision recursion $\Sigma_{t+1}^{-1} = \Sigma_t^{-1} + \sigma^{-2} \sum_{i \in S_t} e_i e_i^\top$ gives $\log \det \Sigma_{t+1}^{-1} - \log \det \Sigma_t^{-1} = \log \det(I + \sigma^{-2} \Sigma_t^{1/2} E_{S_t} \Sigma_t^{1/2}) \geq \log(1 + x_t)$, using $\det(I + M) \geq 1 + \text{tr}(M)$ for $M \succeq 0$. Telescoping, $\sum_t \log(1 + x_t) \leq \log \det(\Sigma_1 \Sigma_{T+1}^{-1}) = 2\gamma_T$. Finally $\|\mathbf{1}_{S_t}\|_{\Sigma_t}^2 \leq m \sum_{i \in S_t} (\Sigma_t)_{ii} = m\sigma^2 x_t$, so $\sum_t \|\mathbf{1}_{S_t}\|_{\Sigma_t}^2 \leq m\sigma^2 C_c \cdot 2\gamma_T$. Combining Steps 1–4, $\mathbb{E}[R(T)] \leq \sqrt{2C_c m\sigma^2 \beta_T T \gamma_T} + m\tau/T$, where γ_T may be replaced by the deterministic bound of Lemma 3 because that bound depends on the play sequence only through $N = mT$. $\square$

A.3 Proof of Lemma 5

The marginal claim is immediate since $\Sigma_{ii} = 1$. Restrict to the coordinates in A and standardize: $Z_i = (\theta_i - \bar{\mu}_i)/\sigma_i$ is centered Gaussian with unit variances and correlations Σ_{ij} . Slepian’s inequality [12] states that for two centered Gaussian vectors with equal variances and $\Sigma_{ij}^{(1)} \leq \Sigma_{ij}^{(2)}$ for all $i \neq j$, $\mathbb{P}(Z_i^{(1)} \leq c_i \forall i) \leq \mathbb{P}(Z_i^{(2)} \leq c_i \forall i)$ for every c ; applied to $-Z^{(1)}, -Z^{(2)}$ (same covariances) it gives

the same monotonicity for $\{Z_i > c_i \forall i\}$. Taking $\Sigma^{(1)} = I_{|A|}$ and $\Sigma^{(2)} = \Sigma_{AA}$, which is admissible because $\Sigma_{ij} \geq 0$, yields the product lower bound; monotonicity in each entry follows from the same comparison. $\square$

A.4 Proof of Lemma 6

(i) is the bilinearity of covariance: $\text{Var}(\theta_i - \theta_j) = \sigma_i^2 + \sigma_j^2 - 2 \text{Cov}(\theta_i, \theta_j)$ and $\text{Cov}(\theta_i, \theta_j) = \sigma_i \sigma_j \Sigma_{ij}$; under independence the same expression has $\Sigma_{ij} = 0$, and for $\sigma_i = \sigma_j = \sigma$ the ratio is $2\sigma^2(1 - \Sigma_{ij})/(2\sigma^2)$. (ii) An equicorrelated vector has the representation $\theta = \bar{\mu} + \sigma(\sqrt{\rho} Z \mathbf{1} + \sqrt{1 - \rho} \epsilon)$ with $Z \sim \mathcal{N}(0, 1)$ independent of $\epsilon \sim \mathcal{N}(0, I_d)$, since $\text{Cov}(\theta) = \sigma^2(\rho \mathbf{1}\mathbf{1}^\top + (1 - \rho)I_d)$. The term $\sigma\sqrt{\rho} Z \mathbf{1}$ adds the same number to every coordinate and therefore leaves the ordering of the coordinates, hence the top- m set, unchanged almost surely (ties have probability zero). Thus $\text{top-}m(\theta) = \text{top-}m(\bar{\mu} + \sqrt{1 - \rho} \sigma \epsilon)$ pathwise, which is the claim. $\square$

A.5 Two consequences of Lemma 3 and Theorem 4

Corollary 8 (Independent prior). *For $\Sigma = I_d$ (CTS-Gaussian under Assumption 2), $\mathbb{E}[R(T)] \leq \sqrt{C_c m \sigma^2 \beta_T T d \log(1 + \tau^2 m T / (d \sigma^2))} + m \tau / T$.*

Proof. Insert the second claim of Lemma 3 into Theorem 4. $\square$

Proposition 9 (Crossover horizon). *Let $T^* = \sup\{T : (1 - \rho)\tau^2 m T / \sigma^2 \leq K \log(1 + \rho\tau^2 m T / (K \sigma^2))\}$. For every $T \leq T^*$, $\gamma_T \leq K \log(1 + \rho\tau^2 m T / (K \sigma^2))$, so the bound of Theorem 4 reads $O(\sqrt{C_c m} \sqrt{KT \log(dT) \log T})$, and*

$$T^* \leq \frac{4\rho K \sigma^2}{(1 - \rho)^2 \tau^2 m}.$$

Proof. For $T \leq T^*$ the second term of Lemma 3 is at most $\frac{1}{2}(1 - \rho)\tau^2 m T / \sigma^2 \leq \frac{K}{2} \log(1 + \rho\tau^2 m T / (K \sigma^2))$, which gives the first claim. For the second, write $x = (1 - \rho)\tau^2 m T^* / \sigma^2$; the defining inequality is $x \leq K \log(1 + \frac{\rho}{(1 - \rho)K} x)$, and $\log(1 + y) \leq 2\sqrt{y}$ gives $x \leq 2K \sqrt{\rho x / ((1 - \rho)K)}$, i.e. $\sqrt{x} \leq 2\sqrt{\rho K / (1 - \rho)}$, i.e. $x \leq 4\rho K / (1 - \rho)$; solving for T^* gives the bound. At the experimental values ($d=50, m=5, K=8, \rho=0.7, \tau^2=1/12, \sigma^2=1/4$) the bound is $T^* \leq 149$ rounds. $\square$

A.6 Proof of Proposition 7

Given a common posterior, the two draws are $\bar{\mu} + \text{diag}(\sigma)\xi$ with $\xi \sim \mathcal{N}(0, \Sigma_w)$ and $\xi \sim \mathcal{N}(0, I_d)$ respectively; total variation is invariant under the common affine map, so $d_{\text{TV}} = d_{\text{TV}}(\mathcal{N}(0, \Sigma_w), \mathcal{N}(0, I_d)) \leq \frac{3}{2} \|\Sigma_w - I_d\|_F$ by Devroye et al. [6, Thm. 1.1]. Now $\Sigma_w - I_d = w^2(\Sigma - I_d) - (2w - w^2)I_d$, $\|\Sigma - I_d\|_F \leq \rho_{\max} \sqrt{d(d-1)} \leq \rho_{\max} d$ and $\|(2w - w^2)I_d\|_F \leq 2w\sqrt{d}$, giving $d_{\text{TV}} \leq \frac{3}{2}w(\rho_{\max} d + 2\sqrt{d})$ since $w \leq 1$. Under zero overlap at every check, w equals 0.7^k during the Δ_{check} rounds following the k -th check, so $\sum_t w_t \leq \Delta_{\text{check}} \sum_{k \geq 0} 0.7^k = \frac{10}{3} \Delta_{\text{check}}$. $\square$

B Pseudocode

Algorithm 1: CORRCTS-FULL

Input: d, m , warmup T_w , kernel parameters $\rho_{\max}, \ell$, number of clusters K
 $\alpha_i \leftarrow 1, \beta_i \leftarrow 1$ for all i ; $L \leftarrow I_d$;
for $t = 1, \dots, T$ **do**
 if $t \leq T_w$ **then**
 $\theta_i \sim \text{Beta}(\alpha_i, \beta_i)$ independently;
 else
 if $t = T_w + 1$ **then**
 $\mathcal{C} \leftarrow \text{ORACLECLUSTERQUERY}(\hat{\mu}, K)$; build Σ from Equation (1);
 $L \leftarrow \text{Cholesky}(\Sigma + 10^{-3}I_d)$;
 end
 $\theta \leftarrow \bar{\mu} + \sigma \odot (L\epsilon)$ with $\epsilon \sim \mathcal{N}(0, I_d)$;
 end
 play $S_t = \text{top-}m(\theta)$; observe $X_{i,t}$; $\alpha_i \leftarrow \alpha_i + X_{i,t}, \beta_i \leftarrow \beta_i + 1 - X_{i,t}$ for $i \in S_t$;
end

C Alternatives: GP-CTS(Σ) directly, and corraling

GP-CTS(Σ) (as run in Section 4.2) uses prior $\mathcal{N}(\frac{1}{2}\mathbf{1}, \tau^2\Sigma)$ with $\tau^2 = 1/12$ (the variance of a Beta(1, 1) mean) and $\sigma^2 = 1/4$ (the Bernoulli variance bound), a CTS warmup of T_w rounds, and thereafter the exact Gaussian posterior computed from all counts. It is the algorithm Theorem 4 is about. Its posterior correlation between two arms decays at rate $1/n$ in their pull counts, which is why the structural effect is transient; CORRCTS-FULL freezes the correlation and thereby trades Bayesian coherence for a persistent change in the sampling dynamics. A corralled combination [1] of CTS and GP-CTS(Σ) would give a consistency–robustness guarantee of the learning-augmented type [10]: regret within $\tilde{O}(\sqrt{T})$ of the better base algorithm. The credibility gate of ROBUSTCORRCTS is a heuristic with the same purpose and only the total-variation guarantee of Proposition 7; corraling is the right baseline for a follow-up and was not benchmarked here.

D Data-adaptive variants

CORRCTS-PRUNE keeps the initial kernel and, every 500 rounds, zeroes Σ_{ij} whenever $|\hat{\mu}_i - \hat{\mu}_j| > 2\sqrt{1/n_i + 1/n_j}$, then refactorizes. CORRCTS-KM blends the initial kernel with a kernel built from k -means on $\hat{\mu}$ at $t \in \{2,000, 10,000\}$ with weight $w(t) = \max(0.2, 1 - t/12,000)$ on the LLM kernel. Both make the decay of the correlation part of the dynamics rather than leaving it fixed.

E Experimental details

E1 instances. For $(d, m) \in \{25, 50\} \times \{3, 5\}$: *uniform-large* (top m at 0.7, rest 0.55), *uniform-small* (rest 0.62), *hard-gap* (top m at 0.7, next m at 0.65, rest 0.5), *staggered* (means linearly spaced from 0.7 to 0.3); $T = 2,500$, 20 seeds.

E2. The 16 E1 families plus 16 with fresh environment seeds; $T = 25,000$, 15 seeds.

E3. MIND-small [14], 17,226 articles with empirical CTRs; each instance samples $d = 200$ articles; three user profiles (“diverse”, “entertainment”, “tech”) reweight CTRs by category with weights fixed before any run; environment seeds $\{100, 200, 300\}$; $T = 10,000$, 15 seeds; oracle Claude Sonnet 4.6 (AWS Bedrock, temperature 0.3) receiving titles and $\hat{\mu}$ and asked for $K = 8$ topic clusters.

Hyperparameters. $T_w = 30$, $K = 8$, $\rho_{\max} = 0.7$, $\ell = 0.5$; ROBUSTCORRCTS $\Delta_{\text{check}} = 100$; RandomCorr within-cluster $\rho = 0.6$; recommendation mixing $\tau = 100$, re-query every 100 rounds, one pseudo-success per recommended arm at each query; warm start 5 pseudo-successes.

$\rho_{\max}, \ell, K, T_w$ were selected on $(d, m) \in \{(25, 3), (50, 5)\} \times \{\text{uniform-large, hard-gap}\}$ with environment seeds 700–704, disjoint from all test seeds. Sensitivity: reductions on E1 stay within 15–20% for $\rho_{\max} \in [0.5, 0.8]$ and $\ell \in [0.5, 1.0]$; $\rho_{\max} = 0.9$ and $\ell = 0.25$ or 2.0 lose 2–4 points; $T_w \in \{10, 100\}$ loses 2–4 points; $K \in \{6, \dots, 12\}$ is flat.

Simulated-oracle ablation. Same instances and seeds as E1; oracle label noise $\varepsilon \in \{0, 1/4, 1/2, 3/4, 1\}$; the adversarial partition places the true top- m and the m worst arms in one cluster and splits the remaining arms into $K - 1$ quality-sorted groups, with cluster order by $\hat{\mu}$ as in Equation (1) (in the E1 adversarial test the merged cluster was additionally forced to rank first); all samplers share the CTS warmup and the kernel of Equation (1) with the E1 hyperparameters. The tempered control draws $\theta = \bar{\mu} + c\sigma \odot \epsilon$ with independent ϵ . Each trial takes 0.02–0.2 s on one CPU core; the full sweep is about two minutes on ten cores.

Compute. E1–E3 ran on a 16-core machine (3 s, 25 s and 12 s per trial respectively); total LLM cost about \$12; LLM responses were cached.

F Tables for the LLM-free ablation

Table 3 checks the re-implementation, Table 4 is the kernel-shape ablation with the tempered controls, Table 6 tabulates Figure 2a, Table 5 tabulates Figure 2b, and Table 2 gives the per- (d, m) breakdown.

Table 2: Per- (d, m) mean regret in the LLM-free ablation ($T=2,500$; 4 gap families $\times$ 20 seeds = 80 trials per cell).

Sampler	(25, 3)	(25, 5)	(50, 3)	(50, 5)
CTS	256.7	284.4	442.2	492.6
RBF kernel, random partition/order	221.4	257.2	363.9	429.8
CORRCTS-FULL ($\varepsilon=0$)	225.0	247.3	364.7	434.3
CORRCTS-FULL ($\varepsilon=0.5$)	223.4	233.8	338.9	413.4
CORRCTS-FULL ($\varepsilon=1$)	221.7	240.5	353.2	427.3
GP-CTS(Σ) ($\varepsilon=0$)	250.6	257.2	421.5	488.4
ROBUSTCORRCTS ($\varepsilon=0$)	250.6	247.9	390.3	438.6
Warm start ($\varepsilon=0$)	256.1	280.4	452.7	496.1
Rec.-mixing ($\varepsilon=0$)	254.2	239.1	431.2	466.0
Tempered CTS-Gaussian ($c=0.6$)	196.8	241.3	281.6	356.5

Table 3: Re-implementation check of the LLM-free rows of Experiment 1 (16 configurations $\times$ 20 seeds, $T=2,500$). “Reported” are the numbers from the LLM experiment harness; “re-run” are from an independent re-implementation of the same instance family and algorithms. Cells are mean cumulative regret; the four right-hand columns are the per- (d, m) means.

Algorithm	Source	Overall	(25, 3)	(25, 5)	(50, 3)	(50, 5)
CTS	reported	330.7	224.5	223.8	415.7	458.8
	re-run	369.0	256.7	284.4	442.2	492.6
CUCB	reported	776.3	546.4	560.6	933.2	1065.2
	re-run	735.8	539.0	583.4	800.2	1020.7
RandomCorr (block, $\rho=0.6$)	reported	307.2	210.9	216.7	386.8	414.4
	re-run	351.2	253.2	261.4	407.8	482.4

G Additional tables for the LLM experiments

E2 by split. Mean regret on the original / held-out families ($n=240$ each): CORRCTS-PRUNE 558.0 / 598.9, CORRCTS-FULL 577.5 / 598.3, CTS 614.7 / 617.1; CORRCTS-KM was run on both splits jointly (573.8). Pilot alternatives were at or below CTS at this horizon: per-arm damping 613.6 (+0.4%), kernel decay toward I_d 623.8 (−1.3%), a hierarchical-Bayes model over cluster means 630.5 (−2.4%).

Table 4: Kernel-shape versus partition-information ablation (LLM-free; $n=320$ paired trials per row, $T=2,500$). All RBF rows use the CORRCTS-FULL kernel with identical $(\rho_{\max}, \ell, K) = (0.7, 0.5, 8)$ and differ only in what the partition and cluster order contain. “vs CTS” is the paired mean regret reduction; W/L are paired wins/losses; p is a two-sided Wilcoxon signed-rank test.

Sampler	Partition	Cluster order	Regret	vs CTS	W/L (p)
CTS	—	—	369.0 ± 6.6	—	—
Block-diagonal (RandomCorr)	random	none	351.2 ± 6.7	+4.8%	202/118 ($<10^{-4}$)
RBF kernel	random	random	318.1 ± 6.6	+13.8%	239/81 ($<10^{-4}$)
RBF kernel	random	by $\hat{\mu}$	309.9 ± 6.4	+16.0%	248/72 ($<10^{-4}$)
RBF kernel	oracle, $\varepsilon=1$	by $\hat{\mu}$	310.7 ± 6.6	+15.8%	250/70 ($<10^{-4}$)
RBF kernel	oracle, $\varepsilon=0.5$	by $\hat{\mu}$	302.4 ± 6.2	+18.0%	252/68 ($<10^{-4}$)
RBF kernel	oracle, $\varepsilon=0$	by $\hat{\mu}$	317.8 ± 6.7	+13.9%	239/81 ($<10^{-4}$)
RBF kernel	adversarial	by $\hat{\mu}$	316.4 ± 6.6	+14.2%	242/78 ($<10^{-4}$)
RBF + credibility gate	adversarial	by $\hat{\mu}$	338.3 ± 6.7	+8.3%	208/112 ($<10^{-4}$)
Tempered CTS-Gaussian ($c=0.8$)	none	none	302.8 ± 6.1	+17.9%	247/73 ($<10^{-4}$)
Tempered CTS-Gaussian ($c=0.6$)	none	none	269.0 ± 7.1	+27.1%	273/47 ($<10^{-4}$)
Tempered CTS-Gaussian ($c=0.4$)	none	none	316.0 ± 11.0	+14.4%	220/100 ($<10^{-4}$)

Table 5: Horizon dependence (LLM-free; 16 configurations $\times$ 10 seeds, $T=25,000$). Paired regret reduction versus CTS at intermediate horizons, read off the cumulative-regret curves of the same runs; the last column is the final W/L.

Sampler	$T=2.5k$	5k	10k	20k	25k	W/L
RBF kernel, random partition/order	+15.1%	+14.2%	+10.1%	+4.3%	+3.1%	114/46
CORRCTS-FULL ($\varepsilon=0$)	+14.9%	+12.6%	+9.0%	+5.0%	+3.1%	113/47
CORRCTS-FULL ($\varepsilon=0.5$)	+19.6%	+18.8%	+15.5%	+11.1%	+9.6%	122/38
GP-CTS(Σ) ($\varepsilon=0$)	+2.2%	+0.5%	-1.9%	-4.0%	-4.6%	65/94
ROBUSTCORRCTS ($\varepsilon=0$)	+10.8%	+5.1%	-1.6%	-5.8%	-6.6%	82/78
Tempered CTS-Gaussian ($c=0.6$)	+29.1%	+25.8%	+19.2%	+6.6%	+1.3%	112/48

Table 6: Structure versus belief injection at matched oracle quality (LLM-free; $n=320$ paired trials per cell, $T=2,500$). Every column receives the *same* simulated partition; ε is the fraction of arms whose cluster label is replaced by a uniformly random one ($\varepsilon=0$: quality-sorted partition; $\varepsilon=1$: random). NMI is against the top- m /rest partition (for reference, GPT-5.4 clusters in Experiment 1 had NMI 0.155; random 0.10). Cells are paired regret reduction versus CTS (positive is better); W/L in brackets.

ε	NMI	CORRCTS-FULL	GP-CTS(Σ)	ROBUSTCORRCTS	Warm-start (belief)	Rec.-mixing (belief)
0.00	0.218	+13.9% [239/81]	+3.9% [189/131]	+10.1% [216/104]	-0.6% [155/165]	+5.8% [210/110]
0.25	0.172	+15.3% [249/71]	+2.0% [175/145]	+12.1% [224/96]	-2.2% [143/177]	-0.1% [171/149]
0.50	0.139	+18.0% [252/68]	+1.7% [179/141]	+11.9% [231/89]	-0.8% [153/167]	-2.7% [143/177]
0.75	0.104	+16.6% [257/63]	-0.0% [152/168]	+12.4% [223/96]	-0.3% [158/162]	-8.6% [105/215]
1.00	0.097	+15.8% [250/70]	+1.8% [171/149]	+11.3% [231/89]	-0.6% [160/160]	-10.5% [88/232]

Endogenous oracle quality. On the $d=25, m=5$ family (20 seeds), the LLM’s best cluster covered 4 of 5 true top- m slots when $\hat{\mu}$ came from 30 rounds of CTS warmup, but 0.15 of 5 on average when the same 30 rounds were round-robin, with only the numbers in the prompt differing; oracle quality is therefore endogenous to the learner’s trajectory, which is why we do not model it as an exogenous constant.

E1 ablations. CORRCTS-FULL beats RandomCorr by 12.7% ($p < 10^{-4}$, 228/92); a block-diagonal kernel on the LLM partition (within-cluster correlation only, no rank ordering) does not beat RandomCorr ($p = 0.34$). Under adversarial clusters (true top- m grouped with the m worst arms, ranked first) CORRCTS-FULL incurs 385.2 against 330.7 for CTS (+16.5% regret) and ROBUSTCORRCTS 336.8 (+1.8%). Section 4.2 shows that the margin over RandomCorr is explained by kernel shape.

Table 7: E1 mean regret by (d, m) (80 trials per cell) and E3 regret reduction by user profile ($n=45$ per cell).

E1	(25, 3)	(25, 5)	(50, 3)	(50, 5)
CORRCTS-FULL	192.0	195.1	312.5	373.3
ROBUSTCORRCTS	191.3	198.6	351.0	395.8
Rec.-mixing	214.1	219.8	393.3	543.8
RandomCorr	210.9	216.7	386.8	414.4
CTS	224.5	223.8	415.7	458.8
CUCB	546.4	560.6	933.2	1065.2
E3	Diverse	Entmt.	Tech	
CORRCTS-FULL	+41.4%	+41.5%	+40.4%	
CORRCTS-KM	+41.2%	+41.5%	+40.3%	
CORRCTS-PRUNE	+40.5%	+39.6%	+40.0%	
Warm-start	−1.6%	−0.3%	−1.4%	

Cluster quality. NMI between the LLM partition and the top- m /rest partition over the 1,280 E1 queries was 0.155 ± 0.002 (random: 0.10), ranging from 0.09 at $(50, 3)$ to 0.23 at $(25, 5)$. NMI rose with model capability on the validation grid (GPT-4.1-mini 0.41, GPT-5-mini 0.44, GPT-5.4 0.52). Given Section 4.2, the E1 gain is not read as evidence of partition quality; the NMI numbers are reported for completeness.

Cost. One cluster query is about 620 tokens (\$0.003 on GPT-5.4); recommendation mixing used about 25 queries per E1 trial. Per round, CORRCTS-FULL costs one $O(d^2)$ matrix–vector product plus a one-time $O(d^3)$ Cholesky (< 50 ms at $d = 200$); ESCB-style index maximization over $\binom{d}{m}$ subsets is infeasible at $d = 200, m = 5$ without additional structure.

H Additional figures

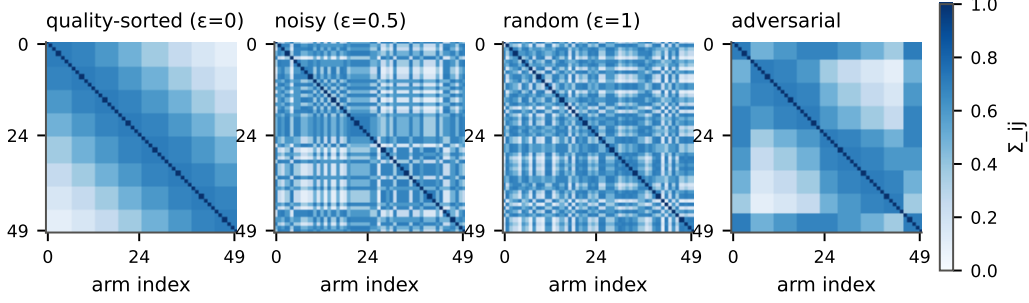


Figure 3: The correlation matrix Σ of Equation (1) ($\rho_{\max}=0.7$, $\ell=0.5$, $K=8$) on a $d=50$ staggered instance, arms sorted by true mean, for a quality-sorted partition, a noisy one ($\varepsilon=0.5$), a random one, and the adversarial construction. Every matrix has unit diagonal and $\lambda_{\min} \geq 1 - \rho_{\max}$; only the placement of the off-diagonal mass changes, which is what Section 4.2 shows matters little on unstructured instances.

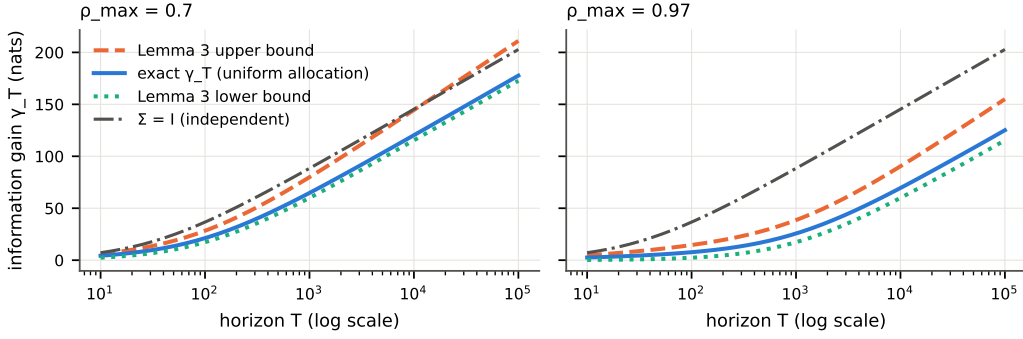


Figure 4: Numerical check of Lemma 3: exact information gain γ_T of the kernel of Figure 3 under a uniform allocation ($\tau^2=1/12$, $\sigma^2=1/4$, $m=5$) against the two-term upper bound, the lower bound, and the independent-prior value. At $\rho_{\max}=0.7$ the curves merge with the independent case within a few hundred rounds; at $\rho_{\max}=0.97$ the structural regime is visibly longer, as Proposition 9 predicts.

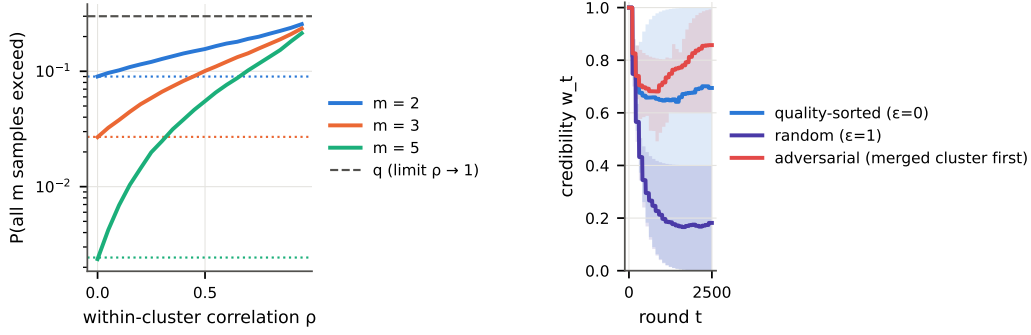


Figure 5: Left: Lemma 5 numerically. Probability that all m equicorrelated standard Gaussian samples exceed the $q=0.3$ upper quantile, as a function of the correlation ρ (2×10^5 Monte Carlo draws); dotted lines are the independent values q^m , the dashed line the limit q . Right: credibility w_t of ROBUSTCORRCTS on the $d=50, m=5$ hard-gap family (10 seeds; band is the min–max range) for a quality-sorted partition, a random one, and the adversarial construction with the merged cluster ranked first. The decay is geometric at the checks every $\Delta_{\text{check}}=100$ rounds as in Proposition 7, but the overlap statistic compares the oracle’s best cluster with the top- m of the *raw* empirical means, which rarely pulled arms with $\hat{\mu}_i \approx 1$ enter; the gate therefore also decays under a good partition (the cost visible in Tables 4 and 6), and it does not fire when the merged adversarial cluster contains the true top- m . A lower-confidence-bound statistic would remove both effects.

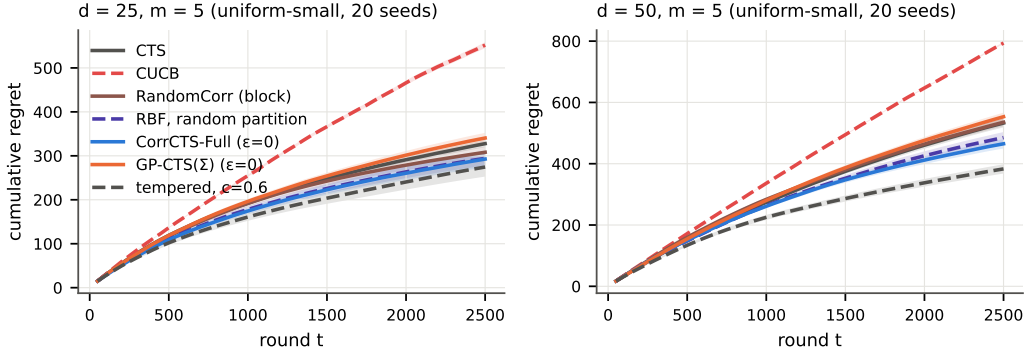


Figure 6: LLM-free cumulative regret on the uniform-small family (20 seeds; bands are one standard error) for the independent re-implementation of Section 4.2: the RBF kernel on a random partition and CORRCTS-FULL on the quality-sorted partition are nearly indistinguishable, the tempered sampler is lowest at this horizon, and the exact Bayesian GP-CTS(Σ) tracks CTS.

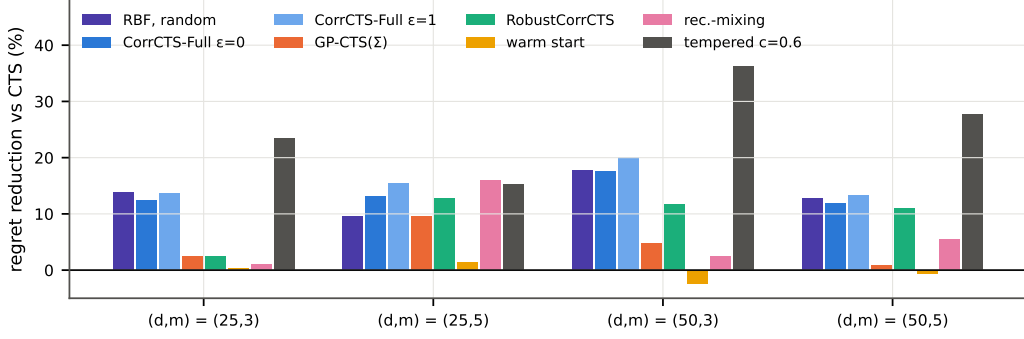


Figure 7: Per- (d, m) paired regret reduction versus CTS at $T=2,500$ for the samplers of Section 4.2 (80 trials per cell). The ordering is the same in every cell: tempering and the RBF samplers (whatever the partition) gain the most, the exact Bayesian sampler and the belief-injection baselines gain little or nothing.

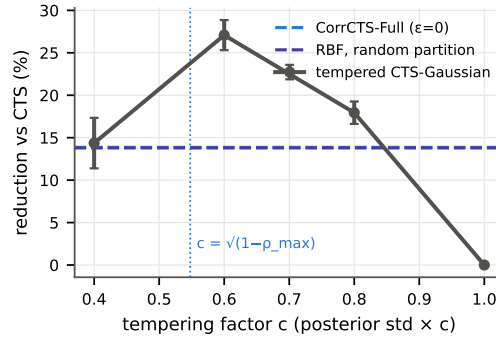


Figure 8: Tempering sweep. Paired regret reduction of independent CTS-Gaussian with posterior standard deviation scaled by c ($n=320$; $c=1$ is CTS-Gaussian), with the two RBF samplers as dashed references and the value $c = \sqrt{1 - \rho_{\max}}$ of Lemma 6(ii) marked.

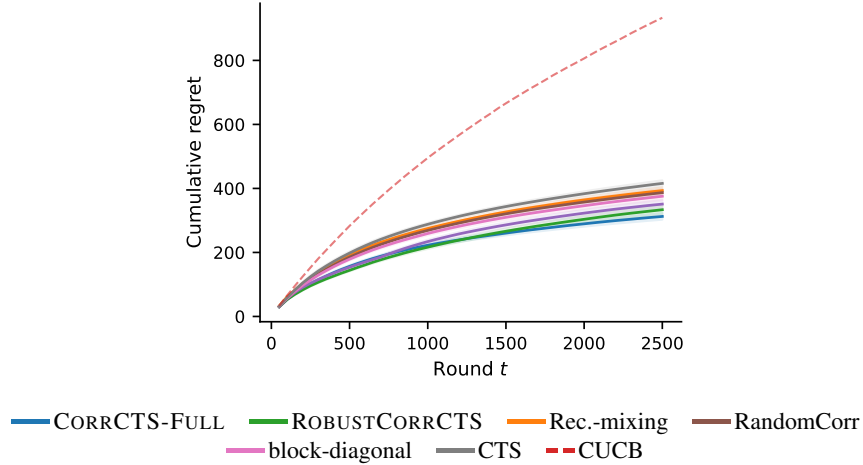


Figure 9: E1 cumulative regret (LLM partitions) on the uniform-small family at $d=50$, $m=3$ (20 seeds; bands are one standard error), complementing Figure 1.