A. Appendix

A.1. Cases of several arms having the same expectation

Up to now, we have assumed that all arms have distinct expectations. Here, we consider cases in which some arms have the same expectations. Without loss of generality, we assume $\mu_1 \geq \mu_2 \geq, \dots, \geq \mu_K$. Let us call arms with a larger expectation than μ_L "strictly optimal" arms, arms with the same expectation as μ_L "marginal" arms, and arms with a smaller expectation than μ_L "strictly suboptimal" arms. Each arm is either strictly optimal, marginal, or strictly suboptimal.

Case 1: Assume that all strictly optimal arms are distinct, that there is only one marginal arm, and that there are several strictly suboptimal arms with the same expectation. In this case, the regret bound of Theorem 1 holds because our analysis deals with each suboptimal arm separately.

Case 2: Assume that there is only one marginal arm, that all strictly suboptimal arms are distinct, and that there are several strictly optimal arms with the same expectation. The regret bound also holds in this case since there is a gap between each strictly suboptimal arm and each strictly optimal arm.

Case 3: Assume that all strictly optimal arms and strictly suboptimal arms are distinct and that there are several marginal arms with the same expectation. Unfortunately, we were unable to perform a meaningful analysis in this case. Intuitively, as stated by Agrawal and Goyal (Agrawal & Goyal, 2012) for SP-MAB, adding an additional marginal arm appears to require some extra exploration, which slightly increases the regret. However, the regret structure is more complex than the SP-MAB because several marginal arms can be drawn simultaneously.

In summary, our Theorem 1 holds when the marginal arm is distinct. That is, $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{L-1} > \mu_L > \mu_{L+1} \geq \cdots \geq \mu_K$.

A.2. Cascade model and position-dependent MP-MAB problem

In the main paper, we assumed that the rewards of arms are independently and identically drawn from individual distributions. In this section, we relax this assumption and consider a wider class of the MP-MAB problem. Remember that, one of our primary applications is multiple advertisement placement in the online advertising problem (c.f., Example 1). In this section, we interchangeably use the terms an advertisement (ad) and an arm. It is known that the CTR of an ad depends on the environment where the ad is placed, especially on the position of the ad. Among several models that explain this dependency on the position, the model that explains human behavior and agrees well with real data (Craswell et al., 2008) is the *cascade* model (Kempe & Mahdian, 2008; Aggarwal et al., 2008), with which it is assumed that the user scans the ads from top to bottom. Following Gatti et al. (2012), we define the discount factor $\gamma_l(i)$ for $l \geq 2$ as the probability that a user observing ad i in position $l-1$ will observe the ad in the next position. Namely, the MP-MAB problem with a discount factor is defined as a MP-MAB problem in which the arm at position l yields reward 1 with probability $\left(\prod_{l'=2}^{l} \gamma_{l'}(I_{l'-1}(t))\right) \mu_{I_l(t)}$, where $I_l(t)$ be the arm placed at the l-th position at round t. Note that, when we set $\gamma_l(i) = 1$ for any position $l \in [L]$ and ad i, this model is reduced to the model we considered in the main paper. In the MP-MAB problem in the main paper, the order of the L arms does not matter. Whereas, under a position-dependent discount factor smaller than 1, the order of L arms matters: the problem is not the selection of an L-set of arms, but an L-sequence of arms.

Algorithm 2 Bias-Corrected Multiple-play Thompson sampling (BC-MP-TS) for binary rewards

```
Input: # of arms $K$, # of positions $L$, discount factors $\{\gamma_l(i)\}$
for $i = 1, 2, \dots, K$ do
  $A_i, N_i = 1, 2$
end for
$t \leftarrow 1.$
for $t = 1, 2 \dots, T$ do
  for $i = 1, 2, \dots, K$ do
    $B_i \leftarrow \max(N_i - A_i, 1)$
    $\theta_i(t) \sim \mathrm{Beta}(A_i, B_i)$
  end for
  Select $I_l(t)$ $(l = 1, \dots, L)$ in accordance with Section A.2.2.
  for $l \in 1, 2, \dots, L$ do
    if $X_i(t) = 1$ then
      $A_i \leftarrow A_i + 1$
    end if
    $N_i \leftarrow N_i + \prod_{l'=2}^{l} \gamma_{l'}(I_{l'-1}(t))$
  end for
end for
```

A.2.1. Thompson sampling for cascade model

In the cascade model, there is some probability that the arm at position $l > 1$ is not drawn. The probability that the arm at position l is drawn, $\prod_{l'=2}^{l} \gamma_{l'}(I_{l'-1}(t))$, can be considered as the *effective number of the draws* at position i. MP-TS (Algorithm 1) keeps A_i and B_i, which respectively correspond to the number of rewards 1 and 0. The number of draws on the arm i is $N_i = A_i + B_i$. When we consider the cascade model, we need to take the effective number of draw into consideration. We introduce Bias-corrected MP-

TS (BC-MP-TS, Algorithm 2). The crux of BC-MP-TS is that, for each arm that is selected, N_i should be increased not by 1, but by the effective number of draw for each position. Note that, when $\gamma_l(i) = 1$, BC-MP-TS is essentially the same as MP-TS.

A.2.2. Optimal arm selection and the regret

In general discount factor $\gamma_l(i)$, even if we have perfect information over the expectation of all arms $\{\mu_i\}_{i=1}^K$, the computation of the optimal sequence of L-arms at each round t (optimal arm selection) appears to be computationally intractable when K is large because we need to search all the possible allocation of K ads over L positions. Kempe & Mahdian (2008) proposed a polynomial-time approximation of the optimal arm selection. We can obtain the arm selection strategy for BC-MP-TS by using this approximation algorithm as an oracle and plugging $\{\theta_i(t)\}_{i=1}^L$ as estimated expected rewards.

Ad-independent discount factor: when the discount factor is independent of the ad at that position (i.e., $\gamma_l(i) = \gamma_l$), the optimal arm selection is easy: just select μ_l (i.e., l-th best arm) on the l-th position. We define the arm selection strategy of BC-MP-TS as placing the arm of the l-th largest θ_i (i.e., $I_l(t) = \max^{(l)}_{i\in[K]} \theta_i$) on the l-th position.

Regret: naturally, the regret per round is defined as the difference between the expected reward of the optimal arm selection and that of an algorithm. Namely,

$$\mathrm{Reg}(T) = \sum_{t=1}^{T}\sum_{l=1}^{L}\Bigg(\prod_{l'=2}^{l}\gamma_{l'}(I_{\mathrm{opt}}(l'-1))\mu_{I_{\mathrm{opt}}(l)} \\ - \underbrace{\prod_{l'=2}^{l}\gamma_{l'}(I_{l'-1}(t))}_{\text{effective number of draw at position } l} \times \mu_{I_l(t)}\Bigg),$$

where $(I_{\mathrm{opt}}(1), \dots, I_{\mathrm{opt}}(L))$ is the optimal arm selection. In the case of the ad-independent discount factor, we conjecture that the regret lower bound should be identical to the case of no-discount factor that we analysed in the main paper (i.e., inequality (7)). Although we do not prove any regret bound for this cascade model, the conjecture is supported by the fact that (i) by identifying the top-L arm we immediately obtain the optimal arm selection, (ii) algorithms should require $\log T/d(\mu_i, \mu_L)$ number of effective draws to convince that suboptimal arm $i > L$ is not as good as arm L, and (iii) the best situation is that the simultaneous draw of several optimal arms rarely occurs: arm L is pushed out instead of arm i, and the regret increase per an effective draw is $\mu_L - \mu_i$. In the case of the general discount factor, the problem is subtler because a slight difference in $\{\mu_i\}$ can change the optimal arm selection.

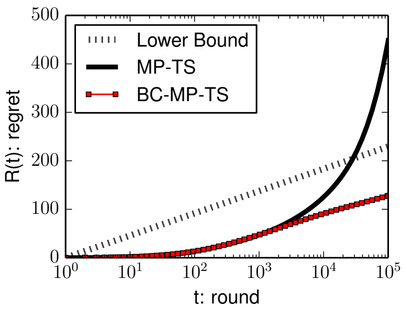

Figure 4. Simulation with a discount factor. Lower Bound is the leading $\Omega(\log T)$ term of the RHS of inequality (7), which we have conjectured to be the lower bound for the cascade model with the ad-independent discount factor in Section A.2.2. The regret is averaged over $10,000$ runs.

A.2.3. Experiment of cascade model

This simulation adapts the cascade model and involves a constant discount factor $\gamma_l(i) = 0.7$ for any position and arm. There are 9 Bernoulli arms with $\mu_1 = 0.24, \mu_2 = 0.21, \dots, \mu_9 = 0.00$ and $L = 3$. In this case the optimal arm selection strategy is to choose $\{I_1(t), I_2(t), I_3(t)\} = \{\mu_1, \mu_2, \mu_3\}$ (c.f., Section A.2.2). The regret of the algorithms is shown in 4. On one hand, MP-TS failed to have a small regret due to its ignorance to the discount factors. On the other hand, the slope of BC-MP-TS quickly approaches the conjectured Lower Bound, which is empirical evidence of the ability of BC-MP-TS to correct the position-dependent bias.

A.3. Key fact and lemmas

Fact 5. (Chernoff bound for binary random variables)

Let $X_1, \dots, X_n$ be i.i.d. binary random variables. Let $\hat{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\mu = \mathbb{E}[X_i]$. Then, for any $\epsilon \in (0, 1-\mu)$,

$$\Pr(\hat{X} \geq \mu + \epsilon) \leq \exp\left(-d(\mu+\epsilon, \mu)n\right).$$

and, for any $\epsilon \in (0, \mu)$,

$$\Pr(\hat{X} \leq \mu - \epsilon) \leq \exp\left(-d(\mu-\epsilon, \mu)n\right).$$

Fact 6. (Beta-Binomial equality) *Let $F^{\mathrm{beta}}_{\alpha,\beta}(y)$ be the cdf of the beta distribution with integer parameters α and β. Let $F^{\mathrm{B}}_{n,p}(\cdot)$ be the cdf of the binomial distribution with parameters n, p. Then,*

$$F^{\mathrm{beta}}_{\alpha,\beta}(y) = 1 - F^{\mathrm{B}}_{\alpha+\beta-1,y}(\alpha - 1),$$

Fact 7. (Pinsker's inequality for binary random variables) *For $p, q \in (0, 1)$, the KL divergence between two Bernoulli distributions is bounded as:*

$$d(p, q) \geq 2(p - q)^2.$$

Lemma 8. (Lemma 2 in Agrawal & Goyal (2013b)) *Let $k \in [K]$, $n \geq 0$ and $x < \mu_k$. Let $\hat{\mu}_{k,n}$ be the empirical average of n samples from* $\mathrm{Bernoulli}(\mu_k)$. *Let $p_{k,n}(x) = 1 - F^{\mathrm{beta}}_{\hat{\mu}_{k,n}n+1,(1-\hat{\mu}_{k,n})n+1}(y)$ be the probability that the posterior sample from the Beta distribution with its parameter $\hat{\mu}_{k,n}n+1, (1-\hat{\mu}_{k,n})n+1$ exceeds x. Then, its average over runs is bounded as:*

$$\mathbb{E}\left[\frac{1}{p_{k,n}(x)}\right] \leq \begin{cases} 1 + \frac{3}{\Delta_k(x)} & (n < 8/\Delta_k(x)) \\ 1 + \Theta\Bigg(e^{-\Delta_k(x)^2 n/2} + \frac{1}{(n+1)\Delta_k(x)^2} e^{-D_k(x)n} \\ \qquad\qquad + \frac{1}{e^{\Delta_k(x)^2 n/4} - 1}\Bigg) & (n \geq 8/\Delta_k(x)), \end{cases}$$

where $\Delta_k(x) = \mu_k - x$, $D_k(x) = d(x, \mu_k)$.

In the proof of Lemma 3 we use the following Lemmas 9, 10, and 11 several times. Lemma 9 is essentially the combination of the existing techniques of Agrawal & Goyal (2013b) and Honda & Takemura (2014). Lemmas 10 and 11 are also existing techniques that appear in several previous analyses in Bayesian bandits with Bernoulli arms.

Lemma 9. *Let $k \in [K]$, $z < \mu_k$ be arbitrary, $\mathcal{S}(t)$, $\mathcal{T}(t)$, and $\mathcal{U}(t)$ be events such that*

(i) if $\{\theta_k(t) \geq z\}$, $\mathcal{S}(t)$, and $\mathcal{T}(t)$ occurred then the arm k is drawn at round t,

(ii) $\theta_k(t)$, $\mathcal{S}(t)$ and $\mathcal{T}(t)$ are mutually independent given $\{\hat{\mu}_i(t)\}_{i=1}^K$ and $\{N_i(t)\}_{i=1}^K$.

(iii) The event $\mathcal{U}(t)$ is deterministic given $\{\hat{\mu}_i(t)\}_{i=1}^K$ and $\{N_i(t)\}_{i=1}^K$.

(iv) Given $\{\hat{\mu}_i(t)\}_{i=1}^K$ and $\{N_i(t)\}_{i=1}^K$ such that $\mathcal{U}(t)$ holds, $\mathcal{T}(t)$ occurs with probability at least $q > 0$.

Then

$$\mathbb{E}\left[\sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t)\}\right] = O\left(\frac{1}{q(\mu_k - z)^2}\right).$$

In particular, by setting $\mathcal{T}(t)$ and $\mathcal{U}(t)$ the trivial events that always hold ($q = 1$), we obtain the following inequality:

$$\mathbb{E}\left[\sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t)\}\right] = O\left(\frac{1}{(\mu_k - z)^2}\right). \quad (17)$$

Proof. First we have

$$\begin{aligned} &\sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t)\} \\ &\leq \sum_{n=0}^{T-1} \sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t), N_k(t) = n\} \\ &\leq \sum_{n=0}^{T-1} \sum_{m=1}^T \mathbf{1}\left[m \leq \sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t), N_k(t) = n\}\right]. \end{aligned} \quad (18)$$

Here note that the event

$$m \leq \sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t), N_k(t) = n\} \quad (19)$$

implies that the event

$$\{\mathcal{S}(t), \mathcal{U}(t), N_k(t) = n\} \quad (20)$$

occurred for at least m rounds and $\{\theta_k(t) < z\}$ or $\mathcal{T}^c(t)$ occurred for the first m rounds such that (20) occurred. Thus, by using the mutual independence of $\{\theta_k(t) < z\}$, $\mathcal{S}(t)$, and $\mathcal{T}(t)$, we have

$$\begin{aligned} \Pr\left[m \leq \sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t), N_k(t) = n\} \middle| \hat{\mu}_{k,n}\right] \\ \leq (1 - p_{k,n}(z)q)^m \end{aligned} \quad (21)$$

and therefore

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=1}^T \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t)\} \middle| \hat{\mu}_{k,n}\right] \\ &\leq \sum_{n=0}^{T-1} \sum_{m=1}^T (1 - p_{k,n}(z)q)^m \qquad \text{(by (18) and (21))} \\ &\leq \sum_{n=0}^{T-1} \frac{1 - p_{k,n}(z)q}{p_{k,n}(z)q} \leq \frac{1}{q} \sum_{n=0}^{T-1} \left(\frac{1}{p_{k,n}(z)} - 1\right), \end{aligned} \quad (22)$$

where we used $q \leq 1$ in the last transformation. By using Lemma 8, we obtain

$$\begin{aligned} &\mathbb{E}\left[\sum_{n=0}^{T-1} \left(\frac{1}{p_{k,n}(z)} - 1\right)\right] \\ &\leq \frac{24}{\Delta_k(z)^2} + \sum_{n=\lceil 8/\Delta_k(z) \rceil}^{T-1} O\Bigg(e^{-\Delta_k(z)^2 n/2} \\ &\qquad + \frac{e^{-D_k(z)n}}{(n+1)\Delta_k(z)^2} + \frac{1}{e^{\Delta_k(z)^2 n/4} - 1}\Bigg). \end{aligned} \quad (23)$$

By using the fact that $D_k(z) = d(z, \mu_k) = \Omega(1/(\mu_k - z)^2)$ (from the Pinsker's inequality), it is easy to verify that the RHS of (23) is $O(1/(\mu_k - z)^2)$. By using these facts, we finally obtain

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\theta_k(t) < z, \mathcal{S}(t), \mathcal{U}(t)\}\right] \le \frac{1}{q}\mathbb{E}\left[\sum_{n=0}^{T-1}\left(\frac{1}{p_{k,n}(z)} - 1\right)\right]$$
$$= O\left(\frac{1}{q(\mu_k - z)^2}\right),$$

which concludes the proof of the lemma. □

Lemma 10. (Deviation of empirical averages, Agrawal & Goyal (2013b, Appendix B.1)) *Let $k \in [K]$ and $z > \mu_k$ be arbitrary. Then,*

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \mathbf{1}\{\mathcal{A}_k(t), \hat{\mu}_k(t) > z\}\right] < 1 + \frac{1}{d(z, \mu_k)}.$$

Lemma 11. (Deviation of Beta posteriors) *Let $k \in [K]$, $x_1, x_2 \in [0, 1]$ be arbitrary values such that $x_1 > x_2$, and $n \geq 1$. Then,*

$$\Pr(\theta_k(t) \ge x_1 | \hat{\mu}_k(t) \le x_2, N_k(t) = n)$$
$$\le \exp\left(-d(x_2, x_1)n\right).$$

Proof. Note that, this lemma is essentially the same as the first display in Agrawal & Goyal (2013b, Appendix B.2). While Agrawal & Goyal (2013b) provide a bound for $N_k(t) > n$, the bound in our lemma is for $N_k(t) = n$. For the sake of rigor, we write the proof here.

$$\begin{aligned}
&\Pr(\theta_j(t) \ge x_1 | \hat{\mu}_j(t) \le x_2, N_j(t) = n) \\
&= \Pr\Bigg(\theta \sim \mathrm{Beta}(\hat{\mu}_j(t)n + 1, (1 - \hat{\mu}_j(t))n + 1), \\
&\qquad\qquad \theta \ge x_1 \Bigg| \hat{\mu}_j(t) \le x_2\Bigg) \\
&= 1 - F^{\mathrm{beta}}_{x_2 n+1,(1-x_2)n+1}(x_1) \\
&= F^{\mathrm{B}}_{n+1,x_1}(x_2 n) \\
&\qquad\qquad \text{(by the Beta-Binomial equality)} \\
&\le F^{\mathrm{B}}_{n,x_1}(x_2 n) \le \exp\left(-d(x_2, x_1)n\right) \\
&\qquad\qquad \text{(by the Chernoff bound).}
\end{aligned}$$

□

A.4. Proof of Lemma 3

Evaluation of term (A):

Proof. Here, we prove inequality (13). Recall that

$$(\mathrm{A}) = \sum_{t=1}^{T} \mathbf{1}\{\mathcal{B}^c(t)\} = \sum_{t=1}^{T} \mathbf{1}\{\theta^*(t) < \mu_L^{(-)}\}.$$

Since $\theta^*(t)$ is the L-th largest posterior sample among arms at round t, $\theta^*(t) < \mu_L^{(-)}$ implies that, there exists at least one arm in $[L]$ with its posterior sample smaller than $\mu_L^{(-)}$. Namely,

$$\{\theta^*(t) < \mu_L^{(-)}\} \subset \bigcup_{k \in [L]} \{\theta_k(t) < \mu_L^{(-)}\},$$

and therefore

$$\begin{aligned}
&\{\theta^*(t) < \mu_L^{(-)}\} \\
&= \bigcup_{k \in [L]} \{\theta_k(t) < \mu_L^{(-)}, \theta^*(t) < \mu_L^{(-)}\} \\
&= \bigcup_{k \in [L]} \{\theta_k(t) < \mu_L^{(-)}, \max_{j \in [L]}{}^{(L)} \theta_j(t) < \mu_L^{(-)}\} \\
&\subset \bigcup_{k \in [L]} \{\theta_k(t) < \mu_L^{(-)}, \max_{j \in [L] \setminus \{k\}}{}^{(L)} \theta_j(t) < \mu_L^{(-)}\}.
\end{aligned}$$

By using the union bound, we obtain

$$\begin{aligned}
&\mathbf{1}\{\theta^*(t) < \mu_L^{(-)}\} \\
&\le \sum_{k \in [L]} \mathbf{1}\{\theta_k(t) < \mu_L^{(-)}, \max_{j \in [L] \setminus \{k\}}{}^{(L)} \theta_j(t) < \mu_L^{(-)}\}.
\end{aligned}$$

Note that the event $\max^{(L)}_{j \in [L] \setminus \{k\}} \theta_j(t) < \mu_L^{(-)}$ satisfies the condition for the event $\mathcal{S}(t)$ in (17) in Lemma 9 with $z := \mu_L^{(-)}$. Therefore we obtain from Lemma 9 that

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\theta^*(t) < \mu_L^{(-)}\}\right] \\
&\quad = O\left(\frac{1}{(\mu_k - \mu_L^{(-)})^2}\right) = O\left(\frac{1}{(\mu_L - \mu_L^{(-)})^2}\right),
\end{aligned}$$

which concludes the proof of inequality (13). □

Evaluation of term (B):

Proof. Here, we prove inequality (14). We have,

$$\begin{aligned}
(\mathrm{B}) &= \sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{C}_i^c(t)\} \\
&= \sum_{t=1}^{T} \mathbf{1}\left\{\bigcup_{j \in [K] \setminus ([L-1] \cup \{i\})} \{\mathcal{A}_i(t), \theta^{**}_{\setminus i,j}(t) < \nu\}\right\} \\
&= \sum_{t=1}^{T} \sum_{j \in [K] \setminus ([L-1] \cup \{i\})} \mathbf{1}\left\{\mathcal{A}_i(t), \theta^{**}_{\setminus i,j}(t) < \nu\right\} \\
&= \sum_{t=1}^{T} \sum_{j \in [K] \setminus ([L-1] \cup \{i\})} \\
&\left\{\mathbf{1}\left\{\mathcal{A}_i(t), \hat{\mu}_i(t) > \mu_L\right\} + \mathbf{1}\left\{\mathcal{A}_i(t), \hat{\mu}_i(t) \le \mu_L, \theta^{**}_{\setminus i,j}(t) < \nu\right\}\right\}.
\end{aligned} \tag{24}$$

In the following, we bound the first and the second terms in the inner sum of the last line of (24). From Lemma 10, the first term of (24) is bounded as:

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbf{1}\left\{\mathcal{A}_i(t),\hat{\mu}_i(t)>\mu_L\right\}\right]\le 1+\frac{1}{d(\mu_L,\mu_i)}=O(1).$$

On the other hand, the second term of (24) is transformed as:

$$\begin{aligned}&\sum_{t=1}^{T}\mathbf{1}\left\{\mathcal{A}_i(t),\hat{\mu}_i(t)\le\mu_L,\theta^{**}_{\setminus i,j}(t)<\nu\right\}\\ &\le\frac{1}{d(\mu_L,\nu)}\\ &+\sum_{t=1}^{T}\mathbf{1}\left\{\mathcal{A}_i(t),N_i(t)>\frac{1}{d(\mu_L,\nu)},\hat{\mu}_i(t)\le\mu_L,\theta^{**}_{\setminus i,j}(t)<\nu\right\}\\ &\le\frac{1}{d(\mu_L,\nu)}\\ &\quad+\sum_{t=1}^{T}\mathbf{1}\left\{N_i(t)>\frac{1}{d(\mu_L,\nu)},\hat{\mu}_i(t)\le\mu_L,\theta^{**}_{\setminus i,j}(t)<\nu\right\}.\end{aligned}$$

Since $\theta^{**}_{\setminus i,j}(t)$ is the $(L-1)$-th largest posterior sample among arms except for arms i and j, $\theta^{**}_{\setminus i,j}(t)<\nu$ indicates that, the number of arms excluding i and j with posterior samples larger than or equal to ν is at most $L-2$, and thus at least one arm among $[L-1]$ has its posterior smaller than ν. Namely,

$$\begin{aligned}\{\theta^{**}_{\setminus i,j}(t)<\nu\}&=\{\max_{l\in[K]\setminus\{i,j\}}{}^{(L-1)}\theta_l(t)<\nu\}\\ &=\bigcup_{k\in[L-1]}\{\theta_k(t)<\nu,\max_{l\in[K]\setminus\{i,j\}}{}^{(L-1)}\theta_l(t)<\nu\}\\ &\subset\bigcup_{k\in[L-1]}\{\theta_k(t)<\nu,\max_{l\in[K]\setminus\{i,j,k\}}{}^{(L-1)}\theta_l(t)<\nu\}.\end{aligned}$$

By using this, we have

$$\begin{aligned}&\sum_{t=1}^{T}\mathbf{1}\left\{N_i(t)>\frac{1}{d(\mu_L,\nu)},\hat{\mu}_i(t)\le\mu_L,\theta^{**}_{\setminus i,j}(t)<\nu\right\}\\ &\le\sum_{t=1}^{T}\sum_{k\in[L-1]}\mathbf{1}\Big\{N_i(t)>\frac{1}{d(\mu_L,\nu)},\hat{\mu}_i(t)\le\mu_L,\\ &\qquad\qquad\theta_k(t)<\nu,\max_{l\in[K]\setminus\{i,j,k\}}{}^{(L-1)}\theta_l(t)<\nu\Big\}.\end{aligned}$$

Here, $z:=\nu$, $\mathcal{S}(t):=\{\max^{(L-1)}_{l\in[K]\setminus\{i,j,k\}}\theta_l(t)<\nu\}$, $\mathcal{T}(t):=\{\theta_i(t)\le\nu\}$, and $\mathcal{U}(t):=\{N_i(t)>1/d(\mu_L,\nu),\hat{\mu}_i(t)\le\mu_L\}$ satisfy the condition in Lemma 9. Under $\mathcal{U}(t)$, $\mathcal{T}(t)$ holds with probability at least

$$1-\exp\left(-d(\mu_L,\nu)(\frac{1}{d(\mu_L,\nu)})\right)=1-1/e$$

by Lemma 11. Therefore, by using Lemma 9 we obtain

$$\begin{aligned}&\mathbb{E}\Bigg[\sum_{t=1}^{T}\mathbf{1}\Big\{N_i(t)>\frac{1}{d(\mu_L,\nu)},\hat{\mu}_i(t)\le\mu_L,\\ &\qquad\theta_k(t)<\nu,\max_{l\in[K]\setminus\{i,j,k\}}{}^{(L-1)}\theta_l(t)<\nu\Big\}\Bigg]\\ &\qquad\le O\left(\frac{1}{(1-1/e)(\mu_k-\nu)^2}\right)=O(1).\end{aligned}\tag{25}$$

From (25) and the union bound over $k\in[L-1]$, the second term of (24) is $O(1)$. In summary, term (B) is $O(1)$ in expectation. □

Evaluation of term (C):

Proof. Here, we prove inequality (15). Recall that,

$$\text{(C)}=\sum_{j\in[K]\setminus([L-1]\cup\{i\})}\sum_{t=1}^{T}\mathbf{1}\{\mathcal{A}_i(t),\mathcal{A}_j(t),\mathcal{C}_i(t),\mathcal{D}_i(t)\}.$$

Let $\nu_2=(\nu+\mu_L)/2=(\mu_{L-1}+3\mu_L)/4$. Note that, we defined ν and ν_2 such that $\mu_{L-1}>\nu>\nu_2>\mu_L$, $O(\mu_{L-1}-\nu)=O(\nu-\nu_2)=O(\nu_2-\mu_L)=O(\mu_{L-1}-\mu_L)=O(1)$ as a function of T. Then,

$$\begin{aligned}&\sum_{t=1}^{T}\mathbf{1}\{\mathcal{A}_i(t),\mathcal{A}_j(t),\mathcal{C}_i(t),\mathcal{D}_i(t)\}\\ &=\sum_{t=1}^{T}\mathbf{1}\{\mathcal{A}_i(t),\mathcal{A}_j(t),\mathcal{C}_i(t),\mathcal{D}_i(t),\hat{\mu}_j(t)>\nu_2\}\\ &\quad+\sum_{t=1}^{T}\mathbf{1}\{\mathcal{A}_i(t),\mathcal{A}_j(t),\mathcal{C}_i(t),\mathcal{D}_i(t),\hat{\mu}_j(t)\le\nu_2\}\\ &\le\sum_{t=1}^{T}\mathbf{1}\{\mathcal{A}_j(t),\hat{\mu}_j(t)>\nu_2\}\\ &\quad+\sum_{t=1}^{T}\mathbf{1}\{\mathcal{A}_i(t),\mathcal{A}_j(t),\mathcal{C}_i(t),\mathcal{D}_i(t),\hat{\mu}_j(t)\le\nu_2\}.\end{aligned}\tag{26}$$

By using Lemma 10 with $z:=\nu_2$, the first term in (26) is bounded as:

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=1}^{T}\mathbf{1}\{\mathcal{A}_j(t),\hat{\mu}_j(t)>\nu_2\}\right]\le 1+\frac{1}{d(\nu_2,\mu_j)}\\ &=O\left(\frac{1}{(\nu_2-\mu_j)^2}\right)=O\left(\frac{1}{(\mu_{L-1}-\mu_L)^2}\right)=O(1).\end{aligned}\tag{27}$$

We now bound the second term in (26). Let $\mathcal{C}'_{i,j}(t) = \{\theta^{**}_{\setminus i,j}(t) \geq \nu\} \supset \mathcal{C}_i(t)$. Let $\mathcal{E}_j(t) = \{N_j(t) \geq \epsilon_2 \log T\}$. We have,

$$
\begin{aligned}
&\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_i(t), \mathcal{D}_i(t), \hat{\mu}_j(t) \leq \nu_2\} \\
&\leq \sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}'_{i,j}(t), \mathcal{D}_i(t), \hat{\mu}_j(t) \leq \nu_2\} \\
&\leq \epsilon_2 \log T \\
&\quad + \sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}'_{i,j}(t), \mathcal{D}_i(t), \hat{\mu}_j(t) \leq \nu_2, \mathcal{E}_j(t)\}. \\
&\leq \epsilon_2 \log T + \sum_{n=0}^{N_i^{\mathrm{suf}}(T)-1} \sum_{t=1}^{T} \\
&\qquad \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}'_{i,j}(t), N_i(t) = n, \hat{\mu}_j(t) \leq \nu_2, \mathcal{E}_j(t)\}.
\end{aligned}
$$

In the following, we bound

$$
\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}'_{i,j}(t), N_i(t) = n, \hat{\mu}_j(t) \leq \nu_2, \mathcal{E}_j(t)\}. \tag{28}
$$

Note that, (28) is at most 1 since $\{\mathcal{A}_i(t), N_i(t) = n\}$ occurs at most once. Let τ be the first round (if exists) at which $\{\mathcal{C}'_{i,j}(t), \theta^{**}_{\setminus i,j}(t) \leq \theta_i(t), \mathcal{A}_i(t), N_i(t) = n\}$ is satisfied. It is necessary that $\{\theta_j(\tau) \geq \theta^{**}_{\setminus i,j}(\tau)\}$ for (28) to be 1: this is because, (i) both $\theta_i(\tau)$ and $\theta_j(\tau)$ need to be larger than $\theta^{**}_{\setminus i,j}(\tau)$ for the simultaneous draw of arms i and j, (ii) and if $\theta_j(\tau) < \theta^{**}_{\setminus i,j}(\tau)$ then arm i is drawn and thus $\{N_i(t) = n\}$ is never satisfied after $t > \tau$. Here,

$$
\begin{aligned}
\Pr\{\theta_j(\tau) \geq \theta^{**}_{\setminus i,j}(\tau), \theta^{**}_{\setminus i,j}(\tau) \geq \nu, \hat{\mu}_j(\tau) \leq \nu_2\} \\
\leq \exp\left(-d(\nu_2, \nu) N_j(\tau)\right),
\end{aligned}
$$

by Lemma 11. Therefore, we have

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_i(t), N_i(t) = n, \hat{\mu}_j(t) \leq \nu_2\}\right] \\
\leq \exp\left(-d(\nu_2, \nu)\epsilon_2 \log T\right) = T^{-\epsilon_2 d(\nu_2, \nu)}.
\end{aligned} \tag{29}
$$

In summary, the second term in (26) is bounded as:

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_i(t), \mathcal{D}_i(t), \hat{\mu}_j(t) \leq \nu_2\}\right] \\
&\leq \epsilon_2 \log T + N_i^{\mathrm{suf}}(T) T^{-\epsilon_2 d(\nu_2, \nu)} \\
&\leq \left(\epsilon_2 + \frac{4T^{-\epsilon_2 d(\nu_2, \nu)}}{d(\mu_i, \mu_L)}\right) \log T \qquad \text{(by } (1+\delta)^2 < 4\text{)},
\end{aligned}
$$

and thus,

$$
\begin{aligned}
&\mathbb{E}[\text{(C)}] \\
&\leq \sum_{j \in [K] \setminus ([L-1] \cup \{i\})} \left(\frac{\left(\epsilon_2 + 4T^{-\epsilon_2 d(\nu_2, \nu)}\right) \log T}{d(\mu_i, \mu_L)}\right) + O(1) \\
&\leq \sum_{j \in [K] \setminus ([L-1] \cup \{i\})} \left(\frac{\left(\epsilon_2 + 4T^{-\epsilon_2 \Delta^2_{L,L-1}/8}\right) \log T}{d(\mu_i, \mu_L)}\right) + O(1),
\end{aligned}
$$

where we used the fact that $d(\nu_2, \nu) \geq 2(\nu - \nu_2)^2 = 2 \times ((\mu_{L-1} - \mu_L)/4)^2$ in the last transformation. □

Evaluation of term (D):

Proof. Here, we prove inequality (16). We first divide term (D) into two subterms as:

$$
\begin{aligned}
&\mathbb{E}[\text{(D)}] = \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), N_i(t) \geq N_i^{\mathrm{suf}}(T)\}\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \hat{\mu}_i(t) > \mu_i^{(+)}, N_i(t) \geq N_i^{\mathrm{suf}}(T)\}\right] \\
&+ \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \hat{\mu}_i(t) \leq \mu_i^{(+)}, N_i(t) \geq N_i^{\mathrm{suf}}(T)\}\right].
\end{aligned} \tag{30}
$$

On one hand, the first term in (30) is bounded as:

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \hat{\mu}_i(t) > \mu_i^{(+)}, N_i(t) \geq N_i^{\mathrm{suf}}(T)\}\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \hat{\mu}_i(t) > \mu_i^{(+)}\}\right] \\
&\leq 1 + \frac{1}{d(\mu_i^{(+)}, \mu_i)} \qquad \text{(by Lemma 10).}
\end{aligned} \tag{31}
$$

On the other hand, each component of the second term of