Online Bilateral Trade With Minimal Feedback: Don't Waste Seller's Time

Francesco Bacchiocchi
Politecnico di Milano
francesco.bacchiocchi@polimi.it

Matteo Castiglioni
Politecnico di Milano
matteo.castiglioni@polimi.it

Roberto Colomboni
Politecnico di Milano &
Università degli Studi di Milano
roberto.colomboni@polimi.it

Alberto Marchesi
Politecnico di Milano
alberto.marchesi@polimi.it

Abstract

Online learning algorithms for designing optimal bilateral trade mechanisms have recently received significant attention. This paper addresses a key inefficiency in prior two-bit feedback models, which synchronously query both the buyer and the seller for their willingness to trade. This approach is inherently inefficient as it offers a trade to the seller even when the buyer rejects the offer. We propose an *asynchronous* mechanism that queries the seller only if the buyer has already accepted the offer. Consequently, the mechanism receives one bit of feedback from the buyer and a "censored" bit from the seller—a signal richer than the standard one-bit (trade/no-trade) feedback, but less informative than the two-bit model. Assuming independent valuations with bounded densities—the same distributional conditions underlying the two-bit results of Cesa-Bianchi et al. [2024a]—we design an algorithm that achieves $\tilde{O}(T^{2/3})$ regret against the best fixed price in hindsight. This matches the lower bound for the strictly richer two-bit model, showing that our mechanism elicits the minimal feedback necessary to attain optimal rates.

1 Introduction

In a bilateral trade problem, a broker faces two rational agents—a *seller* and a *buyer*—who wish to trade an object. Each agent has their own private valuation for the object and seeks to maximize their utility. The goal of the broker is to design a mechanism that intermediates between the seller and the buyer, in order to make a trade happen. Ideally, a mechanism for bilateral trade should be *efficient*, *i.e.*, it should maximize the sum of agents' utilities, while also ensuring incentive compatibility and individual rationality. A well-known mechanism that achieves this objective is the VCG mechanism [Vickrey, 1961]. Unfortunately, the VCG mechanism fails to meet *budget balance*, requiring the broker to subsidize the market and incur in financial losses. Indeed, a general impossibility result by Myerson and Satterthwaite [1983] shows that full efficiency cannot be attained while simultaneously maintaining incentive compatibility, individual rationality, and budget balance.

A recent line of research (see, *e.g.*, [Cesa-Bianchi et al., 2021, 2024a, Azar et al., 2022]) circumvents the impossibility result by Myerson and Satterthwaite [1983] through the lens of online learning. This is done by addressing repeated bilateral trade problems, where the broker faces a sequence of sellers and buyers willing to trade objects, over a time horizon T. At each time t, a new seller and a new buyer arrive, each of them with their own private valuation of the object, say S_t and B_t, respectively. The broker then proposes a trading price P_t to both agents. Thus, the seller is willing to trade if $S_t \leq P_t$, while the buyer if $B_t \geq P_t$. The trade happens only if both agents accept the price P_t, with

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

the buyer paying the trading price to the seller and receiving the object, resulting in *strong budget balance* (*i.e.*, the broker neither subsidize nor extract revenue from the market). In such an online learning framework, the full-efficiency requirement is relaxed by comparing the performance of the broker over the T time steps against the *best fixed price in hindsight*. Specifically, the performance is evaluated in terms of the *gain from trade*, which intuitively encodes the net gain in agents' utilities, defined as $(B_t - S_t)\mathbb{I}\{S_t \leq P_t \leq B_t\}$ at each time t.

Previous works have focused on three models that differ in the kind of feedback that the broker receives at the end of each time step t. In the *full-feedback* model, the broker observes the valuations S_t and B_t of the agents. In the *two-bit* model, the broker separately observes whether each of the two agents is willing to trade or not, namely they observe both $\mathbb{I}\{S_t \leq P_t\}$ and $\mathbb{I}\{P_t \leq B_t\}$. Finally, in the *one-bit* model, they only observe whether the trade has occurred or not, namely $\mathbb{I}\{S_t \leq P_t \leq B_t\}$.

While simple, the one-bit model is insufficient for learning optimal strongly-budget-balanced mechanisms [Cesa-Bianchi et al., 2024a]. This motivates the study of richer feedback models, such as the two-bit one, where learning becomes possible under the assumption of independent seller/buyer valuations with bounded densities [Cesa-Bianchi et al., 2024a]. However, such mechanisms rely on synchronous interaction protocols that are inherently inefficient: they query the seller even when the buyer has already rejected the trade. In this paper, we address the following two natural questions:

What is the minimal feedback required to learn optimal mechanisms?
Is it possible to query the seller only when a trade opportunity arises?

From a more application-oriented perspective, these questions raise from the insight that it would generally be more efficient—and more reasonable—to implement *asynchronous* interaction protocols, in which the seller is approached only if the buyer has already agreed to trade at the proposed price. This approach is especially relevant in many practical applications—such as, *e.g.*, online freelance marketplaces (Upwork, Fiverr), ride-sharing platforms (Uber, Lyft, Grab), rental intermediaries platforms (AirBnB)—where sellers are often involved in multiple simultaneous trading scenarios involving different objects. In such settings, requiring the sellers to make a decision each time a potential buyer for any of their objects appears would impose an excessive burden on them. Furthermore, in many cases, sellers prefer to disclose as little information as possible to the broker in order to protect their reputation.

Our Results We study—for the first time, to the best of our knowledge—*online learning in bilateral trade problems with asynchronous interaction protocols*. The key challenge in such a setting is that the broker receives a particular *asymmetric* feedback that is richer than one-bit feedback, but way less informative than two-bit feedback. Specifically, the broker receives one bit of feedback from the buyer, by observing $\mathbb{I}\{B_t \geq P_t\}$ at every time t, while they only observe a "censored" bit of feedback from the seller, as they get to know $\mathbb{I}\{S_t \leq P_t\}$ only when the buyer accepts the trade. The main result of the paper is a strongly-budget-balanced algorithm that attains $\tilde{O}(T^{2/3})$ regret against the best fixed price in hindsight, assuming independent sellers and buyers' valuations with bounded densities. Notice that both assumptions are required, since removing even one of the two assumptions makes the problem not learnable even with two-bit feedback [Cesa-Bianchi et al., 2024a]. We remark that our result matches the regret rate that Cesa-Bianchi et al. [2024a] obtained using the more informative two-bit feedback, under the *same distributional assumptions*. Moreover, it matches the lower bound by Cesa-Bianchi et al. [2024a] for the richer two-bit feedback model, showing that:

Asynchronous protocols are not only efficient, but they also allow the broker to elicit the minimal feedback necessary to attain optimal regret rates!

1.1 Challenges and Techniques

Our algorithm builds upon the idea of *scouting bandits*, which have been originally introduced by Cesa-Bianchi et al. [2024a] for two-bit models. This idea exploits a suitable decomposition of the *expected* gain from trade $\mathrm{g}(p)$ for a fixed price $p \in [0, 1]$, which is defined as follows:

$$\mathrm{g}(p) = \underbrace{\mathbb{P}[S \leq p]}_{(a)} \underbrace{\int_p^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda}_{(b)} + \underbrace{\mathbb{P}[B \geq p]}_{(c)} \underbrace{\int_0^p \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda}_{(d)},$$

where S and B are generic random variables representing the valuations of the seller and the buyer, respectively. Cesa-Bianchi et al. [2024a] leverage the two-bit feedback by first conducting a uniform exploration phase to estimate the integral terms (b) and (d). Using these estimates, they construct a proxy for the expected reward function in which (b) and (d) are replaced with their empirical counterparts. Then, they run a *bandit* algorithm with this proxy, since the two-bit feedback provides, at each time step, unbiased estimates of terms (a) and (c), thereby allowing the learner to reconstruct the proxy reward function.

Our *asymmetric feedback model introduces a new significant challenge*: the feedback received from the seller is "censored", as the learner gets to know the value of $\mathbb{I}\{S_t \leq P_t\}$ only when the buyer accepts the trade, *i.e.*, when $\mathbb{I}\{P_t \leq B_t\} = 1$. This "censored" seller's feedback makes the scouting bandits approach by Cesa-Bianchi et al. [2024a] unsuitable for our setting: the estimator they build for the integral term (d) cannot be recovered due to missing observations, and the bandit feedback needed to estimate term (a) may be "censored" and thus unavailable. Furthermore, our feedback breaks the symmetry in estimating the four components in the decomposition of the expected gain from trade provided by the two-bit feedback. Indeed, in our model, estimating (a) is harder than estimating (c), and estimating (d) is harder than estimating (b). Consequently, *the main challenge faced in this paper is how to effectively estimate (a) and (d) under "censored" seller's feedback*. Our key technical contribution is addressing this challenge.[1]

In order to address the challenge, we provide a lower bound on the number of time steps in which the value of $\mathbb{I}\{S_t \leq p\}$ (*i.e.*, the seller's feedback for p) is observed, for every price p. In particular, if price p is posted for H times (and $\mathbb{P}[B \geq p]$ is large enough), we can lower bound the number of times that seller's feedback is observed as $\Omega\big(H \cdot \mathbb{P}[B \geq p]\big)$, by using Chernoff's concentration bound. Notably, the usual additive concentration bounds are of no help in this setting. Indeed, our random variables are Bernoulli that might have small mean, making additive bounds non-meaningful. Moreover, the lower bound on the number of times the seller's feedback is observed could still be small. As a consequence, we may lack sufficient samples to build precise confidence bounds around terms (a) and (d), thus precluding the direct application of standard UCB-like techniques. However, we observe that when the confidence intervals on seller-related quantities are large, the buyer's probability of accepting the trade is low, resulting in two effects that counterbalance each other. More formally, the error on term (d) is scaled by the buyer's acceptance probability—namely, term (c). A symmetric argument holds for terms (a) and (b).

1.2 Related Works

Our work contributes to the line of research initiated by Cesa-Bianchi et al. [2024a], which studies bilateral trade through the lens of online learning. Among other results, Cesa-Bianchi et al. [2024a] show that strongly-budget-balanced mechanisms are learnable with two-bit feedback when the seller/buyer distributions are independent and admit bounded density, while the same problem is unlearnable under one-bit model.

Subsequent work focuses mainly on adversarial settings. Azar et al. [2022] design algorithms that guarantee no-2-regret. Cesa-Bianchi et al. [2024b] provide sublinear regret guarantees assuming a smoothed adversary. Bernasconi et al. [2024], Chen et al. [2025] remove the smoothness assumption by relaxing the budget balance constraint to hold globally. Lunghi et al. [2026] go even further by analyzing which regret rates are attainable by relaxing the global budget constraint (*i.e.*, by allowing for its violation). Other works study extensions of bilateral trade to multiple buyers [Babaioff et al., 2024, Lunghi et al., 2025], different objectives [Bachoc et al., 2024], contextual settings [Gaucher et al., 2025], divisible items [Bolić et al., 2025], and situations where traders have no predetermined seller and buyer's roles [Bolić et al., 2024, Cesari and Colomboni, 2025, Bachoc et al., 2025a,b].

There is also a rich literature on bilateral trade without learning, focusing on providing approximations of an optimal mechanism [Colini-Baldeschi et al., 2016, 2017, Blumrosen and Mizrahi, 2016, Brustle et al., 2017, Colini-Baldeschi et al., 2020, Babaioff et al., 2020, Dütting et al., 2021, Deng et al., 2022, Kang and Vondrák, 2019, Archbold et al., 2023].

[1] We remark that, from a technical point of view, the roles of the seller and the buyer in our framework are completely interchangeable without requiring additional effort. Indeed, the case in which we query the buyer only when the seller agrees to trade at a given price can be tackled with the same approach presented here.

2 Preliminaries

In this paper, we study online learning in repeated bilateral trade problems. In this section, we introduce the notation and all the definitions needed in the rest of the paper.

2.1 Bilateral Trade

The learner (a broker) repeatedly interacts with the environment. At each time step $t \in [T]$, a new seller and a new buyer arrive with (random) valuation S_t and B_t in $[0,1]$.[2]

The learner offers a (random) price $P_t \in [0,1]$ to the buyer and the seller. A trade happens if and only if the buyer and the seller accept the proposed price, *i.e.*, when $S_t \leq P_t \leq B_t$. This ensures *strong budget balance* is satisfied during learning. The learner's performance is evaluated through the net increase in market value (*i.e.*, the net increase in agents' utilities), also known as *gain from trade*. Specifically, if we define the gain from trade function as

$$\mathrm{gft}\colon [0,1]^3 \to [0,1], \qquad (p,s,b) \mapsto (\underbrace{b-p}_{\substack{\text{buyer's}\\ \text{net gain}}} + \underbrace{p-s}_{\substack{\text{sellers's}\\ \text{net gain}}}) \underbrace{\mathbb{I}\{s \leq p \leq b\}}_{\text{a trade happens}} = (b-s)\mathbb{I}\{s \leq p \leq b\}$$

and the gain from trade (random) function at time t as

$$\mathrm{GFT}_t\colon [0,1] \to [0,1], p \mapsto \mathrm{gft}(p, S_t, B_t),$$

the gain from trade rewarded to the learner by posting P_t is the random variable $\mathrm{GFT}_t(P_t)$.

We assume that the sequence of sellers' valuations $S_1, S_2, \dots$ and the sequence of buyers' valuations $B_1, B_2, \dots$ are i.i.d. sequences, independent of each other. Moreover, for ease of presentation, we introduce two additional random variables S and B, which are distributed as S_t and B_t, respectively. We assume that S and B are independent of each other and of the two sequences $S_1, S_2, \dots$ and $B_1, B_2, \dots$, and that they admit L-Lipschitz continuous cumulative distribution functions (cdf)[3].

For notational convenience, we also define the random function

$$\mathrm{GFT} : [0,1] \to [0,1], \qquad p \mapsto \mathrm{gft}(p, S, B)\ ,$$

and the expected gain from trade function as

$$\mathrm{g} : [0,1] \to [0,1], \qquad p \mapsto \mathbb{E}[\mathrm{GFT}(p)]\ .$$

We notice that [Cesa-Bianchi et al., 2024a, Lemma 2] ensures that g is upper semicontinuous and, consequently, being defined on the compact set $[0,1]$, it admits a maximum. From this point on, we fix a point $p^\star \in [0,1]$ where this maximum is attained.

2.2 Do Not Waste the Seller's Time: An Asynchronous Protocol

It is well known (see [Cesa-Bianchi et al., 2024a]) that learning is impossible with one-bit feedback, *i.e.*, when the learner only observes the outcome of the trade $\mathbb{I}\{S_t \leq P_t \leq B_t\}$ after each time step t. For this reason, previous works focus on two-bit feedback, where the leaner can separately observe seller and buyer's willingness to trade, namely $\mathbb{I}\{S_t \leq P_t\}$ and $\mathbb{I}\{P_t \leq B_t\}$. In this paper, we show that the regret attainable with two-bit feedback can also be obtained with a weaker feedback, under the same assumptions. In particular, we consider an *asynchronous* protocol that first proposes a trading price to the buyer, and, then, it offers the same price to the seller only if the buyer has already expressed their willingness to trade at the proposed price. This protocol introduces a new *asymmetric* feedback model in which the leaner receives one bit of feedback from the buyer, while it gets a "censored" bit from the seller, as their willingness to trade is observed only when the buyer accepts to the trade at the proposed price.

The asynchronous interaction protocol between the learner (a broker) and the environment (the sequence of sellers and buyers) is formally presented in Online Protocol 1.

[2]Let $n \in \mathbb{N}$, we denote with $[n] = \{1, \dots, n\}$ the set of the first n natural numbers.

[3]The cdf Lipschitz assumption is equivalent to the bounded density one of Cesa-Bianchi et al. [2024a]

Online Protocol 1 Asynchronous Repeated Bilateral Trade

for time step $t = 1, 2 \dots$ **do**
  The learner chooses $P_t \in [0, 1]$
  The learner observes $\mathbb{I}\{P_t \leq B_t\}$
  If $P_t \leq B_t$, then the learner observes $\mathbb{I}\{S_t \leq P_t\}$
  The learner gains (but does *not* observe) $\mathrm{GFT}_t(P_t)$

Regret Given a time horizon $T \in \mathbb{N}$, the goal of the learner is to minimize the regret with respect to the gain from trade, defined as

$$R_T = \sum_{t=1}^{T} \Big(\mathrm{g}(p^\star) - \mathbb{E}\big[\mathrm{g}(P_t)\big] \Big),$$

where $p^\star \in \arg\max_{p \in [0,1]} \mathrm{g}(p)$.

3 Algorithm

In this section, we present an online learning algorithm that achieves optimal regret guarantees under the asynchronous interaction protocol introduced in Section 2. Our learning algorithm is restricted to using a finite grid of prices. In Section 4, we show that this results in a small loss that depends on the granularity of the grid and the Lipschitz constant L. Specifically, we denote by $\mathcal{P}_K \subseteq [0, 1]$ the uniform grid of prices $p_k \coloneqq {}^{k-1}\!/_K$ for $k \in [K]$, where K is a suitable parameter. More formally, we let $\mathcal{P}_K \coloneqq \{p_k\}_{k \in [K]}$.

3.1 Additional Notation

Before moving to the description of our algorithm, we need to introduce some additional notation that is useful to deal with the stochastic feedback observed by posting prices on the grid. For each $k \in [K]$ and each $j \in \mathbb{N}$, we denote with $t_B(k, j)$ the j-th time step in which the broker sets price $p_k \in \mathcal{P}_K$ (if this time step exists, otherwise, we set it to $+\infty$), and with $t_S(k, j)$ the j-th time the broker sets price $p_k \in \mathcal{P}_K$ and the buyer accepts to buy (it this time step exists, otherwise we set it to $+\infty$). For technical reasons (*i.e.*, having well-defined random variables for every $k \in [K]$ and $j \in \mathbb{N}$), we also assume given another i.i.d. family of pairs of random variables $(S'_{k,j}, B'_{k,j})_{k \in [K], j \in \mathbb{N}}$, independent of $(S_1, B_1), (S_2, B_2), \dots$ such that, for any $k \in [K]$ and $j \in \mathbb{N}$, the pair $(S'_{k,j}, B'_{k,j})$ shares the same distribution as (S, B). For each $k \in [K]$ and $j \in \mathbb{N}$, we then set

$$B_{k,j} \coloneqq \begin{cases} B_{t_B(k,j)} & \text{if } t_B(k,j) < +\infty \,, \\ B'_{k,j} & \text{otherwise} \,, \end{cases} \qquad S_{k,j} \coloneqq \begin{cases} S_{t_S(k,j)} & \text{if } t_S(k,j) < +\infty \,, \\ S'_{k,j} & \text{otherwise} \,. \end{cases}$$

In this way, the family $(S_{k,j}, B_{k,j})_{k \in [K], j \in \mathbb{N}}$ is a well-defined independent family of pairs of random variables such that, for every $k \in [K]$, the sequence $(S_{k,j}, B_{k,j})_{j \in \mathbb{N}}$ is i.i.d., and, additionally, for every $j \in \mathbb{N}$, the pair $(S_{k,j}, B_{k,j})$ shares the same distribution as (S, B).

3.2 Algorithm Description

We are now ready to introduce our algorithm achieving optimal regret guarantees under the asynchronous interacton protocol (Algorithm 2). The algorithm takes as input a time horizon T, and it builds the uniform grid $\mathcal{P}_K \subseteq [0, 1]$ with $K \coloneqq \lceil T^{1/3} \rceil$ points. At a very high level, our algorithm performs an initial exploration phase in which the broker plays each of the prices in $\mathcal{P}_K$ a number of times equal to $H \coloneqq \lceil T^{1/3} \rceil$. This exploration phase is necessary to compute an optimistic estimate of the expected gain from trade, as defined in eq. (1) in the following. Once this exploration phase is concluded, in the remaining time steps, Algorithm 2 executes a K-armed bandit-style algorithm by selecting, at each time, the price $p_k \in \mathcal{P}_K$ that maximizes a suitable optimistic estimate of the expected gain from trade. Notice that several additional components discussed in the following are needed to deal with the limited feedback.

Algorithm 2 More than one-bit less than two-bit bilateral trade

Input: time horizon T
Set $\delta \leftarrow \frac{1}{T^2}, H \leftarrow K \leftarrow \lceil T^{1/3} \rceil, \mathcal{T}_{k,0} \leftarrow 0, \mathcal{Q}_{k,0} \leftarrow 0, \mathcal{S}_{k,0} \leftarrow 0 \ \ \forall k \in [K]$
Set $\mathcal{P}_K \leftarrow \{p_k\}_{k\in[K]}$ with $p_k \leftarrow {}^{k-1}/_{K} \ \forall k \in [K]$, $K^\diamond \leftarrow \emptyset$
for $t = 1, 2, \ldots, T$ **do**
  if $t \leq HK$ **then** ▷ Exploration Phase
    Set $l \leftarrow \lceil t/H \rceil$ and post price $P_t \leftarrow p_l$
  else ▷ Bandit Phase
    Select $l \in \operatorname{argmax}_{k\in K^\diamond} \mathit{UCB}_{k,t-1}$ (see eq. (3)) and post price $P_t \leftarrow p_l$
  for $k = 1, 2 \ldots K$ **do** ▷ Update Counters
    if $t = HK$ **then**
      $N_k \leftarrow \min_{i\leq k} \mathcal{Q}_{i,KH}$
      $K^\diamond \leftarrow \{k \in [K] \mid \mathcal{Q}_{k,KH} \geq 32 \log(KT^2/\delta)\}$
      $\hat{F}_k \leftarrow \frac{1}{KH} \sum_{i=k}^{K-1} \sum_{j=1}^{H} \mathbb{I}\{B_{i,j} \geq p_i\}, \ \forall k \in K^\diamond$
      $\hat{G}_k \leftarrow \frac{1}{KN_k} \sum_{i=1}^{k} \sum_{j=1}^{N_k} \mathbb{I}\{S_{i,j} \leq p_i\}, \ \forall k \in K^\diamond$
    Set $\mathcal{T}_{k,t} \leftarrow \mathcal{T}_{k,t-1} + \mathbb{I}\{P_t = p_k\}$
    Set $\mathcal{Q}_{k,t} \leftarrow \mathcal{Q}_{k,t-1} + \mathbb{I}\{P_t = p_k\}\mathbb{I}\{P_t \leq B_t\}$
    Set $\mathcal{S}_{k,t} \leftarrow \mathcal{S}_{k,t-1} + \mathbb{I}\{P_t = p_k\}\mathbb{I}\{S_t \leq P_t\}\mathbb{I}\{P_t \leq B_t\}$
    Set $\hat{\nu}_{k,t} \leftarrow \frac{\mathcal{Q}_{k,t}}{\mathcal{T}_{k,t}}$ and $\hat{\mu}_{k,t} \leftarrow \frac{\mathcal{S}_{k,t}}{\mathcal{Q}_{k,t}}$

We now provide a more detailed description of how the algorithm works. Specifically, in the first KH time steps, Algorithm 2 prescribes the broker to play each price on the grid $\mathcal{P}_K$ exactly H times—the first H time steps on p_1, the next H on p_2, and so forth—so that this initial phase has length KH. Furthermore, independently of the round $t \in [T]$, Algorithm 2 maintains counters that track how many times each price p_k has been selected. Precisely, at each round $t \in [T]$: $\mathcal{T}_{k,t}$ counts the number of times the broker has proposed price p_k up to round $t \in [T]$; $\mathcal{Q}_{k,t}$ counts the number of times the buyer has accepted to buy at price p_k; and $\mathcal{S}_{k,t}$ counts the number of times both the buyer and the seller have agreed to trade at price p_k.

Once the exploration phase is completed—*i.e.*, after KH rounds—Algorithm 2 identifies a subset of arms $K^\diamond \subseteq [K]$ such that, for each $k \in K^\diamond$, the number of samples collected from the seller's distribution when the broker has proposed price p_k is larger than a suitable constant. After that, the algorithm ignores all the arms *not* in $K^\diamond$, while for each arm in $K^\diamond$ it designs an upper confidence bound on the gain from trade (refined at each round), by exploiting the decomposition:

$$\mathrm{g}(p) = \underbrace{\mathbb{P}[S \leq p]}_{(a)} \underbrace{\int_p^1 \mathbb{P}[B \geq \lambda] \, \mathrm{d}\lambda}_{(b)} + \underbrace{\mathbb{P}[B \geq p]}_{(c)} \underbrace{\int_0^p \mathbb{P}[S \leq \lambda] \, \mathrm{d}\lambda}_{(d)}, \tag{1}$$

which is formally derived in [Cesa-Bianchi et al., 2024a, Lemma 1].

Furthermore, order to build the upper confidence bounds to be used in the subsequent bandit-style procedure, for each $k \in K^\diamond$, after the exploration phase Algorithm 2 computes two estimates of the integral terms (b) and (d) in eq. (1), which are defined as follows:

$$\hat{F}_k = \frac{1}{KH} \sum_{i=k}^{K-1} \sum_{j=1}^{H} \mathbb{I}\{B_{i,j} \geq p_i\}, \quad \hat{G}_k = \frac{1}{KN_k} \sum_{i=1}^{k} \sum_{j=1}^{N_k} \mathbb{I}\{S_{i,j} \leq p_i\}, \tag{2}$$

where

$$N_k := \min_{i\leq k} \mathcal{Q}_{i,KH}.$$

Let us remark that the definition of the term N_k is necessary to ensure that, in the sum in $\hat{G}_k$, we have the same number of observations of the seller's valuation for each price p_i with $i \leq k$. Indeed, while for every p_k we always observe whether the buyer accepts to buy or not, we do not have analogous

information for the seller, since the observation of seller's feedback is conditioned on the buyer's decision to buy at the given price.

In the remaining rounds $t > KH$, Algorithm 2 performs a UCB-style procedure using $K^\diamond$ as the set of arms. This refinement of the arm set is necessary, as we can guarantee useful concentration properties only for the arms in $K^\diamond$. Specifically, at each round $t > KH$, Algorithm 2 computes optimistic upper confidence bounds $UCB_{k,t}$ on the value of each $\mathrm{g}(p_k)$ for prices p_k such that $k \in K^\diamond$, formally defined as follows:

$$UCB_{k,t} \coloneqq \left(\hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right)\left(\hat{F}_k + \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K}\right) + \left(\hat{\nu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}}\right)\left(\hat{G}_k + \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K}\right). \quad (3)$$

One of the main challenges in our algorithm analysis will be to show that the upper confidence bounds concentrate on the true mean. This is particularly challenging since $\mathcal{Q}_{k,t}$ might be small.

4 Regret Analysis

In this section, we prove the regret guarantees attained by Algorithm 2. In the following, for the ease of presentation and to avoid repetitions, we assume that a time horizon T is given as a parameter, and we set every occurrence of δ to $\frac{1}{T^2}$, and every occurrence of K and H to $\lceil T^{1/3} \rceil$, as in Algorithm 2.

Our algorithm can be easily extended to an anytime version that does not need knowledge of T by using the standard doubling trick [Cesa-Bianchi and Lugosi, 2006].

4.1 Restrict the Set of Candidate Prices

In the following, we will show that our algorithm has no regret with respect to the best price on the grid $\mathcal{P}_K$. This is sufficient since the best price in $\mathcal{P}_K$ performs almost as well as the best price in the interval $[0, 1]$, as we formally show in the following lemma.

Lemma 1. *It holds that:*

$$\max_{p \in \mathcal{P}_K} \mathrm{g}(p) \geq \mathrm{g}(p^\star) - {}^{L}\!/_{K} \,.$$

Proof. Let $k^\star \in [K]$ be such that $p_{k^\star}$ minimizes the distance from $p^\star$ among all the points in $\mathcal{P}_K$. Noticing that a random variable is $1/L$-smooth if an only if its cdf is L-Lipschitz continuous, [Cesa-Bianchi et al., 2024b, Lemma 1] ensures that g is L-Lipschitz continuous, and hence

$$\mathrm{g}(p^\star) - \max_{p \in \mathcal{P}_K} \mathrm{g}(p) \leq \mathrm{g}(p^\star) - \mathrm{g}(p_{k^\star}) \leq L \cdot |p^\star - p_{k^\star}| \leq {}^{L}\!/_{K} \,. \qquad \square$$

4.2 Define a Clean Event

We now introduce some definitions to aid the presentation. First, we define a set $\mathcal{K}$ (unknown to the learner) that intuitively identifies the prices $p_k \in \mathcal{P}_K$ in which feedback about the seller can be observed with sufficiently high probability, under the asynchronous interaction protocol.

Definition 1. *We define $\mathcal{K}$ as the subset of $k \in [K]$ such that:*

$$\mathbb{P}[B \geq p_k] \geq 8T^{-1/3}\log(KT^2/\delta).$$

We will show that for prices in $\mathcal{K}$ our estimates are sufficiently accurate, while prices not in $\mathcal{K}$ can be ignored. Indeed, this is because they achieve a negligible expected gain from trade.

Moreover, we also introduce the following definition of *clean event* $\mathcal{E}$. Intuitively, this is decomposed into four different events, namely $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, and $\mathcal{E}_4$. These events are related to different high-probability concentration bounds. The event $\mathcal{E}_1$ is defined as follows:

$$\mathcal{E}_1 \coloneqq \bigcap_{t=HK}^{T} \bigcap_{k \in \mathcal{K}} \left\{ \mathcal{Q}_{k,t} > \frac{\mathcal{T}_{k,t}}{2}\mathbb{P}[B \geq p_k] \right\}.$$

Intuitively, under this event, the broker receives feedback about the seller at price p_k at least half of the times the price p_k is proposed, weighted by the probability that the buyer accepts it. It can be shown that this event holds with high probability by the aid of Chernoff's inequality. This event guarantees that we observe a constant fraction of the "expected" samples.

The event $\mathcal{E}_2$ is related to the estimates $\hat{G}_k$ and $\hat{F}_k$ of the integrals appearing in the definition of gain from trade. In particular, it requires that the estimates lie in their respective confidence intervals. Formally, we have:

$$\mathcal{E}_2 \coloneqq \bigcap_{k \in \mathcal{K}} \left\{ \left| \hat{G}_k - \int_0^{p_k} \mathbb{P}[S \leq \lambda] \, \mathrm{d}\lambda \right| \leq \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K} \right\} \cap \bigcap_{k \in [K]} \left\{ \left| \hat{F}_k - \int_{p_k}^1 \mathbb{P}[B \geq \lambda] \, \mathrm{d}\lambda \right| \leq \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K} \right\}.$$

The underlying idea is to show that $\hat{G}_k$ is close to its expectation up to $\tilde{\mathcal{O}}\left(1/KN_k\right)$ with high probability due to a combination of Chernoff and Azuma-Hoeffding's inequality. Moreover, the expectation $\mathbb{E}[\hat{G}_k]$ is close to the integral term (d) in eq. (1), by leveraging the rectangle rule to compute integrals and the fact that the integrand function is monotone to obtain telescopic simplifications. An analogous argument holds for the terms $\hat{F}_k$ as well.

Furthermore, $\mathcal{E}_3$ is the event in which the estimates $\hat{\mu}_{k,t}$ and $\hat{\nu}_{k,t}$ of the probabilities of the seller and the buyer accepting the trade, respectively, lie in confidence intervals that shrink as the inverse of the square root of the number of observed samples.

$$\mathcal{E}_3 \coloneqq \bigcap_{t=HK}^{T} \bigcap_{k \in [K]} \left\{ \hat{\mu}_{k,t} - \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}} \leq \mathbb{P}\left[S \leq p_k\right] \leq \hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}} \right\} \cap \bigcap_{t=HK}^{T} \bigcap_{k \in [K]} \left\{ \hat{\nu}_{k,t} - \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}} \leq \mathbb{P}\left[B \geq p_k\right] \leq \hat{\nu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}} \right\}$$

That $\mathcal{E}_3$ holds with high probability is a consequence of Hoeffding's inequality and a union bound, once we realize that, for each $k \in [K]$, the sequence $(S_{k,j}, B_{k,j})_{j \in \mathbb{N}}$ is i.i.d..

Finally, the event $\mathcal{E}_4$ guarantees that if a price p_k has too low a probability of being accepted by the buyer, *i.e.*, $k \notin \mathcal{K}$, then such a price will not belong to the set $K^\diamond$ of candidate optimal prices used by the algorithm in the second phase. Formally, we have:

$$\mathcal{E}_4 \coloneqq \bigcap_{k \notin \mathcal{K}} \left\{ \mathcal{Q}_{k,KH} \leq 32 \log\left(\frac{KT^2}{\delta}\right) \right\}.$$

Intuitively, also the probability of event $\mathcal{E}_4$ can be bounded by using Chernoff inequality. This event is important to avoid running the second phase on prices p_k that have too low a probability of being accepted by the buyer. This would result in a small number of samples and huge confidence intervals. Moreover, such prices can be safely discarded as their expected gain from trade is small.

Finally, by bounding each event separately and applying a union bound, we obtain the following lemma, which establishes the probability with which the clean event $\mathcal{E}$ holds.

Lemma 2. *Let $\mathcal{E} \coloneqq \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4$. Then, we have:*

$$\mathbb{P}\left[\mathcal{E}\right] \geq 1 - \mathcal{O}(KT\delta).$$

We defer the formal proof of this lemma to the Appendix.

4.3 Bound the Regret

Next, we introduce two crucial lemmas that enable us to derive the regret guarantees of Algorithm 2. Intuitively, the first lemma shows that, under the clean event $\mathcal{E}$, the first term of the upper confidence

bound employed by Algorithm 2 concentrates towards the first term in the gain from trade decomposition, *i.e.*, the product of (a) and (b), at the desired rate. Moreover, it shows that our confidence bounds are always optimistic, an essential requirement for UCB-like algorithms. The second lemma shows an analogous result, but for the second terms.

Lemma 3. *Let $k \in \mathcal{K}$. Then, for each $t \geq HK$, conditional on the event $\mathcal{E}$, we have:*

$$0 \leq \left(\hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right)\left(\hat{F}_k + \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K}\right) - \mathbb{P}[S \leq p_k]\int_{p_k}^{1}\mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda \leq \eta,$$

where $\eta \coloneqq C\log\left(\frac{T}{\delta}\right)\left(\frac{1}{T^{1/3}} + \frac{1}{\sqrt{\mathcal{T}_{k,t}}}\right)$ and $C > 0$ is an absolute constant.

The above lemma shows that the difference between the first component of the confidence bound $UCB_{k,t}$ (defined in eq. (3)) and the first component of the expected gain from trade is proportional to $1/\sqrt{\mathcal{T}_{k,t}}$. At a first glance, this is not what we would expect since the empirical mean $\hat{\mu}_{k,t}$ is estimated using $\mathcal{Q}_{k,t}$ samples. Thus, the confidence bound of $\hat{\mu}_{k,t}$ is proportional to $1/\sqrt{\mathcal{Q}_{k,t}}$. However, such a term is multiplied by (an upper bound of) the probability that the buyer accepts the offer, *i.e.*, $\mathbb{P}[B \geq p_k]$. Thus, if the confidence term is large (*i.e.*, when $\mathcal{Q}_{k,t}$ is small), then the probability that the buyer accepts the trade is low. These two effects compensate for each other, and the resulting contribution ends up scaling as $1/\sqrt{\mathcal{T}_{k,t}}$. We defer the formal proof of this lemma to the Appendix.

Lemma 4. *Let $k \in \mathcal{K}$. Then, for each $t \geq HK$, conditional on the event $\mathcal{E}$, we have:*

$$0 \leq \left(\hat{\nu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}}\right)\left(\hat{G}_k + \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K}\right) - \mathbb{P}[B \geq p_k]\int_{0}^{p_k}\mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda \leq \eta,$$

where $\eta \coloneqq C\log\left(\frac{T}{\delta}\right)\left(\frac{1}{T^{1/3}} + \frac{1}{\sqrt{\mathcal{T}_{k,t}}}\right)$ and $C > 0$ is an absolute constant.

Similarly to Lemma 3, the estimate $\hat{G}_k$ is constructed using a number of samples that depends on the probability that the buyer accepts to trade at price p_k. This component is *mirrored* with respect to the one in Lemma 3: rather than having limited samples for the probability estimate $\hat{\nu}_{k,t}$, here we may have few samples to compute $\hat{G}_k$, the term that corresponds to the integral estimates. Nonetheless, a similar argument shows that this does not affect the rate at which the confidence bound shrinks, since the confidence interval around $\hat{G}_k$ is scaled by an upper bound on the buyer's acceptance probability—effectively compensating for the lower sample size. We defer the formal proof of this lemma to the Appendix.

We are now ready to present our main result.

Theorem 1. *Algorithm 2 guarantees regret $R_T = \tilde{\mathcal{O}}(T^{2/3})$.*

We now present the high-level ideas behind the proof, deferring the formal argument to the Appendix.

To derive regret guarantees for our algorithm, we analyze separately the two phases it executes. The exploration phase always incurs a regret of order $\mathcal{O}(T^{2/3})$, as the exploration strategy is deterministic and runs for $\mathcal{O}(T^{2/3})$ rounds. To provide an upper bound on the regret suffered in the second phase of our algorithm, we consider two cases depending on whether $k^\star$ belongs to $K^\diamond$ or not. If $k^\star \notin K^\diamond$, then we can show that $\mathbb{P}[B \geq p_{k^\star}] = \tilde{\mathcal{O}}(T^{-1/3})$ and, consequently, $\mathrm{g}(p_{k^\star}) = \tilde{\mathcal{O}}(T^{-1/3})$. Therefore, for any possible set $K^\diamond$, the regret suffered by our algorithm is $R_T = \tilde{\mathcal{O}}(T^{2/3})$.

Conversely, if $k^\star \in K^\diamond$, *i.e.*, if $p_{k^\star}$ is not removed before the second phase, then, noticing that $K^\diamond \simeq \mathcal{K}$, Lemmas 3 and 4 guarantee that (i) the confidence bounds of each suboptimal arm scale as $1/\min(T^{1/3}, \sqrt{\mathcal{T}_{k,t}})$, and (ii) all the estimates remain optimistic. By applying a UCB-style analysis and using Lemma 2, we obtain that the cumulative regret in this phase is also of order $\tilde{\mathcal{O}}(T^{2/3})$.

5 Conclusions, Limitations, and Future Research

This work introduced a new asynchronous protocol for repeated bilateral trade, where the broker interacts with a seller only after securing agreement from a buyer. Despite the censored nature

of the seller's feedback, we showed that the broker can still achieve the optimal $\tilde{\mathcal{O}}(T^{2/3})$ regret rate previously known only under the richer two-bit feedback model [Cesa-Bianchi et al., 2024a]. Combined with the known impossibility of learning under one-bit feedback [Cesa-Bianchi et al., 2024a], this work suggests that our protocol elicits the minimal amount of information necessary to enable optimal learning.

While our theoretical guarantees are optimal, some limitations suggest interesting directions for future research. First, our analysis focuses on stationary environments. Although adversarial bilateral trade is known to be unlearnable even under full feedback and when competing with the best fixed price in hindsight [Cesa-Bianchi et al., 2024a], it would be valuable to explore intermediate settings lying in between the i.i.d. and fully adversarial dynamics, perhaps by relaxing the notion of learnability to allow for (dynamic) α-regret. Another limitation of our work is the absence of contextual information that the broker might observe before posting a price. Extending the framework to contextual settings—where the broker might have access to side information encoding item characteristics, market conditions, or user profiles—remains a challenging and interesting open problem in our censored framework. Finally, following recent works [Bachoc et al., 2024, Cesari and Colomboni, 2025], an intriguing direction for future research is to study alternative objectives, such as *fair gain from trade* or *trading volume*, which reflect different priorities for the broker.

Acknowledgments and Disclosure of Funding

FB, MC, RC, and AM are partially supported by the FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence) and by the EU Horizon project ELIAS (European Lighthouse of AI for Sustainability, No. 101120237). RC is partially supported by the MIUR PRIN grant 2022EKNE5K (Learning in Markets and Society). AM is partially supported by the Italian MIUR PRIN 2022 project "Targeted Learning Dynamics: Computing Efficient and Fair Equilibria through No-Regret Algorithms".

References

Thomas Archbold, Bart de Keijzer, and Carmine Ventre. Non-obvious manipulability for single-parameter agents and bilateral trade. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pages 2107–2115, 2023.

Yossi Azar, Amos Fiat, and Federico Fusco. An α-regret analysis of adversarial bilateral trade. *Advances in Neural Information Processing Systems*, 35:1685–1697, 2022.

Moshe Babaioff, Kira Goldner, and Yannai A Gonczarowski. Bulow-klemperer-style results for welfare maximization in two-sided markets. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2452–2471. SIAM, 2020.

Moshe Babaioff, Amitai Frey, and Noam Nisan. Learning to maximize gains from trade in small markets. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 195–195, 2024.

François Bachoc, Nicolò Cesa-Bianchi, Tom Cesari, and Roberto Colomboni. Fair online bilateral trade. *Advances in Neural Information Processing Systems*, 37:37241–37263, 2024.

François Bachoc, Tommaso Cesari, and Roberto Colomboni. A contextual online learning theory of brokerage. In *Forty-second International Conference on Machine Learning*, 2025a.

François Bachoc, Tommaso Cesari, and Roberto Colomboni. A tight regret analysis of non-parametric repeated contextual brokerage. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025b.

Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. No-regret learning in bilateral trade via global budget balance. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 247–258, 2024.

Liad Blumrosen and Yehonatan Mizrahi. Approximating gains-from-trade in bilateral trading. In *Web and Internet Economics: 12th International Conference, WINE 2016, Montreal, Canada, December 11-14, 2016, Proceedings 12*, pages 400–413. Springer, 2016.

Natasa Bolić, Tommaso Cesari, and Roberto Colomboni. An online learning theory of brokerage. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, page 216–224, 2024. ISBN 9798400704864.

Natasa Bolić, Tommaso Cesari, Roberto Colomboni, and Christian Paravalos. Online learning in the repeated mediated newsvendor problem. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Johannes Brustle, Yang Cai, Fa Wu, and Mingfei Zhao. Approximating gains from trade in two-sided markets via simple mechanisms. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 589–590, 2017.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, USA, 2006.

Nicolò Cesa-Bianchi, Tommaso R Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. A regret analysis of bilateral trade. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 289–309, 2021.

Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Bilateral trade: A regret minimization perspective. *Mathematics of Operations Research*, 49(1): 171–203, 2024a.

Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Regret analysis of bilateral trade with a smoothed adversary. *Journal of Machine Learning Research*, 25(234):1–36, 2024b.

Tommaso Cesari and Roberto Colomboni. An online learning theory of trading-volume maximization. In *The Thirteenth International Conference on Learning Representations*, 2025.

Houshuang Chen, Yaonan Jin, Pinyan Lu, and Chihao Zhang. Tight regret bounds for fixed-price bilateral trade. *arXiv preprint arXiv:2504.04349*, 2025.

Riccardo Colini-Baldeschi, Bart de Keijzer, Stefano Leonardi, and Stefano Turchetta. Approximately efficient double auctions with strong budget balance. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 1424–1443. SIAM, 2016.

Riccardo Colini-Baldeschi, Paul Goldberg, Bart de Keijzer, Stefano Leonardi, and Stefano Turchetta. Fixed price approximability of the optimal gain from trade. In *Web and Internet Economics: 13th International Conference, WINE 2017, Bangalore, India, December 17–20, 2017, Proceedings 13*, pages 146–160. Springer, 2017.

Riccardo Colini-Baldeschi, Paul W Goldberg, Bart de Keijzer, Stefano Leonardi, Tim Roughgarden, and Stefano Turchetta. Approximately efficient two-sided combinatorial auctions. *ACM Transactions on Economics and Computation (TEAC)*, 8(1):1–29, 2020.

Yuan Deng, Jieming Mao, Balasubramanian Sivan, and Kangning Wang. Approximately efficient bilateral trade. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 718–721, 2022.

Paul Dütting, Federico Fusco, Philip Lazos, Stefano Leonardi, and Rebecca Reiffenhäuser. Efficient two-sided markets with limited information. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1452–1465, 2021.

Solenne Gaucher, Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Vianney Perchet. Feature-based online bilateral trade. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL `https://openreview.net/forum?id=xnF2U0ro7b`.

Zi Yang Kang and Jan Vondrák. Fixed-price approximations to optimal efficiency in bilateral trade. *Available at SSRN 3460336*, 2019.

Anna Lunghi, Matteo Castiglioni, and Alberto Marchesi. Online two-sided markets: Many buyers enhance learning. *arXiv preprint arXiv:2503.01529*, 2025.

Anna Lunghi, Matteo Castiglioni, and Alberto Marchesi. Better regret rates in bilateral trade via sublinear budget violation. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms*, 2026.

Roger B Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of economic theory*, 29(2):265–281, 1983.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

Omitted Proofs

In order to prove Lemma 2, we need to introduce some auxiliary useful lemmas.

Lemma 5. *It holds* $\mathbb{P}[\mathcal{E}_1] \geq 1 - \delta$.

Proof. Let $s \geq H$ be an integer, and let $k \in \mathcal{K}$. By the Chernoff bound and the fact that the variables are all i.i.d., we have:

$$\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^{s} \mathbb{I}\{B_{k,i} \geq p_k\} \leq \frac{s}{2}\mathbb{P}[B \geq p_k]\right] &\leq e^{-\frac{s\mathbb{P}[B \geq p_k]}{8}} \\
&= e^{-\frac{s\mathbb{P}[B \geq p_k]\log\left(KT^2/\delta\right)}{8\log\left(KT^2/\delta\right)}} \\
&= \left(\frac{\delta}{KT^2}\right)^{\frac{s\mathbb{P}[B \geq p_k]}{8\log\left(KT^2/\delta\right)}} \\
&\leq \left(\frac{\delta}{KT^2}\right)^{\frac{T^{1/3}\mathbb{P}[B \geq p_k]}{8\log\left(KT^2/\delta\right)}} \leq \frac{\delta}{KT^2}.
\end{aligned}$$

If $t \geq HK$, then $\mathcal{T}_{k,t} \geq H$ according to the definition of $\mathcal{T}_{k,t}$ and how our algorithm works. It follows that

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{t=HK}^{T}\bigcup_{k\in\mathcal{K}}\left\{\mathcal{Q}_{k,t} \leq \frac{\mathcal{T}_{k,t}}{2}\mathbb{P}[B \geq p_k]\right\}\right] \\
&\qquad = \mathbb{P}\left[\bigcup_{s=H}^{T}\bigcup_{t=HK}^{T}\bigcup_{k\in\mathcal{K}}\left\{\mathcal{Q}_{k,t} \leq \frac{\mathcal{T}_{k,t}}{2}\mathbb{P}[B \geq p_k]\right\} \cap \{\mathcal{T}_{k,t} = s\}\right] \\
&\qquad \leq \sum_{s=H}^{T}\sum_{t=HK}^{T}\sum_{k\in\mathcal{K}}\mathbb{P}\left[\left\{\mathcal{Q}_{k,t} \leq \frac{\mathcal{T}_{k,t}}{2}\mathbb{P}[B \geq p_k]\right\} \cap \{\mathcal{T}_{k,t} = s\}\right] \\
&\qquad = \sum_{s=H}^{T}\sum_{t=HK}^{T}\sum_{k\in\mathcal{K}}\mathbb{P}\left[\left\{\sum_{i=1}^{s}\mathbb{I}\{B_{k,i} \geq p_k\} \leq \frac{s}{2}\mathbb{P}[B \geq p_k]\right\} \cap \{\mathcal{T}_{k,t} = s\}\right] \\
&\qquad \leq \sum_{s=H}^{T}\sum_{t=HK}^{T}\sum_{k\in\mathcal{K}}\mathbb{P}\left[\sum_{i=1}^{s}\mathbb{I}\{B_{k,i} \geq p_k\} \leq \frac{s}{2}\mathbb{P}[B \geq p_k]\right] \\
&\qquad \leq \delta\ .
\end{aligned}$$

Hence,

$$\mathbb{P}\left[\bigcap_{t=HK}^{T}\bigcap_{k\in\mathcal{K}}\left\{\mathcal{Q}_{k,t} > \frac{\mathcal{T}_{k,t}}{2}\mathbb{P}[B \geq p_k]\right\}\right] \geq 1 - \delta\ ,$$

which concludes the proof. □

Lemma 6. *It holds*

$$\mathbb{P}\left[\bigcap_{k\in\mathcal{K}}\left\{\left|\hat{G}_k - \int_0^{p_k}\mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda\right| \leq \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K}\right\}\right] \geq 1 - 2K\delta.$$

Proof. Under the event $\mathcal{E}_1$, for all $i \in [K]$ such that $\mathbb{P}[B \geq p_i] \geq 8T^{-1/3}\log(KT^2/\delta)$, we have:

$$\mathcal{Q}_{i,HK} \geq \frac{1}{2}\mathcal{T}_{i,HK}\mathbb{P}[B \geq p_i] = \frac{1}{2}H\mathbb{P}[B \geq p_i],$$

since $\mathcal{T}_{i,HK} = H$ for every $i \in [K]$, according how our algorithm works. This implies that, under the event $\mathcal{E}_1$, the following holds:

$$N_k = \min_{i\leq k}\mathcal{Q}_{i,HK} \geq \frac{1}{2}\min_{i\leq k}H\mathbb{P}[B \geq p_i] = \frac{1}{2}H\mathbb{P}[B \geq p_k] := n_k.$$

As a first step, we prove that, conditioned to the event $\{N_k = \ell\} \cap \mathcal{E}_1$, the estimates $\hat{G}_k$ concentrate around their expectation. Specifically, whenever $\ell \geq n_k$, the following holds:

$$\mathbb{P}\left[\left|\hat{G}_k - \mathbb{E}[\hat{G}_k]\right| \leq 2\sqrt{\frac{\log(2/\delta)}{K\ell}} \,\middle|\, \{N_k = \ell\} \cap \mathcal{E}_1\right] \geq 1 - \delta. \tag{4}$$

Indeed, by Azuma–Hoeffding inequality, we have:

$$\mathbb{P}\left[\left|\sum_{i=1}^{k}\sum_{j=1}^{\ell} \mathbb{I}\{S_{i,j} \leq p_i\} - \sum_{i=1}^{k}\sum_{j=1}^{\ell} \mathbb{P}[S \leq p_i]\right| \leq \sqrt{2K\ell\log(2/\delta)}\right] \geq 1 - \delta\,.$$

Furthermore, noticing that

$$\left|\hat{G}_k - \mathbb{E}[\hat{G}_k]\right| = \frac{1}{K\ell}\left|\sum_{i=1}^{k}\sum_{j=1}^{\ell} \mathbb{I}\{S_{i,j} \leq p_i\} - \sum_{i=1}^{k}\sum_{j=1}^{\ell} \mathbb{P}[S \leq p_i]\right|\,,$$

and that $\{N_k = \ell\} \cap \mathcal{E}_1$ is $\mathbb{P}$-independent from $S_1, S_2, \ldots$, we have:

$$\begin{aligned}\mathbb{P}\left[\left|\hat{G}_k - \mathbb{E}[\hat{G}_k]\right| \leq \sqrt{\frac{2\log(2/\delta)}{K\ell}} \,\middle|\, \{N_k = \ell\} \cap \mathcal{E}_1\right] &= \mathbb{P}\left[\left|\hat{G}_k - \mathbb{E}[\hat{G}_k]\right| \leq \sqrt{\frac{2\log(2/\delta)}{K\ell}}\right] \\ &\geq 1 - \delta\,,\end{aligned}$$

showing that Equation 4 holds. Therefore, we can prove that:

$$\begin{aligned}&\mathbb{P}\left[\left|\hat{G}_k - \mathbb{E}[\hat{G}_k]\right| \leq \sqrt{\frac{2\log(2/\delta)}{KN_k}} \,\middle|\, \mathcal{E}_1\right] \\ &\geq \sum_{\substack{\ell=0,\\ \ell\geq n_k}}^{H} \mathbb{P}\left[\left|\hat{G}_k - \mathbb{E}[\hat{G}_k]\right| \leq \sqrt{\frac{2\log(2/\delta)}{K\ell}} \,\middle|\, \{N_k = \ell\} \cap \mathcal{E}_1\right] \mathbb{P}\left[N_k = \ell \mid \mathcal{E}_1\right] \\ &\geq \sum_{\substack{\ell=0,\\ \ell\geq n_k}}^{H} (1-\delta)\mathbb{P}\left[N_k = \ell \mid \mathcal{E}_1\right] \\ &= 1 - \delta,\end{aligned}$$

where the first inequality holds because of the law of total probability, noticing that $\mathbb{P}[N_k = \ell \mid \mathcal{E}_1] = 0$ for all $\ell < n_k$, while the second inequality by eq. (4). Thanks to the i.i.d. hypothesis we have:

$$\begin{aligned}\mathbb{E}[\hat{G}_k] - \int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda &= \mathbb{E}\left[\frac{1}{KN_k}\sum_{i=1}^{k}\sum_{j=1}^{N_k} \mathbb{I}\{S_{i,j} \leq p_i\}\right] - \int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda \\ &= \frac{1}{K}\sum_{i=1}^{k} \mathbb{P}[S \leq p_i] - \int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda \\ &= \sum_{i=1}^{k} \int_{\frac{i-1}{K}}^{\frac{i}{K}} \left(\mathbb{P}\left[S \leq \frac{i}{K}\right] - \mathbb{P}[S \leq \lambda]\right)\,\mathrm{d}\lambda =: (\star)\,,\end{aligned}$$

Now, due to the fact that $\lambda \mapsto \mathbb{P}[S \leq \lambda]$ is a non-decreasing function, we have that

$$\begin{aligned}0 \leq (\star) &\leq \sum_{i=1}^{k} \int_{\frac{i-1}{K}}^{\frac{i}{K}} \left(\mathbb{P}\left[S \leq \frac{i}{K}\right] - \mathbb{P}\left[S \leq \frac{i-1}{K}\right]\right)\,\mathrm{d}\lambda \\ &= \frac{1}{K}\sum_{i=1}^{k}\left(\mathbb{P}\left[S \leq \frac{i}{K}\right] - \mathbb{P}\left[S \leq \frac{i-1}{K}\right]\right) = \frac{\mathbb{P}[S \leq p_k] - \mathbb{P}[S \leq 0]}{K} \leq \frac{1}{K}.\end{aligned}$$

Thus, the following holds:

$$\mathbb{P}\left[\left|\hat{G}_k - \int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda\right| \leq \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K} \mid \mathcal{E}_1\right] \geq 1-\delta.$$

Thus, thanks to Lemma 5, we have:

$$\begin{aligned}
&\mathbb{P}\left[\left|\hat{G}_k - \int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda\right| \leq \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K}\right] \\
&\geq \mathbb{P}\left[\left|\hat{G}_k - \int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda\right| \leq \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K} \mid \mathcal{E}_1\right]\mathbb{P}[\mathcal{E}_1] \\
&\geq 1-2\delta.
\end{aligned}$$

Finally, taking a union bound over all possible sets of arms $\mathcal{K}$, we prove the lemma. □

Lemma 7. *It holds*

$$\mathbb{P}\left[\bigcap_{k\in\mathcal{K}}\left\{\left|\hat{F}_k - \int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda\right| \leq \sqrt{\frac{2\log(2/\delta)}{KH}} + \frac{1}{K}\right\}\right] \geq 1-K\delta.$$

Proof. Thanks to the i.i.d. hypothesis and the fact that $\lambda \mapsto \mathbb{P}[B \geq \lambda]$ is a non-increasing function, with an argument analogous to that provided in the proof of Lemma 6, we have that

$$\left|\mathbb{E}[\hat{F}_k] - \int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda\right| \leq \frac{1}{K}.$$

By Azuma–Hoeffding inequality, we have:

$$\left|\sum_{i=1}^{k}\sum_{j=1}^{H}\mathbb{I}\{B_{i,j} \geq p_i\} - \sum_{i=1}^{k}\sum_{j=1}^{H}\mathbb{P}\left[B \geq p_i\right]\right| \leq \sqrt{2kH\log(2/\delta)} \leq \sqrt{2KH\log(2/\delta)}$$

with probability $1-\delta$. Hence, to conclude, it is enough to notice that

$$\left|\hat{F}_k - \mathbb{E}[\hat{F}_k]\right| = \frac{1}{KH}\left|\sum_{i=1}^{k}\sum_{j=1}^{H}\mathbb{I}\{B_{i,j} \geq p_i\} - \sum_{i=1}^{k}\sum_{j=1}^{H}\mathbb{P}\left[B \geq p_i\right]\right|$$

and take a union bound over all possible sets of arms $[K]$. □

Lemma 8. *It holds* $\mathbb{P}[\mathcal{E}_3] \geq 1-2KT\delta$.

Proof. Fix $k \in \mathcal{K}$ and $t \geq HK$. Recall that, by definition,

$$\hat{\mu}_{k,t} \coloneqq \frac{\sum_{\tau=1}^{\mathcal{Q}_{k,t}}\mathbb{I}\{S_{k,\tau} \leq p_k\}}{\mathcal{Q}_{k,t}}.$$

Employing a union bound and the Hoeffding inequality, we have that:

$$\hat{\mu}_{k,t} - \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}} \leq \mathbb{P}\left[S \leq p_k\right] \leq \hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}$$

with probability at least $1-\delta$. Thus, taking a union bound, we have:

$$\mathbb{P}\left[\bigcap_{k\in[K]}\bigcap_{t=HK}^{T}\left\{\hat{\mu}_{k,t} - \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}} \leq \mathbb{P}\left[S \leq p_k\right] \leq \hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right\}\right] \geq 1-KT\delta.$$

With an analogous argument, it is possible to show that:

$$\mathbb{P}\left[\bigcap_{k\in[K]}\bigcap_{t=HK}^{T}\left\{\hat{\nu}_{k,t} - \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}} \leq \mathbb{P}\left[B \geq p_k\right] \leq \hat{\nu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}}\right\}\right] \geq 1-KT\delta.$$

Thus, taking a union bound, we have that the lemma holds. □

Lemma 9. *It holds* $\mathbb{P}[\mathcal{E}_4] \geq 1 - \delta$

Proof. Let $\epsilon = 8T^{-1/3}\log(KT^2/\delta)$ and $\hat{\nu}_k \coloneqq \hat{\nu}_{k,KH}$. If $k \notin \mathcal{K}$, then $\mathbb{P}[B \geq p_k] \leq \epsilon$. Therefore, we can employ the multiplicative Chernoff inequality as follows:

$$\mathbb{P}\big[\hat{\nu}_k \geq (1+c)\mathbb{P}[B \geq p_k]\big] \leq e^{-\frac{c^2 H \mathbb{P}[B \geq p_k]}{2+c}},$$

with $c = \frac{\epsilon}{\mathbb{P}[B \geq p_k]}$. Thus, we get:

$$\begin{aligned}
\mathbb{P}\big[\hat{\nu}_k \geq \mathbb{P}[B \geq p_k] + \epsilon\big] &\leq \exp\left(-\frac{\left(\frac{\epsilon}{\mathbb{P}[B \geq p_k]}\right)^2 H\mathbb{P}[B \geq p_k]}{2 + \frac{\epsilon}{\mathbb{P}[B \geq p_k]}}\right) \\
&\leq \exp\left(-\frac{\frac{\epsilon}{\mathbb{P}[B \geq p_k]}}{2 + \frac{\epsilon}{\mathbb{P}[B \geq p_k]}} \cdot \epsilon H\right) \\
&\leq \left(\frac{\delta}{KT^2}\right)^{8/3} \leq \frac{\delta}{KT^2},
\end{aligned}$$

since $x/(x+2) \geq 1/3$, for every $x \geq 1$. As a result, we have:

$$\mathbb{P}\big[\mathcal{Q}_{k,HK} \leq 2H\epsilon\big] = \mathbb{P}\big[\hat{\nu}_k \leq 2\epsilon\big] \geq \mathbb{P}\big[\hat{\nu}_k \leq \mathbb{P}[B \geq p_k] + \epsilon\big] \geq 1 - \frac{\delta}{KT^2},$$

recalling that $\hat{\nu}_{k,HK} = {}^{\mathcal{Q}_{k,HK}}/_{H}$. Furthermore, we notice that:

$$2H\epsilon \leq 16HT^{-1/3}\log({}^{KT^2}/_{\delta}) = 16\frac{\lceil T^{1/3}\rceil}{T^{1/3}}\log({}^{KT^2}/_{\delta}) \leq 32\log({}^{KT^2}/_{\delta}).$$

Thus, by taking a union bound, we have:

$$\mathbb{P}\left[\bigcap_{k \notin \mathcal{K}} \left\{\mathcal{Q}_{k,HK} \leq 32\log({}^{KT^2}/_{\delta})\right\}\right] \geq 1 - \frac{\delta}{T^2} \geq 1 - \delta,$$

concluding the proof. □

Lemma 2. *Let* $\mathcal{E} \coloneqq \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4$. *Then, we have:*

$$\mathbb{P}\left[\mathcal{E}\right] \geq 1 - \mathcal{O}(KT\delta).$$

Proof. Thanks to Lemmas 5, 6, 7, 8, and 9, by taking a union bound we have:

$$\mathbb{P}\left[\mathcal{E}\right] \geq 1 - (5KT + 2)\delta = 1 - \mathcal{O}(KT\delta),$$

concluding the proof. □

Lemma 3. *Let* $k \in \mathcal{K}$. *Then, for each* $t \geq HK$, *conditional on the event* $\mathcal{E}$, *we have:*

$$0 \leq \left(\hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right)\left(\hat{F}_k + \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K}\right) - \mathbb{P}[S \leq p_k]\int_{p_k}^{1}\mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda \leq \eta,$$

where $\eta \coloneqq C\log\left(\frac{T}{\delta}\right)\left(\frac{1}{T^{1/3}} + \frac{1}{\sqrt{\mathcal{T}_{k,t}}}\right)$ *and* $C > 0$ *is an absolute constant.*

Proof. We first prove that:

$$\left(\hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right)\left(\hat{F}_k + \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K}\right)$$

is the (optimistic) estimator of

$$\mathbb{P}[S \leq p_k]\int_{p_k}^{1}\mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda.$$

To do so, we observe that under the event $\mathcal{E}$, we have:

$$\begin{aligned}
&\left(\hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right)\left(\hat{F}_k + \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K}\right) \; - \; \mathbb{P}[S \leq p_k]\int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda \\
&\qquad \leq \left(\mathbb{P}[S \leq p_k] + \sqrt{\frac{2\log(2T/\delta)}{\mathcal{Q}_{k,t}}}\right)\left(\int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda + \sqrt{\frac{8\log(2/\delta)}{HK}} + \frac{2}{K}\right) \\
&\qquad\quad - \mathbb{P}[S \leq p_k]\int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda \\
&\qquad \leq \sqrt{\frac{2\log(2T/\delta)}{\mathcal{Q}_{k,t}}}\int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda + \mathbb{P}[S \leq p_k]\sqrt{\frac{8\log(2/\delta)}{HK}} + \frac{4\log(2T/\delta)}{\sqrt{HK\mathcal{Q}_{k,t}}} \\
&\qquad\quad + \frac{2}{K}\left(\mathbb{P}[S \leq p_k] + \sqrt{\frac{2\log(2T/\delta)}{\mathcal{Q}_{k,t}}}\right) \\
&\qquad \leq \underbrace{\sqrt{\frac{2\log(2T/\delta)}{\mathcal{Q}_{k,t}}}\mathbb{P}[B \geq p_k]}_{(\star)} \quad + \quad \frac{20\log(2T/\delta)}{T^{1/3}}.
\end{aligned}$$

Now, notice that, since $k \in \mathcal{K}$ by assumption, we have $\mathcal{Q}_{k,t} \geq \frac{1}{2}\mathcal{T}_{k,t}\mathbb{P}\left[B \geq p_k\right]$ for all $t \geq HK$, under the event $\mathcal{E}$. Therefore, the following holds:

$$(\star) = \sqrt{\frac{2\log(2T/\delta)}{\mathcal{Q}_{k,t}}}\mathbb{P}\left[B \geq p_k\right] \leq \sqrt{\frac{4\log(2T/\delta)}{\mathcal{T}_{k,t}\mathbb{P}\left[B \geq p_k\right]}}\mathbb{P}[B \geq p_k] \leq 2\frac{\log(2T/\delta)}{\sqrt{\mathcal{T}_{k,t}}}.$$

Putting all together, we have:

$$\begin{aligned}
&\left(\hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right)\left(\hat{F}_k + \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K}\right) - \mathbb{P}[S \leq p_k]\int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda \\
&\qquad\qquad \leq \log\left(\frac{T}{\delta}\right)\mathcal{O}\left(\frac{1}{\sqrt{\mathcal{T}_{k,t}}} + \frac{1}{T^{1/3}}\right).
\end{aligned}$$

Finally, we notice that:

$$\mathbb{P}[S \leq p_k]\int_{p_k}^1 \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda \leq \left(\hat{\mu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{Q}_{k,t}}}\right)\left(\hat{F}_k + \sqrt{\frac{2\log(2/\delta)}{HK}} + \frac{1}{K}\right),$$

as a direct consequence of being under the clean event $\mathcal{E}$. This concludes the proof. □

Lemma 4. *Let $k \in \mathcal{K}$. Then, for each $t \geq HK$, conditional on the event $\mathcal{E}$, we have:*

$$0 \leq \left(\hat{\nu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}}\right)\left(\hat{G}_k + \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K}\right) - \mathbb{P}[B \geq p_k]\int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda \leq \eta,$$

where $\eta \coloneqq C\log\left(\frac{T}{\delta}\right)\left(\frac{1}{T^{1/3}} + \frac{1}{\sqrt{\mathcal{T}_{k,t}}}\right)$ and $C > 0$ is an absolute constant.

Proof. Notice that the first inequality is trivially given by the fact that, under the event $\mathcal{E}$, the quantity

$$\left(\hat{\nu}_{k,t} + \sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}}\right)\left(\hat{G}_k + \sqrt{\frac{2\log(2/\delta)}{KN_k}} + \frac{1}{K}\right)$$

is an (optimistic) estimator of

$$\mathbb{P}[B \geq p_k]\int_0^{p_k} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda.$$

For the second inequality, we notice that, under the event $\mathcal{E}$:

$$\begin{aligned}
&\left(\hat{\nu}_{k,t}+\sqrt{\frac{\log(2T/\delta)}{2\mathcal{T}_{k,t}}}\right)\left(\hat{G}_k+\sqrt{\frac{2\log(2/\delta)}{KN_k}}+\frac{1}{K}\right)-\mathbb{P}[B\geq p_k]\int_0^{p_k}\mathbb{P}[S\leq\lambda]\,\mathrm{d}\lambda \\
&\quad\overset{\mathcal{E}\subset\mathcal{E}_2\cap\mathcal{E}_3}{\leq}\left(\mathbb{P}[B\geq p_k]+\sqrt{\frac{2\log(2T/\delta)}{\mathcal{T}_{k,t}}}\right)\left(\int_0^{p_k}\mathbb{P}[S\leq\lambda]\,\mathrm{d}\lambda+\sqrt{\frac{8\log(2/\delta)}{KN_k}}+\frac{2}{K}\right) \\
&\qquad-\mathbb{P}[B\geq p_k]\int_0^{p_k}\mathbb{P}[S\leq\lambda]\,\mathrm{d}\lambda \\
&\quad\overset{\mathcal{E}\subset\mathcal{E}_1}{\leq}\sqrt{\frac{2\log(2T/\delta)}{\mathcal{T}_{k,t}}}\int_0^{p_k}\mathbb{P}[S\leq\lambda]\,\mathrm{d}\lambda+4\mathbb{P}[B\geq p_k]\sqrt{\frac{\log(2/\delta)}{KH\mathbb{P}[B\geq p_k]}} \\
&\qquad+\frac{4\sqrt{2}\log(2T/\delta)}{\sqrt{KH\mathbb{P}[B\geq p_k]}\sqrt{\mathcal{T}_{k,t}}}+\frac{2}{K}\left(\mathbb{P}[B\geq p_k]+\sqrt{\frac{2\log(2T/\delta)}{\mathcal{T}_{k,t}}}\right) \\
&\quad\overset{k\in\mathcal{K}}{=}\log\left(\frac{T}{\delta}\right)\cdot\mathcal{O}\left(\frac{1}{\sqrt{\mathcal{T}_{k,t}}}+\frac{1}{T^{1/3}}\right)\,,
\end{aligned}$$

concluding the proof. □

Theorem 1. *Algorithm 2 guarantees regret $R_T=\tilde{\mathcal{O}}(T^{2/3})$.*

Proof. We first notice that, by defining

$$\mathcal{R}_T\coloneqq\sum_{t=1}^{T}\mathrm{g}(p^\star)-\sum_{t=1}^{T}\mathrm{g}(P_t),$$

we have that $R_T=\mathbb{E}[\mathcal{R}_T]$ and

$$\begin{aligned}
R_T&=\mathbb{E}[\mathcal{R}_T\mathbb{I}_{\mathcal{E}}]+\mathbb{E}[\mathcal{R}_T\mathbb{I}_{\mathcal{E}^c}]\\
&\leq\mathbb{E}[\mathcal{R}_T\mathbb{I}_{\mathcal{E}}]+\mathbb{E}[T\mathbb{I}_{\mathcal{E}^c}]\\
&\leq\mathbb{E}[\mathcal{R}_T\mathbb{I}_{\mathcal{E}}]+6KT^2\delta=\mathbb{E}[\mathcal{R}_T\mathbb{I}_{\mathcal{E}}]+\mathcal{O}(T^{1/3}).
\end{aligned}$$

It is then sufficient to control the magnitude of $\mathcal{R}_T$ under the clean event $\mathcal{E}$. Hence, from this point on, we assume we are under the clean event $\mathcal{E}$.

Let $k^\star\in\arg\max_{k\in[K]}\mathrm{g}(p_k)$.

First, notice that, if $\mathbb{P}[p_{k^\star}\leq B]\leq 64T^{-1/3}\log(KT^2/\delta)$, then

$$\mathrm{g}(p_{k^\star})=\mathbb{E}[(B-S)\,\mathbb{I}\{S\leq p_{k^\star}\leq B\}]\leq\mathbb{E}[\mathbb{I}\{B\geq p_{k^\star}\}]=\mathbb{P}[B\geq p_{k^\star}]\leq 64T^{-1/3}\log(KT^2/\delta),$$

where we used $(B-S)\leq 1$ and $\{S\leq p\leq B\}\subseteq\{B\geq p\}$. Thus, when $\mathbb{P}[p_{k^\star}\leq B]\leq 64T^{-1/3}\log(KT^2/\delta)$, we have, due to Lemma 1, that if we pay an additional term whose instantaneous regret is upper bounded by L/K, we can control $\mathcal{R}_T$ by comparing our performance against the performance of the best point in the grid $p_{k^\star}$, from which

$$\mathcal{R}_T=T\cdot\tilde{\mathcal{O}}(T^{-1/3})+\frac{TL}{K}=\tilde{\mathcal{O}}(T^{2/3})\,.$$

Hence, we are left to analyze what happens when $\mathbb{P}[p_{k^\star}\leq B]>64T^{-1/3}\log(KT^2/\delta)$, which we assume being the case from this point on. First, since $\mathbb{P}[p_{k^\star}\leq B]>64T^{-1/3}\log(KT^2/\delta)$, given that $\mathcal{E}\subset\mathcal{E}_1$, it follows that $k^\star\in K^\diamond$.

We now notice that for each $k\in K^\diamond$ we have that $\mathcal{Q}_{k,HK}>32T^{-1/3}\log(KT^2/\delta)$ by definition. In the clean event $\mathcal{E}$, we have that $\mathcal{E}_4$ holds, and hence for each $h\notin\mathcal{K}$ we have that $\mathcal{Q}_{h,HK}\leq 32T^{-1/3}\log(KT^2/\delta)$. It follows that, in the clean event $\mathcal{E}$, $k\in K^\diamond$ implies $k\in\mathcal{K}$, *i.e.*, $K^\diamond\subset\mathcal{K}$.

Now, we recall that Lemma 3 and Lemma 4 imply that, under the event $\mathcal{E}$, for all $t \geq HK+1$ and $k \in \mathcal{K}$:

$$\mathrm{g}(p_k) \leq \mathit{UCB}_{k,t-1} \leq \mathrm{g}(p_k) + \eta_{k,t-1} \ , \tag{5}$$

where $\eta_{k,t-1} \coloneqq \tilde{C} \log\Big(\frac{T}{\delta}\Big)\Big(\frac{1}{T^{1/3}} + \frac{1}{\sqrt{\mathcal{T}_{k,t-1}}}\Big)$ and $\tilde{C} > 0$ is a universal constant. (We use $\mathrm{UCB}_{k,t-1}$ because P_t is chosen at the start of round t based only on information up to time $t-1$.)

If, for every $p \in [0,1]$, we define the quantity $\Delta_p \coloneqq \mathrm{g}(p_{k^\star}) - \mathrm{g}(p)$, then, for each $t \geq HK+1$, if $k_t \in K^\diamond$ is such that $P_t = p_{k_t}$, by eq. (5) we have

$$\mathrm{g}(p_{k^\star}) = \max_{k \in K^\diamond} \mathrm{g}(p_k) \leq \max_{k \in K^\diamond} \mathit{UCB}_{k,t-1} = \mathit{UCB}_{k_t,t-1} \leq \mathrm{g}(P_t) + \eta_{k_t,t-1} \ ,$$

and hence

$$\Delta_{P_t} \leq \eta_{k_t,t-1} \ .$$

In addition, by Lemma 1 and the fact that the instantaneous regret is upper bounded by 1, we have:

$$\mathcal{R}_T \leq HK + \frac{LT}{K} + \sum_{t=HK+1}^{T} \Delta_{P_t}.$$

Now, we have

$$\begin{aligned}
\sum_{t=HK+1}^{T} \Delta_{P_t} &= \sum_{k \in K^\diamond} \sum_{t=HK+1}^{T} \Delta_{P_t} \, \mathbb{I}\{P_t = p_k\} \\
&\leq \sum_{k \in K^\diamond} \sum_{t=HK+1}^{T} \eta_{k,t-1} \, \mathbb{I}\{P_t = p_k\} \\
&\leq \sum_{k \in K^\diamond} \sum_{t=HK+1}^{T} \left[\tilde{C} \, \log\Big(\frac{T}{\delta}\Big)\left(\frac{1}{T^{1/3}} + \frac{1}{\sqrt{\mathcal{T}_{k,t-1}}}\right)\right] \mathbb{I}\{P_t = p_k\} \quad \text{by the definition of } \eta_{k,t-1} \\
&= \tilde{C} \, \log\Big(\frac{T}{\delta}\Big) \left[\frac{1}{T^{1/3}} \sum_{k \in K^\diamond} \sum_{t=HK+1}^{T} \mathbb{I}\{P_t = p_k\} \ + \ \sum_{k \in K^\diamond} \sum_{t=HK+1}^{T} \frac{\mathbb{I}\{P_t = p_k\}}{\sqrt{\mathcal{T}_{k,t-1}}}\right] \\
&\leq \tilde{C} \, \log\Big(\frac{T}{\delta}\Big) \left[\frac{1}{T^{1/3}} \sum_{t=HK+1}^{T} 1 \ + \ \sum_{k \in K^\diamond} \sum_{t=HK+1}^{T} \frac{\mathbb{I}\{P_t = p_k\}}{\sqrt{\mathcal{T}_{k,t-1}}}\right] \\
&\leq \tilde{C} \, \log\Big(\frac{T}{\delta}\Big) \left[T^{2/3} \ + \ \sum_{k \in K^\diamond} \sum_{t=HK+1}^{T} \frac{\mathbb{I}\{P_t = p_k\}}{\sqrt{\mathcal{T}_{k,t-1}}}\right] \\
&\leq \tilde{C} \, \log\Big(\frac{T}{\delta}\Big) \left[T^{2/3} \ + \ \sum_{k \in K^\diamond} \underbrace{\sum_{t=HK+1}^{T} \frac{\mathbb{I}\{P_t = p_k\}}{\sqrt{\mathcal{T}_{k,t-1}}}}_{\leq \ 2\sqrt{\mathcal{T}_{k,T}} \ \text{by (a)}}\right] \\
&\leq \tilde{C} \, \log\Big(\frac{T}{\delta}\Big) \left[T^{2/3} \ + \ 2 \sum_{k \in K^\diamond} \sqrt{\mathcal{T}_{k,T}}\right] \\
&\leq \tilde{C} \, \log\Big(\frac{T}{\delta}\Big) \left[T^{2/3} \ + \ 2 \sqrt{|K^\diamond| \sum_{k \in K^\diamond} \mathcal{T}_{k,T}}\right] \quad \text{(b) Cauchy–Schwarz} \\
&\leq \tilde{C} \, \log\Big(\frac{T}{\delta}\Big) \left[T^{2/3} \ + \ 2\sqrt{K\,T}\right] \\
&= \ \tilde{\mathcal{O}}\big(T^{2/3}\big),
\end{aligned}$$

where (a) and (b) can be proved as follows

$$\text{(a)} \quad \sum_{t=HK+1}^{T} \frac{\mathbb{I}\{P_t = p_k\}}{\sqrt{\mathcal{T}_{k,t-1}}} = \sum_{i=1}^{m_k} \frac{1}{\sqrt{H+(i-1)}} \quad \text{where } m_k \coloneqq \sum_{t=HK+1}^{T} \mathbb{I}\{P_t = p_k\}, \;\; H = \lceil T^{1/3} \rceil$$

$$\leq \sum_{j=H}^{H+m_k-1} \frac{1}{\sqrt{j}} \leq \int_{H-1}^{H+m_k-1} \frac{1}{\sqrt{x}}\,\mathrm{d}x = 2\big(\sqrt{H+m_k-1} - \sqrt{H-1}\big)$$

$$\leq 2\sqrt{m_k} \leq 2\sqrt{\mathcal{T}_{k,T}},$$

$$\text{(b)} \quad \sum_{k \in K^\diamond} \sqrt{\mathcal{T}_{k,T}} \leq \sqrt{|K^\diamond| \sum_{k \in K^\diamond} \mathcal{T}_{k,T}} \leq \sqrt{KT}, \qquad \text{since} \quad \sum_{k \in K^\diamond} \mathcal{T}_{k,T} \leq \sum_{t=1}^{T} \sum_{k} \mathbb{I}\{P_t = p_k\} = T.$$

Hence

$$\mathbb{E}[\mathcal{R}_T \mathbb{I}_{\mathcal{E}}] \leq \tilde{C}' \log\left(\frac{T}{\delta}\right) T^{2/3} = \tilde{\mathcal{O}}(T^{2/3}) \,,$$

concluding the proof. □

Experimental Results

In this section, we present some experimental results obtained on synthetically-generated instances. Specifically, we consider instances where the seller's valuations are sampled from a Beta distribution with parameters α_s and β_s, while the buyer's valuations are sampled from a Beta distribution with parameters α_b and β_b. For each instance, we evaluate the performance of our algorithm and the Scouting Bandits algorithm of Cesa-Bianchi et al. [2021] in terms of cumulative regret. To this end, we run both algorithms on each instance $n = 5$ times and report the mean and standard deviation of the achieved cumulative regret.

Table 1: Comparison between our algorithm and the one of Cesa-Bianchi et al. [2021] in terms of cumulative regret across different instances where buyers and sellers' valuations are distributed according to Beta distributions.

Parameter	**Instance 1**	**Instance 2**	**Instance 3**
Time horizon (T)	10000	50000	10000
(α_s, β_s)	(5.0, 10.0)	(5.0, 10.0)	(10.0, 10.0)
(α_b, β_b)	(15.0, 10.0)	(15.0, 10.0)	(15.0, 10.0)
Regret $\pm$ std (ours)	199.6 ± 17.1	714.4 ± 73.1	135.2 ± 22.1
Regret $\pm$ std (Cesa-Bianchi et al. [2021])	732.0 ± 21.4	2253.8 ± 75.2	548.8 ± 16.4

Table 2: Comparison between our algorithm and the one of Cesa-Bianchi et al. [2021] in terms of cumulative regret across different instances where buyers and sellers' valuations are distributed according to Beta distributions.

Parameter	**Instance 4**	**Instance 5**	**Instance 6**
Time horizon (T)	50000	10000	50000
(α_s, β_s)	(10.0, 10.0)	(2.0, 3.0)	(2.0, 3.0)
(α_b, β_b)	(15.0, 10.0)	(10.0, 10.0)	(10.0, 10.0)
Regret $\pm$ std (ours)	326.5 ± 84.0	168.5 ± 28.8	439.2 ± 136.2
Regret $\pm$ std (Cesa-Bianchi et al. [2021])	2381.0 ± 98.1	583.5 ± 27.6	2502.4 ± 52.2

We observe that the regret incurred by our algorithm is lower than that of Cesa-Bianchi et al. [2021]. While this may appear counterintuitive (since we use less feedback than the Scouting Bandits algorithm of Cesa-Bianchi et al. [2021]), the improvement stems from a key difference. Indeed, after the initial exploration phase, we eliminate arms that are guaranteed to be suboptimal. This pruning

step, absent in Scouting Bandits, allows us to restrict the subsequent bandit phase to the reduced set $K^{\diamond}$, which can be significantly smaller than the original set of arms. In contrast, Cesa-Bianchi et al. [2021] run their algorithm over the full set, whose size is $T^{1/3}$.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes, we state the main theorem of the paper in the abstract/introduction.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Discussed in section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: All our results have a complete proof in the appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: the paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The paper does not include data or code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: There is no societal impact of the work performed, since the work is mainly theoretical.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The paper does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.