Online certification of preference-based fairness
for personalized recommender systems

Virginie Do[1,2], Sam Corbett-Davies[2], Jamal Atif[1], Nicolas Usunier[2]

[1]LAMSADE, Université PSL, Université Paris Dauphine, CNRS, France
[2]Meta AI
virginie.do@dauphine.eu, scd@fb.com, jamal.atif@lamsade.dauphine.fr, usunier@fb.com

Abstract

Recommender systems are facing scrutiny because of their growing impact on the opportunities we have access to. Current audits for fairness are limited to coarse-grained parity assessments at the level of sensitive groups. We propose to audit for *envy-freeness*, a more granular criterion aligned with individual preferences: every user should prefer their recommendations to those of other users. Since auditing for envy requires to estimate the preferences of users beyond their existing recommendations, we cast the audit as a new pure exploration problem in multi-armed bandits. We propose a sample-efficient algorithm with theoretical guarantees that it does not deteriorate user experience. We also study the trade-offs achieved on real-world recommendation datasets.

1 Introduction

Recommender systems shape the information and opportunities available to us, as they help us prioritize content from news outlets and social networks, sort job postings, or find new people to connect with. To prevent the risk of unfair delivery of opportunities across users, substantial work has been done to audit recommender systems (Sweeney 2013; Asplund et al. 2020; Imana et al. 2021). For instance, Datta et al. (2015) found that women received fewer online ads for high-paying jobs than equally qualified men, while Imana et al. (2021) observed different delivery rates of ads depending on gender for different companies proposing similar jobs.

The audits above aim at controlling for the possible acceptable justifications of the disparities, such as education level in job recommendation audits. Yet, the observed disparities in recommendation do not necessarily imply that a group has a less favorable treatment: they might as well reflect that individuals of different groups tend to prefer different items. To strengthen the conclusions of the audits, it is necessary to develop methods that account for user preferences. Audits for equal satisfaction between user groups follow this direction (Mehrotra et al. 2017), but they also have limitations. For example, they require interpersonal comparisons of measures of satisfaction, a notoriously difficult task (Sen 1999).

We propose an alternative approach to incorporating user preferences in audits which focuses on *envy-free recommendations*: the recommender system is deemed fair if each user

prefers their recommendation to those of all other users. Envy-freeness allows a system to be fair even in the presence of disparities between groups as long as these are justified by user preferences. On the other hand, if user B systematically receives better opportunities than user A *from A's perspective*, the system is unfair. The criterion does not require interpersonal comparisons of satisfaction, since it relies on comparisons of different recommendations from the perspective of the same user. Similar fairness concepts have been studied in classification tasks under the umbrella of preference-based fairness (Zafar et al. 2017; Kim et al. 2019; Ustun et al. 2019). Envy-free recommendation is the extension of these approaches to personalized recommender systems.

Compared to auditing for recommendation parity or equal satisfaction, auditing for envy-freeness poses new challenges. First, envy-freeness requires answering counterfactual questions such as "would user A get higher utility from the recommendations of user B than their own?", while searching for the users who most likely have the best recommendations from A's perspective. This type of question can be answered reliably only through active exploration, hence we cast it in the framework of pure exploration bandits (Bubeck et al. 2009). To make such an exploration possible, we consider a scenario where the auditor is allowed to replace a user's recommendations with those that another user would have received in the same context. Envy, or the absence thereof, is estimated by suitably choosing whose recommendations should be shown to whom. While this scenario is more intrusive than some black-box audits of parity, auditing for envy-freeness provides a more compelling guarantee on the wellbeing of users subject to the recommendations.

The second challenge is that active exploration requires randomizing the recommendations, which in turn might alter the user experience. In order to control this cost of the audit (in terms of user utility), we follow the framework of conservative exploration (Wu et al. 2016; Garcelon et al. 2020), which guarantees a performance close to the audited system. We provide a theoretical analysis of the trade-offs that arise, in terms of the cost and duration of the audit (measured in the number of timesteps required to output a certificate).

Our technical contributions are twofold. **(1)** We provide a novel formal analysis of envy-free recommender systems, including a comparison with existing item-side fairness criteria and a probabilistic relaxation of the criterion. **(2)** We

cast the problem of auditing for envy-freeness as a new pure exploration problem in bandits with conservative exploration constraints, and propose a sample-efficient auditing algorithm which provably maintains, throughout the course of the audit, a performance close to the audited system.

We discuss the related work in Sec. 2. Envy-free recommender systems are studied in Sec. 3. In Sec. 4, we present the bandit-based auditing algorithm. In Sec. 5, we investigate the trade-offs achieved on real-world datasets.

2 Related work

Fair recommendation The domain of fair machine learning is organized along two orthogonal axes. The first axis is whether fairness is oriented towards groups defined by protected attributes (Barocas and Selbst 2016), or rather oriented towards individuals (Dwork et al. 2012). The second axis is whether fairness is a question of *parity* (predictions [or prediction errors] should be invariant by group or individual) (Corbett-Davies and Goel 2018; Kusner et al. 2017), or *preference-based* (predictions are allowed to be different if they faithfully reflect the preferences of all parties) (Zafar et al. 2017; Kim et al. 2019; Ustun et al. 2019). Our work takes the perspective of envy-freeness, which follows the preference-based approach and is aimed towards individuals.

The literature on fair recommender systems covers two problems: *auditing* existing systems, and *designing* fair recommendation algorithms. Most of the *auditing* literature focused on group parity in recommendations (Hannak et al. 2014; Lambrecht and Tucker 2019), and equal user utility (Mehrotra et al. 2017; Ekstrand et al. 2018), while our audit for envy-freeness focuses on whether personalized results are aligned with (unknown) user preferences. On the *designing* side, Patro et al. (2020); Ilvento, Jagadeesan, and Chawla (2020) cast fair recommendation as an allocation problem, with criteria akin to envy-freeness. They do not address the partial observability of preferences, so they cannot guarantee user-side fairness without an additional certificate that the estimated preferences effectively represent the true user preferences. Our work is thus complementary to theirs.

While we study fairness for users, recommender systems are multi-sided (Burke 2017; Patro et al. 2020), thus fairness can also be oriented towards recommended items (Celis, Straszak, and Vishnoi 2017; Biega et al. 2018; Geyik, Ambler, and Kenthapadi 2019).

Multi-armed bandits In *pure exploration* bandits (Bubeck et al. 2009; Audibert and Bubeck 2010), an agent has to identify a specific set of arms after exploring as quickly as possible, without performance constraints. Our setting is close to threshold bandits (Locatelli, Gutzeit, and Carpentier 2016; Kano et al. 2019) where the goal is to find arms with better performance than a given baseline. Outside pure exploration, in the *regret minimization* setting, conservative exploration (Wu et al. 2016) enforces the anytime average performance to be not too far worse than that of a baseline arm.

In our work, the baseline is *unknown* – it is the current recommender system – and the other "arms" are other users' policies. The goal is to make the decision as to whether an arm is better than the baseline, while not deteriorating

performance compared to the baseline. We thus combine pure exploration and conservative constraints.

Existing work on fairness in exploration/exploitation (Joseph et al. 2016; Jabbari et al. 2017; Liu et al. 2017) is different from ours because unrelated to personalization.

Fair allocation Envy-freeness was first studied in fair allocation (Foley 1967) in social choice. Our setting is different because: a) the same item can be given to an unrestricted number of users, and b) true user preferences are unknown.

3 Envy-free recommendations
3.1 Framework

There are M users, and we identify the set of users with $[M] = \{1, \ldots, M\}$. A personalized recommender system has one stochastic recommendation policy π^m per user m. We denote by $\pi^m(a|x)$ the probability of recommending item $a \in \mathcal{A}$ for user $m \in [M]$ in context $x \in \mathcal{X}$. We assume that $\mathcal{X}$ and $\mathcal{A}$ are finite to simplify notation, but this has no impact on the results. We consider a synchronous setting where at each time step t, the recommender system observes a context $x_t^m \sim q^m$ for each user, selects an item $a_t^m \sim \pi^m(.|x_t^m)$ and observes reward $r_t^m \sim \nu^m(a_t^m|x_t^m) \in [0, 1]$. We denote by $\rho^m(a|x)$ the expected reward for user m and item a in context x, and, for any recommendation policy π, $u^m(\pi)$ is the utility of m for π:

$$u^m(\pi) = \mathbb{E}_{x \sim q^m} \mathbb{E}_{a \sim \pi(.|x)} \mathbb{E}_{r \sim \nu^m(a|x)} [r]$$
$$= \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} q^m(x) \pi(a|x) \rho^m(a|x) \quad (1)$$

We assume that the environment is *stationary*: the context and reward distributions q^m and ν^m, as well as the policies π^m are fixed. Even though in practice policies evolve as they learn from user interactions and user needs change over time, we leave the study of non-stationarities for future work. The stationary assumption approximately holds when these changes are slow compared to the time horizon of the audit, which is reasonable when significant changes in user needs or recommendation policies take e.g., weeks. Our approach applies when items a are single products as well as when items are structured objects such as rankings. Examples of (context x, item a) pairs include: x is a query to a search engine and a is a document or a ranking of documents, or x is a song chosen by the user and a a song to play next or an entire playlist. Remember, our goal is *not* to learn the user policies π^m, but rather to audit existing π^ms for fairness.

3.2 ϵ-envy-free recommendations

Existing audits for user-side fairness in recommender systems are based on two main criteria:

1. *recommendation parity*: the distribution of recommended items should be equal across (groups of) users,
2. *equal user utility*: all (groups of) users should receive the same utility, i.e. $\forall m, n, u^m(\pi^m) = u^n(\pi^n)$.

There are two ways in which these criteria conflict with the goal of personalized recommender systems to best accomodate user preferences. First, recommendation parity does not

control for disparities that are aligned with user preferences. Second, equal user utility drives utility down as soon as users have different best achievable utilities. To address these shortfalls, we propose envy-freeness as a complementary diagnosis for the fairness assessment of personalized recommender systems. In this context, envy-freeness requires that users prefer their recommendations to those of any other user:

Definition 3.1. *Let $\epsilon \geq 0$. A recommender system is ϵ-envy-free if:* $\forall m, n \in [M]: \quad u^m(\pi^n) \leq \epsilon + u^m(\pi^m).$

Envy-freeness, originally studied in fair allocation (Foley 1967) and more recently fair classification (Balcan et al. 2018; Ustun et al. 2019; Kim et al. 2019), stipulates that it is fair to apply different policies to different individuals or groups as long as it benefits everyone. Following this principle, we consider the personalization of recommendations as fair only if it better accommodates individuals' preferences. In contrast, we consider unfair the failure to give users a better recommendation when one such is available to others.

Unlike parity or equal utility, envy-freeness is in line with giving users their most preferred recommendations (see Sec. 3.3). Another improvement from equal user utility is that it does not involve interpersonal utility comparisons.

Envy can arise from a variety of sources, for which we provide concrete examples in our experiments (Sec. 5.1).

Remark. We discuss an immediate extension of envy-freeness from individuals to groups of users in App. B, in the special case where groups have homogeneous preferences and policies. Defining group envy-free recommendations in the general case is nontrivial and left for future work.

3.3 Compatibility of envy-freeness

Optimal recommendations are envy-free[1] Let $\pi^{m,*} \in \arg\max_\pi u^m(\pi)$ denote an optimal recommendation policy for m. Then *the optimal recommender system $(\pi^{m,*})_{m \in M}$ is envy-free* since: $u^m(\pi^{m,*}) = \max_\pi u^m(\pi) \geq u^m(\pi^{n,*})$. In contrast, achieving equal user utility in general can only be achieved by decreasing the utility of best-served users for the benefit of no one. It is also well-known that achieving parity in general requires to deviate from optimal predictions (Barocas, Hardt, and Narayanan 2018).

Envy-freeness vs. item-side fairness Envy-freeness is a user-centric notion. Towards multisided fairness (Burke 2017), we analyze the compatibility of envy-freeness with item-side fairness criteria for rankings from Singh and Joachims (2018), based on sensitive categories of items (denoted $\mathcal{A}_1, ..., \mathcal{A}_S$). *Parity of exposure* prescribes that for each user, the exposure of an item category should be proportional to the number of items in that category. In *Equity of exposure*[2], the exposure of item categories should be proportional to their average relevance to the user.

The optimal policies under parity and equity of exposure constraints, denoted respectively by $(\pi^{m,\mathrm{par}})_{m=1}^M$ and

_____

[1]App.A shows the difference between envy-freeness and optimality certificates.

[2]Singh and Joachims (2018) use the terminology of demographic parity (resp. disparate treatment) for what we call parity (resp. equity) of exposure. Our use of "equity" follows Biega et al. (2018).

$(\pi^{m,\mathrm{eq}})_{m=1}^M$, are defined given user m and context x as:

$$(parity) \quad \pi^{m,\mathrm{par}}(.|x) = \underset{\substack{p:\mathcal{A}\to[0,1] \\ \sum_a p(a)=1}}{\arg\max} \sum_{a \in \mathcal{A}} p(a)\rho^m(a|x)$$

$$\text{u.c. } \forall s \in [S], \sum_{a \in \mathcal{A}_s} p(a) = \frac{|\mathcal{A}_s|}{|\mathcal{A}|}. \quad (2)$$

Optimal policies under equity of exposure are defined similarly[3], but the constraints are $\forall s, \sum_{a \in \mathcal{A}_s} p(a) = \frac{\sum_{a \in \mathcal{A}_s} \rho^m(a|x)}{\sum_{a \in \mathcal{A}} \rho^m(a|x)}$. We show their relation to envy-freeness:

Proposition 1. *With the above notation:*
- *the policies $(\pi^{m,\mathrm{par}})_{m=1}^M$ are envy-free, while*
- *the policies $(\pi^{m,\mathrm{eq}})_{m=1}^M$ are not envy-free in general.*

Optimal recommendations under parity of exposure are envy-free because the parity constraint (2) is the same for all users. Given two users m and n, $\pi^{m,\mathrm{par}}$ is optimal for m under (2) and $\pi^{n,\mathrm{par}}$ satisfies the same constraint, so we have $u^m(\pi^{m,\mathrm{par}}) \geq u^m(\pi^{n,\mathrm{par}})$.

In contrast, the optimal recommendations under equity of exposure are, in general, not envy-free. A first reason is that less relevant item categories reduce the exposure of more relevant categories: a user who prefers item a but who also likes item b from another category envies a user who only liked item is a. Note that *amortized* versions of the criterion and other variants considering constraint averages over user/contexts (Biega et al. 2018; Patro et al. 2020) have similar pitfalls unless envy-freeness is explictly enforced, as in Patro et al. (2020) who developed an envy-free algorithm assuming the true preferences are known. For completeness, we describe in App.A a second reason why equity of exposure constraints create envy, and an edge case where they do not.

3.4 Probabilistic relaxation of envy-freeness

Envy-freeness, as defined in Sec. 3.2, (a) compares the recommendations of a target user to those of *all* other users, and (b) these comparisons must be made for *all* users. In practice, as we show, this means that the sample complexity of the audit increases with the number of users, and that all users must be part of the audit.

In practice, it is likely sufficient to relax both conditions on all users to give a guarantee for most recommendation policies and most users. Given two small probabilities λ and γ, the relaxed criterion we propose requires that for at least $1 - \lambda$ fraction of users, the utility of users for their own policy is in the top-$\gamma\%$ of their utilities for anyone else's policy. The formal definition is given below. The fundamental observation, which we prove in Th. 2 in Sec. 4.5, is that the sample complexity of the audit and the number of users impacted by the audit are now *independent on the total number of users*. We believe that these relaxed criteria are thus likely to encourage the deployment of envy-free audits in practice.

_____

[3]The original criterion (Singh and Joachims 2018, Eq. 4) would be written in our case as $\forall s, s' \in [S], \frac{1}{|\mathcal{A}_s|} \sum_{a \in \mathcal{A}_s} p(a) = \frac{1}{|\mathcal{A}_{s'}|} \sum_{a \in \mathcal{A}_{s'}} p(a)$, which is equivalent to (2). A similar remark holds for the equity constraint.
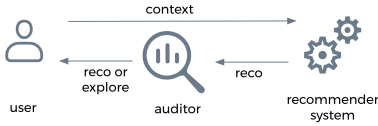
Figure 1: Auditing scenario: the auditor either shows the user their recommendation in the current rec. system, or explores by showing the recommendation given to another user.

Definition 3.2. *Let* $\epsilon, \gamma, \lambda \geq 0$. *Let* U_M *denote the discrete uniform distribution over* $[M]$. *A user* m *is* (ϵ, γ)-*envious if:*

$$\mathbb{P}_{n \sim U_M}\left[u^m(\pi^m) + \epsilon < u^m(\pi^n)\right] > \gamma.$$

A recommender system is $(\epsilon, \gamma, \lambda)$-*envy-free if at least a* $(1 - \lambda)$ *fraction of its users are not* (ϵ, γ)-*envious.*

4 Certifying envy-freeness

4.1 Auditing scenario

The envy-freeness auditor must answer the counterfactual question: "had user m been given the recommendations of user n, would m get higher utility?". The main challenge is that the answer requires to access to user preferences, which are only partially observed since users only interact with recommended items. There is thus a need for an active exploration process that recommends items which would not have been recommended otherwise.

To make such an exploration possible, we consider the following auditing scenario: at each time step t, the auditor chooses to either (a) give the user a "normal" recommendation, or (b) explore user preferences by giving the user a recommendation from another user (see Fig. 1) . This scenario has the advantage of lightweight infrastructure requirements, since the auditor only needs to query another user's policy, rather than implementing a full recommender system within the operational constraints of the platform. Moreover, this interface is sufficient to estimate envy because envy is defined based on the performance of other user's policies. This type of internal audit (Raji et al. 2020) requires more access than usual external audits that focus on recommendation parity, but this is necessary to explore user preferences.

We note that the auditor must make sure that this approach follows the relevant ethical standard for randomized experiments in the context of the audited system. The auditor must also check that using other users' recommendation policies does not pose privacy problems. From now on, we assume these issues have been resolved.

4.2 The equivalent bandit problem

We now cast the audit for envy-freeness as a new variant of pure exploration bandit problems. We first focus on auditing envy for a single target user and define the corresponding objectives, then we present our auditing algorithm. Finally we specify how to use it for the certification of either the exact or probabilistic envy-freeness criteria.

For a target user m, the auditor must estimate whether $u^m(\pi^m) + \epsilon \geq u^m(\pi^n)$, for n in a subset $\{n_1, ..., n_K\}$ of K users from $[M]$ (where K is specified later, depending

---

Algorithm 1: OCEF algorithm. ξ_t (line 4) evaluates the conservative exploration constraint and is defined in (4). Values for $\beta_k(t)$ and confidence bounds $\underline{\mu}_k$ and $\overline{\mu}_k$ are given in Lemma 4.

input : Confidence parameter δ, conservative exploration parameter α, envy parameter ϵ

output : envy or ϵ-no-envy

1 $S_0 \leftarrow [K]$ // all arms except 0
2 **for** $t=1, \ldots$ **do**
3   Choose ℓ_t from S_{t-1} // e.g., unif.sample
4   **if** $\beta_0(t-1) > \min_{k \in S_{t-1}} \beta_k(t-1)$ **or** $\xi_t < 0$ **then** $k_t \leftarrow 0$
5   **else** $k_t \leftarrow \ell_t$
6   Observe context $x_t \sim q$, show $a_t \sim \pi^{k_t}(.|x_t)$ and observe $r_t \sim \nu(a_t|x_t)$ // i.e., pull arm k_t and update conf.intervals with Lem.4
7      $S_t \leftarrow \{k \in S_{t-1} : \overline{\mu}_k(t) > \underline{\mu}_0(t) + \epsilon\}$
8   **if** $\exists k \in S_t, \underline{\mu}_k(t) > \overline{\mu}_0(t)$ **then return** envy
9   **if** $S_t = \emptyset$ **then return** ϵ-no-envy
10 **end**

---

on the criterion). As we first focus on auditing envy for one target user m, we drop all superscripts m to simplify notation. We identify $\{n_1, ..., n_K\}$ with $[K]$ and rename $\left(u^m(\pi^{n_1}), ..., u^m(\pi^{n_K})\right)$ as $(\mu_1, ..., \mu_K)$. To estimate μ_k, we obtain samples by making recommendations using the policy π^k and observing the reward. The remaining challenge is to choose which user k to sample at each time step while not deteriorating the experience of the target user too much. Index 0 represents the target user: we use μ_0 for the utility of the user for their policy (i.e., $u^m(\pi^m)$). Because the audit is a special form of bandit problem, following the bandit literature, an index of a user is called an *arm*, and arm 0 is the *baseline*.

Objectives and evaluation metrics We present our algorithm OCEF (Online Certification of Envy-Freeness) in the next subsection. Given $\epsilon > 0$ and $\alpha \geq 0$, OCEF returns either envy or ϵ-no-envy and has two objectives:

1. Correctness: if OCEF returns envy, then $\exists k, \mu_k > \mu_0$. If OCEF returns ϵ-no-envy then $\max_{k \in [K]} \mu_k \leq \mu_0 + \epsilon$.

2. Recommendation performance: during the audit, OCEF must maintain a fraction $1 - \alpha$ of the baseline performance. Denoting by $k_s \in \{0, \ldots, K\}$ the arm (group index) chosen at round s, this requirement is formalized as a conservative exploration constraint (Wu et al. 2016):

$$\forall t, \frac{1}{t} \sum_{s=1}^{t} \mu_{k_s} \geq (1 - \alpha)\mu_0. \tag{3}$$

We focus on the *fixed confidence* setting, where given a confidence parameter $\delta \in (0, 1)$ the algorithm provably satisfies both objectives with probability $1 - \delta$. In addition, there are two criteria to assess an online auditing algorithm:

1. Duration of the audit: the number of time-steps before the algorithm stops.

2. Cost of the audit: the cumulative loss of rewards incurred. Denoting the duration by τ, the cost is $\tau\mu_0 - \sum_{s=1}^{\tau} \mu_{k_s}$.

   It is possible that the cost is negative when there is envy. In that case, the audit increased recommendation performance by finding better recommendations for the group.

We note the asymmetry in the return statements of the algorithm: envy does not depend on ϵ. This asymmetry is necessary to obtain finite worst-case bounds on the duration and the cost of audit, as we see in Theorem 1.

Our setting had not yet been addressed by the pure exploration bandit literature, which mainly studies the identification of (ϵ-)optimal arms (Audibert and Bubeck 2010). Auditing for envy-freeness requires proper strategies in order to efficiently estimate the arm performances compared to the unknown baseline. Additionally, by making the cost of the audit a primary evaluation criterion, we also bring the principle of conservative exploration to the pure exploration setting, while it had only been studied in regret minimization (Wu et al. 2016). In our setting, conservative constraints involve nontrivial trade-offs between the duration and cost of the audit. We now present the algorithm, and then the theoretical guarantees for the objectives and evaluation measures.

4.3 The OCEF algorithm

OCEF is described in Alg. 1. It maintains confidence intervals on arm performances $(\mu_k)_{k=0}^{K}$. Given the confidence parameter δ, the lower and upper bounds on μ_k at time step t, denoted by $\underline{\mu}_k(t)$ and $\overline{\mu}_k(t)$, are chosen so that with probability at least $1 - \delta$, we have $\forall k, t, \mu_k \in [\underline{\mu}_k(t), \overline{\mu}_k(t)]$. In the algorithm, $\beta_k(t) = (\overline{\mu}_k(t) - \underline{\mu}_k(t))/2$. As Jamieson et al. (2014), we use anytime bounds inspired by the law of the iterated logarithm. These are given in Lem. 4 in App. E.

OCEF maintains an active set S_t of all arms in $[K]$ (i.e., excluding the baseline) whose performance are not confidently less than $\mu_0 + \epsilon$. It is initialized to $S_0 = [K]$ (line 1). At each round t, the algorithm selects an arm $\ell_t \in S_t$ (line 3). Then, depending on the state of the conservative exploration constraint (described later), the algorithm pulls k_t, which is either ℓ_t or the baseline (lines 4-6). After observing the reward r_t, the confidence interval of μ_{ℓ_t} is updated, and all active arms that are confidently worse than the baseline plus ϵ are de-activated (line 7). The algorithm returns envy if an arm k is confidently better than the baseline (line 8), returns ϵ-no-envy if there are no more active arms, (line 9) or continues if neither of these conditions are met.

Conservative exploration To deal with the conservative exploration constraint (3), we follow (Garcelon et al. 2020). Denoting $A_t = \{s \leq t : k_s \neq 0\}$ the time steps at which the baseline was not pulled, we maintain a confidence interval such that with probability $\geq 1 - \delta$, we have $\forall t > 0, \left| \sum_{s \in A_t} (\mu_{k_s} - r_s) \right| \leq \Phi(t)$. The formula for Φ is given in Lem. 6 in App. E. This confidence interval is used to estimate whether the conservative constraint (3) is met at round t as follows. First, let us denote by $N_k(t)$ the number of times arm k has been pulled until t, and notice that (3) is equivalent to $\sum_{s \in A_t} \mu_{k_s} - ((1 - \alpha)t - N_0(t))\mu_0 \geq 0$. After choosing ℓ_t (line 3), we use the lower bound on $\sum_{s \in A_t} \mu_{k_s}$

and the upper bound for μ_0 to obtain a conservative estimate of (3). Using $\tau = t - 1$, this leads to:

$$\xi_t = \sum_{s \in A_\tau} r_s - \Phi(t) + \underline{\mu}_{\ell_t}(\tau) + (N_0(\tau) - (1-\alpha)t)\overline{\mu}_0(\tau). \quad (4)$$

Then, as long as the confidence intervals hold, pulling ℓ_t does not break the constraint (3) if $\xi_t \geq 0$. The algorithm thus pulls the baseline arm when $\xi_t < 0$. To simplify the theoretical analysis, OCEF also pulls the baseline if it does not have the tightest confidence interval (lines 4-6).

4.4 Analysis

The main theoretical result of the paper is the following:

Theorem 1. *Let* $\epsilon \in (0, 1]$, $\alpha \in (0, 1]$, $\delta \in (0, \frac{1}{2})$ *and* $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$ *and* $h_k = \max(1, \frac{1}{\eta_k})$. *Using* $\underline{\mu}, \overline{\mu}$ *and* Φ *given in Lemmas 4 and 6 (App. E), OCEF achieves the following guarantees with probability* $\geq 1 - \delta$:

- *OCEF is correct and satisfies the conservative constraint on the recommendation performance* (3).

- *The duration is in* $O\left(\sum_{k=1}^{K} \dfrac{h_k \log\left(\frac{K \log(Kh_k/\delta\eta_k)}{\delta} \right)}{\min(\alpha\mu_0, \eta_k)} \right)$.

- *The cost is in* $O\left(\sum_{k:\mu_k < \mu_0} \dfrac{(\mu_0 - \mu_k)h_k}{\eta_k} \log\left(\frac{K \log(Kh_k/\delta\eta_k)}{\delta} \right) \right)$.

The important problem-dependent quantity η_k is the gap between the baseline and other arms k. It is asymmetric depending on whether the arm is better than the baseline $(\mu_k - \mu_0)$ or the converse $(\mu_0 - \mu_k + \epsilon)$ because the stopping condition for envy does not depend on ϵ. This leads to a worst case that only depends on ϵ, since $\eta_k = \max(\mu_k - \mu_0, \mu_0 - \mu_k + \epsilon) \geq \frac{\epsilon}{2}$, while if the condition was symmetric, we would have possibly unbounded duration when $\mu_k = \mu_0 + \epsilon$ for some $k \neq 0$. Overall, ignoring log terms, we conclude that when $\alpha\mu_0$ is large, the duration is of order $\sum_k \frac{1}{\eta_k^2}$ and the cost is of order $\sum_k \frac{1}{\eta_k}$. This becomes $\sum_k \frac{1}{\alpha\mu_0\eta_k}$ and $\sum_k \frac{1}{\eta_k}$ when $\alpha\mu_0$ is small compared to η_k. This means that the conservative constraint has an impact mostly when it is strict. It also means that when either $\alpha\mu_0 \ll \eta_k$ or $\eta_k^2 \ll \eta_k$ the cost can be small even when the duration is fairly high.

4.5 Full audit

Exact criterion To audit for envy-freeness on the full system, we apply OCEF to all M users simultaneously and with $K = M$, meaning that the set of arms corresponds to all the users' policies. By the union bound, using $\delta' = \frac{\delta}{M}$ instead of δ in OCEF's confidence intervals, the guarantees of Theorem 1 hold simultaneously for all users.

For recommender systems with large user databases, the duration of OCEF thus becomes less manageable as M increases. We show how to use OCEF to certify the probabilistic criterion with guarantees that do not depend on M.

Probabilistic criterion The AUDIT algorithm for auditing the full recommender system is described in Alg. 2. AUDIT samples a subset of users and a subset of arms for each sampled user. Then it applies OCEF to each user simultaneously

Algorithm 2: AUDIT algorithm. The algorithm either outputs a probabilistic certificate of $(\epsilon, \gamma, \lambda)$-envy-freeness, or evidence of envy.

input : Confidence parameter δ, conservative exploration parameter α, envy parameters $(\epsilon, \gamma, \lambda)$

output : $(\epsilon, \gamma, \lambda)$-`envy-free` or `not-envy-free`

1 Draw a sample $\tilde{S}$ of $\tilde{M} = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ users from $[M]$

2 **for** *each user $m \in \tilde{S}$ in parallel* **do**

3   Sample $K = \left\lceil \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))} \right\rceil$ arms from $[M] \setminus \{m\}$

4   Run OCEF$\left(\frac{\delta}{3\tilde{M}}, \alpha, \epsilon\right)$ for user m with the K arms

5   **if** *OCEF outputs* `envy` **then**

       return `not-envy-free`

6 **end**

7 **return** $(\epsilon, \gamma, \lambda)$-`envy-free`

with their sampled arms. It stops either upon finding an envious user, or when all sampled users are certified with ϵ-no envy. Again there is a necessary asymmetry in the return statements of AUDIT to obtain finite worst-case bounds whether or not the system is envy-free.

The number of target users $\tilde{M}$ and arms K in Alg. 2 are chosen so that ϵ-envy-freeness *w.r.t.* the sampled users and arms translates into $(\epsilon, \gamma, \lambda)$-envy-freeness. Combining these random approximation guarantees with Th. 1, we get:

Theorem 2. *Let* $\tilde{M} = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ *and* $K = \left\lceil \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))} \right\rceil$. *With probability* $1 - \delta$, *AUDIT is correct, it satisfies the conservative constraint* (3) *for all* $\tilde{M}$ *target users, and the bounds on duration and cost from Th. 1 (using* $\frac{\delta}{3\tilde{M}}$ *instead of* δ*) are simultaneously valid.*

Importantly, in contrast to naively using OCEF to compare all users against all, the audit for the probabilistic relaxation of envy-freeness only requires to query a constant number of users and policies that *does not depend on the total number of users M*. Therefore, the bounds on duration and cost are also independent of M, which is a drastic improvement.

5 Experiments

We present experiments describing sources of envy (Sec. 5.1) and evaluating the auditing algorithm OCEF on two recommendation tasks (Sec. 5.2).

We create a music recommendation task based on the Last.fm dataset from Cantador et al. (2011), which contains the music listening histories of 1.9k users. We select the 2500 items most listened to, and simulate ground truth user preferences by filling in missing entries with a popular matrix completion algorithm for implicit feedback data[4]. We also address movie recommendation with the MovieLens-1M dataset (Harper and Konstan 2015), which contains ratings of movies by real users, and from which we extract the top 2000 users and 2500 items with the most ratings. We binarize

---

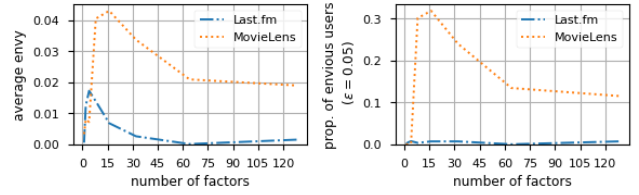[4]Using the Python library Implicit: https://github.com/benfred/implicit (MIT License).

Figure 2: Envy from model mispecification on MovieLens and Lastfm: envy is high when the latent factor model is mispecified, but it decreases as the number of factors increases.

ratings by setting those < 3 to zero, and as for Last.fm we complete the matrix to generate ground truth preferences.

For both recommendation tasks, the simulated recommender system estimates relevance scores using low-rank matrix completion (Bell and Sejnowski 1995) on a training sample of 20% of the ground truth preferences, where the rated / played items are sampled uniformly at random. Recommendations are given by a fixed-temperature *softmax* policy over the predicted scores. We generate binary rewards using a Bernoulli distribution with expectation given by our ground truth preferences.

5.1 Sources of envy

We consider two measures of the degree of envy. Denoting $\Delta^m = \max(\max_{n \in [M]} u^m(\pi^n) - u^m(\pi^m), 0)$, these are:

- the average envy experienced by users: $\frac{1}{M} \sum_{m \in [M]} \Delta^m$,

- the proportion of ϵ-envious users: $\frac{1}{M} \sum_{m \in [M]} \mathbb{1}_{\{\Delta^m > \epsilon\}}$.

Envy from model mispecification We demonstrate that envy arises from a standard recommendation model when the modeling assumptions are too strong. We vary the number of latent factors of the matrix completion model and evaluate a softmax policy with inverse temperature set to 5. In Fig. 2, with one latent factor we observe no envy. This is because all users receive the same recommendations since matrix completion is then equivalent to a popularity-based recommender system. With enough latent factors, preferences are properly captured by the model and the degree of envy decreases. For intermediate number of latent factors, envy is visible.

Envy from equal user utility We show that in contrast to envy-freeness, enforcing equal user utility (EUU) degrades user satisfaction and creates envy between users. We compute optimal EUU policies and unconstrained optimal policies (OPT) on the ground truth preferences of Last.fm and MovieLens. Our results in Table 1 confirm the pitfalls of EUU, while illustrating that OPT policies are always envy-free.

We discuss more sources of envy and provide the details of these computations in App. C.

5.2 Evaluation of the auditing algorithm

Our goal is now to answer for OCEF: in practice, what is the interplay between the required sample size per user, the cost of exploration and the conservative exploration parameter?

	Last.fm		MovieLens	
	EUU	OPT	EUU	OPT
Total utility	1552	**1726**	1671	**1761**
Average envy	0.10	**0**	0.04	**0**
Prop. 0.05-envious	0.61	**0**	0.13	**0**

Table 1: Optimal policies with equal user utility penalty (EUU) vs. Unconstrained optimal policies (OPT), computed on ground truth preferences: EUU deteriorates total utility and creates envy between users.
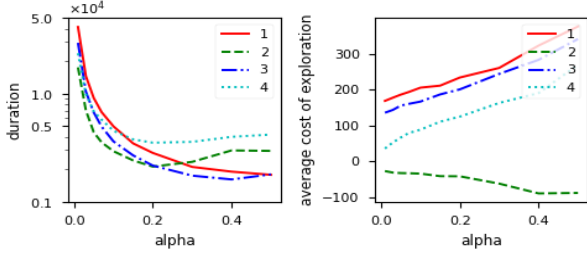

Figure 3: Effect of the conservative exploration parameter α on the duration and cost of auditing on Bandit experiments.

Bandit experiments We first study the trade-off between duration and cost of the audit on 4 bandit problems with Bernoulli rewards and 10 arms. In Problem 1, the baseline is the best arm and all other arms are equally bad. In Prob. 2, arm 1 is best and all other arms are as bad as the baseline. In Prob.3 the baseline is best and the means of arms from best to worst decrease rapidly. Prob. 4 uses the same means as Prob. 3, but the means of the baseline and arm 1 are swapped, making the baseline second-to-best. We set $\delta = \epsilon = 0.05$ and report results averaged over 100 trials. The details of the bandit configurations are given in Appendix D.1.

Figure 3 plots the duration and the cost of exploration (8) as a function of the conservative constraint parameter α (smaller α means more conservative). The curves show that for Problems 2, 3, and 4, duration is minimal for a non-trivial α. This is because when α is large, all arms are pulled as much as the baseline, so their confidence intervals are similar. When α decreases, the baseline is pulled more, which reduces the length of the relevant confidence intervals $\beta_0(t) + \beta_k(t)$ for *all* arms k. This, in turn, shortens the audit because non-baseline arms are more rapidly discarded or declared better. When α becomes too small, however, the additional pulls of the baseline have no effect on $\beta_0(t) + \beta_k(t)$ because it is dominated by $\beta_k(t)$, so the duration only increases. This subtle phenomenon is not captured by our analysis (Th. 1), because the ratios $\beta_0(t)/\beta_k(t)$ are difficult to track formally.

The sign of the cost of exploration depends on whether there is envy. In Prob. 2 where the baseline has the worst performance, exploration is beneficial to the user and so the cost is negative. On all other instances however, the cost is positive. The cost of exploration is closest to 0 when α becomes small because then $\beta_0(t) + \beta_k(t)$ is the smallest possible for a given number of pulls of k. For instance, in Prob. 4, the cost is close to 0 when α is very small and increases with
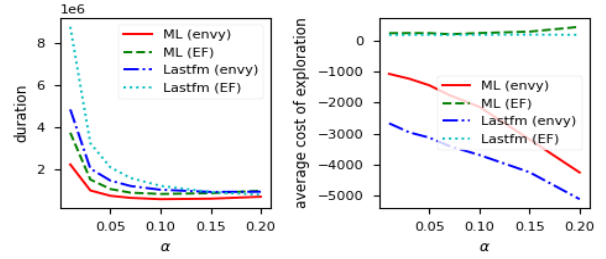

Figure 4: Scaling w.r.t. α on MovieLens (ML) and Last.fm, for recommender systems that are either envy-free (EF) or with envy. There are 41 target users and 75 arms.

α. It is the case where the baseline is not the best arm but is close to it, and there are many bad arms. When the algorithm is very conservative, bad arms are discarded rapidly thanks to the good estimation of the baseline performance. In this "low-cost" regime however, the audit is significantly longer.

Appendix D.1 contains additional results when varying the number of arms and the confidence parameter δ.

MovieLens and Last.fm experiments We now evaluate the certification of the (absence of) envy of recommendation policies on MovieLens (ML) and Last.fm. We consider two recommendation policies which are softmax functions over predicted relevance scores with inverse temperature set to either 5 or 10. These scores were obtained by matrix completion with 48 latent factors. On both datasets, with inverse temperature equal to 5, the softmax recommender system is envy-free, whereas there is envy when it is set to 10. We use AUDIT with OCEF to certify the probabilistic criterion. The envy parameters are set to $\epsilon = \delta = 0.05$ and $\lambda = \gamma = 0.1$, therefore we have $\tilde{M} = 41$ target users and $K = 75$ arms, independently on the number of users in each dataset.

The results of applying OCEF on each dataset (ML or Last.fm) with each policy (envy-free or with envy) are shown in Fig. 4. For the $(\epsilon, \gamma, \lambda)$-envy-free policies, results are averaged over 20 trials and over all the non-(ϵ, γ)-envious users, whereas when there is envy, results are averaged over the target users who are ϵ-envious. We observe clear tendencies similar to those of the previous section, although the exact sweet spots in terms of α depends on the specific configuration. In particular, on envy-free configurations, the cost of the audit is positive and grows when relaxing the conservative constraint, while it is negative and decreasing with α when there is envy. More details are provided in App. D.2.

6 Conclusion

We proposed the audit of recommender systems for user-side fairness with the criterion of envy-freeness. The auditing problem requires an explicit exploration of user preferences, which leads to a formulation as a bandit problem with conservative constraints. We presented an algorithm for this problem and analyzed its performance experimentally.

Acknowledgments

References

Asplund, J.; Eslami, M.; Sundaram, H.; Sandvig, C.; and Karahalios, K. 2020. Auditing race and gender discrimination in online housing markets. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 24–35.

Audibert, J.-Y.; and Bubeck, S. 2010. Best arm identification in multi-armed bandits.

Balcan, M.-F.; Dick, T.; Noothigattu, R.; and Procaccia, A. D. 2018. Envy-free classification. *arXiv preprint arXiv:1809.08700*.

Barocas, S.; Hardt, M.; and Narayanan, A. 2018. *Fairness and Machine Learning*. fairmlbook.org. http://www.fairmlbook.org.

Barocas, S.; and Selbst, A. D. 2016. Big Data's Disparate Impact. *Calif. L. Rev.*, 104: 671–769.

Bell, A. J.; and Sejnowski, T. J. 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6): 1129–1159.

Biega, A. J.; Gummadi, K. P.; Weikum, G.; and . 2018. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*, 405–414.

Bubeck, S.; Munos, R.; Stoltz, G.; and . 2009. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, 23–37. Springer.

Burke, R. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4): 331–370.

Burke, R. 2017. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093*.

Cantador, I.; Brusilovsky, P.; Kuflik, T.; and . 2011. 2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011). In *Proceedings of the 5th ACM conference on Recommender systems*, RecSys 2011. New York, NY, USA: ACM.

Celis, L. E.; Straszak, D.; and Vishnoi, N. K. 2017. Ranking with fairness constraints. *arXiv preprint arXiv:1704.06840*.

Chevaleyre, Y.; Endriss, U.; and Maudet, N. 2017. Distributed fair allocation of indivisible goods. *Artificial Intelligence*, 242: 1–22.

Corbett-Davies, S.; and Goel, S. 2018. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*.

Datta, A.; Tschantz, M. C.; Datta, A.; and . 2015. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1): 92–112.

Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214–226. ACM.

Ekstrand, M. D.; Tian, M.; Azpiazu, I. M.; Ekstrand, J. D.; Anuyah, O.; McNeill, D.; and Pera, M. S. 2018. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on Fairness, Accountability and Transparency*, 172–186. PMLR.

Foley, D. K. 1967. Resource allocation and the public sector.

Frank, M.; and Wolfe, P. 1956. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2): 95–110.

Freedman, D. A. 1975. On tail probabilities for martingales. *the Annals of Probability*, 100–118.

Garcelon, E.; Ghavamzadeh, M.; Lazaric, A.; and Pirotta, M. 2020. Improved algorithms for conservative exploration in bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3962–3969.

Geyik, S. C.; Ambler, S.; and Kenthapadi, K. 2019. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2221–2231.

Hannak, A.; Soeller, G.; Lazer, D.; Mislove, A.; and Wilson, C. 2014. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*, 305–318.

Harper, F. M.; and Konstan, J. A. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4): 1–19.

Hoeffding, W. 1963. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301): 13–30.

Hu, Y.; Koren, Y.; and Volinsky, C. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*, 263–272. Ieee.

Ilvento, C.; Jagadeesan, M.; and Chawla, S. 2020. Multi-category fairness in sponsored search auctions. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 348–358.

Imana, B.; Korolova, A.; Heidemann, J.; and . 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *Proceedings of the Web Conference 2021*, 3767–3778.

Jabbari, S.; Joseph, M.; Kearns, M.; Morgenstern, J.; and Roth, A. 2017. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 1617–1626. JMLR. org.

Jamieson, K.; Malloy, M.; Nowak, R.; and Bubeck, S. 2014. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, 423–439.

Jamieson, K.; and Nowak, R. 2014. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, 1–6. IEEE.

Johnson, C. C. 2014. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems*, 27(78): 1–9.

Joseph, M.; Kearns, M.; Morgenstern, J. H.; and Roth, A. 2016. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, 325–333.

Kano, H.; Honda, J.; Sakamaki, K.; Matsuura, K.; Nakamura, A.; and Sugiyama, M. 2019. Good arm identification via bandit feedback. *Machine Learning*, 108(5): 721–745.

Kim, M. P.; Korolova, A.; Rothblum, G. N.; and Yona, G. 2019. Preference-Informed Fairness. *arXiv preprint arXiv:1904.01793*.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37.

Kusner, M. J.; Loftus, J.; Russell, C.; and Silva, R. 2017. Counterfactual Fairness. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30*, 4069–4079. Curran Associates, Inc.

Lambrecht, A.; and Tucker, C. 2019. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science*, 65(7): 2966–2981.

Langford, J. 2005. Tutorial on Practical Prediction Theory for Classification. *Journal of machine learning research*, 6(3).

Liu, Y.; Radanovic, G.; Dimitrakakis, C.; Mandal, D.; and Parkes, D. C. 2017. Calibrated fairness in bandits. *arXiv preprint arXiv:1707.01875*.

Locatelli, A.; Gutzeit, M.; and Carpentier, A. 2016. An optimal algorithm for the thresholding bandit problem. *arXiv preprint arXiv:1605.08671*.

Mehrotra, R.; Anderson, A.; Diaz, F.; Sharma, A.; Wallach, H.; and Yilmaz, E. 2017. Auditing search engines for differential satisfaction across demographics. In *Proceedings of the 26th international conference on World Wide Web companion*, 626–633.

Patro, G. K.; Biswas, A.; Ganguly, N.; Gummadi, K. P.; and Chakraborty, A. 2020. FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms. In *Proceedings of The Web Conference 2020*, 1194–1204.

Raji, I. D.; Smart, A.; White, R. N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; and Barnes, P. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44.

Schölkopf, B.; and Smola, A. J. 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Sen, A. 1999. The possibility of social choice. *American economic review*, 89(3): 349–378.

Singh, A.; and Joachims, T. 2018. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2219–2228. ACM.

Suresh, H.; and Guttag, J. V. 2019. A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*.

Sweeney, L. 2013. Discrimination in online ad delivery. *Queue*, 11(3): 10.

Ustun, B.; Liu, Y.; Parkes, D.; and . 2019. Fairness without harm: Decoupled classifiers with preference guarantees. In *International Conference on Machine Learning*, 6373–6382.

Wang, M.; Gong, M.; Zheng, X.; and Zhang, K. 2018. Modeling dynamic missingness of implicit feedback for recommendation. *Advances in neural information processing systems*, 31: 6669.

Wu, Y.; Shariff, R.; Lattimore, T.; and Szepesvári, C. 2016. Conservative bandits. In *International Conference on Machine Learning*, 1254–1262.

Zafar, M. B.; Valera, I.; Rodriguez, M.; Gummadi, K.; and Weller, A. 2017. From parity to preference-based notions of fairness in classification. In *Advances in Neural Information Processing Systems*, 228–238.

		item cat. 1		item cat. 2			utilities	
(item idx)		1	2	3	4		u^1	u^2
(rewards)	ρ^1	1	0	0.8	0.7			
	ρ^2	0.8	0.7	1	0			
(policies)	$\pi^{1,\text{eq}}$	0.4	0	0.6	0		**0.88**	**0.92**
	$\pi^{2,\text{eq}}$	0.6	0	0.4	0		**0.92**	**0.88**

Table 2: Example where the optimal recommendations under item-side equity of exposure constraints are not user-side fair because both users envy each other. There are 4 items, 2 item categories and 2 users. User 1 envies user 2 since $u^1(\pi^{2,\text{eq}}) > u^1(\pi^{1,\text{eq}})$. Also, $u^2(\pi^{1,\text{eq}}) > u^2(\pi^{2,\text{eq}})$.

A   (In-)Compatibility of envy-freeness

A.1   Envy-freeness vs. optimality certificates

We showed in Section 3.3 that envy-freeness is compatible with optimal predictions. To understand the differences between a certificate of envy-freeness and a certificate of optimality, let us denote by $\Pi^* = \{\pi : \exists u \text{ satisfying } (1), \pi \in \arg\max_{\pi'} u(\pi')\}$ the set of potentially optimal policies. If the set of users policies approximately covers the set of potentially optimal policies Π^*, then an envy-free system is also optimal. Formally, let $D(\pi, \pi')$ such that $|u(\pi) - u(\pi')| \leq D(\pi, \pi')$. It is easy to see that if $\max_{\pi \in \Pi^*} \min_{m \in M} D(\pi, \pi^m) \leq \tilde{\epsilon}$, then ϵ-envy-freeness implies $\epsilon + \tilde{\epsilon}$-optimality.

In practice, the space of optimal policies is much larger than the number of users (for instance, there are $|\mathcal{A}|^{|\mathcal{X}|}$ optimal policies in our setting), so that auditing for envy is tractable in cases where auditing for optimality is not.

A.2   Envy-freeness vs. equity of exposure

We remind the definition of optimal policies with equity of exposure constraints from Section 3.3:

$$(\textit{equity}) \quad \pi^{m,\text{eq}}(.|x) = \underset{\substack{p:\mathcal{A}\to[0,1]\\\sum_a p(a)=1}}{\arg\max} \sum_{a\in\mathcal{A}} p(a)\rho^m(a|x)$$

$$\text{u.c. } \forall s \in [S], \quad \sum_{a\in\mathcal{A}_s} p(a) = \frac{\sum_{a\in\mathcal{A}_s}\rho^m(a|x)}{\sum_{a\in\mathcal{A}}\rho^m(a|x)}$$

The constraints should be ignored when $\sum_{a\in\mathcal{A}}\rho^m(a|x) = 0$.

Following Proposition 1 from Section 3.3, we describe here a second source of envy when using optimal policies with equity of exposure constraints. By the linearity of the optimization problem for $\pi^{m,\text{eq}}$, the policy assigns to the best item in a category the exposure of the entire category. It implies that categories with high average relevance have more exposure than categories with few but highly relevant items. Table 2 gives an example with two users and two categories of items where both users envy each other with the optimal recommendations under equity of exposure constraints.

In some degenerate cases though, equity of exposure policies are envy-free.

Lemma 3. *If for all contexts $x \in \mathcal{X}$, each user $m \in [M]$ only likes a single item category $\mathcal{A}_{s_m}$, i.e. $\forall a \in \mathcal{A} \setminus \mathcal{A}_{s_m}, \rho^m(a|x) = 0$, then the policies $(\pi^{m,\text{eq}})_{m=1}^M$ are envy-free.*

Proof. We set contexts x aside to simplify notation, but the generalization is straightforward.

We actually prove a stronger result than the lemma: if each user m only likes a single item, then $(\pi^{m,\text{eq}})_{m=1}^M = (\pi^{m,*})_{m=1}^M$, where $\pi^{m,*}$ is the optimal unconstrained policy for m.

Let $a_s^m = \arg\max_{a\in\mathcal{A}_s} \rho^m(a)$ be the favorite item in category $\mathcal{A}_s$ for user m, then the optimal equity of exposure constrained policies has the following analytical expression:

$$\forall s \in S, \forall a \in \mathcal{A}_s, \quad \pi^{m,\text{eq}}(a) = \mathbb{1}_{\{a=a_s^m\}} \frac{\sum_{a\in\mathcal{A}_s}\rho^m(a')}{\sum_{a'\in\mathcal{A}}\rho^m(a')},$$

and we thus have:

$$u^m(\pi^{m,\text{eq}}) = \sum_{s\in[S]} \rho^m(a_s^m) \frac{\sum_{a\in\mathcal{A}_s}\rho^m(a)}{\sum_{a\in\mathcal{A}}\rho^m(a)}. \quad (5)$$

If each user $m \in [M]$ only likes a single item category $s_m \in [S]$, i.e. $\forall a \in \mathcal{A}\setminus\mathcal{A}_{s_m}, \rho^m(a) = 0$, then $\frac{\sum_{a\in\mathcal{A}_s}\rho^m(a)}{\sum_{a\in\mathcal{A}}\rho^m(a)} = \mathbb{1}_{\{s=s_m\}}$.

Then $u^m(\pi^{m,\text{eq}}) = \rho^m(a_{s_m}^m) = \max_{a\in\mathcal{A}}\rho^m(a)$.

Then $\pi^{m,\text{eq}}$ is the optimal unconstrained policy for user m, meaning the whole system is envy-free (cf. Sec 3.2).

From Eq. 5, we actually note that $(\pi^{m,\text{eq}})_{m=1}^M = (\pi^{m,*})_{m=1}^M$ if and only if each user m equally values their favorite items in each category they like, i.e. $\forall m, \exists \kappa > 0, \forall s \in S, \rho^m(a_s^m) > 0 \Rightarrow \rho^m(a_s^m) = \kappa$. $\square$

B   Extension to group envy-freeness

We briefly discuss an extension of envy-free recommendation to groups, since most of the literature on fair machine learning focuses on systematic differences between groups. Certifying envy-freeness at the level of groups rather than individuals also relaxes the criterion because it requires less exploration. Let us assume we are given a partition G of the users into disjoint groups. For $g, g' \in G$, we define the group utility of g with respect to g' as:

$$U(g, g') = \frac{1}{|g|} \sum_{m\in g} u^m\left(\frac{1}{|g'|}\sum_{n\in g'}\pi^n\right). \quad (6)$$

Definition B.1. *Given $\epsilon \geq 0$, the recommender system is ϵ-group-envy-free if:* $\forall g, g' \in G, \quad U(g, g') \leq U(g, g) + \epsilon$.

Group envy-freeness is equivalent to envy-freeness when each group is a singleton. When we have prior knowledge that user preferences and policies are homogeneous within each group, ϵ-envy-freeness translates to ϵ'-group envy-freeness, with $\epsilon' \approx \epsilon$, and the reciprocal is also true:

Proposition 2. *Let $\epsilon, \tilde{\epsilon} > 0$, and assume that for all groups and all pairs of users m, n in the same group g, we have $\sup_{x \in \mathcal{X}} \|\pi^m(.|x) - \pi^n(.|x)\|_1 \leq \tilde{\epsilon}$ and $\sup_{x \in \mathcal{X}} \|\rho^m(.|x) - \rho^n(.|x)\|_1 \leq \tilde{\epsilon}$. Then, ϵ-group envy-freeness implies $(\epsilon + 4\tilde{\epsilon})$-envy-freeness.*

The result is natural since when all groups have users with homogeneous preferences and policies, groups and users are a similar entity as regards the assessment of envy-freeness. The proof is straightforward and omitted. When groups have heterogeneous policies, the "average policy" $\frac{1}{|g|}\sum_{n \in g}\pi^n$ is uninformative because it does not represent any user's policy. Defining a notion of group utility in the general case is thus nontrivial and left for future work.

C   Sources of envy

In this section, we first list a few possible sources of envy in recommender systems. Then we provide the details of experiments[5] which showcase one of these sources, namely model mispecification (App. C.2).

C.1   Examples of sources of envy

Model mispecification   Recommender systems often rely on strong modeling assumptions and multi-task learning, with methods such as low-rank matrix factorization (Koren, Bell, and Volinsky 2009). The limited capacity of the models (e.g., a rank that is too low) or incorrect assumptions might leave aside users with less common preference patterns. Appendix C.2 gives a more detailed example on two simulated recommendation tasks.

Misaligned incentives   A recommender system might have incentives to recommend some items to specific users, e.g., sponsored content. Envy appears when there is a mismatch between users who like these items and users to whom they are recommended.

Measurement bias   Many hybrid recommender systems rely on user interactions together with user-side data (Burke 2002). This includes side-information such as browsing history on third-party, partner websites. Envy arises in these settings if there is measurement bias (Suresh and Guttag 2019), e.g., if the side information is unevenly collected for all users (e.g., browsing patterns are different across users and partners are aligned with the patterns of a user groups only).

Operational constraints   Regardless of incentives, recommendations might need to obey additional constraints. As described in Proposition 1, the item-side fairness constraint of equity of exposure is an example of possible source of (user-side) envy. The user-side fairness constraint of equal utility also creates envy, as we showed in Sec. 5.1.

In the following, we provide the details of our experiments from Sec. 5.1 where we showcase examples of environments with envy based on movie and music recommendation tasks.

In these experiments, we measure envy based on the quantity:

$$\Delta^m = \max\left(\max_{n \in [M]} u^m(\pi^n) - u^m(\pi^m), 0\right)$$

In line with (Chevaleyre, Endriss, and Maudet 2017), we consider two ways of measuring the degree of envy:

- the average envy experienced by users: $\frac{1}{M}\sum_{m \in [M]}\Delta^m$,

- the proportion of ϵ-envious users: $\frac{1}{M}\sum_{m \in [M]}\mathbb{1}_{\{\Delta^m > \epsilon\}}$.

C.2   Setup of the experiments on envy from model mispecification

We describe in this section the details of the experiments on envy from mispecification presented in Section 5.1. We used Lastfm-2k (Cantador et al. 2011), a dataset from the online music service Last.fm[6] which contains real play counts of $2k$ users for $19k$ artists, and was used by Patro et al. (2020) who also study envy-freeness as a user-side fairness criterion. We filter the top $2,500$ items most listened to. Following (Johnson 2014), we pre-process the raw counts with log-transformation. We split the dataset into train/validation/test sets, each including $70\%/10\%/20\%$ of the user-item listening counts. We create three different splits using three random seeds. We estimate relevance scores for the whole user-item matrix using the standard matrix factorization algorithm[7] of Hu, Koren, and Volinsky (2008) trained on the train set, with hyperparameters selected on the validation set by grid search. The number of latent factors is chosen in $[16, 32, 64, 128]$, the regularization in $[0.01, 0.1, 1., 10.]$, and the confidence weighting parameter in $[0.1, 1., 10., 100.]$. The resulted matrix of estimated relevance scores serves as the ground truth preferences.

We also address movie recommendation using the MovieLens-1M dataset (Harper and Konstan 2015), which contains 1 million ratings on a 5-star scale from approximately 6000 users and 4000 movies. We extract a 2000×2500 user $\times$ items matrix, keeping users and items with the most rating. We transform MovieLens ratings into an implicit feedback dataset similar to Last.fm. Since setting ratings < 3 are usually considered as negative (Wang et al. 2018), we set ratings < 3 to zero, resulting in a dataset with preference values among $\{0, 3, 3.5, 4, 4.5, 5\}$. We then use the same algorithm as for Last.fm to obtain relevance scores that we use to simulate ground truth preferences.

We then simulate a recommender system's estimation of preferences using low-rank matrix completion (Bell and Sejnowski 1995) on a training sample of 70% of the whole "ground truth" preferences, with hyperparameter selection on a 10% validation sample. Here, the regularization is chosen

---

[5]For all our experiments, we used Python and a machine with Intel Xeon Gold 6230 CPUs, 2.10 GHz, 1.3 MiB of cache.

[6]http://www.lastfm.com

[7]Using the Python library Implicit: https://github.com/benfred/implicit (MIT License).

in $[0.001, 0.01, 0.1, 1.]$, and the confidence weighting parameter in $[0.1, 1., 10., 100.]$. The estimated preference scores are given as input to the recommendation policies.

The recommendation policies we consider are softmax distributions over the predicted scores with fixed inverse temperature. These policies recommend a single item, drawn from the softmax distribution.

We generate binary rewards using a Bernoulli distribution with expectation given by our ground truth. We consider no context in these experiments, so that the policies and rewards only depend on the user and the item.

Figure 2 in Sec. 5.1 was generated by varying the number of latent factors in the recommender system's preference estimation model. For each number of latent factors in the range $[1, 2, 4, 8, 16, 32, 64, 128, 256]$, a new model was trained on the train set with hyperparameter selection on the validation set. The degrees of envy are measured on the whole ground truth preference matrix.

C.3 Envy from equal user utility constraints

We provide the full details of the experiments on envy from equal user utility presented in Sec. 5.1 from the main paper. The goal of these experiments is to show that in contrast to envy-freeness, enforcing equal user utility (EUU) degrades user satisfaction and creates envy between users. We remind from Sec. 3.2 that the fairness constraint of EUU is defined as:

$$\forall m, n \in [M], u^m(\pi^m) = u^n(\pi^n),$$

or equivalently:

$$\forall m \in [M], u^m(\pi^m) = \frac{1}{M} \sum_{n \in [M]} u^n(\pi^n).$$

Equal user utility is enforced by adding a penalty to the maximization of user utilities. Optimal EUU policies are found by maximizing the following concave objective function, where the parameter $b > 0$ controls the strength of the penalty:

$$(EUU) \quad \pi_b^{\text{euu}} = \operatorname*{argmax}_{\substack{p: \mathcal{A} \to [0,1]^M \\ \forall m, \sum_a p^m(a)=1}} \sum_{m \in [M]} u^m(p^m) - b\sqrt{D(p)}$$

$$\text{with} \quad D(p) = \sum_{m \in [M]} \left(u^m(p^m) - \frac{1}{M} \sum_{n \in [M]} u^n(p^n) \right)^2.$$

$$(7)$$

We infer EUU policies using the Frank-Wolfe algorithm (Frank and Wolfe 1956) with the ground truth preferences given as input. The parameter of the penalty is set to $b = 50$. We also generate the unconstrained optimal policies (OPT) based on the ground truth (recall that these are $u^m(\pi^{m,*}) = \max_\pi u^m(\pi) \geq u^m(\pi^{n,*})$).

A comparison of EUU and OPT is provided in Table 1 in Sec. 5.1, with the following evaluation measures : total utility (higher is better), average envy and proportion of 0.05-envious users (lower is better). The results on both dataset confirm the claim that enforcing EUU penalties deteriorates total utility and creates envy between users, while illustrating the known property that OPT policies are compatible with envy-freeness.
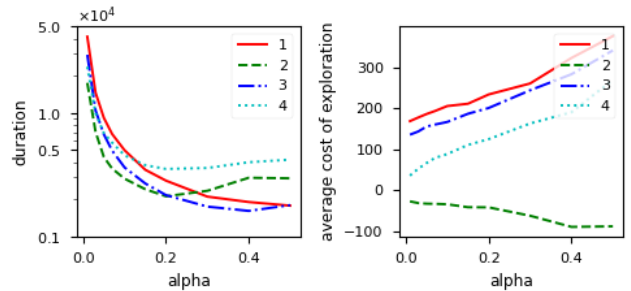

Figure 5: Effect of the conservative exploration parameter α on the duration and cost of auditing on Bandit experiments.

D OCEF experiments

D.1 Bandit experiments

We performed experiments on toy bandit environments to assess the performance of our algorithm OCEF on various configurations, which were also considered in (Jamieson and Nowak 2014). The four bandits instances have 10 arms. They are Bernoulli variables with means equal to

1) $\mu_0 = 0.6$ and $\mu_k = 0.3$ for $k \in [9]$,
2) $\mu_0 = 0.3$, $\mu_1 = 0.6$ and $\mu_k = 0.3$ for $k = 2..9$,
3) $\mu_k = 0.7 - 0.7 * \left(\frac{k}{10}\right)^{0.6}, k = 0, ..., 9$, and the baseline is μ_0,
4) same as 3), but permuting μ_0 and μ_1.

Fig. 5 shows the result of applying OCEF on the various configurations, where we set $\delta = \epsilon = 0.05$ and report results averaged over 100 trials. We observe clear tendencies similar to those presented in Section 5.2, although the exact sweet spots in terms of α depends on the specific configuration.

The cost of exploration follows similar patterns as in in Section 5.2. In Prob. 2, the baseline has the worst performance, so exploration is beneficial to the user and the cost is negative. On the other hand, for instance in Prob. 4, the cost is close to 0 when α is very small and increases with α. It is the case where the baseline is not the best arm but is close to it, and there are many bad arms. When the algorithm is very conservative, bad arms are discarded rapidly thanks to the good estimation of the baseline performance. In this "low-cost" regime however, the audit is significantly longer.

We show additional results when varying δ in Figure 6. Results are averaged over 100 simulations and the conservative exploration parameter is set to $\alpha = 0.05$. The duration decreases as δ increases, i.e. a lower confidence certificate requires fewer samples per user. The duration for Problem 1 is longer than for the other instances. This is because with α set to 0.05 and the baseline mean being much higher than non-baseline arms, the conservative constraint 3 enforces many pulls of the baseline, since each exploration round is very costly. As a consequence, too little data is collected on the non-baseline arms to conclude that they are below $\mu_0 + \epsilon$. Since all non-baseline arms have equal means, the size of the active set remains the same for a long time, while in Problem 3, where the baseline is also the best arm, arms are eliminated one at a time.
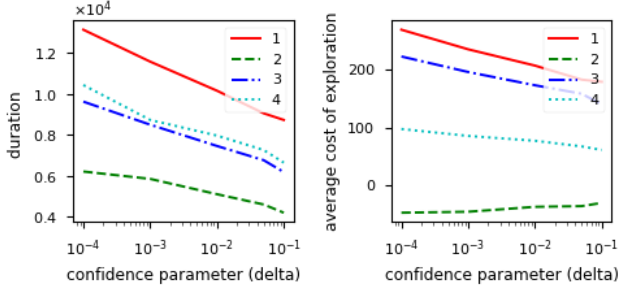
Figure 6: Effect of the confidence parameter δ on the duration and cost on 4 different bandit instances.
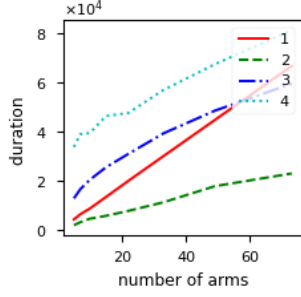

Figure 7: Effect of the number of arms on the duration on 4 different bandit instances.

We show how OCEF scales with the number of arms in Figure 7, for fixed values $\alpha = \delta = \epsilon = 0.05$. We set $K_{max} = 100$ and define 4 instances as in the list above, except that $K = K_{max}$ instead of $K = 9$. We run OCEF on the instances $\mu_{0:K'}$ and vary the value of $K' \leq K_{max}$. The duration increases for all problems, and the slope depends on the gaps between μ_0 and the μ_k.

D.2 Setup of the MovieLens and Last.fm experiments

We now provide additional details on the experimental evaluation of OCEF on MovieLens and Last.fm presented in Sec. 5.2. The protocole to generate the recommendation task is the same as the one described in App. C for the experiments on sources of envy. The policies are softmax distributions over scores predicted by the matrix factorization model with a number of factors equal to 48.

In these experiments, the auditor interacts with the audited users. Rewards are drawn from Bernoulli distributions with expectation equal to the ground truth preferences.

Two recommendation policies are audited. The first one is a softmax with inverse temperature equal to 5. Since the inverse temperature is small, the softmax distribution is closer to random, which means users get more similar recommendations: the recommender system is thus envy-free. The second one is a softmax with inverse temperature equal to 15. With higher inverse temperature, the distribution is more peaked, which exacerbates differences between policies. Since the

model with 48 factors is mispecified (see Sec.5.1), envy is visible.

E Proofs

E.1 Theoretical results

Useful lemmas Recall that OCEF considers a single audited group m, therefore we do not use superscripts m in the following (e.g., $\mu_k, r_t...$).

The algorithm relies on valid confidence intervals. As in (Jamieson et al. 2014), we use anytime bounds inspired by the law of the iterated algorithm (LIL), and a union bound.

We say that a random variable is σ-subgaussian if it is subgaussian with variance proxy σ^2. Since we assume the rewards for each user are bounded, more precisely $r_t \in [0, 1]$, they are $\frac{1}{2}$-subgaussian.

Throughout the paper, we assume that rewards for each user are independent conditionally to the arm played.

Lemma 4. *Let $\delta \in (0, 1)$. Assume the rewards are σ-subgaussian.*

$$Let\ \omega \in (0, 1), \quad \theta = \log(1 + \omega)\left(\frac{\omega \delta}{2(2+\omega)}\right)^{\frac{1}{1+\omega}}.$$

$$Let \quad N_k(t) = \sum_{s=1}^{t} \mathbb{1}_{\{k_s = k\}} \quad \widehat{\mu}_k(t) = \frac{\sum_{s=1}^{t} r_s \mathbb{1}_{\{k_s = k\}}}{N_k(t)}$$

$$\beta_k(t) = \sqrt{\frac{2\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)}{N_k(t)}}$$

$$\times \sqrt{\log\left(\frac{2(K + 1)}{\theta} \log((1 + \omega)N_k(t))\right)}$$

$$\underline{\mu}_k(t) = \widehat{\mu}_k(t) - \beta_k(t) \qquad \overline{\mu}_k(t) = \widehat{\mu}_k(t) + \beta_k(t)$$

Then,

$$\mathbb{P}\left[\forall t > 0, \forall k \in [K], \ \mu_k \in [\underline{\mu}_k(t); \overline{\mu}_k(t)]\right] \geq 1 - \frac{\delta}{2}.$$

Notice that the choice of θ makes sure that β_k is well defined as long as $N_k(t) > 0$. We use the convention that when $N_k(t) = 0$, $\beta_k(t)$ is strictly larger than when $N_k(t) = 1$ to ensure β_k is strictly decreasing with N_k. Also, when $N_k(t) = 0$, we set $\widehat{\mu}_k(t) = 0$.

Following (Garcelon et al. 2020), our lower bound on the conservative constraint relies on Freedman's martingale inequality (Freedman 1975).

Lemma 5. *Assume all rewards are σ-subgaussian. Let $A_t = \{s \leq t : k_s \neq 0\}$ be the number of times a non-baseline arm $k \neq 0$ has been pulled up to time t. Let*
$$\phi(t) = \sigma\sqrt{2|A_{t-1}|\log\left(\frac{6|A_{t-1}|^2}{\delta}\right)} + \frac{2}{3}\log\left(\frac{6|A_{t-1}|^2}{\delta}\right).$$
Then, $\forall \delta > 0$,

$$\mathbb{P}\left[\forall t > 0, \left|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)\right| \leq \phi(t)\right] \geq 1 - \frac{\delta}{2}.$$

As in Lemma 4, we use the convention $\phi(t) = 0$ when $|A_{t-1}| = 0$.

Lemma 6. *Let $\delta \in (0,1)$.*

Let $\Phi(t) = \min\left(\sum_{k=1}^{K} \beta_k(t-1)N_k(t-1), \phi(t)\right)$, with $\phi(t)$ defined in Lemma 5. Let $\mathcal{E}$ be the event under which all confidence intervals are valid, i.e.:

$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$ *with*

$\mathcal{E}_1 = \left\{ \forall k \in \{0, \ldots, K\}, \forall t > 0, \mu_k(t) \in [\underline{\mu}_k(t); \overline{\mu}_k(t)] \right\}$

$\mathcal{E}_2 = \left\{ \forall t > 0, \left| \sum_{s \in A_{t-1}} (\mu_{k_s} - r_s) \right| \leq \Phi(t) \right\}.$

Then $\mathbb{P}\left[\mathcal{E}\right] \geq 1 - \delta$.

Proof. By Lemma 4, $\mathbb{P}\left[\mathcal{E}_1\right] \geq 1 - \frac{\delta}{2}$. By the lemma above, with probability $1 - \frac{\delta}{2}$, we have for all $t > 0$, $\left|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)\right| \leq \phi(t)$.

Then, notice that

$$\left| \sum_{s \in A_{t-1}} (\mu_{k_s} - r_s) \right| = \left| \sum_{k=1}^{K} N_k(t-1)(\mu_k - \widehat{\mu}_k(t-1)) \right|.$$

Hence under $\mathcal{E}_1$ we also have:

$$\left| \sum_{s \in A_{t-1}} (\mu_{k_s} - r_s) \right| \leq \sum_{k=1}^{K} N_k(t-1)\beta_k(t-1).$$

Therefore,

$$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 = \mathcal{E}_1 \cap \left\{ \left| \sum_{s \in A_{t-1}} (\mu_{k_s} - r_s) \right| \leq \phi(t) \right\},$$

and thus, by a union bound, we have: $\mathbb{P}\left[\mathcal{E}\right] \geq 1 - \delta$. $\qquad\square$

Theorems We now provide our complete theoretical guarantees for correctness (Theorem 7), duration (Theorem 8) and cost (Theorem 9), which we then prove in App. E.2 and E.3. From these results, we derive Theorem 1 in the main paper, which we prove in App. E.4.

Theorem 7 (Correctness)**.** *With probability at least $1 - \delta$:*

1. *OCEF satisfies the safety constraint (3) at every time step,*
2. *if OCEF outputs ϵ-no-envy then the user m is not ϵ-envious, and if it outputs* envy*, then m is envious.*

We denote $\log^+(.) = \max(1, \log(.))$.

Theorem 8 (Duration)**.** *Let $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$, $\delta \in (0,1)$, $\theta = \log(2)\sqrt{\frac{\delta}{6}}$, and*

$$\forall k \neq 0, \quad H_k = 1 + \frac{64}{\eta_k^2} \log\left(\frac{2(K+1)\log^+\left(\frac{128(K+1)}{\theta \eta_k^2}\right)}{\theta} \right),$$

$$H_0 = \max\left(\max_{k \in [K]} H_k, \frac{6K+2}{\alpha \mu_0} \right.$$
$$\left. + \sum_{k=1}^{K} \frac{256 \log\left(\frac{2(K+1)\log(2H_k)}{\theta} \right)}{\alpha \mu_0 \eta_k} \right).$$

With probability at least $1 - \delta$, OCEF stops in at most τ steps, with

$$\tau \leq \sum_{k=0}^{K} H_k.$$

Finally, we define the *cost of exploration* as the potential reward lost because of exploration actions, in our case the cumulative reward lost, on average over users in the group:

$$C_t = t\mu_0 - \sum_{s=1}^{t} \mu_{k_s}. \tag{8}$$

In the worst case, the following bound holds:

Theorem 9 (Cost of exploration)**.** *Under the assumptions and notation of Theorem 8, let τ be the time step where OCEF stops. With probability $1 - \delta$, we have:*

$$C_\tau \leq \sum_{k:\mu_k < \mu_0} (\mu_0 - \mu_k)H_k \tag{9}$$

Certification of the exact criterion for all users The audit of the full system for the exact envy-freeness criterion consists in running OCEF for every user. Since we are making multiple tests, we need to use a tighter confidence parameter for each user so that the confidence intervals simultaneously hold for all users.

Corollary 9.1 (Online certification)**.** *With probability at least $1 - \delta$, running OCEF simultaneously for all M users, each with confidence parameter $\delta' = \frac{\delta}{M}$, we have:*

1. *for all $m \in [M]$ OCEF satisfies the constraints (3),*
2. *all users for which OCEF returns ϵ-NO ENVY are not ϵ-envious of any other users, and all users for which OCEF returns ENVY are envious of another user.*
3. *For every user, the bounds on the duration of the experiment and the cost of exploration given by Theorems 8 and 9 (using δ/M instead of δ) are simultaneously valid.*

For the certification of the probabilistic envy-freeness criterion, we refer to Theorem 2 in the main paper, which we prove in App. E.5.

E.2 Proof of Theorem 7

Proof. We assume that event $\mathcal{E}$ holds true. Then all confidence intervals are valid, i.e., for all $k = 0, \ldots, K$, $\underline{\mu}_k(t) \leq \mu_k \leq \overline{\mu}_k(t)$, and $\sum_{s \in A_{t-1}} \mu_{k_s} \geq \sum_{s \in A_{t-1}} r_s - \Phi(t)$.

Let Z_t be the safety budget, defined as $Z_t = \sum_{s=1}^{t} \mu_{k_s} - (1-\alpha)\mu_0 t$, so that the conservative constraint (3) is equivalent to $\forall t, Z_t \geq 0$. We have $Z_t = \sum_{s \in A_{t-1}} \mu_{k_s} + \mu_{k_t} + (N_0(t-1) - (1-\alpha)t)\mu_0$. Therefore, ξ_t (eq. (4)) is a lower bound on the safety budget Z_t if ℓ_t is played. By construction of the algorithm, the safety constraint (3) is immediately satisfied since a pull that could violate it is not permitted.

By the validity of confidence intervals under $\mathcal{E}$, if OCEF stops because of the first condition, then $\exists k, \mu_k > \mu_0$. Therefore 0 is not ϵ-envious of k and OCEF is correct.

If OCEF stops because of the second condition, i.e., $\forall k, \overline{\mu}_k(t) \leq \underline{\mu}_0(t) + \epsilon$, then $\forall k, \mu_k \leq \mu_0 + \epsilon$. Therefore 0 is not envious and OCEF is correct.

Since $\mathbb{P}\left[\mathcal{E}\right] \geq 1 - \delta$, OCEF satisfies the safety constraint and is correct with probability $\geq 1 - \delta$.

$\qquad\square$

E.3 Proofs of Theorem 8 and Theorem 9

Notation For conciseness, we use $\tilde{K} = K + 1$, and

$$\psi_k(t) = 2\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)\log\left(\frac{2\tilde{K}}{\theta}\log((1 + \omega)N_k(t))\right),$$

so that $\beta_k(t) = \sqrt{\dfrac{\psi_k(t)}{N_k(t)}}$.

We shall also use $\Gamma_\omega = 2\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)$. We use the convention $\psi_k(t) = 0$ when $N_k(t) = 0$, and set $\beta_k(t)$ to some value strictly larger than when $N_k(t) = 1$.

We remind that $\omega \in (0, 1)$, $\quad \theta = \log(1 + \omega)\left(\frac{\omega\delta}{2(2+\omega)}\right)^{\frac{1}{1+\omega}}$ and $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$. We denote by $\eta_{\min} = \min_{k \in [K]} \eta_k$.

Finally, we notice that under event $\mathcal{E}$ (as defined in Sec. E.1), we have for all $k \in \{0, \ldots, K\}$ and all t:

$$\mu_k + 2\beta_k(t) \geq \overline{\mu}_k(t) \geq \mu_k \geq \underline{\mu}_k(t) \geq \mu_k - 2\beta_k(t). \quad (10)$$

Lemma 10. *Under event $\mathcal{E}$, for every $k \in [K]$, if k is pulled at round t, then $4\beta_k(t) \geq \eta_k$.*

Proof of Lemma 10. Since k is pulled at t, the two following inequalities hold:

$$\overline{\mu}_k(t - 1) > \underline{\mu}_0(t - 1) + \epsilon \quad (11)$$

$$\underline{\mu}_k(t - 1) \leq \overline{\mu}_0(t - 1) \quad (12)$$

We prove them by contradiction. If (11) does not hold, then k should be discarded from the active set at time $t - 1$, and therefore cannot be pulled at t. Likewise, if (12) does not hold, then the algorithm stops at $t - 1$, so k cannot be pulled at t.

Using (11) and (10), we have:

$$\mu_k + 2\beta_k(t-1) \geq \overline{\mu}_k(t-1) > \underline{\mu}_0(t-1) + \epsilon \geq \mu_0 - 2\beta_0(t-1) + \epsilon.$$

Since 0 was not pulled at time t, we also have $\beta_0(t - 1) \leq \beta_k(t - 1)$, hence $4\beta_k(t - 1) \geq \mu_0 + \epsilon - \mu_k$.

Using (12) and (10) we have $\mu_k - 2\beta_k(t) \leq \mu_0 + 2\beta_0(t)$ and since $\beta_0(t) \leq \beta_k(t)$, we obtain $4\beta_k(t - 1) \geq \mu_k - \mu_0$. $\quad\square$

In the following lemma, we recall that we denote $\log^+(.) = \max(1, \log(.))$.

Lemma 11. *Under event $\mathcal{E}$, $\forall \tau > 0, \forall k \in [K]$, we have*

$$N_k(\tau) \leq H_k \quad \text{with}$$

$$H_k = 1 + \frac{32\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)}{\eta_k^2} \times$$

$$\log\left(\frac{2(K + 1)\log^+\left(\frac{64(K+1)\sigma^2(1+\sqrt{\omega})^2(1+\omega)^2}{\theta\eta_k^2}\right)}{\theta}\right)$$

Proof. Let $\tau > 0$, $k \in [K]$, and let $t \leq \tau$ be last time step before τ at which k was pulled. If such a t does not exist, then $N_k(\tau) = 0$ and the result holds. In all cases, we have $N_k(t) = N_k(\tau)$.

We consider $t > 0$ from now on.

By Lemma 10, we have $4\beta_k(t-1) \geq \eta_k$, and thus $N_k(t-1) \leq \frac{16\psi_k(t-1)}{\eta_k^2}$, which writes, if $N_k(t) > 0$:

$$N_k(t - 1) \leq \frac{16\psi_k(t - 1)}{\eta_k^2}$$

$$\leq \frac{16\Gamma_\omega}{\eta_k^2}\log\left(\frac{2\tilde{K}}{\theta}\log\left((1 + \omega)N_k(t - 1)\right)\right).$$

$$(13)$$

Using $\frac{1}{t}\log\left(\frac{\log((1+\omega)t)}{\Omega}\right) \geq c \Rightarrow t \leq \frac{1}{c}\log\left(\frac{\log((1+\omega)/c\Omega)}{\Omega}\right)$ (see Equation (1) in (Jamieson et al. 2014)) with $\Omega = \frac{\theta}{2\tilde{K}}$ and $c = \frac{\eta_k^2}{16\Gamma_\omega}$, we obtain

$$N_k(t - 1) \leq \frac{16\Gamma_\omega}{\eta_k^2}\log\left(\frac{2\tilde{K}}{\theta}\log\left(\frac{(1 + \omega)32\tilde{K}\Gamma_\omega}{\theta\eta_k^2}\right)\right) \quad (14)$$

Since $N_k(t) = N_k(t - 1) + 1$, using $\log^+$ instead of $\log$ inside to deal with the case $N_k(t - 1) = 0$ gives the desired result. $\quad\square$

Lemma 12. *Under event $\mathcal{E}$, at every time step τ, we have*

$$N_0(\tau) \leq \max\left(\max_{k \in [K]} H_k, \frac{6K + 2}{\alpha\mu_0}\right.$$

$$\left. + \sum_{k=1}^{K}\frac{64\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)\log\left(\frac{2(K+1)\log((1+\omega)H_k)}{\theta}\right)}{\alpha\mu_0\eta_k}\right)$$

Proof. Let $\tau > 0$ and $t \leq \tau$ the last time 0 was pulled before τ. We assume $t > 0$.

Case 1: 0 was pulled because $\beta_0(t-1) > \min_{k \in [K]} \beta_k(t-1)$.
Then $N_0(\tau) = N_0(t - 1) + 1 \leq 1 + \max_{k \neq 0} N_k(t - 1)$.

By lemma 10, we thus have $N_0(\tau) \leq \max_{k \in [K]} H_k$.

Case 2: 0 was pulled because $\xi_t < 0$. Here the proof follows similar steps as that of Theorem 5 in (Wu et al. 2016).

$$\sum_{s \in A_{t-1}} r_s - \Phi(t) + \underline{\mu}_{\ell_t}(t-1)$$

$$+ (N_0(t-1) - (1 - \alpha)t)\overline{\mu}_0(t-1) < 0$$

We drop $\underline{\mu}_{\ell_t}(t-1)$, replace t by $\sum_{k=0}^{K} N_k(t - 1) + 1$ and rearrange terms to obtain:

$$\alpha N_0(t - 1)\overline{\mu}_0(t - 1) \leq (1 - \alpha)\overline{\mu}_0(t - 1)$$

$$+ (1 - \alpha)\sum_{k=1}^{K} N_k(t - 1)\overline{\mu}_0(t - 1) - \sum_{s \in A_{t-1}} r_s + \Phi(t)$$

$$(15)$$

Since we have $\beta_0(t - 1) \leq \beta_k(t - 1)$ (otherwise we would be in case 1), and $A_{t-1} = \sum_{k=1}^{K} N_k(t - 1)$, we bound the

the sum over arms in (15):

$$\sum_{k=1}^{K} N_k(t-1)\overline{\mu}_0(t-1)$$

$$\leq \sum_{k=1}^{K} N_k(t-1)(\mu_0 + 2\beta_0(t-1))$$

$$\leq \sum_{k=1}^{K} N_k(t-1)(\mu_0 + 2\beta_k(t-1))$$

$$= \sum_{s \in A_{t-1}} \mu_0 + \sum_{k=1}^{K} 2\beta_k(t-1)N_k(t-1).$$

Using Lemma 6, we also bound $-\sum_{s \in A_{t-1}} r_s \geq \sum_{s \in A_{t-1}} \mu_s + \Phi(t)$ (under $\mathcal{E}$).

Plugging this into (15) gives:

$$\alpha N_0(t-1)\overline{\mu}_0(t-1) \leq (1-\alpha)\overline{\mu}_0(t-1)$$

$$+ 2(1-\alpha)\sum_{k=1}^{K} N_k(t-1)\beta_k(t-1)$$

$$+ \sum_{s \in A_{t-1}} ((1-\alpha)\mu_0 - \mu_{k_s}) + 2\Phi(t).$$

Recall that $\Phi(t) = \min(\sum_{k=1}^{K} N_k(t-1)\beta_k(t-1), \phi(t))$, and therefore $\Phi(t) \leq \sum_{k=1}^{K} N_k(t-1)\beta_k(t-1)$.

Using $\mu_0 - \mu_{k_s} \leq \eta_{k_s}$ and $\sum_{s \in A_{t-1}} \eta_{k_s} = \sum_{k=1}^{K} N_k(t-1)\eta_k$, we obtain:

$$\alpha N_0(t-1)\overline{\mu}_0(t-1) \leq (1-\alpha)\overline{\mu}_0(t-1)$$

$$+ \sum_{k=1}^{K} \Big((\eta_k - \alpha\mu_0)N_k(t-1)$$

$$+ 4\sqrt{\Psi_k(t-1)N_k(t-1)} \Big).$$

We bound $f_k := (\eta_k - \alpha\mu_0)N_k(t-1) + 4\sqrt{\Psi_k(t-1)N_k(t-1)}$.

Since (13) $N_k(t-1) \leq \frac{16\psi_k(t-1)}{\eta_k^2} + 1$, and $\eta_k - \alpha\mu_0 \leq \eta_k$, we have

$$f_k \leq \frac{16\psi_k(t-1)}{\eta_k} + \eta_k + 4\sqrt{\frac{16\psi_k(t-1)^2}{\eta_k^2} + \psi_k(t-1)}$$

Using $\sqrt{(\frac{x}{z})^2 + x} \leq \frac{x}{z} + \frac{z}{2}$ for $x \geq 0, z > 0$, with $x = 4\psi_k(t-1)$ and $z = \eta_k$, we obtain:

$$f_k \leq \frac{16\psi_k(t-1)}{\eta_k} + \frac{16\psi_k(t-1)}{\eta_k} + 3\eta_k$$

$$\leq \frac{32\psi_k(t-1)}{\eta_k} + 3\eta_k. \tag{16}$$

Using $\psi_k(t-1) = \Gamma_\omega \log\left(\frac{2\tilde{K}}{\theta}\log((1+\omega)N_k(t-1))\right)$ if $N_k(t) > 0$ and $N_k(t-1) \leq H_k$ by Lemma 11, we obtain

$$f_k \leq \frac{32\Gamma_\omega}{\eta_k}\log\left(\frac{2\tilde{K}}{\theta}\log((1+\omega)H_k)\right) + 3\eta_k.$$

This bound is also valid when $N_k(t) > 0$.

Going back to (15), and since $\mu_0 \leq \overline{\mu}_0(t-1)$ under $\mathcal{E}$, we have (notice $\eta_k \leq 2$ since $\mu_k \in [0,1]$ and $\epsilon \in [0,1]$):

$$\alpha N_0(t-1)\mu_0 \leq (1-\alpha)\overline{\mu}_0(t-1) + 6K$$

$$+ \sum_{k=1}^{K} \frac{32\Gamma_\omega}{\eta_k}\log\left(\frac{2\tilde{K}}{\theta}\log((1+\omega)H_k)\right). \tag{17}$$

To bound the first term of the right-hand side, let us first notice that the final result holds if $N_0(t-1) \leq \max_{k \in [K]} H_k$. So we can assume $N_0(t-1) > \max_{k \in [K]} H_k$ from now on. By the definition of the H_ks (see above (13)), this implies $N_0(t-1) > \frac{16\psi_0(t-1)}{\eta_{\min}^2}$, which in turn implies $4\beta_0(t-1) \leq \eta_{\min}$.

We thus use $\overline{\mu}_0(t-1) \leq \mu_0 + 2\beta_0(t-1) \leq \mu_0 + \frac{\eta_{\min}}{2} \leq 2$, which gives the final result.

The result directly follows from (17). $\qquad\square$

The proof of Theorem 8 follows from $\tau = \sum_{k=1}^{K} N_k(\tau) + N_0(\tau)$, by setting $\omega = 1$ for ease of reading, and $\sigma = \frac{1}{2}$ since Bernoulli variables are $\frac{1}{2}$-subgaussian (using Hoeffding's inequality (Hoeffding 1963)).

We prove Corollary 9.1 from Theorem 7 and Theorem 8.

We now prove Theorem 9:

Proof. Since playing the baseline is neutral in the cost of exploration, it can be re-written as:

$$C_\tau = \sum_{k=1}^{K} (\mu_0 - \mu_k)N_k(\tau) \leq \sum_{k:\mu_k < \mu_0} (\mu_0 - \mu_k)N_k(\tau),$$

where τ is the time the algorithm stops. Using Lemma 11 to upper bound $N_k(\tau)$, we obtain the result. $\qquad\square$

Corollary 9.1 simply follows from the fact that by applying each algorithm with confidence δ/M, the confidence intervals are then simultaneously valid for all users with probability $1-\delta$, so all the correctness/duration/cost proofs holds for all groups simultaneously with probability $1-\delta$. For the statistical guarantees on certifying the probabilistic envy-freeness criterion, we provide the proof of Theorem 2 in App. E.5.

E.4   Proof of Theorem 1

Theorems 7, 8, and 9 are summarized in Theorem 1 in the main paper. We restate Theorem 1 and prove it below:

Theorem. *Let $\epsilon \in (0,1]$, $\alpha \in (0,1], \delta \in (0, \frac{1}{2})$ and*

$$\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k) \text{ and } h_k = \max(1, \frac{1}{\eta_k}).$$

Using $\mu, \overline{\mu}$ and Φ given in Lemmas 4 and 6, OCEF achieves the following guarantees with probability at least $1-\delta$:

- *OCEF is correct and satisfies the conservative constraint on the recommendation performance* (3).

- *The duration is in* $O\Big(\sum_{k=1}^{K} \frac{h_k \log\left(\frac{K\log(\frac{Kh_k}{\delta\eta_k})}{\delta}\right)}{\min(\alpha\mu_0, \eta_k)} \Big).$

- *The cost is in* $O\bigg(\sum\limits_{k:\mu_k<\mu_0} \frac{(\mu_0-\mu_k)h_k}{\eta_k} \log\big(\frac{K\log(\frac{Kh_k}{\delta\eta_k})}{\delta}\big)\bigg).$

Proof. With $\delta \in (0,\frac{1}{2})$, let $\theta = \log(2)\sqrt{\frac{\delta}{6}}$. Then Theorems 8 and 9 hold for (δ, θ).

Duration We first show that:

$$H_k = O\bigg(\frac{h_k}{\eta_k} \log\big(\frac{Kh_k}{\delta\eta_k}\big)\bigg), \tag{18}$$

$$\log(H_k) = O\bigg(\log\big(\frac{Kh_k}{\delta\eta_k}\big)\bigg). \tag{19}$$

Recall from Th. 8 that H_k is defined as:

$$H_k = 1 + \frac{64}{\eta_k^2} \log\left(\frac{2(K+1)\log^+\big(\frac{256(K+1)}{\theta\eta_k^2}\big)}{\theta}\right)$$

We replace the $\log^+$ term from Th. 8 by $\log\big(\frac{Kh_k}{\delta\eta_k}\big) > 0$, because $\frac{Kh_k}{\delta} \geq 3$ as soon as $K \geq 2$. We thus have

$$H_k = 1 + O\bigg(\underbrace{\frac{1}{\eta_k^2} \log\bigg(\frac{K}{\delta} \log\big(\frac{Kh_k}{\delta\eta_k}\big)\bigg)}_{=B}\bigg), \tag{20}$$

Using $\log(x) \leq x \Rightarrow x\log(x) \leq x^2$ for $x \geq 0$, and the fact that $\log\big(\frac{Kh_k}{\delta\eta_k}\big) \geq 0$, we have:

$$B \leq \log\bigg(\frac{Kh_k}{\delta\eta_k} \log\big(\frac{Kh_k}{\delta\eta_k}\big)\bigg) \leq 2\log\big(\frac{Kh_k}{\delta\eta_k}\big).$$

Since $1 + \frac{1}{\eta_k^2} \leq 2\frac{h_k}{\eta_k}$, eq. (18) holds.

We now bound $\log(H_k)$:

$$\log(H_k) = O\bigg(\log\bigg(\frac{h_k}{\eta_k} \log\big(\frac{Kh_k}{\delta\eta_k}\big)\bigg)\bigg) \tag{21}$$

$$= O\bigg(\log\bigg(\frac{Kh_k}{\delta\eta_k} \log\big(\frac{Kh_k}{\delta\eta_k}\big)\bigg)\bigg) \tag{22}$$

$$= O\bigg(\log\big(\frac{Kh_k}{\delta\eta_k}\big)\bigg) \tag{23}$$

where the last line comes from $\frac{Kh_k}{\delta\eta_k} \log\big(\frac{Kh_k}{\delta\eta_k}\big) \leq \big(\frac{Kh_k}{\delta\eta_k}\big)^2$.

Therefore, eq. (19) holds.

Now, let

$$\Gamma = \frac{6K+2}{\alpha\mu_0} + \sum_{k=1}^{K} \frac{128\log\left(\frac{2(K+1)\log(2H_k)}{\theta}\right)}{\alpha\mu_0\eta_k},$$

so that $H_0 = \max(\max_{k\in[K]} H_k, \Gamma)$.

We have:

$$\Gamma = O\bigg(\frac{K}{\alpha\mu_0} + \sum_{k=1}^{K} \frac{h_k}{\alpha\mu_0} \log\big(\frac{K\log(H_k)}{\delta}\big)\bigg)$$

$$= O\bigg(\sum_{k=1}^{K} \frac{h_k}{\alpha\mu_0} \log\big(\frac{K\log(H_k)}{\delta}\big)\bigg)$$

$$= O\bigg(\sum_{k=1}^{K} \frac{h_k}{\alpha\mu_0} \log\big(\frac{K\log(\frac{Kh_k}{\delta\eta_k})}{\delta}\big)\bigg),$$

where the second equality is because $K = \sum_{k=1}^{K} 1 \leq \sum_{k=1}^{K} h_k$, and the last equality uses eq. (19). Combining this with eq. (18) we have:

$$H_0 = O\bigg(\sum_{k=1}^{K} \frac{h_k}{\min(\alpha\mu_0, \eta_k)} \log\big(\frac{K\log(\frac{Kh_k}{\delta\eta_k})}{\delta}\big)\bigg).$$

Using eq. (18) again to bound $\tau = H_0 + \sum_{k=1}^{K} H_k$, , we get the desired bound for duration.

Cost For the cost, we remind the bound given in Th. 9:

$$C_\tau \leq \sum_{k:\mu_k<\mu_0} (\mu_0 - \mu_k)H_k$$

$$= O\bigg(\sum_{k:\mu_k<\mu_0} \frac{(\mu_0-\mu_k)h_k}{\eta_k} \log\big(\frac{K}{\delta} \log\big(\frac{Kh_k}{\delta\eta_k}\big)\big)\bigg) \tag{24}$$

using (20) and $1 + \frac{1}{\eta_k^2} = O(\frac{h_k}{\eta_k})$.

$\square$

E.5 Proof of Theorem 2

We restate Theorem 2 which summarizes the guarantees for the audit of the probabilistic envy-freeness criterion with AUDIT, and we prove it below:

Theorem. *Let* $\epsilon, \gamma, \lambda \in (0,1], \delta \in (0,\frac{1}{2})$. *Let* $\tilde{M} = \lceil\frac{\log(3/\delta)}{\lambda}\rceil$ *and* $K = \lceil\frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))}\rceil$. *With probability at least* $1 - \delta$,

- *AUDIT satisfies the conservative constraint* (3) *for all* $\tilde{M}$ *audited users,*
- *the bounds on duration and cost from Th. 1 (using* $\frac{\delta}{3\tilde{M}}$ *instead of* δ*) are simultaneously valid,*
- *if AUDIT outputs* $(\epsilon, \gamma, \lambda)$-`envy-free`, *then the recommender system is* $(\epsilon, \gamma, \lambda)$-*envy-free, and if it outputs* `not-envy-free`, *then* $\exists(m,n), u^m(\pi^m) < u^m(\pi^n)$.

Proof. The first point is a consequence of Theorem 7 and the second point is a consequence of Theorems 8 and 9. Since we apply OCEF to each target user with confidence $\frac{\delta}{3\tilde{M}}$, by the union bound the confidence intervals are simultaneously valid for all $\tilde{M}$ target users with probability $1 - \frac{\delta}{3}$. Therefore, with probability at least $1 - \frac{\delta}{3}$, the conservative constraint is satisfied for all $\tilde{M}$ users *and* the bounds on cost and duration hold simultaneously for all $\tilde{M}$ users.

We now prove the third bullet point in two steps.

Step 1 We show that the value of $K = \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))}$ is chosen to guarantee the following result: with probability $1 - \frac{\delta}{3\tilde{M}}$, if for a user we have $\mu_0 + \epsilon \geq \max_{k\in[K]} \mu_k$, then the user is not (ϵ, γ)-envious.

First, we apply the theorem on random subset selection from (Schölkopf and Smola (2002), Theorem 6.33), which guarantees that with probability $1 - (1-\gamma)^K$, the arm with maximal reward among the K arms is in the $(1-\gamma)$-quantile range of all possible M arms. Solving for $(1-\gamma)^K = \frac{\delta}{3\tilde{M}}$,

we get that when $K = \left\lceil \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))} \right\rceil$, the arm with maximal reward among the K is in the $(1-\gamma)$ quantile range with probability $1 - \frac{\delta}{3\tilde{M}}$. This means that if for a target user m, we have $u^m(\pi^m) + \epsilon = \mu_0 + \epsilon \geq \max_{k \in [K]} \mu_k$, then with probability $1 - \frac{\delta}{3\tilde{M}}$, we also have:

$$\mathbb{P}_{n \sim U_M}[u^m(\pi^m) + \epsilon \geq u^m(\pi^n)] \geq 1 - \gamma,$$

meaning the user is not (ϵ, γ)-envious. By a union bound over the $\tilde{M}$ target users, the property holds simultaneously for all $\tilde{M}$ target users with probability $1 - \frac{\delta}{3}$.

Step 2 We now show that the number of users to audit $\tilde{M} = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ is chosen to guarantee that if none of the $\tilde{M}$ sampled users are (ϵ, γ)-envious, then this holds true for an $(1-\lambda)$ fraction of the whole population with probability $1 - \frac{\delta}{3}$.

Let $\delta' = \frac{\delta}{3}$. Denoting q the probability that a user is not (ϵ, γ)-envious, we want to guarantee that $q \geq 1 - \lambda$ with probability at least $1 - \delta'$, using $\tilde{M}$ Bernoulli trials where $p := 1 - q$ is the probability of success.

Let $\bar{B}(\tilde{M}, k, \delta')$ denote the largest p' such that the probability of observing k or more successes is at least $1 - \delta'$ (i.e., $\bar{B}(\tilde{M}, k, \delta')$ is the binomial tail inversion). By definition, we have $p \leq \bar{B}(\tilde{M}, 0, \delta')$. Using the property that $\bar{B}(\tilde{M}, 0, \delta') \leq \frac{\log(1/\delta')}{\tilde{M}}$ (see e.g., (Langford 2005)), we can guarantee that $p \leq \lambda$ as soon as $\frac{\log(1/\delta')}{\tilde{M}} \leq \lambda$. Solving for $\tilde{M}$, we obtain that $\tilde{M} = \left\lceil \frac{\log(1/\delta')}{\lambda} \right\rceil = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ is sufficient to guarantee $p \leq \lambda$, or equivalenly $q \geq 1 - \lambda$ with probability $1 - \frac{\delta}{3}$.

We combining Step 1 and 2 by a union bound: if for $\tilde{M}$ users and K arms, we have $\mu_0 + \epsilon \geq \max_{k \in [K]} \mu_k$, then with probability $1 - \frac{2\delta}{3}$, an $(1-\lambda)$ fraction of the whole population is not (ϵ, γ)-envious – or equivalently, the recommender system is $(\epsilon, \gamma, \lambda)$-envy-free. Since OCEF is correct with probability $1 - \frac{\delta}{3}$ when outputting that $\mu_0 + \epsilon \geq \max_{k \in [K]} \mu_k$ (i.e., ϵ-no-envy), the union bound guarantees with probability $1 - \delta$ that AUDIT is correct when outputting $(\epsilon, \gamma, \lambda)$-envy-free. Since OCEF is correct with probability $\geq 1 - \delta$ when outputting envy, then so is AUDIT when outputting not-envy-free, which achieves the proof of the third bullet point.

$\square$