Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Sister Conferences Best Papers Track

Online Certification of Preference-Based Fairness for Personalized Recommender Systems* (Extended Abstract)

Virginie Do[1,2], **Sam Corbett-Davies**[2], **Jamal Atif**[1] and **Nicolas Usunier**[2]
[1]LAMSADE, Université PSL, Université Paris Dauphine, CNRS, France
[2]Meta AI
virginie.do@dauphine.eu, scd@meta.com, jamal.atif@lamsade.dauphine.fr, usunier@meta.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 influence the information and opportunities we encounter, by prioritizing content from news outlets and social networks, and sorting job postings. To prevent the risk of unfair delivery of opportunities across users, substantial work has been done to audit recommender systems [Sweeney, 2013; 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.

Yet, observed disparities in recommendation do not necessarily imply a less favorable treatment: they might reflect differences of preferences across user groups. 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 are limited by the difficulty of interpersonal comparisons of measures of satisfaction [Sen, 1999].

In this paper, we propose an alternative approach focused 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.

Auditing for envy-freeness poses new challenges. First, it 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. We consider a scenario where the auditor is allowed to replace a user's recommendations with those of another user. 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 the potential impact of randomizing recommendations on user experience. To control this cost of the audit (in terms of user utility), we follow the framework of conservative exploration [Wu *et al.*, 2016], 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 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.

2 Envy-Free Recommendations

2.1 Framework

We identify the set of users with $[M] = \{1, \ldots, M\}$. A personalized recommender system has one stochastic recom-

*Published as a long paper at AAAI-2022. A complete version with more results, details and technical appendix is available on arXiv: https://arxiv.org/abs/2104.14527.

mendation 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 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:

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

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. 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.

2.2 Envy-Free Recommendations

Existing audits for user-side fairness in recommender systems are based on parity criteria that either do not control for disparities that are aligned with user preferences, or drive user utility down. 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 all users prefer their recommendations to those of any other user:

Definition 2.1. *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] stipulates that it is fair to apply different policies to different people as long as it benefits everyone. Following this principle, we consider the personalization of recommendations as fair only if it better accommodates user preferences. In contrast, we consider unfair the failure to give users a better recommendation when one such is available to others.

Optimal recommendations are envy-free. Unlike parity criteria, envy-freeness is compatible with giving users their most preferred recommendations. Let $\pi^{m,*} \in \operatorname{argmax}_\pi u^m(\pi)$ denote an optimal recommendation policy for m. Then *the optimal recommender system* $(\pi^{m,*})_{m\in M}$ *is envy-free* since: $u^m(\pi^{m,*}) = \max_\pi u^m(\pi) \geq u^m(\pi^{n,*})$.

Probabilistic relaxation of envy-freeness. Envy-freeness, as defined in Def. 2.1, (a) compares the recommendations of a target user to those of *all* other users, and (b) these comparisons must be made for *all* users. 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-γ% 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. 3.4, 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-freeness audits in practice.

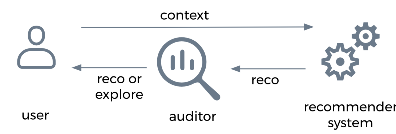

Figure 1: Auditing scenario: the auditor either shows the user their recommendation in the current recommender system, or explores by showing the recommendation given to another user.

Definition 2.2. *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.

3 Certifying Envy-Freeness

3.1 Envy-Freeness Audit as a Bandit Problem

Auditing scenario. The envy-freeness auditor faces the challenge of answering the counterfactual question: "had user m been given the recommendations of user n, would m get higher utility?". However, accessing user preferences is difficult as they are only partially observed through interactions with recommended items. To enable active exploration of user preferences, we propose the following auditing scenario: at each time step t, the auditor can either (a) give the user a "normal" recommendation or (b) explore user preferences by giving a recommendation from another user (Fig. 1).

The equivalent bandit problem. We 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, then we present our full auditing algorithm.

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 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 $\big(u^m(\pi^{n_1}), ..., u^m(\pi^{n_K})\big)$ 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, 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

Algorithm 1: OCEF algorithm. ξ_t (line 4) evaluates the conservative exploration constraint and is defined in (3). Values for $\beta_k(t)$ and confidence bounds $\underline{\mu}_k$ and $\overline{\mu}_k$ are given in the full paper.

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, ...* **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`
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**

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 $k_s \in \{0, \dots, K\}$ the arm (group index) chosen at round s, this requirement is formalized as a conservative exploration constraint [Wu *et al.*, 2016]:

$$\forall t, \frac{1}{t} \sum_{s=1}^{t} \mu_{k_s} \geq (1-\alpha)\mu_0 \,. \tag{2}$$

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.

Our setting had not yet been addressed by the pure exploration bandit literature, which mainly studies the identification of (ϵ-)optimal arms [Audibert *et al.*, 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.

3.2 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, $\beta_k(t) = (\overline{\mu}_k(t) - \underline{\mu}_k(t))/2$.

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 (2), 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 an upper bound $\Phi(t), \forall t > 0$ on $\left|\sum_{s \in A_t} (\mu_{k_s} - r_s)\right|$. This confidence interval is used to estimate whether the conservative constraint (2) 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 (2) 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 (2). Using $\tau = t-1$, this leads to:

$$\xi_t = \sum_{s \in A_\tau} r_s - \Phi(t) + \underline{\mu}_{\ell_t}(\tau) + (N_0(\tau) - (1-\alpha)t)\overline{\mu}_0(\tau) \,. \tag{3}$$

Then, as long as the confidence intervals hold, pulling ℓ_t does not break the constraint (2) 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).

3.3 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* $h_k = \max(1, \frac{1}{\eta_k})$. *With a choice of* $\underline{\mu}, \overline{\mu}$ *and* Φ *(given in the full paper), OCEF achieves the following guarantees with probability* $\geq 1 - \delta$*:*

- *OCEF is correct and satisfies the conservative constraint on the recommendation performance* (2).
- *The duration is in* $O\left(\sum_{k=1}^{K} \frac{h_k \log\left(\frac{K \log(K h_k / \delta \eta_k)}{\delta}\right)}{\min(\alpha\mu_0, \eta_k)}\right)$.
- *The cost is in* $O\left(\sum_{k:\mu_k < \mu_0} \frac{(\mu_0 - \mu_k) h_k}{\eta_k} \log\left(\frac{K \log(K h_k / \delta \eta_k))}{\delta}\right)\right)$.

The important problem-dependent quantity η_k is the gap between the baseline and other arms k, leading to a worst case that only depends on ϵ, since $\eta_k = \max(\mu_k - \mu_0, \mu_0 - \mu_k + \epsilon) \geq \frac{\epsilon}{2}$. 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.

3.4 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 the main paper. It samples a subset of users, and a subset of arms for each sampled user. Then it applies OCEF to each user simultaneously with their sampled arms. It stops either upon finding an envious user, or when all sampled users are certified with ϵ-no envy.

The number of target users $\tilde{M}$ and arms K in AUDIT 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* (2) *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.

4 Experiments

We present experiments evaluating the auditing algorithm OCEF on two recommendation tasks. Our goal is to answer: *in practice, what is the interplay between the required sample size per user, the cost of exploration and the conservative exploration parameter?*

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 2.5k 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. We use the same protocole with the MovieLens-1M (ML) dataset.

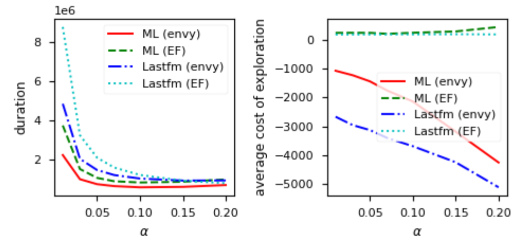

Figure 2: Scaling w.r.t. α on MovieLens (ML) and Last.fm, for recommender systems that are either envy-free (EF) or with envy. There are 41 target users and 75 arms.

For both tasks, the simulated recommender system estimates relevance scores using low-rank matrix completion [Bell and Sejnowski, 1995] on a training sample of 70% of the ground truth preferences, where the rated / played items are sampled uniformly at random. We consider two recommendation policies which are softmax functions over predicted relevance scores. 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 generate binary rewards using a Bernoulli distribution with expectation given by our ground truth preferences.

We use AUDIT with OCEF to certify the probabilistic criterion. The 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 from 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. 2. It plots the duration and the cost of exploration as a function of the conservative constraint parameter α (smaller α means more conservative). On MovieLens, duration is minimal for a non-trivial α. A large α leads to similar confidence intervals for all arms, reducing the duration. As α decreases, the baseline is pulled more, shortening the relevant confidence intervals for all arms and accelerating the audit. However, if α becomes too small, the additional pulls of the baseline have no effect, and the duration increases.

The cost of exploration depends on whether there is envy. On envy-free configurations, it is positive and grows when relaxing the conservative constraint. When there is envy, exploration is beneficial to users, therefore the cost is negative and decreasing with α.

5 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 refer to the full paper for more detailed explanations and results.

References

[Audibert *et al.*, 2010] Jean-Yves Audibert, Sébastien Bubeck, and Remi Munos. Best arm identification in multi-armed bandits. In *COLT 2010 - The 23rd Conference on Learning Theory*, pages 41–53, 11 2010.

[Bell and Sejnowski, 1995] Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.

[Cantador *et al.*, 2011] Iván Cantador, Peter Brusilovsky, Tsvi Kuflik, and . 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, 2011. ACM.

[Datta *et al.*, 2015] Amit Datta, Michael Carl Tschantz, Anupam Datta, and . Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1):92–112, 2015.

[Foley, 1967] Duncan Karl Foley. *Resource allocation and the public sector*. Yale University, 1967.

[Garcelon *et al.*, 2020] Evrard Garcelon, Mohammad Ghavamzadeh, Alessandro Lazaric, and Matteo Pirotta. Improved algorithms for conservative exploration in bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3962–3969, 2020.

[Imana *et al.*, 2021] Basileal Imana, Aleksandra Korolova, John Heidemann, and . Auditing for discrimination in algorithms delivering job ads. In *Proceedings of the Web Conference 2021*, pages 3767–3778, 2021.

[Mehrotra *et al.*, 2017] Rishabh Mehrotra, Ashton Anderson, Fernando Diaz, Amit Sharma, Hanna Wallach, and Emine Yilmaz. Auditing search engines for differential satisfaction across demographics. In *Proceedings of the 26th international conference on World Wide Web companion*, pages 626–633, 2017.

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

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

[Wu *et al.*, 2016] Yifan Wu, Roshan Shariff, Tor Lattimore, and Csaba Szepesvári. Conservative bandits. In *International Conference on Machine Learning*, pages 1254–1262, 2016.