Ethical Bandits

by

Bob Rombach

Gui Liberali

and

Yang Li

July 3rd, 2025

Bob Rombach is a Ph.D. Candidate at Rotterdam School of Management, Erasmus University. Mailing address: Mandeville Building T10, Postbus 1738, 3000 DR Rotterdam, The Netherlands, rombach@rsm.nl.

Gui Liberali is the Professor of Digital Marketing at Rotterdam School of Management, Erasmus University. Mailing address: Mandeville Building T10-14, 3000 DR Rotterdam, The Netherlands, liberali@rsm.nl.

Yang Li is Associate Professor of Marketing at Cheung Kong Graduate School of Business, yangli@ckgsb.edu.cn. Mailing address: Oriental Plaza, Tower E2, Room 211, 1 East Chang An Avenue, Beijing 100738, China.

The authors gratefully acknowledge the helpful comments participants at seminars at the 2023 Choice Symposium at INSEAD, participants at the Marketing Science Conference (2023, 2024, and 2025) and at the lunchclub at RSM/Erasmus. This research was supported by the Erasmus Research Institute of Management (ERIM), and the Erasmus Centre for Optimization of Online Experiments (http://www.erim.eur.nl/ecode).

Ethical Bandits

Abstract

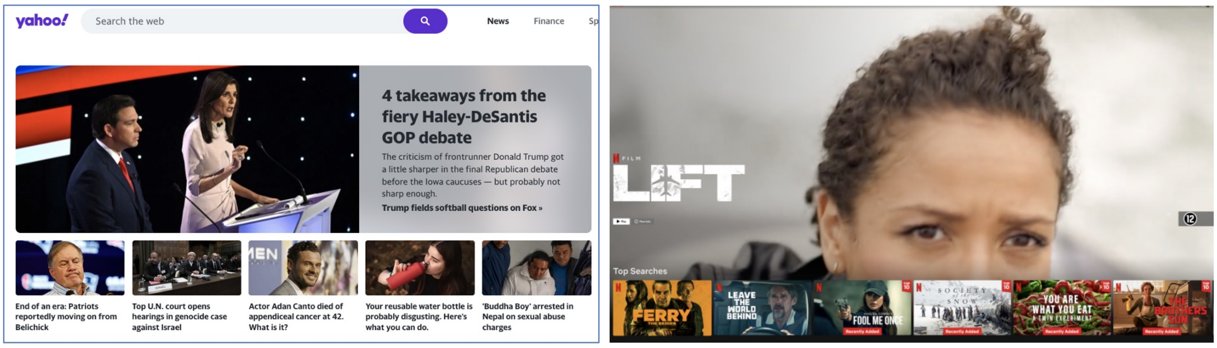
Multi-armed bandits (MABs) have been successfully applied in a range of marketing contexts. However, in finite resource allocation settings such as online product featuring, outright bans on the use of sensitive demographic information can lead MABs to systematically disadvantage minority consumer segments. While the trade-off between exploration and exploitation in MABs is well known, we show that in these settings, MABs also face another important trade-off: the balance between maximizing revenue and mitigating discrimination. We propose an ethical MAB framework to capture both trade-offs simultaneously. Three well-established ethical principles – egalitarian, utilitarian, and proportional – are incorporated to rebalance MAB outcomes in real time. To reduce discrimination without compromising consumer privacy, we introduce a stochastic blinding mechanism that obscures sensitive user information during learning. We demonstrate our approach using randomized controlled trial clickstream data from Yahoo! Research. Our analysis evaluates the extent to which standard MABs can induce discrimination against minorities and how ethical MABs can mitigate such effects. We compare policies in terms of consumer utility, firm reward, and a novel ethical regret metric. The results provide empirical support for recent calls to "myth-bust" prevailing assumptions in data privacy law, quantify the cost of ethical considerations in MAB environments, and reveal the relative performance of conventional versus ethically-guided MAB models.

Keywords: Multi-Armed Bandit, Finite Resource Allocation, Privacy Regulation, Stochastic Blinding.

1 Introduction

The integration of machine learning models into marketing decision making, such as those based on multi-armed bandits (MAB; e.g., Hauser et al. 2009), has become increasingly popular in the field of marketing. Alongside their impressive benefits, such development has also given rise to concerns about customer discrimination. For instance, an algorithm might target high-price products exclusively to populations of a specific ethnicity or gender based on past data patterns, excluding others from these opportunities. Also, certain user demographics might experience an unequal exposure to either negative or positive campaigns. Such discriminatory practices not only risk alienating potential customers and affecting brand reputation, but also pose ethical and legal issues. The current enactment of data protection laws across the world, e.g., the General Data Protection Regulation (GDPR) from the European Union, explicitly contains an in-principle prohibition on using sensitive consumer data, such as racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and genetic or biometric data. It is generally believed that data protection will lead to consumer protection as reflected by, for example, the new EU AI Pact.

While we understand the concerns regarding potential biases associated with the use of sensitive consumer data, we show in this paper that categorically prohibiting its use can, in fact, exacerbate discrimination. We argue that, in numerous cases within digital marketing (e.g., MAB applications), outright bans of sensitive demographic information might lead to unjust treatment of minority consumer segments. This could happen not only because various unprotected variables (e.g., user zip codes) might be correlated with the protected sensitive characteristics in a complex manner, therefore resulting in discrimination (Ascarza and Israeli 2022), but also because the inherent mechanism of an algorithmic learning model like MAB will progressively prioritize items with the highest conversion or sales, thereby diminishing the chances of personalization that may be preferred by smaller or niche consumer segments.



4 takeaways from the fiery Haley-DeSantis GOP debate

The criticism of frontrunner Donald Trump got a little sharper in the final Republican debate before the Iowa caucuses — but probably not sharp enough.

Trump fields softball questions on Fox »

End of an era: Patriots reportedly moving on from Belichick

Top U.N. court opens hearings in genocide case against Israel

Actor Adan Canto died of appendiceal cancer at 42. What is it?

Your reusable water bottle is probably disgusting. Here's what you can do.

'Buddha Boy' arrested in Nepal on sexual abuse charges

Figure 1: Finite Resource Allocation in Featuring News Articles and Movies

Consider the problem of product featuring on the internet. As most users now access news online, often via small-screen devices, platforms face a "real estate" constraint: they generate much more content than can be displayed to readers at any given moment. To manage this, firms commonly use personalization algorithms such as MABs to select and highlight a limited number of news articles or products on their front pages (Coenen 2019; Figure 1). Typically, the number of available slots is several orders of magnitude smaller than the pool of eligible content. In the absence of sensitive consumer data, MAB algorithms would learn to feature products that cater to the preferences of the majority of web visitors, resulting in the marginalization of minority groups. The more the MAB algorithm learns, the more it will benefit large or affluent customer segments while neglecting smaller or low-income ones, which contradicts the intention of policymakers. A small but growing view among law scholars even argues that there is a "need to 'myth bust' the notion that data protection legislation should preclude the processing of equality data" (Bekkum 2023).

One way for a MAB system to mitigate discrimination is by granting it access to the data attributes that distinguish consumer groups and "hard-wire" it to not discriminate. However, such attributes are often classified as sensitive and are protected by privacy regulations like the GDPR, so companies are legally restricted from using them. Another approach is to correct for discrimination based on periodic auditing. However, a MAB that balances discrimination and revenue in real time, as opposed to relying on post-hoc corrections, would be preferable for both firms and consumers.

In this paper, we propose a novel MAB framework to achieve this goal by exploring the interplay between algorithmic learning and equity in finite resource allocation settings, a context with broad applications from economics to engineering (e.g., Amanatidis et al. 2023;

Bertsimas et al. 2011; Li et al. 2025). We add to this research stream by offering a real-time method that simultaneously maximizes the sum of expected rewards while controlling for discrimination against minorities. Central to our methodology are two key principles. First, we employ a *stochastic blinding* strategy to avoid revealing sensitive attributes to the MAB engine, thereby preventing direct access to confidential information by the algorithm.

Second, we implement *ethical rebalancing* by integrating alternative ethical rules of finite resource allocation within the MAB framework. Ethical principles play a crucial role in determining the extent to which particular consumer groups will be favored or neglected by the MAB, and the mechanism behind these outcomes. We illustrate our method with three prevalent allocation rules — *utilitarian*, *egalitarian*, and *proportional*. We analyze and contrast the results in terms of the rewards to firms and the utilities to consumers across segments under each allocation rule.

We intend to make three contributions. First, we show why and how MAB may discriminate against minorities under the current ban of sensitive data, providing empirical support for the emerging myth-busting efforts. In doing so, our research contributes to a growing body of legal and ethical scholarship that calls for a reevaluation of data protection strategies (e.g., Habbal et al. 2024).

Second, we propose a dynamic rebalancing mechanism for MABs by incorporating specific principles of finite resource allocation. This flexibility enables firms to implement policies that adhere to current social norms or the ethics upheld by the organization. The resulting MAB framework is well-suited for digital marketing, where evolving consumer expectations and regulatory landscapes calls for a proactive and flexible approach to algorithmic ethics.

Third, we introduce *stochastic blinding* to the marketing discipline as a convenient form to protect consumer privacy. This means that our MAB does not directly access the original protected attributes, but rather utilizes a set of “ethical weights” given an intended ethical principle. These weights mask the attributes from the recommendation system, ensuring compliance with current data protection laws and balancing discrimination and revenue. The functional form used to calculate these weights is determined by the choice of ethical principles by an organization (e.g., egalitarian or utilitarian). Our model employs these weights to generate probability distributions from which personalization decisions are drawn.

This process also aligns with privacy-preserving technologies like federated learning (Zhang et al. 2021). By introducing stochastic blinding, we offer a novel solution that bridges the gap between ethical algorithm design and practical implementation, reducing discrimination without compromising privacy.

The rest of the paper is structured as follows. First, we review the relevant literature. Next, we elaborate on our modeling framework and explain its key components in detail. We then describe the data and present the primary simulations to illustrate our methods. This is followed by the results obtained from applying the proposed framework to online advertising data. We also discuss the ethical learning process, the resulting ethical regrets, model convergence, and robustness checks. Finally, we conclude by summarizing our contributions, acknowledging the limitations of our approach, and suggesting directions for future research.

2 Literature Review

In this section, we briefly review research streams in marketing, economics, and computer science that are relevant to our work and discuss our contributions to these areas.

2.1 Discrimination and Privacy Research in Marketing

Researchers in marketing have extensively studied the impacts of fairness on consumers and firms in various managerial settings, such as pricing, behavioral decision-making, channel coordination, and policy intervention. Several behavioral studies have established that price fairness is a widely understood concept and have developed frameworks to explain perceptions of unjust pricing (Bolton and Alba 2006; Bolton et al. 2003, 2010; Campbell 1999; Cappelen et al. 2007; Haws and Bearden 2006; Xia et al. 2004). Other behavioral research has provided evidence that equitable concerns are intrinsic to human decision-making (Koenigs et al. 2007; Sanfey et al. 2003). Analytic modelers have shown how fairness can affect channel coordination (Cui et al. 2007; Cui and Mallucci 2016; Ho et al. 2014), proposed strategic actions to mitigate unjust consequences (Allender et al. 2021; Diao et al. 2023), examined equitable selling and welfare implications with buyer inequity aversion (Guo and Jiang 2016), investigated different definitions of fairness and their impacts (Cohen et al. 2022), and highlighted the strategic behavior of firms investing in machine learning under different equity requirements (Fu et al. 2022). Field experiments, such as Lambrecht and Tucker (2019), demonstrated that algorithms optimizing for cost-effectiveness in ad delivery can inadvertently exhibit discriminatory behavior due to crowding out. Much of this literature is conceptual, experimental, or theoretical, with limited empirical modeling of ethical actions in marketing. Our work contributes to this literature by proposing a data-driven framework to address equitable marketing within the MAB context.

Our approach centers on empirically determining when and how to keep sensitive information private, such that discrimination is minimized. A growing body of research on differential privacy (DP) has examined the inherent trade-off between privacy and utility (Dinur and Nissim 2003; Dwork and Roth 2014; Dwork et al. 2019). In DP, privacy is protected by adding noise to the individual user information so that it becomes harder to infer specific information about individuals (Ponte et al. 2024). The added noise, while without direct interpretation in itself, does provide a provable guarantee that the individual information is protected. Our work contributes to this literature by developing a mathematically tractable discrimination-reducing noise element that obscures user sensitive information.

2.2 Discrimination in Algorithmic Personalization

Personalization algorithms have been used for resource allocation decisions across many domains, and discriminatory outcomes for users and products have been extensively examined (e.g., Ascarza and Israeli 2022). For instance, Sun et al. (2024) use controlled field experiments to evaluate the value of personal data in e-commerce. They find a ban on the use of personal data inadvertently leads to disproportionately more negative outcomes for niche merchants and customers. In the MAB research, studies have shown that conventional bandits can lead to distortionary, winner-takes-all resource allocations and have proposed solutions to mitigate this issue. On the product side, Wang et al. (2021) investigate equitable product exposure in contextual bandits, proposing that each arm (i.e., product) should receive exposure proportional to its merit. On the user side, Patil et al. (2020) study the trade-off between predictive accuracy and equal treatment, focusing on the cost of achieving unbiased outcomes. Celis et al. (2020) propose a bandit framework to mitigate polarization and echo-chamber effects in personalized recommendations. Barocas et al. (2023) examine discriminatory concerns in different machine learning contexts and assess various equity interventions. Li et al. (2019) propose equity-constrained MABs where multiple arms can be played simultaneously, and some arms can become "sleeping" (unavailable).

2.3 Ethical Principles in Resource Allocation

Ethical resource allocation in economic contexts can be conceptualized in multiple ways, each grounded in different normative criteria aimed at preventing discrimination and ensuring equitable outcomes (Thomson 2011). For instance, Cappelen et al. (2007) explore distributive justice ideals and focus on the egalitarian principle that employs the max-min rule, which maximizes the minimum utility among all agents to safeguard the most disadvantaged. Bertsimas et al. (2011) examine proportional allocation, a principle extending the Nash bargaining solution from two-player scenarios by maximizing the product of individual utilities. This principle ensures allocations of finite resource that are both Pareto efficient and invariant to scale transformations. In contrast, utilitarianism emphasizes maximizing the aggregate welfare or total utility across the population, promoting efficiency but without explicitly addressing disparities among individuals (Rigobon 2023).

These ethical principles can be placed along a continuum reflecting different distributional priorities (Roberts 1980). At one extreme lies the max-max rule, prioritizing the enhancement of the most advantaged individual's outcome. At the other extreme, the egalitarian max-min principle prioritizes improving outcomes for the least advantaged. Between these extremes, the utilitarian principle aims to maximize the average utility, reflecting a balance focused on total welfare (Harsanyi 1955). The proportional rule similarly occupies an intermediate position, combining aspects of both egalitarian and utilitarian approaches. Our study concentrates on these three primary ethical principles – utilitarian (Bertsimas and Dunn 2019), egalitarian (Binns 2020), and proportional (Banerjee et al. 2022) – to illustrate and analyze how each influences outcomes in the context of online real-time personalization using MAB algorithms.

3 Ethical MAB Model

In this study, we apply MAB algorithms to real-time product featuring on the internet. Typically, product-featuring decisions are targeted on segments, based on individual characteristics such as gender and age. As in standard MAB settings, the firm aims to maximize the sum of discounted rewards (e.g., clicks) by balancing two objectives: exploiting current beliefs about the most profitable product for a given group and exploring uncertain options to improve future decisions.

In our context, two key sources of uncertainty affect the product-featuring decisions. First, direct access to individual characteristics is restricted or prohibited by privacy regulations such as the GDPR, rendering traditional segmentation of web visitors infeasible for the MAB. Second, users' true preferences are not known a priori. The new MAB must therefore explore different arms (i.e., product options) to learn these preferences, while simultaneously balancing revenue maximization and discrimination mitigation.

3.1 Model Setup

Consider K distinct consumer groups and A available products. Our bandit framework thus consists of $K \times A$ arms. We assume that consumers inherently choose products that best match their preferences, i.e., a news reader is more likely to click on an article that aligns with their specific reading interests. This is captured by a consumer utility function $v_{kt}(a)$ for a consumer in group $k = 1, \dots, K$ consuming product $a = 1, \dots, A$ at time $t = 1, \dots, T$.

The firm's primary objective is to maximize revenue (i.e., the total number of clicks) by featuring products to every web visitor with the highest probability of eliciting clicks, or click-through rate (CTR). To estimate this probability, we use Thompson Sampling (TS; Thompson 1933; Scott 2010; Russo and Van Roy 2014) to solve all the MABs studied in this paper. In TS, each arm tends to be selected with a probability proportional to the belief that it is the optimal choice, while these beliefs are informed by a vector of consumer utilities corresponding to each arm. In our case, the consumer utility $v_{kt}(a)$ is sampled from a Beta distribution with parameters $\alpha_{kt}(a)$ and $\beta_{kt}(a)$, and for every visitor, TS selects a product to feature on the webpage with the maximal normalized sampled utilities[1]. After the exposure,

[1] Normalization across products ensures a consistent utility scale across consumer segments.

the firm observes consumer response through a binary variable $\delta_{kt}(a)$, indicating whether a click occurred. Thanks to the conjugacy of the Beta distribution, the updated distribution for the next visitor can be expressed as $\mathrm{Beta}(\alpha_{kt}(a) + \delta_{kt}(a), \beta_{kt}(a) + 1 - \delta_{kt}(a))$.

We choose TS to solve our MAB for several reasons. First, its simplicity allows for *ethical rebalancing*, a straightforward integration of ethical principles into MAB by directly imposing constraints on the utility vector based on the adopted ethical rule. Second, the sampling step in TS can be easily adapted to mask sensitive customer information, as we detail in the following sections, enabling *stochastic blinding* to protect consumer privacy. Third, TS in general is a computationally efficient method that has demonstrated strong empirical performance (Chapelle and Li 2011; Kasy and Sautmann 2021) and is widely used by Internet companies such as Amazon (Hill et al. 2017), LinkedIn (Agarwal et al. 2014), and Google (Scott 2010).

3.2 Ethical Principles

In a purely revenue-maximizing context, a firm uses a MAB to balance exploration and exploitation, i.e., featuring products with uncertain consumer interest alongside products known to attract high consumer interest. However, in settings involving finite resource allocation, such as decisions about which items to feature, conventional MAB algorithms learn to feature products favored by the majority, leading to an *additional* trade-off: maximizing overall rewards through majority preferences versus promoting non-discriminatory outcomes by prioritizing minority groups. In this paper, we explicitly incorporate ethical considerations into the MAB framework, enabling firms to integrate various ethical principles into their allocation strategies. We quantify how and when this ethical trade-off emerges, measure the revenue firms must forgo to achieve more equitable outcomes, and reveal the mechanisms influencing which consumer groups ultimately bear the costs for reducing discrimination. We now turn to how these ethical considerations can be integrated into a standard MAB model.

Specifically, we introduce a mechanism of multiplicative ethical weighting, denoted as $\psi_t(a)$ for each product at time t, satisfying the conditions $\sum_{a=1}^{A} \psi_t(a) = 1$ and $0 \leq \psi_t(a) \leq 1$. Rather than using $v_{kt}(a)$, we consider $\psi_t(a)$ as the sampling probability in TS to determine which arm to pull (i.e., which product to feature). This reflects the legal constraints on accessing user identity k. Also, as consumers now receive a weighted utility $\mu_{kt}(a) = \psi_t(a)v_{kt}(a)$, altering the value of $\psi_t(a)$ allows the firm to implement product featuring policies that align with its ethical principle. Moreover, $\psi_t(a)$ is dynamically updated based on consumer responses, which facilitates an ethical rebalancing between arms with lower and higher expected rewards and diverges from the path a traditional MAB might take.

The derivation of the product-specific ethical weights $\psi_t(a)$ is informed by the ethical principles upheld by the firm. We explore three prevalent ethical principles:

Utilitarian Rule Resources are allocated to maximize the collective utility across all groups:

$$\boldsymbol{\psi}_t^* = \underset{\psi}{\text{argmax}} \sum_{k=1}^{K} \sum_{a=1}^{A} \mu_{kt}(a). \tag{1}$$

The utilitarian approach has been widely studied in resource allocation settings due to its efficiency-focused orientation (Bentham 2024). By prioritizing the maximization of overall utility, this framework ensures that the collective satisfaction across all user groups is optimized. Its appeal lies in its ability to deliver the highest aggregate benefits, which aligns with the goal of improving revenue (Fleurbaey et al. 2023). In the MAB context, the utilitarian rule directs exploration and exploitation efforts toward arms with the highest expected utility, thereby ensuring rapid convergence to optimal performance.

Recent work has emphasized the strengths of the utilitarian framework in dynamic and uncertain environments. For instance, Mnih et al. (2015) demonstrate the efficiency of utility-maximizing algorithms in reinforcement learning settings, where resource allocation strategies significantly influence outcomes. Similarly, Bertsimas and Dunn (2019) highlight the scalability of utilitarian principles in large-scale decision-making problems, underscoring their practicality in real-world applications. However, when applied in MAB, the utilitarian rule may favor high-performing arms, leaving low-performing arms under-explored and potentially perpetuating a certain level of discriminatory biases (Joseph et al. 2016; Heidari et al. 2019).

Egalitarian Rule Resources are distributed equitably across all groups, maximizing the utility of the worst-off:

$$\boldsymbol{\psi}_t^* = \underset{\psi}{\text{argmax}} \ \min_{k} \left(\sum_{a=1}^{A} \mu_{kt}(a) \right). \tag{2}$$

The egalitarian principle stems from the philosophy of justice proposed by Rawls (1971). This approach advocates for resource allocation strategies that improve the welfare of the most disadvantaged, thereby reducing discrimination. In the context of MAB, this principle reallocates resources to arms that historically perform poorly, addressing biases against underrepresented consumer groups. Recent literature has highlighted its performance in algorithmic decision-making. For example, Heidari et al. (2019) discuss how egalitarian principles reduce group-level disparities, while Binns (2020) explore its role in mitigating inequities within dynamic systems.

The egalitarian framework provides a contrast to the utilitarian approach by explicitly addressing the needs of minorities. While the utilitarian rule focuses on maximizing collective utility, the egalitarian approach ensures that minorities are not overlooked. Although this approach may lead to some inefficiencies by sacrificing aggregate utility gains, these trade-offs are increasingly viewed as necessary for protecting minorities (Žliobaitė 2017). Furthermore, ethical resource allocation algorithms inspired by egalitarian principles have been shown to achieve better inter-group equity without significantly compromising revenue (Holm 2023).

Proportional Rule In resource allocation contexts, the proportional rule serves as a method to achieve a balanced allocation that account for both revenue and discrimination. Specifically, when utility sets are convex, the proportional rule can be formally defined as (Bertsimas et al. 2011):

$$\boldsymbol{\psi}_t^* = \underset{\psi}{\operatorname{argmax}} \prod_{k=1}^{K} \sum_{a=1}^{A} \mu_{kt}(a). \tag{3}$$

By aggregating group-level utilities multiplicatively, the proportional rule offers a compromise between the utilitarian and the egalitarian approaches. This multiplicative objective function ensures each consumer group's welfare is proportionally represented, mitigating extreme disparities that might arise from either purely utilitarian or egalitarian frameworks. Consequently, the proportional rule provides an alternative that explicitly balances revenue and discrimination in resource allocation.

Recent research highlights the performance of proportional allocation in practical applications. For instance, Nicosia et al. (2017) demonstrate that multiplicative aggregation of utilities can enhance equity without substantially compromising revenue. Similarly, Banerjee et al. (2022) illustrate that proportionality is a particularly robust ethical criterion in the context of online public goods allocation, showing substantial performance improvements when predictive information is incorporated into decision-making processes.

With the three prevailing ethical frameworks, firms can select their preferred allocation rule and calculate the optimal ethical weights to guide the operation of their bandit algorithms. The proposed ethical weighting scheme introduces two key features. First, it facilitates ethical rebalancing between revenue and discrimination during the exploration phase of the MAB. Second, it enables the use of *stochastic blinding*, a probabilistic sampling method to mask sensitive consumer attributes, which we describe next.

3.3 Stochastic Blinding and Learning

Figure 2 illustrates the stochastic blinding framework, where information flows between a module that generates MAB recommendations for product featuring and a module handling sensitive consumer information. When a consumer arrives at time t, the MAB module selects a product a_t by sampling from a categorical distribution defined by the ethical weight vector $p(a_t) \propto \psi_t^*(a)$. After the product is featured, the consumer decides whether to click on it. The outcome $\delta_{kt}(a)$ is passed to the module handling sensitive consumer information, which updates the parameters of the Beta distribution used in TS. Group-level utility estimates are then sampled from the updated Beta distributions. These sampled utilities inform the computation of the new ethical weights $\boldsymbol{\psi}_{t+1}^*$ for the next consumer, based on the chosen ethical rule specified in Equations (1), (2), and (3) respectively. In this way, the updated ethical weights reflect the most recent consumer responses, allowing the MAB to adapt dynamically while upholding the chosen ethical principle.

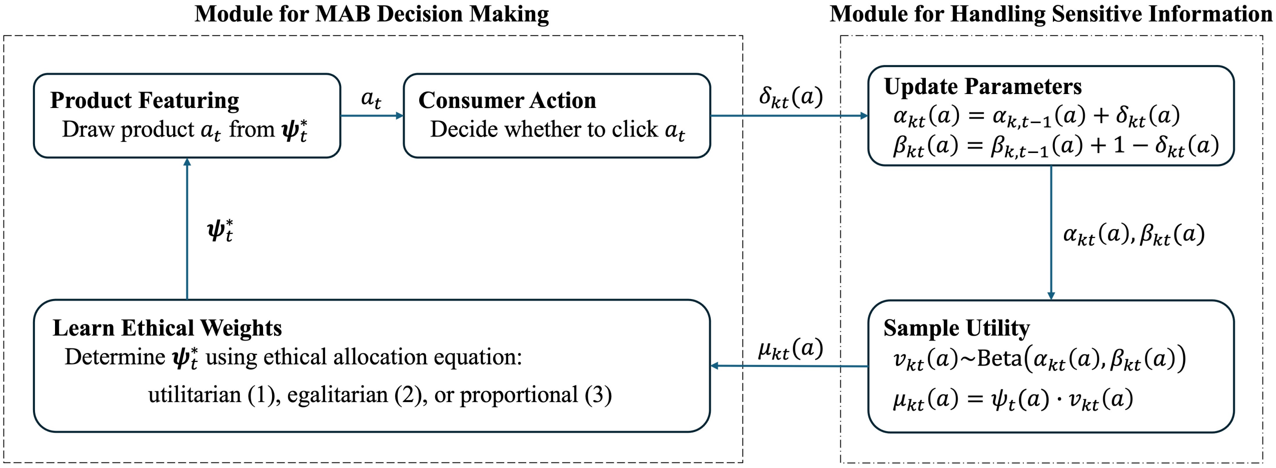

Module for MAB Decision Making
Module for Handling Sensitive Information
Product Featuring
Draw product a_t from $\boldsymbol{\psi}_t^*$
a_t
Consumer Action
Decide whether to click a_t
$\delta_{kt}(a)$
Update Parameters
$\alpha_{kt}(a) = \alpha_{k,t-1}(a) + \delta_{kt}(a)$
$\beta_{kt}(a) = \beta_{k,t-1}(a) + 1 - \delta_{kt}(a)$
$\boldsymbol{\psi}_t^*$
$\alpha_{kt}(a), \beta_{kt}(a)$
Learn Ethical Weights
Determine $\boldsymbol{\psi}_t^*$ using ethical allocation equation:
utilitarian (1), egalitarian (2), or proportional (3)
$\mu_{kt}(a)$
Sample Utility
$v_{kt}(a) \sim \text{Beta}(\alpha_{kt}(a), \beta_{kt}(a))$
$\mu_{kt}(a) = \psi_t(a) \cdot v_{kt}(a)$

Figure 2: Stochastic Blinding in Ethical MAB

As shown in Figure 2, sensitive information indicating group membership k of the current web visitor is handled exclusively within the dashed blocks, which represent internal organizational structures such as a company's Data Ethics Committee, Data Protection Office, or Algorithmic Accountability Board. The computation of group-level ethical-weighted consumer utilities, $\mu_{kt}(a)$, masks sensitive attributes across the K groups.

3.4 Ethical Weight Optimization

Determining the ethical weights involves solving a complex optimization problem guided by the ethical principle specified in (1), (2), or (3). The ill-posed numerical properties, such as the non-smoothness in the egalitarian function (2), pose computational challenges for optimization. These difficulties are especially pronounced in the real-time MAB context, where rapid output is critical. To address these issues, we adapt the Nelder-Mead optimization (NMO; Nelder and Mead 1965), a robust and widely used gradient-free, direct-search method for fast unconstrained optimization. Since our problem is constrained by the requirement $\sum_{a=1}^{A} \psi_t(a) = 1$, we reparameterize by $\psi_t(a) = e^{-\tilde{\psi}_t(a)} / \sum_{a=1}^{A} e^{-\tilde{\psi}(a)}$, transforming the original problem into an unconstrained optimization over $\tilde{\psi}_t(a)$.

The simplex-based NMO algorithm is well-suited to this context as it avoids costly gradient computations, evaluating the objective function $\mathcal{F}(\tilde{\boldsymbol{\psi}})$ only a few times per iteration. The optimization process begins with a simplex of $A + 1$ dimensions, iteratively refining the search area by updating vertex positions and moving the simplex to converge toward an optimal solution. This is achieved through a sequence of geometric transformations designed to explore the objective function space effectively.

To maximize the objective function, each NMO iteration includes the following steps:

- **Ordering**: Identify the best, worst, and second-worst vertices of the current simplex, denoted as $\tilde{\boldsymbol{\psi}}_h$, $\tilde{\boldsymbol{\psi}}_l$, and $\tilde{\boldsymbol{\psi}}_s$, respectively, based on their objective function values.
- **Centroid**: Compute the centroid $\boldsymbol{c}$ of all vertices excluding the worst vertex $\tilde{\boldsymbol{\psi}}_l$.
- **Reflection**: Reflect the worst vertex $\tilde{\boldsymbol{\psi}}_l$ across the centroid to generate a reflected candidate $\tilde{\boldsymbol{\psi}}_r = \boldsymbol{c} + \eta(\boldsymbol{c} - \tilde{\boldsymbol{\psi}}_l)$, where $\eta > 0$ is the reflection coefficient. Depending on the quality of the reflected candidate:
  - ⋆ If $\mathcal{F}(\tilde{\boldsymbol{\psi}}_s) < \mathcal{F}(\tilde{\boldsymbol{\psi}}_r) \leq \mathcal{F}(\tilde{\boldsymbol{\psi}}_h)$, the reflection is sufficiently good but not the best, so the worst vertex is replaced with the reflected candidate, and the algorithm proceeds to the next iteration.
  - ⋆ **Expansion**: If $\mathcal{F}(\tilde{\boldsymbol{\psi}}_r) > \mathcal{F}(\tilde{\boldsymbol{\psi}}_h)$, the reflection indicates a promising direction. We can stretch further in this direction to compute an expanded candidate $\tilde{\boldsymbol{\psi}}_e = \boldsymbol{c} + \gamma(\tilde{\boldsymbol{\psi}}_r - \boldsymbol{c})$, where $\gamma > \eta$ is the expansion coefficient. The worst vertex is replaced with the better of the reflected and expanded candidates, and the algorithm proceeds to the next iteration.
  - ⋆ **Contraction**: If $\mathcal{F}(\tilde{\boldsymbol{\psi}}_r) \leq \mathcal{F}(\tilde{\boldsymbol{\psi}}_s)$, the contraction step is applied using $0 < \kappa < \eta$ as the contraction coefficient:
    - **Outside Contraction**: If $\mathcal{F}(\tilde{\boldsymbol{\psi}}_r) > \mathcal{F}(\tilde{\boldsymbol{\psi}}_l)$, we try a shorter move towards the reflected direction by computing an outside contracted candidate $\tilde{\boldsymbol{\psi}}_c = \boldsymbol{c} + \kappa(\tilde{\boldsymbol{\psi}}_r - \boldsymbol{c})$. If this contraction yields a better value than the worst vertex, the worst vertex is replaced by $\tilde{\boldsymbol{\psi}}_c$, and the algorithm proceeds to the next iteration.
    - **Inside Contraction**: If $\mathcal{F}(\tilde{\boldsymbol{\psi}}_r) \leq \mathcal{F}(\tilde{\boldsymbol{\psi}}_l)$, we attempt a move in the opposite direction of the reflection by computing an inside contracted candidate $\tilde{\boldsymbol{\psi}}_c = \boldsymbol{c} + \kappa(\tilde{\boldsymbol{\psi}}_l - \boldsymbol{c})$. Should this point improve the objective compared to the worst vertex, the worst vertex is replaced by $\tilde{\boldsymbol{\psi}}_c$, and the algorithm proceeds to the next iteration.

- ⋆ **Shrinkage**: If none of the above transformations improves the worst vertex, the simplex is shrunk toward the best vertex $\tilde{\boldsymbol{\psi}}_h$, i.e., $\tilde{\boldsymbol{\psi}}_m = \tilde{\boldsymbol{\psi}}_h + \rho(\tilde{\boldsymbol{\psi}}_m - \tilde{\boldsymbol{\psi}}_h)$ for $m = 1, \ldots, A+1$, $m \neq h$, where $0 < \rho < 1$ is the shrinkage coefficient.

The hyperparameters for NMO applications are commonly set as $\eta = 1$, $\gamma = 2$, $\kappa = 0.5$, and $\rho = 0.5$ (Lagarias et al. 1998), striking a balance between exploration and convergence efficiency for reliable optimization performance.

In addition, from a rational perspective, none of the ethical principles would suggest featuring a product that is inferior to another for every consumer group. This insight allows us to further improve computational efficiency by reducing the dimensionality of the optimization problem, which is equal to the number of products. In particular, we can exclude all Pareto-dominated products. That is, we eliminate a product a if there exists another product a' such that $v_k(a) \leq v_k(a')$, for all $k = 1, \ldots, K$, with at least one strict inequality.

4 Application

We demonstrate our method through an empirical application using data from a randomized controlled trial (RCT). This dataset, made publicly available through Yahoo! Research, provides a detailed clickstream of user interactions with featured articles on Yahoo! News (Figure 1)[2]. It is well-suited for our analysis for several reasons.

First, it includes sensitive user characteristics that can be used to capture consumer heterogeneity and enable discriminatory scenarios. For instance, Chu and Park (2009) show that featuring articles based on consumer gender and age in this dataset leads to more than a twofold increase in CTR, compared to featuring articles without considering these user characteristics. Second, the randomized nature of the trials provides a reliable setting for offline simulations of policy interventions (Li et al. 2011). Third, this dataset has been successfully used in multiple studies to showcase the performance of novel MAB methods (e.g., Vanchinathan et al. 2014, 2015).

Below, we describe the Yahoo dataset and the simulation process used to evaluate the performance of various ethical MAB strategies against established benchmarks. We then discuss the experimental outcomes and their broader implications.

[2] https://www.yahooinc.com/research/research-areas/machine-learning

4.1 Data

The Yahoo! Front Page attracts millions of daily visitors, generating a substantial number of clicks on its featured articles. The data was collected over the first 10 days of May 2009, during which approximately 45 million users visited the site[3]. These observations are organized into consecutive batches, each containing 500,000 observations. For every visit during the trial period, an article was *randomly* selected from a daily pool of approximately 20-25 available articles. The data includes user variables containing sensitive information related to age and gender. To protect user privacy, this user-level information can not be identified exactly. However, this does not affect their analytical relevance for the purpose of our study. For the illustration of our methodology, we select one of the sensitive variables for the simulation analyses.

For the analyses, we define majority, intermediate, and minority consumer segments based on the sensitive user variable. A minority group comprises individuals who differ from the dominant population in terms of race, ethnicity, religion, sexual orientation, or other factors (*Cambridge Learner's Dictionary* 2024). Using the selected sensitive variable, we set the 1/3 and 2/3 points of its range in the data as the cutoff thresholds to split all the users into three distinct subsets. This three-group partition balances analytical simplicity with sufficient complexity to capture realistic inter-group preference relationships. In contrast, a binary grouping would oversimplify the dynamics, reducing preferences to either opposition (each group favors different items) or alignment (both groups favor the same items).

Table 1 presents summary statistics for the majority, intermediate, and minority groups. The table reveals that the three consumer segments differ notably in size and CTR, implying substantial variation in the potential rewards a firm can obtain across these groups. Under a conventional MAB framework that has no access to consumer sensitive information, the minority group would appear the least appealing. Despite exhibiting higher CTR, it represents only 12.22% of the population. Consequently, a conventional MAB lacking sensitivity to these distinctions would likely disadvantage the minority group, especially in contexts where consumer preferences exhibit significant heterogeneity between groups.

Given the large size of the dataset and the limitations of our computational resources, we

[3] As user identities are not disclosed, each visit is assumed to originate from a unique user.

Table 1: Majority vs. Minority Consumer Groups

	Sample Size		CTR	
Group	# Users	% Users	mean	s.d.
Majority	32,691,572	71.36	0.03	0.01
Intermediate	7,522,984	16.42	0.05	0.02
Minority	5,597,327	12.22	0.07	0.02

focus our analysis on a subset of the original data to illustrate our proposed ethical MAB framework and report results for other data subsets in robustness checks and in the online appendix[4]. Since the relevance and appeal of news articles typically diminish after one or two days (i.e., they are no longer considered "new"), we structure the MAB learning process within individual days rather than spanning multiple days. To ensure comparability across all analyses, we consistently use the same sensitive variable throughout the samples.

4.2 Benchmark Models

We evaluate the extent to which discrimination against the minority consumer segment can be (1) induced by a standard MAB and (2) mitigated by the proposed ethical MAB schemes. To this end, we perform simulation analyses to estimate the CTR outcomes had different MAB policies been applied to the RCT Yahoo dataset.

We examine the stochastic blinding MAB model with the three ethical principles : *Utilitarian* MAB (Equation 1), *Egalitarian* MAB (Equation 2), and *Proportional* MAB (Equation 3). For comparison, we consider two benchmark settings that vary in the availability of information for product featuring:

- *Oracle* MAB: This benchmark bandit is fully aware of each user's identity and uses this sensitive information to make product featuring decisions.
- *Blind* MAB: This benchmark bandit represents the *status quo*. It replicates the restrictive scenario under the current privacy regulation like the GDPR, where firms cannot use sensitive user information. Consequently, this MAB cannot differentiate minority users from majority ones.

[4] https://sites.google.com/view/guiliberali/publications?authuser=0

4.3 Simulation Procedure

We conduct counterfactual policy simulations to evaluate the proposed as well as the benchmark MAB models had they been implemented in real time. The data-generating process proceeds as follows.

Data Generating Process At each time step t, a user visits the website with probability π_k of belonging to group k. The MAB system selects an article a_t to feature, and the user then decides whether to click on it based on the CTR p_{ka}. Here, π_k is estimated as the proportion of users in group k relative to the total user base, as reported in Table 1. The CTR p_{ka} represents the empirical click probability for group k on article a, calculated as the average click rate across all impressions where the article was shown. Given the RCT nature of the dataset, we treat the empirical CTR as a reliable estimate of the true CTR for each group-article pair.

Using the values of π_k and p_{ka}, we simulate how each MAB model would have performed. After observing the click result at each time step, the parameters of the MAB policy get updated. For simplicity, we assume that during the period of investigation, the CTR distribution remains constant over time, and group identity is independent of the CTR distribution, i.e., $\pi_k \perp p_{ka}$.

Based on the simulated outcomes, i.e., the articles selected by each MAB model and the resulting user clicks, we evaluate performance from both the user and firm perspectives. For each consumer group, we calculate average utility by summing the true CTRs of the simulated featuring products across time periods and dividing by the total number of arrivals for group k. On the firm side, we compute the expected reward from each segment by counting the number of clicks generated by users in that group. Aggregating across all segments yields the overall expected reward (total clicks) for the firm under each MAB policy.

MAB Structures The alternative MAB policies differ in their structural design and how they incorporate consumer group-level information into decision-making. Using sensitive consumer attributes, the oracle MAB maintains one arm for every combination of article and consumer group, explicitly modeling variations in user behavior across distinct consumer segments. This full-information structure allows the oracle bandit to tailor its article selection to maximize expected rewards based on differences across consumer groups. By contrast, without sensitive consumer information, the blind MAB employs a simplified approach by maintaining only one arm per article, ignoring consumer grouping altogether. Consequently, the blind MAB does not incorporate any group-specific CTR heterogeneity into its decisions, thereby sacrificing performance and potentially inducing inadvertent discrimination. Finally, the ethical MABs adopt an intermediate design and maintain group-level estimates, but utilize this information exclusively for estimating the ethical weight $\boldsymbol{\psi}_t^*$. Critically, these detailed group-level arms do not directly influence the selection of featuring articles. Instead, article selection occurs solely based on sampling from the aggregate ethical weights $\psi_t^*(a)$, which are article-specific rather than consumer group-specific. These ethical bandits intentionally decouple the estimation of consumer group differences from the direct decision-making process, thereby limiting discrimination, while still accounting for consumer heterogeneity indirectly.

4.4 Results

To illustrate the proposed methodology, Table 2 reports the consumer utility and firm-side reward outcomes from the simulation.[5] Group-level CTRs and the utilities associated with each article are provided in the Appendix. All reported values represent averages over 10 simulation replicates and are calculated based on the final 10% of the 500,000 simulation iterations.

Table 2: Empirical Results: Ethical vs. Benchmark MABs

	Consumer Utility					Firm Reward			
	Majority	Interm.	Minority	Welfare	Discrim. Ratio	Majority	Interm.	Minority	Total
MAB Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Oracle	0.99	0.97	0.99	2.95	0.00	1,760.30	602.60	654.20	3,017.10
Utilitarian	0.70	0.97	0.97	2.64	0.25	1,330.40	596.00	639.20	2,565.60
Egalitarian	0.80	0.93	0.80	2.53	0.03	1,464.60	591.70	565.40	2,621.70
Proportional	0.70	0.98	0.98	2.66	0.26	1,327.70	613.60	630.80	2,572.10
Blind	0.98	0.99	0.46	2.43	0.50	1,745.40	616.90	346.10	2,708.40

[5] Here we use Batch 9 of Day 7 to illustrate our method. Results for other data batches are included in the robustness checks and the online appendix.

In Table 2, the first three columns show the utilities that consumers derive from featured news under each MAB scheme, segmented by majority, intermediate, and minority groups. Column (4) presents the unweighted sum of utilities across the three groups, representing the normative social welfare (Sen 1976). Column (5) reports a discrimination ratio, which captures the relative absolute gap between the minority's utility and the group size-weighted average utility of the population, with higher values indicating greater disparity. The final four columns present the corresponding firm-side rewards. We now turn to the interpretation of these results.

4.4.1 Cost of Privacy Compliance: Myth Busting

The oracle MAB assumes *full access* to sensitive user characteristics and can tailor news featuring to each consumer group. Thus, it achieves the highest levels of both consumer welfare (2.95) and firm reward (3,017.10). Notably, this customized decision-making also provides strong support for the minority consumers, achieving the highest level of parity with the discrimination ratio close to zero. However, as data privacy regulations increasingly limit access to sensitive user information, implementing such an oracle bandit is becoming practically infeasible.

In contrast, a MAB algorithm that is *blind* to sensitive user characteristics – thus complying with data protection legislation – learns to systematically discriminate against the minority. Under this policy, utility for the majority (0.98) is more than twice of the minority (0.46), and the minority's utility is about 50% lower than the population-weighted average, indicating a substantial level of discrimination. Consumer welfare also declines significantly, from 2.95 under the oracle to the lowest level of 2.43. A direct comparison between the oracle and blind bandits suggests that blanket prohibitions on the use of sensitive attributes may unintentionally harm the very groups they aim to protect, even in cases where the minority group exhibits a high average CTR (Table 1). While such privacy restrictions align with current legal frameworks, they may produce unintended adverse outcomes for marginalized customers and reduce overall welfare in MAB-driven online environments. These findings reinforce the growing argument among legal scholars that call for "myth-busting" the assumption that data protection laws should categorically prohibit the processing of equality-related data (Bekkum 2023).

4.4.2 Cost of Ethical Considerations

While the status quo policy (blind bandit) tends to favor the majority by predominantly featuring products aligned with mainstream preferences, incorporating ethical principles into the MAB framework substantially reduces this disparity. Specifically, the minority's utility rises from 0.46 (blind) to 0.80 (egalitarian), 0.97 (utilitarian), and 0.98 (proportional). However, these improvements cannot come without a cost. The majority's utility declines from 0.98 (blind) to 0.80 (egalitarian), 0.70 (utilitarian), and 0.70 (proportional), and firm-side rewards generated from majority consumers drop from 1,745.40 (blind) to 1,464.60 (egalitarian), 1,330.40 (utilitarian), and 1,327.70 (proportional). These results highlight that reducing discrimination against minorities inevitably imposes costs on the majority. In ethical MAB settings, gains for the minority are largely offset by losses to the majority, reflecting the intrinsic tension between equity and revenue. This finding is consistent with theoretical insights on ethical resource allocation and fairness-aware learning (Nikzad and Strack 2024; Li et al. 2025).

4.4.3 A Spectrum of Ethical Allocations

Given the shortcomings of the blind and oracle policies at opposite ends of the spectrum, the three proposed ethical MAB schemes warrant closer examination. These ethical alternatives introduce a *new* trade-off that balances firm rewards and discrimination, in addition to the classic exploration-exploitation trade-off inherent in the MAB framework.

The egalitarian MAB substantially increases utility for the minority group to 0.80, compared to just 0.46 under the blind policy. It also achieves a very low discrimination ratio of 0.03, indicating that the minority receives utility close to the population-weighted average and highlighting this parity-focused bandit's strong support for marginalized customers. While the utilitarian approach further improves minority utility to 0.97, it does so at the expense of other groups, resulting in a higher discrimination ratio of 0.25. The proportional bandit yields the highest normative welfare (2.66), but also exhibits a disparity of 0.26, similar to the utilitarian case. These empirical findings align with theoretical expectations that, while all three ethical MAB frameworks can promote equity, they do so through distinct mechanisms and entail different trade-offs. The utilitarian and proportional rules tend to maximize aggregate utility (as reflected in normative welfare in Column 4), whereas the egalitarian bandit appears to prioritize the well-being of the minorities while preserving overall system efficiency (Bertsimas et al. 2011; Nicosia et al. 2017).

From the firm's perspective, revenue generation follows a predictable pattern. With full access to consumer information, the oracle policy achieves the highest reward at 3,017.10. The blind MAB comes next at 2,708.40, as it optimizes for population-level engagement while ignoring consumer heterogeneity. In comparison, the three ethical bandits result in lower rewards at 2,621.70 (egalitarian), 2,565.60 (utilitarian), and 2,572.10 (proportional), highlighting the inherent cost of promoting parity. These results emphasize that ensuring equitable treatment for disadvantaged consumer segments requires firms to bear additional costs beyond those incurred under a blind policy. Therefore, firms and regulators should carefully weigh these trade-offs when designing online reinforcement learning systems such as the MABs.

The above discussion highlights that the ethical MAB models offer a structured framework for navigating the tension between revenue maximization and consumer protection, shaped by the explicit choice of underlying ethical principles.

4.5 Ethical Regret

While conventional regret in a MAB setting is defined on the firm side, we are interested in a different type of regret on the consumer side. Specifically, we aim to understand how the proposed ethical rebalancing in MABs evolves over time from the consumer's point of view. Whereas Table 2 provides a summary of ethical performance after learning, we now investigate how that performance develops throughout the learning process. To do so, we formally define *ethical regret*. In the MAB literature, regret is traditionally used to evaluate the optimality of policies, computed as the difference between the reward obtained by the chosen arms and the reward that would have been achieved had the MAB known the optimal arm to choose in each period (Lattimore and Szepesvari 2021). In the context of ethical MABs, however, the notion of "optimality" extends beyond firm-side reward maximization to include consumer-side considerations. Akin to conventional regret measures, ethical regret compares the arm pulled by a MAB to the arm that would have been optimal to pull under a specific ethical principle. For a given principle, the optimal arm is the one that maximizes ethical outcomes according to that principle. For example, under the utilitarian rule, it is the arm that maximizes the sum of group-level utilities.

To compute ethical regret, we need to obtain the average utility resulting from the articles featured by the MAB and those that would have resulted from a policy that always selects the utility maximizing article. For that, we first calculate the *true* utility associated with each article for each group by normalizing the expected CTR, $p_{ka} = E[\delta_{kt}(a)]$, based on the data used for the simulation. We refer to the normalized value of $E[\delta_{kt}(a)]$ as $v_k^{true}(a)$. Next, for a featuring policy π that has featured articles $a_1, ..., a_T$, we compute the average utility received by members of group k as $v_k^{true}(\pi) = \frac{\sum_{t=1}^{T} I_{k_t=k} v_k^{true}(a_t)}{\sum_{t=1}^{T} I_{k_t=k}}$, where $I_{k_t=k} = 1$ if the article featured by policy at time t was to a member of group k and zero otherwise. Given the group-level utilities attained by some policy π, and a policy π^* that always selects the utility maximizing arm, we calculate the principle-dependent ethical regret as follows:

$$\text{Utilitarian regret} = \sum_{k=1}^{K} v_k^{true}(\pi^*) - \sum_{k=1}^{K} v_k^{true}(\pi), \tag{4}$$

$$\text{Proportional regret} = \prod_{k=1}^{K} v_k^{true}(\pi^*) - \prod_{k=1}^{K} v_k^{true}(\pi), \tag{5}$$

$$\text{Egalitarian regret} = \min_k v_k^{true}(\pi^*) - \min_k v_k^{true}(\pi). \tag{6}$$

Notably, the upper bounds of ethical regret differ across ethical principles. This can be seen by considering a policy that consistently selects the article with the lowest utility. Such a policy yields an ethical regret of 1 under both the proportional and egalitarian principles, but results in an ethical regret of K under the utilitarian rule.

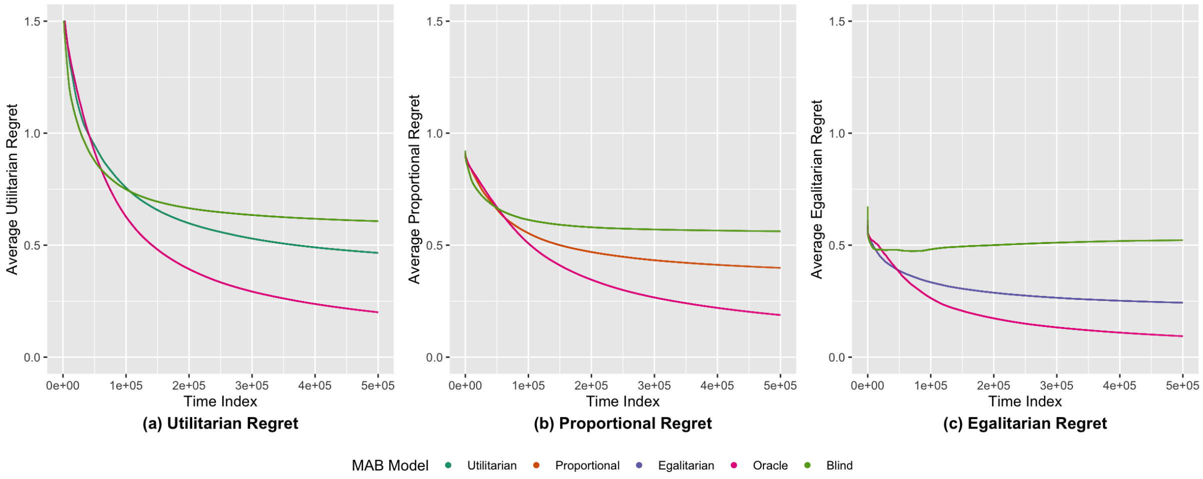


Figure 3: Ethical Regret over Time

Figure 3 presents the evolution of ethical regret over time. Each subfigure displays a specific type of ethical regret for three MAB models: (1) the oracle, (2) the blind, and (3) the ethical MAB corresponding to the regret type being plotted. For example, Subfigure (a) shows utilitarian regret for the oracle MAB, the blind MAB, and the utilitarian MAB. Across all subfigures, the regret trajectories initially follow similar paths before diverging. Notably, the blind MAB consistently exhibits the highest regret and is the first to stabilize, reflecting its rapid convergence to a policy that disregards ethical considerations. The oracle MAB, which can tailor its offerings to each consumer group, quickly outperforms the others in terms of ethical performance. Finally, although the ethical MABs optimize multi-dimensional objective functions, their convergence rates are not far behind the two benchmark MABs that optimize expected rewards.

4.6 Model Convergence

We now assess how each MAB learns the CTRs of different arms over time, using root mean squared error (RMSE) to quantify the discrepancy between true and estimated CTRs. The true values are derived from the RCT Yahoo dataset, while the estimated CTRs are calculated as the expected values of the Beta distributions associated with each arm in the TS procedure.

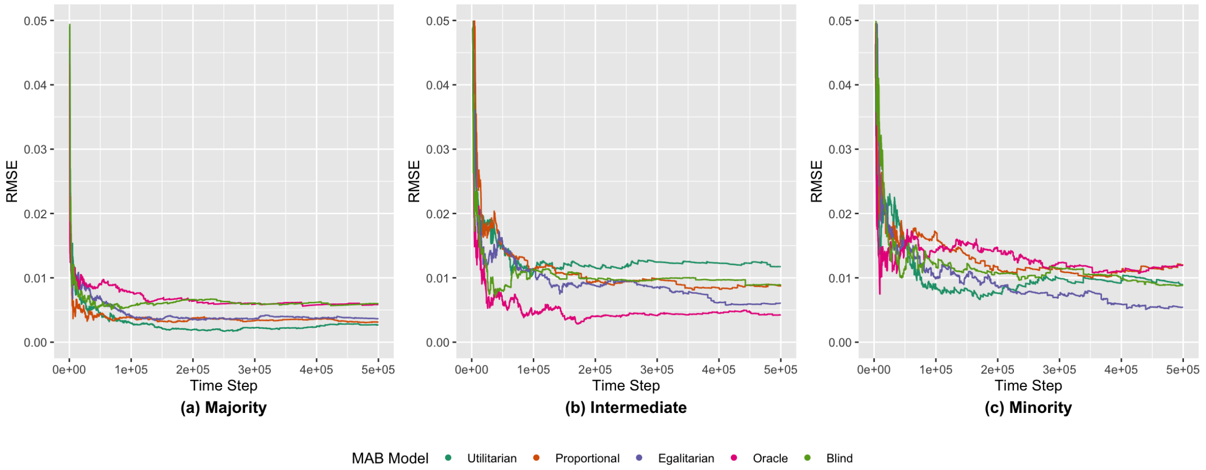


Figure 4: RMSE over Time

Figure 4 shows the evolution of RMSE during the MAB simulation on the Yahoo dataset across the three user groups. CTR estimates converge more quickly for larger groups. For example, the RMSE for majority consumers stabilizes rapidly around its steady-state value, whereas the minority group exhibits the slowest convergence and the highest final RMSE. By the end of the simulation, the utilitarian MAB yields the highest RMSE at 0.007, while the egalitarian MAB achieves the lowest at 0.005. The proportional MAB lies in between, with a final RMSE of 0.006.

4.7 Robustness Check: Different Group Sizes

We now examine how our results are influenced by the relative sizes of the three user groups. To do so, we replicate our main analysis using the same data sample but set the group sizes uniformly to 1/3. The simulation results are reported in Table 3. Although the groups are now equal in size, we retain the original group labels for consistency and ease of interpretation.

Table 3: Ethical vs. Benchmark MABs for Equal Group Size with Heterogeneous Preferences

	Consumer Utility					Firm Reward			
	Majority	Interm.	Minority	Welfare	Discrim. Ratio	Majority	Interm.	Minority	Total
MAB Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Oracle	0.98	0.98	1.00	2.96	0.01	822.40	1165.40	1859.60	3847.40
Utilitarian	0.69	0.97	0.99	2.65	0.12	628.30	1130.30	1832.30	3590.90
Egalitarian	0.78	0.92	0.82	2.52	0.02	675.40	1084.80	1571.40	3331.6
Proportional	0.70	0.97	0.98	2.65	0.11	619.50	1152.70	1812.70	3584.90
Blind	0.69	0.97	1.00	2.66	0.13	627.30	1148.20	1830.20	3605.70

Note: Group sizes are artificially set to $(1/3, 1/3, 1/3)$.

From Table 3, we observe that when group sizes are equal, consumer utility remains largely unchanged for both the oracle and the ethical MABs. However, firm rewards increase notably across all MABs. This is because the intermediate and minority groups, both of which exhibit higher CTRs, now account for a larger share of the population. Also, the consumer utility under the blind MAB changes substantially due to the shift in group sizes. Since the blind MAB selects the article that maximizes the population-level CTR, the most profitable article changes as the minority group becomes more influential. Finally, the results confirm the intuition that discrimination induced by the blind MAB diminishes under equal group sizes, reinforcing the point that group size plays a critical role in driving discriminatory outcomes.

4.8 Robustness Check: Homogeneous Consumer Preferences

When consumer preferences are largely uniform, we expect that MABs are less likely to discriminate against the minorities. To investigate this, we isolate a subset of data characterized by homogeneous preferences, where users predominantly click on the same featured articles. We identify such cases using a CTR-based classification. First we calculate the CTR for each article within a batch. Next we label a batch as "homogeneous" if the ratio of the lowest to highest CTR among articles is less than a threshold of 0.7. Finally all homogeneous batches are combined into a separate dataset. Table 4 presents the simulation results for the dataset with homogeneous preferences.

Table 4: Ethical vs. Benchmark MABs for Homogeneous Preferences

	Consumer Utility					Firm Reward			
	Majority	Interm.	Minority	Welfare	Discrim. Ratio	Majority	Interm.	Minority	Total
MAB Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Oracle	0.99	0.98	0.99	2.97	0.00	1502.40	539.60	580.50	2622.50
Utilitarian	1.00	1.00	1.00	2.99	0.00	1498.40	529.20	587.70	2615.30
Egalitarian	0.97	0.98	0.98	2.92	0.00	1459.90	527.80	587.20	2574.90
Proportional	1.00	1.00	1.00	2.99	0.00	1521.60	517.40	576.90	2615.90
Blind	1.00	1.00	1.00	2.99	0.00	1512.00	521.10	579.70	2612.80

When all users have the same preferences, all MAB models exhibit similar performance. In other words, when users display largely uniform preferences for certain products, the revenue-maximizing product also tends to be the most equitable choice. Although the discrimination measure changes sign between the benchmark and the ethical MABs, the differences remain minimal. Notably, the egalitarian policy that achieves the greatest reduction in discrimination also yields the lowest firm reward, once again highlighting the inherent trade-off between revenue and discrimination.

5 Conclusion

The integration of machine learning into marketing, particularly through the use of MABs, is widely recognized for its superior performance. In this paper, however, we demonstrate that in product-featuring applications, learning can also lead to increased discrimination against minority groups, especially when firms are prohibited from using sensitive demographic data. Specifically, we reveal a fundamental trade-off between revenue and equity in MAB applications, and we propose an adaptation of the TS algorithm to manage this trade-off by incorporating well-established ethical principles: utilitarian, proportional, and egalitarian. For each principle, we examine how the cost of reducing discrimination impacts both consumer utility and firm revenue.

Through the use of stochastic blinding and ethical rebalancing, our findings demonstrate the effectiveness of the proposed strategies in mitigating discriminatory biases, enabling firms to pursue more inclusive marketing practices without violating regulatory mandates. We encourage future research to extend stochastic blinding to other popular MAB formulations, such as Upper Confidence Bound, Gittins index, and related approaches.

This research is not without limitations. First, the effectiveness of our approach is bounded by the degree of preference heterogeneity across consumer groups. If all consumers share similar preferences, there is no need to mitigate discrimination, as MABs will naturally learn to serve the universally preferred products. Second, the scope of our study is limited by the difficulty of obtaining large-scale RCT data from real-world internet platforms. As a result, our analysis is confined to the context of news featuring on the Yahoo platform. Future research could expand the application of our framework by testing it in real-world online environments, thereby evaluating its scalability and adaptability across different industries and marketing settings. Finally, our current analysis considers only mutually exclusive ethical strategies. Future work could explore the effectiveness of mixed strategies, such as a hybrid ethical principle combining, for example, 50% egalitarian and 50% utilitarian.

References

Agarwal, Deepak, Bo Long, Jonathan Traupman, Doris Xin and Liang Zhang (2014), Laser: A scalable response prediction platform for online advertising, *in* 'Proceedings of the 7th ACM international conference on Web search and data mining', pp. 173–182.

Allender, William J., Jura Liaukonyte, Sherif Nasser and Timothy J. Richards (2021), 'Price fairness and strategic obfuscation', *Marketing Science* **40**(1), 122–146.

Amanatidis, Georgios, Haris Aziz, Georgios Birmpas, Aris Filos-Ratsikas, Bo Li, Hervé Moulin, Alexandros A Voudouris and Xiaowei Wu (2023), 'Fair division of indivisible goods: Recent progress and open questions', *Artificial Intelligence* **322**, 103965.

Ascarza, Eva and Ayelet Israeli (2022), 'Eliminating unintended bias in personalized policies using bias-eliminating adapted trees (beat)', *Proceedings of the National Academy of Sciences* **119**(11), e2115293119.

Banerjee, Siddhartha, Vasilis Gkatzelis, Safwan Hossain, Billy Jin, Evi Micha and Nisarg Shah (2022), 'Proportionally fair online allocation of public goods with predictions', *arXiv preprint arXiv:2209.15305* .

Barocas, Solon, Moritz Hardt and Arvind Narayanan (2023), *Fairness and Machine Learning: Limitations and Opportunities*, MITPress.

Bekkum, Borgesius (2023), 'Using sensitive data to prevent discrimination by artificial intelligence. does the gdpr need a new exception?', *Computer Law and Security Review* **48**(4), 2–12.

Bentham, Jeremy (2024), From an introduction to the principles of morals and legislation. printed in the year 1780, and now first published, *in* 'Literature and Philosophy in Nineteenth Century British Culture', Routledge, pp. 261–268.

Bertsimas, Dimitris and Jack Dunn (2019), *Machine learning under a modern optimization lens*, Dynamic Ideas LLC Waltham.

Bertsimas, Dimitris, Vivek F Farias and Nikolaos Trichakis (2011), 'The price of fairness', *Operations Research* **59**(1), 17–31.

Binns, Reuben (2020), On the apparent conflict between individual and group fairness, *in* 'Proceedings of the 2020 conference on fairness, accountability, and transparency', pp. 514–524.

Bolton, Lisa E., Hean Tat Keh and Joseph W. Alba (2010), 'How do price fairness perceptions differ across culture?', *Journal of Marketing Research* **47**(3), 564–576.

Bolton, Lisa E. and Joseph W. Alba (2006), 'Price Fairness: Good and Service Differences and the Role of Vendor Costs', *Journal of Consumer Research* **33**(2), 258–265.

Bolton, Lisa E., Luk Warlop and Joseph W. Alba (2003), 'Consumer Perceptions of Price (Un)Fairness', *Journal of Consumer Research* **29**(4), 474–491.

Cambridge Learner's Dictionary (2024).

Campbell, Margaret C. (1999), 'Perceptions of price unfairness: Antecedents and consequences', *Journal of Marketing Research* **36**(2), 187–199.

Cappelen, Alexander W., Astri D. Hole, Erik Ø. Sørensen and Bertil Tungodden (2007), 'The pluralism of fairness ideals: An experimental approach', *American Economic Review* **97**(3), 818–827.

Celis, L. Elisa, Lingxiao Huang, Vijay Keswani and Nisheeth K. Vishnoi (2020), 'Classification with fairness constraints: A meta-algorithm with provable guarantees'.

Chapelle, Olivier and Lihong Li (2011), 'An empirical evaluation of thompson sampling', *Advances in neural information processing systems* **24**.

Chu, Xiaofeng and Seok-Woo Park (2009), ‘A case study of behavior-driven conjoint analysis on yahoo! front page today module’, *ACM Transactions on Information Systems* **27**(3), 10:1–10:31.

Coenen, Anna (2019), ‘How the new york times is experimenting with recommendation algorithms’, *The New York Times* .
URL: *https://open.nytimes.com/how-the-new-york-times-is-experimenting-with-recommendation-algorithms-562f78624d26*

Cohen, Maxime C., Adam N. Elmachtoub and Xiao Lei (2022), ‘Price Discrimination with Fairness Constraints’, *Management Science* **68**(12), 8536–8552.

Cui, Tony Haitao, Jagmohan S. Raju and Z. John Zhang (2007), ‘Fairness and channel coordination’, *Management Science* **53**(8), 1303–1314.

Cui, Tony and Paola Mallucci (2016), ‘Fairness ideals in distribution channels’, *Journal of Marketing Research* **53**, 969–987.

Diao, Wen, Mushegh Harutyunyan and Baojun Jiang (2023), ‘Consumer Fairness Concerns and Dynamic Pricing in a Channel’, *Marketing Science* **42**(3), 569–588.

Dinur, Irit and Kobbi Nissim (2003), Revealing information while preserving privacy, *in* ‘Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems’, Association for Computing Machinery, pp. 202–210.

Dwork, Cynthia and Aaron Roth (2014), *The algorithmic foundations of differential privacy*, Vol. 9 of *Foundations of Trends in Theoretical Computer Science.*

Dwork, Cynthia, Nishanth Kohli and David Mulligan (2019), ‘Differential privacy in practice: Expose your epsilons!’, *Journal of Privacy and Confidentiality* **9**(2).

Fleurbaey, Marc, SM Ravi Kanbur and Dennis J Snower (2023), *An Analysis of Moral Motives in Economic and Social Decisions*, Centre for Economic Policy Research.

Fu, Runshan, Manmohan Aseri, Param Vir Singh and Kannan Srinivasan (2022), ‘“un” fair machine learning algorithms’, *Management Science* **68**(6), 4173–4195.

Guo, Xiaomeng and Baojun Jiang (2016), ‘Signaling through price and quality to consumers with fairness concerns’, *Journal of Marketing Research* **53**(6), 988–1000.

Habbal, Adib, Mohamed Khalif Ali and Mustafa Ali Abuzaraida (2024), ‘Artificial intelligence trust, risk and security management (ai trism): Frameworks, applications, challenges and future research directions’, *Expert Systems with Applications* **240**, 122442.

Harsanyi, John C (1955), ‘Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility’, *Journal of political economy* **63**(4), 309–321.

Hauser, John R, Glen L Urban, Guilherme Liberali and Michael Braun (2009), ‘Website morphing’, *Marketing Science* **28**(2), 202–223.

Haws, Kelly L. and William O. Bearden (2006), ‘Dynamic Pricing and Consumer Fairness Perceptions’, *Journal of Consumer Research* **33**(3), 304–311.

Heidari, Hoda, Michele Loi, Krishna P Gummadi and Andreas Krause (2019), A moral framework for understanding fair ml through economic models of equality of opportunity, *in* ‘Proceedings of the conference on fairness, accountability, and transparency’, pp. 181–190.

Hill, Daniel N, Houssam Nassif, Yi Liu, Anand Iyer and SVN Vishwanathan (2017), An efficient bandit algorithm for realtime multivariate optimization, *in* ‘Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining’, pp. 1813–1821.

Ho, Teck-Hua, Xuanming Su and Yaozhong Wu (2014), ‘Distributional and peer-induced fairness in supply chain contract design’, *Production and Operations Management* **23**(2), 161–175.

Holm, Sune (2023), 'Egalitarianism and algorithmic fairness', *Philosophy & Technology* **36**(1), 6.

Joseph, Matthew, Michael Kearns, Jamie H Morgenstern and Aaron Roth (2016), 'Fairness in learning: Classic and contextual bandits', *Advances in neural information processing systems* **29**.

Kasy, Maximilian and Anja Sautmann (2021), 'Adaptive treatment assignment in experiments for policy choice', *Econometrica* **89**(1), 113–132.

Koenigs, Michael, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser and Antonio Damasio (2007), 'Damage to the prefrontal cortex increases utilitarian moral judgements', *Nature* **446**(7138), 908–911.

Lagarias, Jeffrey C, James A Reeds, Margaret H Wright and Paul E Wright (1998), 'Convergence properties of the nelder–mead simplex method in low dimensions', *SIAM Journal on optimization* **9**(1), 112–147.

Lambrecht, Anja and Catherine Tucker (2019), 'Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads', *Management science* **65**(7), 2966–2981.

Lattimore, Tor and Csaba Szepesvari (2021), *Bandit Algorithms*, Cambridge University Press.

Li, Fengjiao, Jia Liu and Bo Ji (2019), 'Combinatorial sleeping bandits with fairness constraints'.

Li, Lihong, Wei Chu, John Langford and Xuanhui Wang (2011), Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms, *in* 'Proceedings of the fourth ACM international conference on Web search and data mining', pp. 297–306.

Li, Xiaolong, Ying Rong, Renyu Zhang and Huan Zheng (2025), 'Online advertisement allocation under customer choices and algorithmic fairness', *Management Science* **71**(1), 825–843.

Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg and Demis Hassabis (2015), 'Human-level control through deep reinforcement learning', *Nature* **518**(7540), 529–533.
URL: *http://dx.doi.org/10.1038/nature14236*

Nelder, John A and Roger Mead (1965), 'A simplex method for function minimization', *The computer journal* **7**(4), 308–313.

Nicosia, Gaia, Andrea Pacifici and Ulrich Pferschy (2017), 'Price of fairness for allocating a bounded resource', *European Journal of Operational Research* **257**(3), 933–943.

Nikzad, Afshin and Philipp Strack (2024), 'Equity and efficiency in dynamic matching: Extreme waitlist policies', *Management Science* **70**(8), 5187–5207.

Patil, Indrajeet, Micaela Zucchelli, Wouter Kool, Stephanie Campbell, Federico Fornasier, Marta Calò, Giorgia Silani, Mina Cikara and Fiery Cushman (2020), 'Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures', *Journal of Personality and Social Psychology* **120**.

Ponte, Gilian R., Jaap E. Wieringa, Tom Boot and Peter C. Verhoef (2024), 'Where's waldo? a framework for quantifying the privacy-utility trade-off in marketing applications', *International Journal of Research in Marketing* .

Rawls, John (1971), *A Theory of Justice*, Harvard University Press, Cambridge, MA.

Rigobon, Daniel E. (2023), 'From utilitarian to rawlsian designs for algorithmic fairness', *arXiv preprint arXiv:2302.03567* .

Roberts, Kevin W. S. (1980), 'Interpersonal comparability and social choice theory', *The Review of Economic Studies* **47**(2), 421–439.

Russo, Daniel and Benjamin Van Roy (2014), 'Learning to optimize via posterior sampling', *Mathematics of Operations Research* **39**(4), 1221–1243.

Sanfey, Alan G., James K. Rilling, Jessica A. Aronson, Leigh E. Nystrom and Jonathan D. Cohen (2003), 'The neural basis of economic decision-making in the ultimatum game', *Science* **300**(5626), 1755–1758.

Scott, Steven L. (2010), 'A modern bayesian look at the multi-armed bandit', *Applied Stochastic Models in Business and Industry* **26**(6), 639–658.

Sen, Amartya (1976), 'Welfare inequalities and rawlsian axiomatics', *Theory and decision* **7**(4), 243–262.

Sun, Tianshu, Zhe Yuan, Chunxiao Li, Kaifu Zhang and Jun Xu (2024), 'The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy', *Management Science* **70**(4), 2645–2660.

Thompson, W. R. (1933), 'On the likelihood that one unknown probability exceeds another in view of the evidence of two samples', *Biometrika* **25**(3-4), 285–294.

Thomson, William (2011), Fair allocation rules, *in* 'Handbook of social choice and welfare', Vol. 2, Elsevier, pp. 393–506.

Vanchinathan, Hastagiri P, Andreas Marfurt, Charles-Antoine Robelin, Donald Kossmann and Andreas Krause (2015), Discovering valuable items from massive data, *in* 'Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 1195–1204.

Vanchinathan, Hastagiri P, Isidor Nikolic, Fabio De Bona and Andreas Krause (2014), Explore-exploit in top-n recommender systems via gaussian processes, *in* 'Proceedings of the 8th ACM Conference on Recommender systems', pp. 225–232.

Wang, Lequn, Yiwei Bai, Wen Sun and Thorsten Joachims (2021), Fairness of exposure in stochastic bandits, *in* 'International Conference on Machine Learning', PMLR, pp. 10686–10696.

Xia, Lan, Kent Monroe, Jennifer Cox, B Kent, Kent Monroe and J Jones (2004), 'The price is unfair! a conceptual framework of price fairness perceptions', *Journal of Marketing* **68**, 1–15.

Zhang, Chen, Yu Xie, Hang Bai, Bin Yu, Weihong Li and Yuan Gao (2021), 'A survey on federated learning', *Knowledge-Based Systems* **216**, 106775.

Žliobaitė, Indrė (2017), 'Measuring discrimination in algorithmic decision making', *Data Mining and Knowledge Discovery* **31**(4), 1060–1089.

APPENDICES

Appendix 1: CTR and Utility Results

To help interpret the main results in Table 2, we provide the group-level CTRs and utilities associated with each article in our simulation in Table 5.

Table 5: Click Through Rates and Consumer Utility

	Click Through Rate			Consumer Utility		
	Majority	Interm.	Minority	Majority	Interm.	Minority
Article	(1)	(2)	(3)	(4)	(5)	(6)
1	0.015	0.020	0.033	0.270	0.000	0.109
2	0.040	0.056	0.077	0.713	0.605	0.778
3	0.027	0.040	0.058	0.423	1.000	0.474
4	0.047	0.058	0.074	0.949	0.630	0.641
5	0.027	0.041	0.071	0.504	0.666	0.828
6	0.036	0.052	0.067	0.799	0.488	0.805
7	0.030	0.044	0.069	0.686	0.531	0.620
8	0.036	0.045	0.067	0.716	0.425	0.659
9	0.018	0.025	0.042	0.800	0.161	0.077
10	0.019	0.033	0.051	0.525	0.838	0.336
11	0.010	0.024	0.043	0.087	0.157	0.000
12	0.025	0.041	0.071	0.207	0.407	0.580
13	0.030	0.038	0.049	0.753	0.325	0.506
14	0.032	0.058	0.073	0.611	0.784	0.915
15	0.030	0.043	0.048	0.467	0.386	0.340
16	0.008	0.013	0.017	0.000	0.009	0.126
17	0.017	0.035	0.051	0.250	0.969	0.360
18	0.051	0.070	0.058	1.000	0.897	0.531
19	0.032	0.039	0.062	0.827	0.431	0.418
20	0.032	0.031	0.033	0.735	0.918	0.311
21	0.025	0.029	0.055	0.575	0.412	0.352
22	0.020	0.032	0.056	0.214	0.217	0.330
23	0.020	0.029	0.052	0.307	0.299	0.532
24	0.037	0.069	0.111	0.713	0.793	1.000
25	0.031	0.052	0.072	0.519	0.915	0.586
26	0.032	0.048	0.062	0.482	0.504	0.770