Ethical Thompson Sampling for Marketing Resource Allocation

by

Bob Rombach

Yang Li

Gui Liberali*

November 21st, 2026

*Corresponding author: Gui Liberali. All authors contributed equally to this work.

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.

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

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.

The authors gratefully acknowledge the helpful comments of participants at the 2023 Choice Symposium at INSEAD, at the Marketing Science Conference (2023, 2024, and 2025) and at the Marketing Lunchclub at Rotterdam School of Management, Erasmus University. The authors also thank Paulo Albuquerque, Bruce Hardie, Ricardo Montoya, Gilian Ponte, and Francesca Sotgiu for the detailed and insightful comments. This research was supported by the Erasmus Research Institute of Management (ERIM), and the Erasmus Centre for Optimization of Online Experiments.

Ethical Thompson Sampling for Marketing Resource Allocation

Abstract

Firms often rely on adaptive algorithms to allocate scarce marketing resources. Maximizing aggregate performance can disadvantage smaller consumer segments, while regulations such as the GDPR constrain personalization as a remedy. We develop an ethical Thompson Sampling (TS) framework that learns segment preferences in a restricted, demographic-aware module while allocating resources without using the arriving consumer's segment, and that integrates utilitarian, proportional, and egalitarian principles. Under standard assumptions, ethical allocations approach their full information targets and ethical regret grows sublinearly. Our central contribution is the dynamic price of fairness (PoF), which measures expected click losses relative to standard TS ("blind") that ignores segments over the learning horizon. We decompose it into the static cost of fairness and the difference in learning costs between ethical and blind policies. Simulations calibrated with randomized exposure data from Yahoo and ZOZO show that ethical policies improve minority utility when segment preferences conflict, but firms pay most of the cost early while benefits arrive gradually. In sparse feedback environments, like ZOZO's, dynamic PoF can exceed its static counterpart and only a fraction of the anticipated minority gains is delivered. We show that accounting for when costs and benefits are realized is critical to evaluating ethical allocation.

Keywords: Multiarmed Bandits; Algorithmic Fairness; Ethical Principles; Price of Fairness; Product Featuring; Consumer Privacy.

Ethical Thompson Sampling for Marketing Resource Allocation

1 Introduction

Marketing resources such as featured slots, promotions, and ad impressions are finite, and how firms allocate them affects more than aggregate revenue. When a firm selects products or promotions for heterogeneous customers, its choices also determine whose preferences get accommodated. An algorithm maximizing aggregate response may favor alternatives preferred by larger consumer segments, leaving smaller segments underserved even when all consumers face the same allocation distribution. Such disparities can arise without any intent to discriminate, from the interaction between resource constraints, heterogeneous preferences, and an objective giving greater influence to consumers who contribute more heavily to aggregate performance.

Consider a news platform deciding which articles to feature on its front page (Coenen 2019; Figure 1). Prominent display positions are limited, and choosing one article for a position necessarily excludes others. The platform initially does not know which articles different readers prefer. It must learn through successive interactions, balancing exploration of articles with uncertain appeal against exploitation of those already believed to perform well. Yet as it improves at maximizing clicks, a standard bandit that pools feedback across all readers, which we call blind TS, may learn to serve majority preferences while giving insufficient attention to articles valued by smaller segments. The challenge also includes deciding whose preferences should guide featuring decisions as learning unfolds.

The standard remedy is a bandit that uses segment membership as context, in effect running a separate bandit for each segment, which we call personalized TS. It serves each segment's preferences, but it conditions each featuring decision on the consumer's sensitive attributes, and firms must respect restrictions on processing such information. For example, the European Union's General Data Protection Regulation (GDPR) generally prohibits processing specified

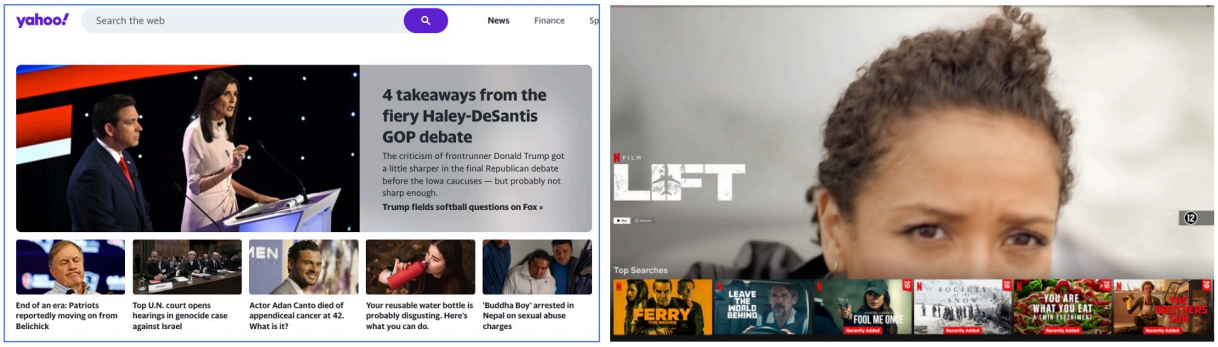


Figure 1: Finite Resource Allocation in Featuring News and Movies

categories of sensitive data, including racial or ethnic origin and religious beliefs, unless an applicable exception permits it. Removing demographic information from a decision system, however, does not necessarily remove unequal outcomes. This tension motivates our research, i.e., how can a firm learn to allocate scarce resources more equitably when sensitive identity information is available for restricted learning but cannot be used to personalize individual allocation decisions?

We address this challenge by developing an ethical multiarmed bandit (MAB) based on Thompson Sampling (TS; Thompson 1933; Scott 2010; Russo and Van Roy 2014; Schwartz et al. 2017). This bandit decouples demographic-aware posterior preference learning from demographic-blind resource allocation. A restricted learning module uses sensitive customer information from completed interactions to learn segment-specific preferences, whereas a separate allocation module uses these accumulated beliefs to sample products from a common, ethically optimized probability distribution without accessing the current consumer's segment identity. Sensitive information about the current consumer informs subsequent learning rather than the allocation this consumer receives.

Building on this separation of information, the firm uses learned historical preferences to guide allocation decisions under three established ethical principles, i.e., utilitarian, which prioritizes aggregate utility (Bentham 2024 [1789]); proportional, which focuses on Nash social welfare (Kaneko and Nakamura 1979); and egalitarian, which maximizes the utility of the worst off segment (Heidari et al. 2019). Personalized TS and blind TS serve as our benchmarks: personalized TS uses segment information in both learning and allocation, and blind TS uses it in neither. We also establish theoretical guarantees for the ethical bandits and extend the *static* price of fairness (PoF) from Bertsimas et al. (2011) to a sequential learning and allocation setting.

This framework requires processing sensitive segment information to identify and audit util-

ity disparities across demographic groups and to learn how to mitigate them. Recent research also incorporates sensitive information into equitable allocation. For instance, Chohlas-Wood et al. (2026) allow policies to condition directly on protected attributes while acknowledging restrictions on their use. We study a more restrictive information regime in which segment labels are used solely for posterior preference learning, while the current consumer's identity is excluded from the allocation decision.

We evaluate ethical TS via offline simulations calibrated with randomized exposure data from Yahoo! News and the fashion ecommerce platform ZOZO. We compare consumer utility and firm rewards across ethical and benchmark policies, examine learning dynamics and the decomposition of fairness costs, and assess sensitivity to the number and relative sizes of consumer segments. Both applications show that demographic-blind allocation policies can leave minority consumers underserved. Ethical policies can mitigate these disparities, but their benefits and click costs vary across objectives and data environments. In Yahoo, minority improvements are substantial with relatively modest click losses, whereas proportional and egalitarian policies in ZOZO deliver greater minority improvements than utilitarian TS at higher costs. Robustness analyses show that gains are constrained when minorities provide too little feedback or are already well served by blind TS.

Importantly, sequential learning shapes both the realized price of fairness and the timing of fairness benefits. We define dynamic PoF as the percentage reduction in cumulative expected clicks under ethical TS relative to blind TS over the same learning horizon. After the first 10% of Yahoo arrivals, for example, the ethical policies deliver only 13%–20% of their full information minority utility gains, while their dynamic PoFs already reach 71%–132% of their respective static values. The gap between costs and benefits is more pronounced in ZOZO's sparser feedback environment. Over 1.37 million arrivals, egalitarian TS has a dynamic PoF of 25.1%, compared with a static PoF of 23.3%, but delivers only 41% of its static minority utility gain. Moreover, utilitarian and proportional TS spend much of the ZOZO learning horizon performing worse than blind TS under their own welfare criteria. These findings show that substantial click costs can precede both minority benefits and improvements in the ethical objective the policy seeks to optimize. Evaluating fairness requires considering the benefits actually delivered alongside the costs incurred during learning.

We make three main contributions. First, we propose a novel ethical bandit framework for product featuring that integrates fairness considerations into sequential learning and allocation. Beyond the conventional exploration-exploitation tradeoff, our model explicitly addresses the

tradeoffs among firm revenue, consumer welfare, and restrictions on the use of sensitive privacy information. By incorporating alternative ethical principles into a common learning framework, our approach provides a unified method for evaluating and comparing allocation policies according to firms' fairness objectives and the marketing contexts in which they operate.

Second, central to our study, we contribute a dynamic perspective on the efficiency-equity tradeoff by distinguishing the cost of pursuing fairness from the fairness actually delivered during learning. Extending the PoF literature, we show both theoretically and empirically that sequential learning changes both the magnitude and timing of this tradeoff. A key information constraint arises because ethical policies treat segments symmetrically, while observations accumulate in proportion to segment size and sparse clicks further slow learning. As a result, substantial efficiency costs can emerge well before a policy delivers much of its full information fairness benefit. Our analysis thus underscores the importance of evaluating ethical policies by the costs and benefits realized along the learning path, in addition to their ultimate fairness targets.

Third, we establish statistical guarantees for the proposed ethical bandits. We introduce *ethical regret*, a welfare-based measure of learning performance that captures the cumulative shortfall relative to the full information allocation maximizing the selected ethical objective. Under standard regularity conditions, we show that ethical regret is sublinear and the ethical weights approach their full information optimal set. We also bound ethical regret by the preference estimation error and use martingale concentration to show that average product sampling errors vanish asymptotically. These results confirm that the ethical objectives can be integrated into TS without losing its desirable convergence guarantees.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces the ethical allocation principles. Section 4 proposes the ethical bandit model and introduces the dynamic price of fairness. Section 5 provides the theoretical guarantees of ethical TS. Section 6 presents empirical applications, examines the dynamics of learning, and reports robustness checks. Section 7 concludes with directions for future research.

2 Literature Review

Our research builds on several intersecting streams of literature in marketing, computer science, and operations research, including fairness in bandit algorithms, personal data restrictions and privacy, fairness research in marketing, and ethical principles for allocating scarce

resources. While each stream illuminates an important aspect of our problem, none jointly addresses its defining information structure and allocation mechanism. Below, we briefly review these literatures and highlight our contributions to each.

2.1 Fairness in Bandits

Research on fair bandits has developed several directions to model equitable treatment under sequential learning and decision making. One line of research focuses on how selection opportunities are distributed across arms. Joseph et al. (2016) introduce *meritocratic fairness*, requiring that an arm with a lower expected reward not be selected with greater probability than a higher-reward arm. Because expected rewards are initially unknown, this requirement constrains exploration and creates a tradeoff between fairness and learning efficiency. Liu et al. (2017) broaden this perspective through *smooth fairness*, under which arms with similar reward distributions receive similar selection probabilities, and *calibrated fairness*, under which selection probabilities reflect the likelihood that an arm generates the highest reward. Gillen et al. (2018) further consider individual fairness when the relevant similarity metric is unknown and must be learned alongside the allocation policy. Other approaches impose explicit *allocation guarantees*. For instance, Li et al. (2019) study minimum long run selection fractions in combinatorial sleeping bandits, while Patil et al. (2021) develop a general stochastic bandit framework with prespecified minimum selection shares. Connecting allocation shares to merit, Wang et al. (2021) propose the FairX framework for *fairness of exposure*, which allocates exposure in proportion to merit and distinguishes reward regret from fairness regret measuring deviations from the target exposure distribution. These approaches conceptualize arm level fairness through merit- or similarity-based treatment, minimum allocation shares, or proportional exposure.

The literature has also extended fairness from individual arms to groups and multiple stakeholders. At the group level, Schumann et al. (2019) model group-specific biases in contextual bandit rewards, while Grazzi et al. (2022) define *group meritocratic fairness* by evaluating candidates relative to the reward distributions of their own groups. In multiagent settings, Hossain et al. (2021) use welfare objectives, including Nash social welfare, to distribute rewards among agents with heterogeneous arm valuations. A related maximin fairness stream protects the worst off player. Bistritz et al. (2020) develop an algorithm for learning maximin fair matchings between players and arms under collisions, and Wang et al. (2025) derive a decentralized algorithm with order optimal regret. Fairness may also concern the costs of learning. Baek and Farias (2021) use Nash bargaining to distribute exploration costs across groups relative to their

standalone learning outcomes, whereas Caiata et al. (2026) distinguish outcome-based fairness from procedural fairness in the distribution of decision making power. These studies broaden fair bandits from arm level treatment to the distribution of rewards, opportunities, learning costs, and decision authority across groups and agents. A related stream evaluates fairness across rounds rather than across groups: Sarkar et al. (2026) characterize the minimax regret cost of p-mean welfare over the sequence of per-round expected rewards, a price of fairness that arises when early participants bear the burden of exploration. In contrast, we study fairness across consumer segments under a common allocation and compare the efficiency cost realized during learning with its full information benchmark.

Closely related to our work, Chohlas-Wood et al. (2026) learn equitable allocations by repeatedly solving a policy optimization problem within a contextual bandit. Their policies can condition on individual covariates, including protected attributes, and they acknowledge that legal restrictions on using such attributes remain an open issue. In contrast, we directly address these information use restrictions through an information separation mechanism, which confines sensitive attributes to posterior preference learning while ensuring the allocation decision is independent of the arriving consumer's demographic identity. Moreover, whereas their theoretical sample complexity analysis focuses on nonadaptive data collection, we establish asymptotic convergence and sublinear ethical regret for adaptive ethical bandits under standard regularity conditions. More broadly, we contribute to the fairness in bandits literature by comparing utilitarian, proportional, and egalitarian objectives within a *unified* framework and examining the *dynamic* price of fairness as preferences are learned. Our empirical analysis reveals a temporal mismatch between efficiency costs and fairness benefits, and assesses the extent to which full information benchmarks mischaracterize these tradeoffs over finite learning horizons.

2.2 Personal Data Restrictions and Privacy

Another literature examines the consequences of limiting data availability for marketing decisions. Rafieian and Yoganarasimhan (2021) investigate the relationship between consumer targeting and privacy in mobile advertising, highlighting the importance of consumer information for advertising decisions. Using field experiments, Sun et al. (2024) show that restrictions on personal data reduce the profitability of internet commerce and disproportionately harm niche merchants and their customers. Korganbekova and Zuber (2023) similarly study browser-imposed privacy restrictions and document adverse results for price sensitive consumers. Re-

lated work demonstrates that apparently neutral personalization policies can result in unequal outcomes when observable covariates proxy for protected characteristics or when optimization favors large, profitable, or easily reached populations (Ascarza and Israeli 2022). These studies suggest that restricting personal data does not necessarily eliminate distributional harm and may instead shift who bears it.

While existing studies document the consequences of data restrictions, they do not directly address how an adaptive decision maker can learn heterogeneous consumer preferences while excluding the arriving consumer's demographic information from the allocation rule. In our model, sensitive privacy data are solely used within a restricted learning module to estimate heterogeneous preferences before constructing a common population level ethical policy. This differs from differential privacy, which aims to limit information disclosure through formal privacy guarantees (Dinur and Nissim 2003; Dwork and Roth 2014; Dwork et al. 2019; Ponte et al. 2024). Our paper thus complements empirical evidence on the costs of data restrictions with a *dynamic* learning model of how firms can manage their allocation consequences.

2.3 Fairness Research in Marketing

Marketing research has extensively examined fairness as both a consumer judgment and a strategic consideration for firms. Behavioral studies develop frameworks for price fairness perceptions and their consequences (Bolton and Alba 2006; Bolton et al. 2003, 2010; Campbell 1999; Cappelen et al. 2007; Haws and Bearden 2006; Xia et al. 2004), while related work shows that distributional concerns enter individual decision making (Koenigs et al. 2007; Sanfey et al. 2003). Analytic modelers study fairness in channel coordination (Cui et al. 2007; Cui and Mallucci 2016; Ho et al. 2014), equitable selling and inequity aversion (Guo and Jiang 2016), strategic responses to fairness constraints (Cohen et al. 2022; Fu et al. 2022), and interventions intended to reduce unfair outcomes (Allender et al. 2021; Diao et al. 2023). Empirical work further shows that optimization can create unequal delivery even without an explicit discriminatory objective. For example, Lambrecht and Tucker (2019) document how cost effective advertising can generate gender-skewed exposure via market competition and crowding out.

While much of the fairness in marketing literature is conceptual, experimental, or theoretical, it leaves open how a firm should learn about uncertain consumer preferences and make fair sequential decisions under restrictions on consumer information. We address this gap by developing a novel MAB framework that integrates consumer heterogeneity, ongoing experimentation, restrictions on privacy data, and explicit fairness objectives into one unified model.

2.4 Ethical Principles in Resource Allocation

The ethical objectives applied in this paper are established principles of distributive justice and resource allocation rather than new fairness definitions. The utilitarian allocation maximizes aggregate utility, the egalitarian or maximin policy prioritizes the utility of the worst off party, and the proportional rule optimizes the Nash social welfare (Thomson 2011; Cappelen et al. 2007; Bertsimas and Dunn 2019; Rigobon 2023; Bentham 2024 [1789]; Heidari et al. 2019; Binns 2020). These objectives encode different judgments about the relative importance of aggregate gains versus interpersonal disparities.

Operations research has made these efficiency implications explicit. In particular, Bertsimas et al. (2011) use the utilitarian solution as an efficiency benchmark and characterize the price of adopting proportional and maximin fairness. Bertsimas et al. (2012) place utilitarian, proportional, and maximin allocations within a constant elasticity welfare and analyze the price of fairness versus efficiency. Related work studies online public goods allocation under predictive information (Banerjee et al. 2022). Thus, this literature establishes that ethical allocations result in an efficiency-equity tradeoff and that stronger protection of the worst off would lead to a loss in aggregate welfare (Roberts 1980).

We use these established results as the static foundation of our model and formalize the three ethical criteria via a generalized power mean family. Our contribution is to incorporate these criteria in a sequential bandit setting and extend the literature on the static price of fairness (Bertsimas et al. 2011) to a dynamic decision environment with learning.

3 Ethical Allocation Principles

Before introducing our ethical MAB framework, we first review its static foundation by considering allocation rules under full information, where product preferences are known. This setting allows us to distinguish between demographic-blind and demographic-aware allocations and to summarize established ethical criteria that motivate the development of our ethical TS models in subsequent sections.

3.1 Static Allocation Setting

At a featuring opportunity, the firm chooses among A available products, or arms, indexed by $a = 1, \ldots, A$, to feature to its visitor. Each consumer belongs to one of K mutually exclusive

demographic segments, indexed by $k = 1, \ldots, K$. Let $\pi_k > 0$ denote the population share of segment k, with $\sum_{k=1}^{K} \pi_k = 1$. Segment membership is defined using a sensitive demographic attribute whose use in individual featuring decisions may be subject to legal or organizational restrictions.

When product a is featured to a consumer of type k, the observed click response follows Bernoulli(p_{ka}), where $p_{ka} \in (0, 1)$ is the segment-product CTR. Let $P = \{p_{ka}\}$ denote the CTR matrix. Because CTRs reflect both baseline click propensities and relative product preferences, we scale them within each segment to measure how well an allocation accommodates that segment's preferences

$$\mathcal{N}_k(p)_a = \frac{p_{ka}}{\max_b p_{kb}}. \tag{1}$$

Thus, a scaled utility of 0.8 means achieving 80% of that segment's best available CTR. Denote $v_{ka} = \mathcal{N}_k(P)_a \in (0, 1]$ the scaled utility of product a for segment k, and $V = \{v_{ka}\}$ the resulting utility matrix. A common featuring policy using the same product distribution for every consumer, regardless of segment membership, is represented by $\psi = (\psi_1, \ldots, \psi_A) \in \Delta_A$, where $\Delta_A = \left\{\psi \in \mathbb{R}_+^A : \sum_{a=1}^{A} \psi_a = 1\right\}$ is the probability simplex and ψ_a the probability of featuring product a. Under policy ψ, the expected utility of a consumer of segment k is

$$U_k(\psi; V) = \sum_{a=1}^{A} \psi_a v_{ka}. \tag{2}$$

Let $U(\psi; V) = \{U_1, \ldots, U_K\}$ denote utility allocation across consumers from different segments. The set of all utility allocations under a common featuring policy is

$$\mathcal{U}(V) = \{U(\psi; V) : \psi \in \Delta_A\} = \operatorname{conv}\{v(1), \ldots, v(A)\}, \tag{3}$$

where $v(a) = \{v_{1a}, \ldots, v_{Ka}\}$. $\mathcal{U}(V)$ is a compact convex polytope. Selecting a static allocation policy is equivalent to selecting a point in this feasible utility set.

We distinguish static benchmark policies that differ in whether the featuring decision may depend on the consumer's private demographic information. Under a *demographic-aware* policy (henceforth, "personalized"), the firm has no restriction on demographic conditioning, so the revenue maximizing policy assigns each segment to its most preferred product

$$a_k^* \in \operatorname*{argmax}_a p_{ka}. \tag{4}$$

At the opposite extreme, a *demographic-blind* policy (henceforth, "blind") treats every consumer the same and chooses a product to feature such that

$$a^* \in \operatorname*{argmax}_a \sum_{k=1}^{K} \pi_k p_{ka}. \tag{5}$$

Because the objective in (5) is weighted by segment sizes, large segments exert greater influence on the selected product mix. While demographic-aware allocation accommodates segment-specific preferences, demographic-blind revenue maximization may yield little utility to smaller segments when segment sizes and preferences are heterogeneous. Such tension motivates the ethical allocations that maintain a common featuring distribution across consumers while explicitly accounting for disparities in segment utility.

3.2 Ethical Allocation Objectives

Similar to the blind policy, the ethical allocation retains a common action distribution at the population level, but replaces the segment-weighted objective in (5) with a social welfare function of $U(\psi; V)$ that assigns each segment *equal* weight. This reflects a normative choice in defining customer welfare to prevent differences in segment size alone from determining allocation outcomes. The specification of ethical criteria draws on established principles from welfare economics and fair resource allocation. In particular, Bertsimas et al. (2011) use the sum of utilities as an efficiency benchmark and analyze the efficiency costs associated with proportional and maximin fairness, while Bertsimas et al. (2012) embed these principles in a one-parameter family of constant elasticity welfare functions and characterize how the selected objective trades off efficiency against the utility guaranteed to the worst off party.

We summarize these ethical principles using the generalized power mean of order ρ,

$$W_\rho(U_1, \ldots, U_K) = \left(\frac{1}{K} \sum_{k=1}^{K} U_k^\rho \right)^{1/\rho}, \tag{6}$$

where parameter ρ determines the allocation mechanism. In fact, the constant elasticity, or α-fair, welfare family studied by Bertsimas et al. (2012) can be written as

$$W_\alpha(U_1, \ldots, U_K) = \begin{cases} \displaystyle\sum_{k=1}^{K} \frac{U_k^{1-\alpha}}{1-\alpha}, & \alpha \neq 1, \\ \displaystyle\sum_{k=1}^{K} \log U_k, & \alpha = 1. \end{cases} \tag{7}$$

Setting $\rho = 1 - \alpha$, it is straightforward to see that maximizing (6) and maximizing (7) induce the same allocation, as these are special cases of the generalized power mean family, such that

$$\rho \to 1 \quad (\alpha \to 0): \qquad W_1(U) = \frac{1}{K}\sum_{k=1}^{K} U_k, \qquad \text{(utilitarian)}, \tag{8}$$

$$\rho \to 0 \quad (\alpha \to 1): \qquad W_0(U) = \left(\prod_{k=1}^{K} U_k\right)^{1/K} \qquad \text{(proportional)}, \tag{9}$$

$$\rho \to -\infty \quad (\alpha \to \infty): \qquad W_{-\infty}(U) = \min_{1\le k\le K} U_k, \qquad \text{(egalitarian)}. \tag{10}$$

Derivation of these limits and additional special cases are provided in Web Appendix A.1.

Utilitarian Policy ($\rho = 1$, $\alpha = 0$) Let $W_{\mathfrak{U}}(\psi; V) = \sum_{k=1}^{K} U_k(\psi; V)$ denote the utilitarian welfare (Fleurbaey et al. 2023). The utilitarian ethical policy ψ guides resource allocation to maximize the aggregate utility across segments,

$$\Psi^*_{\mathfrak{U}}(V) = \operatorname*{argmax}_{\psi\in\Delta_A} W_{\mathfrak{U}}(\psi; V) = \operatorname*{argmax}_{\psi\in\Delta_A} \sum_{k=1}^{K} U_k(\psi; V). \tag{11}$$

This is the efficiency benchmark in Bertsimas et al. (2011). Its reward increases in total utility but is neutral toward how that total is distributed. In the present setting, it should not be conflated with firm revenue. (11) uses scaled utilities and gives equal weight to segments, whereas the firm's CTR objective in (5) uses raw conversion rates and population shares.

Proportional Policy ($\rho = 0$, $\alpha = 1$) Let $W_{\mathfrak{P}}(\psi; V) = \prod_{k=1}^{K} U_k(\psi; V)$ denote the proportional welfare (a.k.a. Nash social welfare; Kaneko and Nakamura 1979). The proportional ethical policy ψ optimizes the geometric mean, or the product of utilities across segments,

$$\Psi^*_{\mathfrak{P}}(V) = \operatorname*{argmax}_{\psi\in\Delta_A} W_{\mathfrak{P}}(\psi; V) = \operatorname*{argmax}_{\psi\in\Delta_A} \prod_{k=1}^{K} U_k(\psi; V). \tag{12}$$

Its multiplicative structure penalizes allocations that provide very low utility to any segment while retaining sensitivity to aggregate gains.

Egalitarian Policy ($\rho \to -\infty$, $\alpha \to \infty$) Let $W_{\mathfrak{E}}(\psi; V) = \min_{k=1,\ldots,K} U_k(\psi; V)$ denote the egalitarian welfare (Holm 2023). The egalitarian ethical policy ψ distributes resources equitably

across all segments by maximizing the utility of the worst off,

$$\Psi^*_{\mathfrak{E}}(V) = \underset{\psi \in \Delta_A}{\operatorname{argmax}}\, W_{\mathfrak{E}}(\psi; V) = \underset{\psi \in \Delta_A}{\operatorname{argmax}} \min_{k=1,\ldots,K} U_k(\psi; V). \tag{13}$$

This Rawlsian maximin objective prioritizes the worst off segment, seeking to improve its welfare even when doing so may come at the expense of higher aggregate utility (Rawls 1971).

The three ethical objectives correspond to different points on the Pareto frontier of $\mathcal{U}(V)$, as a function of ρ. When inequality aversion α increases (ρ decreases), the resulting allocation generally shifts away from aggregate efficiency toward stronger protection of the worst off segment (Bertsimas et al. 2012). Thus, no single ethical objective uniformly dominates the others, and the choice among them reflects a managerial and normative judgment about the balance between aggregate utility and fairness. Specifically, the utilitarian criterion prioritizes aggregate performance and is well suited to settings such as advertising, where maximizing overall engagement is the primary goal. In contrast, the egalitarian criterion prioritizes the least-served segment and is more appropriate when avoiding severe disparities is essential, such as in allocating access to critical information or opportunities. The proportional criterion provides a middle ground, balancing improvements across segments without exclusively prioritizing either aggregate efficiency or the worst off segment.

The allocation policies considered so far serve as static and complete information baselines that have been established in the literature. They do not address how the firm learns the unknown product preferences p_{ka}, how uncertainty arising from learning affects the resulting ethical allocation, or how sensitive segment information can be decoupled from individual level featuring decisions. The next section addresses these questions by proposing our ethical MAB framework to incorporate the utilitarian, proportional, and egalitarian objectives within a common sequential learning and allocation process.

4 Ethical Bandit Model

We now adapt the static, complete information ethical principles from (11)–(13) into a sequential learning and decision making context. Unlike the full information setting, segment-specific product preferences are initially unknown and must be learned over time through click observations within a MAB framework. As in the full information setting, the proposed sequential ethical bandit occupies an intermediate information regime. By separating learning

and allocation, it is demographic-aware when learning consumer preferences inside a restricted module, but demographic-blind when choosing what to show to the arriving customer.

4.1 Information Separation and Restricted Learning

Consumer arrivals are assumed to be independent and identically distributed across periods $t = 1, \ldots, T$, with one consumer arriving in each period and belonging to segment k with probability π_k. Because P is unknown a priori, we employ Thompson Sampling to implement the bandit. TS is particularly well suited to our setting as its posterior sampling framework provides a mathematically simple mechanism for balancing exploration and exploitation, while allowing the three ethical criteria to be incorporated directly into the allocation process.

We first define the sequential counterparts of the two static benchmarks from the preceding section. The *personalized TS* benchmark observes consumer type k at t before choosing an action and maintains a separate Beta distribution for every segment-product pair. Upon a consumer arrival, it draws from the prior and selects the product with the highest sampled preference

$$\widetilde{p}_{ka,t} \sim \text{Beta}\big(\alpha_{ka,t}, \beta_{ka,t}\big), \qquad a_t \in \arg\max_a \widetilde{p}_{ka,t}. \tag{14}$$

Thus, the demographic attribute of current customer is used in both learning and allocation. Also, in this TS specification, the current Beta distribution parameters provide sufficient statistics for previous segment identities, featured products, and click outcomes. Denote the prior parameters at the beginning of period t as $\mathcal{H}_t = \{\alpha_{ka,t}, \beta_{ka,t}\}_{k=1,\ldots,K,\, a=1,\ldots,A}$. As a result, the impact of past interactions has all been incorporated into $\mathcal{H}_t$.

The *blind TS* benchmark does not use segment information in either stage. It aggregates observations to product level and maintains one Beta distribution for each product across segments,

$$\widetilde{p}_{a,t} \sim \text{Beta}\big(\alpha_{a,t}, \beta_{a,t}\big), \qquad a_t \in \arg\max_a \widetilde{p}_{a,t}. \tag{15}$$

The blind benchmark learns the revenue maximizing product common to all segments, potentially disadvantaging segments with smaller population shares.

In contrast, the proposed *ethical TS* occupies an intermediate information regime and combines segment-specific posterior learning with a common allocation distribution. Figure 2 presents the information flow of this framework. Access to the arriving customer's segment label is confined to the restricted learning module, whereas the resource allocation module receives only the prior parameters $\mathcal{H}_t$ from the past, which do not incorporate the arriving

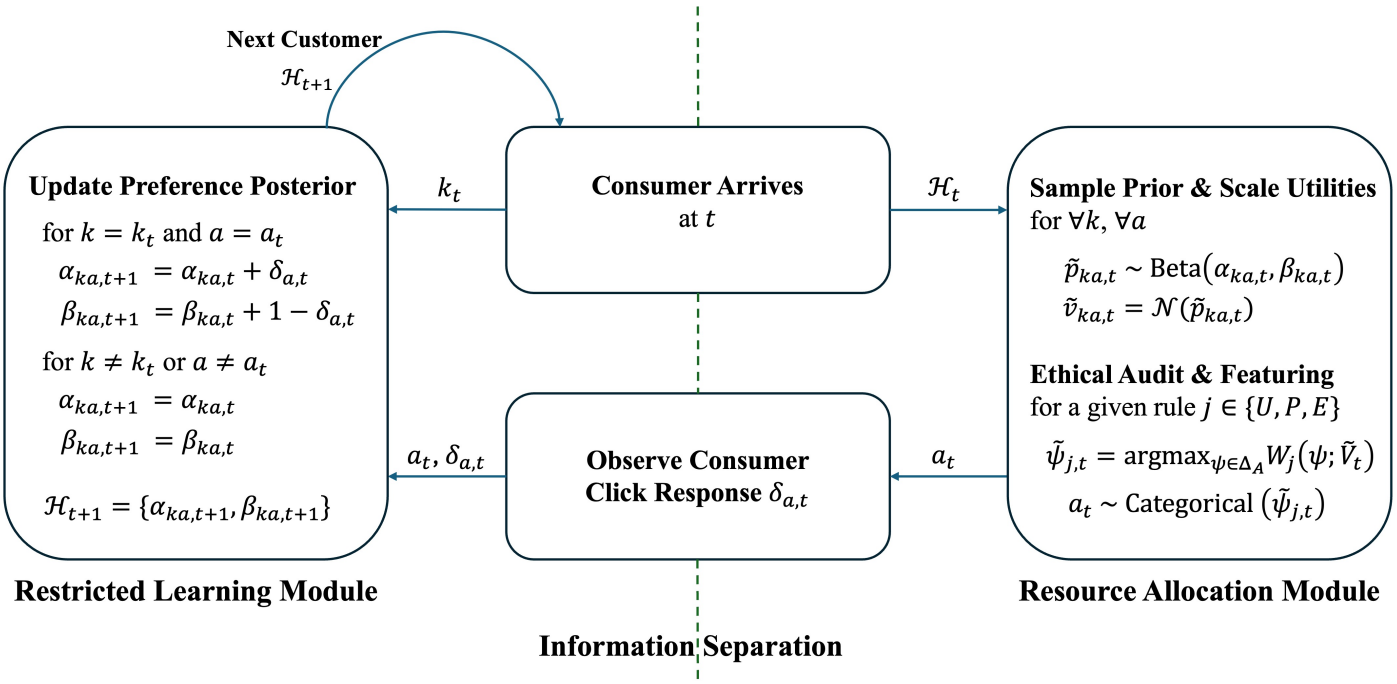


Figure 2: Information Flow in Ethical Thompson Sampling

customer's sensitive information. According to the figure, each period t proceeds as follows:

(1) A consumer arrives and segment identity k_t is realized. This information is available only to the restricted learning module for the subsequent posterior update and is not disclosed to the resource allocation module.

(2) The resource allocation module is notified of the arrival and receives only the prior parameters $\mathcal{H}_t$, which summarize the past, without access to the current sensitive information k_t.

(3) Given $\mathcal{H}_t$, the resource allocation module draws $\widetilde{P}_t$ and scales it to produce $\widetilde{V}_t$.

(4) Based on $\widetilde{V}_t$, the allocation module optimizes the population level ethical weight vector

$$\widetilde{\psi}_{j,t} \in \arg\max_{\psi \in \Delta_A} W_j(\psi; \widetilde{V}_t), \tag{16}$$

where $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$ is the ethical principle adopted by the firm, and draws product

$$a_t \sim \text{Categorical}(\widetilde{\psi}_{j,t}) \tag{17}$$

to feature to the consumer.

(5) The consumer's click response $\delta_{a,t} \in \{0, 1\}$ is observed and transmitted to the restricted

learning module.

(6) The restricted learning module maintains a separate Beta distribution for each segment-product pair $\{k, a\}$. Upon receiving the realized tuple $\{k_t, a_t, \delta_{a,t}\}$, it updates the posterior parameters $\mathcal{H}_{t+1} = \{\alpha_{ka,t+1}, \beta_{ka,t+1}\}$, which also serves as the prior for the next period, such that

$$\alpha_{ka,t+1} = \alpha_{ka,t} + \delta_{a,t} \ , \qquad \beta_{ka,t+1} = \beta_{ka,t} + 1 - \delta_{a,t}, \tag{18}$$

for $k = k_t$ and $a = a_t$. Only the posterior corresponding to the realized segment-product pair gets updated, while all other posterior parameters remain unchanged.

The sequence of events illustrates the *information separation* mechanism underlying the ethical TS framework. Conditional on $\mathcal{H}_t$, the allocation distribution in (16) is independent of the arriving consumer's segment identity at t, which is revealed in $\mathcal{H}_{t+1}$ rather than $\mathcal{H}_t$, i.e., $\widetilde{\psi}_{j,t} \perp k_t \mid \mathcal{H}_t$. Consequently, the consumer faces the same conditional allocation distribution regardless of his/her type k, and differences in featured products arise solely from random sampling. The current consumer's identity influences future decisions only *retrospectively* through the posterior update in (18), without affecting the product shown to him/her.

Such information separation does not imply that sensitive attributes are never processed or provide a formal differential privacy guarantee. Rather, the framework presumes that segment labels are available exclusively to the restricted posterior learning. If an organization neither collects nor is permitted to process such attributes, the ethical bandits cannot be implemented exactly as specified. Organizationally, the restricted learning module could be housed within a company's data ethics or algorithmic accountability function.

Existing organizational practices suggest the feasibility of such separation. For instance, LinkedIn's privacy-preserving fairness measurement system separates responsible-AI and product functions, sharing aggregate results without exposing individual race or ethnicity estimates to product owners (Badrinarayanan et al. 2024). Airbnb's Project Lighthouse analyzes booking disparities using perceived-race information that cannot be linked to user profiles (Airbnb 2024). Meta's Variance Reduction System uses aggregate, privacy-protected demographic measurements to adjust advertising delivery without exposing individual demographic attributes to its controller (Meta 2023). Although these systems do not implement our algorithm, they demonstrate how marketing decision making can be separated from sensitive data access.

4.2 Static vs. Dynamic Efficiency-Equity Tradeoff

To evaluate the tradeoff between efficiency and equity, we adopt the *price of fairness* concept from Bertsimas et al. (2011), which measures the relative loss in aggregate efficiency from adopting a fair allocation. In our setting, the firm's efficiency is measured by expected clicks. Under a common allocation ψ, the expected number of clicks per arrival is $C(\psi) = \sum_{k=1}^{K} \pi_k \sum_{a=1}^{A} p_{ka}\psi_a$, which aggregates expected clicks across customer segments, weighted by their arrival probabilities. This quantity is maximized by the demographic-blind product a^* defined in (5). We use the firm reward of the static blind, $C_B = C(a^*)$, as the efficiency benchmark and specify the *static* price of fairness associated with ethical rule $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$ as

$$\mathrm{PoF}_j = \frac{C_B - C(\Psi_j^*)}{C_B}. \tag{19}$$

A distinction from Bertsimas et al. (2011) is that the utilitarian objective $W_{\mathfrak{U}}$ in (11) is not the efficiency benchmark because it aggregates scaled utilities v_{ka} without weighting and thus does not coincide with the demographic-blind welfare in (5). As a result, the utilitarian policy is treated as an ethical allocation rather than the efficiency baseline, with a generally positive price of fairness, $\mathrm{PoF}_{\mathfrak{U}} \geq 0$, since (5) maximizes expected clicks among all common allocations.

The static price of fairness in (19) compares allocations that are optimal under known preferences. In practice, however, firms must learn consumer preferences over time, incurring learning costs under both ethical and blind bandit policies. To compare these bandits over the same horizon t, let $C_{B,t} = \sum_{s=1}^{t} C(\widetilde{\psi}_{B,s})$ and $C_{j,t} = \sum_{s=1}^{t} C(\widetilde{\psi}_{j,s})$ denote the cumulative expected clicks under blind TS and ethical TS, respectively, and define the *dynamic* price of fairness as

$$\mathrm{PoF}_j(t) = \frac{C_{B,t} - C_{j,t}}{C_{B,t}}. \tag{20}$$

For $\psi_j^* \in \Psi_j^*$, we further define the learning cost of each policy relative to its respective full information target

$$L_{B,t} = t \cdot C_B - C_{B,t}, \qquad L_{j,t} = t \cdot C(\psi_j^*) - C_{j,t}. \tag{21}$$

Because the blind solution maximizes expected clicks among all common allocations, $L_{B,t} \geq 0$. In contrast, $L_{j,t}$ may be negative, as ethical TS can temporarily favor products that generate more clicks than its ethical target at the expense of the segment the target seeks to protect.

5 Theoretical Analyses

We now establish theoretical guarantees that ethical TS preserves the desirable asymptotic properties of standard TS while incorporating ethical objectives. We then decompose the dynamic price of fairness to provide a theoretical foundation for subsequent empirical comparisons. Finally, we discuss the optimization of ethical weights.

5.1 Statistical Guarantees of Ethical Bandits

We begin by laying out the required regularity conditions and establishing posterior concentration for the Beta-Bernoulli model. We then prove ethical consistency and establish asymptotic guarantees on ethical regret, followed by a corresponding finite sample analysis. Finally, we show that the randomness induced by categorical action sampling constitutes a zero-mean martingale fluctuation whose average magnitude vanishes asymptotically. We start with the regularity conditions underlying these results.

Assumption 1 (Stationary Preferences and Independent Arrivals)**.** *The segment-product preferences p_{ka} are time invariant. Segment arrivals are independent across time periods, with each segment k arriving with probability π_k.*

Assumption 2 (Sufficient Exploration)**.** *For every segment-product pair $\{k, a\}$, the number of times product a is featured to customers from segment k,*

$$N_{ka}(t) = \sum_{s=1}^{t} \mathbb{I}_s\{k, a\} \longrightarrow \infty \qquad \textit{almost surely.} \tag{22}$$

Assumption 2 serves as a simplifying regularity condition ensuring all arms remain in the asymptotic information pool (Russo and Van Roy 2014), allowing us to isolate the convergence property of ethical TS from the risk that some products stop being shown before their CTRs are learned for every segment. Note that it assumes infinitely many observations per segment-product pair, *not* a positive limiting exposure share.[1] Under these regularity conditions, we can now show that the posterior Beta samples used by the ethical bandit concentrate around the true segment specific preferences, establishing statistical consistency of consumer preference learning.

[1]The condition is not automatic because an ethical allocation may assign zero weight to most products, so a product is shown only when a posterior draw places it in the optimal allocation. The condition can be ensured by mixing each allocation with an exploration floor that assigns probability ε_t/A to each product, where $\varepsilon_t \to 0$ and $\sum_t \varepsilon_t = \infty$ (e.g., $\varepsilon_t = t^{-1/2}$). This preserves the asymptotic results while adding at most $\sum_{t \le T} \varepsilon_t = O(\sqrt{T})$ to the regret. Our simulations do not use an exploration floor.

Lemma 1 (Beta-Bernoulli Posterior Consistency). *Under Assumptions 1 and 2, for every segment-product pair* $\{k, a\}$,

$$\widetilde{p}_{ka,t} \longrightarrow p_{ka} \qquad \textit{in probability.} \tag{23}$$

Because K *and* A *are finite,*

$$\|\widetilde{P}_t - P\|_\infty \longrightarrow 0 \qquad \textit{in probability.} \tag{24}$$

Web Appendix A.2 details the proof. The next lemma extends posterior consistency from the Beta samples of raw CTRs to the scaled utilities used in the ethical optimization.

Lemma 2 (Consistency of Scaling). *Define* $M_{\min} = \min_k \max_a p_{ka} > 0$, *where positivity follows from* $p_{ka} \in (0, 1)$. *For any CTR matrix* $Q \in (0, 1)^{K \times A}$,

$$\|\mathcal{N}(Q) - \mathcal{N}(P)\|_\infty \leq \frac{2}{M_{\min}} \|Q - P\|_\infty. \tag{25}$$

Consequently,

$$\|\widetilde{V}_t - V\|_\infty \longrightarrow 0 \qquad \textit{in probability.} \tag{26}$$

The proof of Lemma 2 is reported in Web Appendix A.3. We now demonstrate that the ethical weights learned by the ethical bandit converge to their perfect information counterparts.

Lemma 3 (Uniform Convergence of Ethical Objectives). *For each* $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$,

$$\sup_{\psi \in \Delta_A} \left| W_j(\psi; \widetilde{V}_t) - W_j(\psi; V) \right| \longrightarrow 0 \qquad \textit{in probability.} \tag{27}$$

Web Appendix A.4 provides the detailed proof. With posterior consistency and convergence of the ethical objective established, we can now show the convergence of the ethical weights.

Theorem 1 (Convergence of Ethical Weights). *Under Assumptions 1 and 2, for each ethical principle* $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$, *any measurable maximizer* $\widetilde{\psi}_{j,t}$ *of the sampled objective in* (16) *satisfies*

$$\inf_{\phi \in \Psi_j^*} \|\widetilde{\psi}_{j,t} - \phi\| \longrightarrow 0 \qquad \textit{in probability.} \tag{28}$$

Moreover, for any $\psi_j^* \in \Psi_j^*$,

$$W_j(\widetilde{\psi}_{j,t}; V) \longrightarrow W_j(\psi_j^*; V) \qquad \textit{in probability.} \tag{29}$$

The proof is located in Web Appendix A.5. Theorem 1 establishes that, as segment-product preferences are learned, the ethical weight vector approaches the set of full information optimal allocations, and the resulting ethical welfare converges to the optimal value. When multiple allocations attain the same optimal value, the weights need not converge to a unique allocation. We next turn to the finite sample analysis and examine the cumulative learning performance of ethical TS over a finite horizon T. Analogous to revenue regret in conventional MAB settings, we define *ethical regret* to measure the opportunity cost of inequitable treatment in terms of consumer welfare to the firm.

Definition 1 (Ethical Regret)**.** *Define the policy level cumulative ethical regret as*

$$\mathcal{R}_{j,T} = \sum_{t=1}^{T} \left[W_j(\psi_j^*; V) - W_j(\widetilde{\psi}_{j,t}; V) \right]. \tag{30}$$

Thus, the ethical regret measures how the ethical allocation evolves over time. In the bandit literature, regret is defined to evaluate policy performance by comparing the reward from the chosen arm with that from the optimal arm under full information, typically from a revenue perspective (Lattimore and Szepesvári 2020). For an ethical bandit, however, optimality is defined in terms of consumer welfare under the selected ethical principle. Accordingly, ethical regret compares the arm selected by the bandit with the arm that would maximize consumer welfare under that principle, as specified in (30).

Corollary 1 (Vanishing Ethical Regret)**.** *Under the conditions of Theorem 1, for* $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$*,*

$$\frac{\mathcal{R}_{j,T}}{T} \longrightarrow 0 \qquad \textit{in probability.} \tag{31}$$

Web Appendix A.6 provides the proof. Corollary 1 implies that the ethical regret is *sublinear* and the per-period consumer welfare loss asymptotically vanishes as observations accumulate, even though the limiting policies associated with the three principles may differ. Thus far, we have characterized the *asymptotic* behavior of the ethical bandit. In any finite learning round, however, the firm does not know the true utility V and must compute featuring decisions by optimizing over a noisy posterior draw $\widetilde{V}_t$. As this optimization induces suboptimality under true consumer preferences, we analyze how severely posterior noise degrades optimized welfare across the three ethical criteria. Define the utility estimation error bound as

$$b_t = \mathbb{E}\left[\left\| \widetilde{V}_t - V \right\|_\infty \right]. \tag{32}$$

The following theorem translates this error bound into the bound on ethical regret.

Theorem 2 (Finite Sample Ethical Regret)**.** *For every horizon T,*

$$\mathbb{E}\left[\mathcal{R}_{\mathfrak{U},T}\right] \leq 2K \sum_{t=1}^{T} b_t, \tag{33}$$

$$\mathbb{E}\left[\mathcal{R}_{\mathfrak{P},T}\right] \leq 2K \sum_{t=1}^{T} b_t, \tag{34}$$

$$\mathbb{E}\left[\mathcal{R}_{\mathfrak{E},T}\right] \leq 2 \sum_{t=1}^{T} b_t. \tag{35}$$

The bounds are stated in terms of the estimation error b_t and do not impose a rate. When every segment-product pair receives a positive share of exposure, $N_{ka}(t)$ grows linearly in t, the posterior error satisfies $b_t = O(t^{-1/2})$, and the bounds imply ethical regret of order $O(\sqrt{T})$. The proof is located in Web Appendix A.7. This theorem relates ethical regret to the accuracy of the learned utility matrix without imposing a particular posterior contraction rate. Because scaled utilities are bounded, the posterior consistency in Lemma 2 implies $b_t \to 0$. Consequently, $\mathbb{E}[\mathcal{R}_{j,T}]/T \to 0$ for each ethical principle. Thus, the analysis establishes vanishing average ethical regret while leaving its precise convergence rate dependent on the information acquired by the sequential policy.

Finally, because the featuring decision in (17) is sampled categorically from the ethical weight vector, realized customer utility may deviate from its policy implied expectation. For segment k, define the sampling error $\xi_{k,t} = v_{ka_t} - U_k(\widetilde{\psi}_t; V)$, where $U_k(\widetilde{\psi}_t; V) = \sum_{a=1}^{A} \widetilde{\psi}_{t,a} v_{ka}$ is the expected utility under the ethical policy, and let $N_k(T) = \sum_{t=1}^{T} \mathbb{I}_t\{k\}$ denote the number of arrivals from segment k over horizon T.

Proposition 1 (Action Sampling Concentration)**.** *Conditional on the realized arrival sequence and $N_k(T) \geq 1$, for every $\lambda \in (0, 1)$,*

$$\Pr\left(\left|\frac{1}{N_k(T)} \sum_{t=1}^{T} \mathbb{I}_t\{k\}\xi_{k,t}\right| > \sqrt{\frac{2\log(2/\lambda)}{N_k(T)}}\right) \leq \lambda.$$

Equivalently, with probability at least $1 - \lambda$,

$$\left|\frac{1}{N_k(T)} \sum_{t=1}^{T} \mathbb{I}_t\{k\}\xi_{k,t}\right| \leq \sqrt{\frac{2\log(2/\lambda)}{N_k(T)}}. \tag{36}$$

The proof is located in Web Appendix A.8. Since $N_k(T)/T \to \pi_k$, the average sampling

error decreases at the rate $O(1/\sqrt{\pi_k T})$ for a fixed confidence level. Therefore, the randomness introduced by action sampling vanishes as more consumer interactions are observed. From a managerial perspective, Proposition 1 implies that the information separation mechanism does not create a persistent loss of ethical performance. Consequently, the long run behavior of the algorithm is governed solely by the learned ethical policy.

These theoretical results establish convergence in ethical welfare and bound the welfare losses incurred during learning. We next examine the firm's click loss relative to blind TS and how learning changes the price of fairness over a finite horizon.

5.2 Decomposition of Price of Fairness

The dynamic PoF in (20) depends on the relative learning costs of ethical and blind TS in (21). Ethical policies weight segments equally, although observations arrive in proportion to segment shares π_k, whereas blind TS pools observations across segments to learn product-level CTRs. The following proposition separates the static fairness cost from these learning effects.

Proposition 2 (Dynamic PoF Decomposition)**.** *For every horizon T and principle $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$,*

1. *Loss in expected clicks:* $C_{B,T} - C_{j,T} = T \cdot \big(C_B - C(\psi_j^*)\big) + (L_{j,T} - L_{B,T})$.
2. *Price of fairness:* $\mathrm{PoF}_j(T) > \mathrm{PoF}_j$ *if and only if* $L_{j,T} > (1 - \mathrm{PoF}_j)\, L_{B,T}$.
3. *When Assumptions 1 and 2 hold and $C(\psi)$ is constant on Ψ_j^*,* $\mathrm{PoF}_j(T) \to \mathrm{PoF}_j$ *in probability as* $T \to \infty$.

Web Appendix A.9 provides the proof. Item 1 is an accounting identity. The click loss of ethical TS relative to blind TS equals the static loss under known preferences plus the difference in learning costs. Thus, the realized click loss exceeds its static counterpart precisely when ethical TS incurs greater learning loss than blind TS. Item 2 shows why the percentage comparison differs. As the dynamic PoF is normalized by the clicks of blind TS, which are themselves reduced due to learning, it can exceed the static PoF even when ethical TS incurs less learning loss than blind TS, $(1 - \mathrm{PoF}_j)L_{B,T} < L_{j,T} < L_{B,T}$. Item 3 establishes that both learning losses vanish per arrival asymptotically and the dynamic PoF converges to its static counterpart.

Relative learning losses depend on segment sizes, preference heterogeneity, and click frequencies, so neither blind nor ethical TS consistently incurs greater learning loss. Learning also affects the benefits of fairness, as minority utility may fall short of its full information ethical

target over finite horizons. Thus, the costs and benefits of fairness can accrue at different rates. Section 6 examines these effects empirically using two real world datasets.

5.3 Ethical Weight Optimization

Each time it updates the allocation, the ethical bandit evaluates the selected objective $W_j(\psi; \widetilde{V}_t)$ using the sampled utilities. The utilitarian objective is linear in ψ, $W_{\mathfrak{U}}(\psi; \widetilde{V}_t) = \sum_{a=1}^{A} \psi_a \sum_{k=1}^{K} \widetilde{v}_{ka,t}$. Consequently, except in the event of a tie, its maximizing allocation is one-hot,

$$\widetilde{\psi}_{\mathfrak{U},t} = \nu_{a_{\mathfrak{U},t}}, \qquad a_{\mathfrak{U},t} \in \arg\max_{a} \sum_{k=1}^{K} \widetilde{v}_{ka,t}, \tag{37}$$

where ν_a is the ath unit vector. If multiple products attain the same maximum, any distribution supported on the maximizing set is optimal.[2]

The proportional and egalitarian objectives can yield nondegenerate mixtures and do not generally reduce to a single product comparison. The proportional problem admits an equivalent concave log-welfare formulation, while the egalitarian problem can be expressed as a linear program. Computational procedures, the allocation update schedule, and numerical accuracy checks are provided in Web Appendix B.

6 Data Applications

Having analyzed the theoretical properties and the optimization solvers for the ethical bandit, we now evaluate its empirical performance using randomized controlled trial (RCT) data from Yahoo! News and ZOZO, a large Japanese fashion ecommerce platform. In both applications, we classify consumers into majority, intermediate, and minority segments based on sensitive user attributes. Broadly, a minority segment refers to a group that differs from the dominant population in characteristics such as race, ethnicity, religion, or sexual orientation (*Cambridge Learner's Dictionary* 2024). Using context-specific segmentations, we compare ethical TS against benchmark policies in terms of consumer utility, firm rewards, and the efficiency-equity tradeoff. We then examine the learning dynamics of ethical policies, including the evolution of ethical regret and the timing of fairness benefits and costs. Finally, the robustness of our findings is assessed against the variations in the number of consumer segments and the relative size of minority segment.

[2]The implementation uses a fixed tie-breaking rule, e.g., the lowest product index.

6.1 Application I: Yahoo News

Our first application uses the Yahoo! Front Page News data, publicly available through Yahoo Research (Yahoo Research 2009). This dataset provides a natural setting for studying the featuring problem for three reasons. First, the platform must select a limited number of articles to feature from a much larger pool of available content. Second, the Yahoo! Front Page attracts millions of daily visitors, generating substantial user engagement with featured articles. Third, this RCT dataset has been widely used to evaluate online recommendation and bandit algorithms (Chu et al. 2009; Li et al. 2011; Vanchinathan et al. 2015).

The dataset covers the first 10 days of May 2009, during which the Yahoo! Front Page recorded approximately 45 million visits.[3] For each visit, an article was *randomly* selected for display from a daily pool of approximately 20–25 available articles. The dataset records article exposures, click outcomes, and user characteristics, which are used to estimate consumer preferences and construct segments. Importantly, the randomized article assignments allow us to estimate the segment-article preference matrix from observed CTRs, without confounding from a preexisting recommendation algorithm.

Given the short lived relevance of news articles, which typically lose their appeal within one or two days, we focus on observations from a single day. To construct a manageable sample for offline policy simulations, we divide these observations into consecutive batches of 500,000 and use one batch for analysis. This sampling strategy balances computational tractability with a sufficiently large number of user interactions to evaluate the proposed framework.

For the illustration of our methodology, we select one of the user features (feature 5) as the sensitive attribute throughout our analysis. Specifically, we divide this feature's numerical range into three equal width intervals, yielding majority, intermediate, and minority segments that respectively account for 70.21%, 17.79%, and 12.00% of the observed population, with corresponding mean CTRs of 0.027, 0.040, and 0.058. Compared with a binary classification that distinguishes only between aligned and conflicting preferences, this three-segment partition allows us to capture more nuanced intersegment preference relationships.

To measure how minority utility differs from the population, we define relative utility disparity (RUD) as the absolute relative deviation of minority utility from the population-weighted average

$$\mathrm{RUD} = \frac{\left| U_{\text{Minority}} - \sum_{k=1}^{K} \pi_k U_k \right|}{\sum_{k=1}^{K} \pi_k U_k}. \tag{38}$$

[3]Because user identities are not disclosed, each visit is treated as originating from a unique user.

As a symmetric measure of disparity rather than a one-sided measure of harm, RUD captures deviations in both directions. A high value may indicate that the minority segment is either substantially underserved or overserved relative to the population benchmark.

Table 1 reports the performance of ethical bandits and benchmarks from both consumer and firm perspectives, measured by consumer utility and expected clicks, respectively. Consumer utility is defined as the scaled expected CTR for each segment, averaged across 500,000 arrivals and 30 independent runs of each bandit. Firm reward measures the expected number of clicks generated by each segment, summed over all arrivals within each run and then averaged across the 30 runs. The 95% confidence intervals reported beneath each estimate are computed as mean $\pm$ $t_{0.975,29} \times \text{SD}/\sqrt{30}$, where SD denotes the standard deviation across the 30 runs. Bootstrap confidence intervals are applied to the RUD. These intervals capture the randomness inherent in the bandit experiments.

Table 1: Ethical vs. Benchmark Thompson Sampling Policies in Yahoo Application

	Consumer Utility				Firm Reward			
MAB Model	Majority	Interm.	Minority	RUD	Majority	Interm.	Minority	Total
Personalized	0.967	0.923	0.936	0.020	17,244	5,762	6,238	29,244
	[0.965, 0.969]	[0.919, 0.927]	[0.933, 0.939]	[0.017, 0.024]	[17,206, 17,281]	[5,735, 5,789]	[6,214, 6,263]	[29,194, 29,293]
Utilitarian	0.749	0.942	0.924	0.149	13,360	5,880	6,149	25,389
	[0.745, 0.753]	[0.940, 0.943]	[0.917, 0.931]	[0.138, 0.159]	[13,294, 13,427]	[5,870, 5,889]	[6,103, 6,195]	[25,361, 25,417]
Proportional	0.743	0.942	0.933	0.165	13,248	5,882	6,210	25,339
	[0.739, 0.746]	[0.941, 0.944]	[0.925, 0.941]	[0.153, 0.176]	[13,178, 13,317]	[5,872, 5,892]	[6,158, 6,262]	[25,309, 25,370]
Egalitarian	0.796	0.918	0.819	0.002	14,200	5,729	5,447	25,377
	[0.793, 0.799]	[0.914, 0.922]	[0.812, 0.825]	[0.000, 0.011]	[14,142, 14,258]	[5,703, 5,756]	[5,405, 5,489]	[25,336, 25,418]
Blind	0.957	0.964	0.556	0.389	17,067	6,016	3,701	26,784
	[0.952, 0.961]	[0.959, 0.968]	[0.548, 0.564]	[0.378, 0.398]	[16,984, 17,150]	[5,985, 6,046]	[3,648, 3,754]	[26,724, 26,844]

Notes: These are based on Day 7 (batch 9). Results for other days are reported in Web Appendix C. "Personalized" is demographic-aware. "Blind" is demographic-blind.

The blind TS illustrates how ignoring segment heterogeneity can favor the majority, resulting in utilities of 0.957, 0.964, and 0.556 for the majority, intermediate, and minority, respectively. In contrast, the personalized TS tailors article selection to each segment, yielding respective utilities of 0.967, 0.923, and 0.936. Among the ethical policies, proportional TS gives the minority higher utility than egalitarian TS, whereas egalitarian TS achieves the lowest RUD of 0.002. The majority is the worst off segment under all three ethical policies, and egalitarian TS gives it the highest utility of 0.796 among these policies. Thus, protecting the worst off segment and maximizing the minority segment's benefit are distinct goals.

These results illustrate the central tension motivating our research. The demographic-blind

benchmark can generate substantial disparities when segment preferences differ,[4] whereas the personalized benchmark can achieve higher consumer welfare and firm reward but may be subject to legal or regulatory restrictions on the use of sensitive attributes. The proposed ethical policies offer an intermediate alternative by incorporating segment-specific preferences during learning while maintaining demographic-blind allocation decisions. Although these policies can mitigate disparities, their effects on the distribution of consumer utility depend on the underlying ethical principle. Moreover, all three policies entail some loss in firm reward relative to the personalized benchmark, highlighting the tradeoff between efficiency and equity.

Price of Fairness As specified in (19) and (20), PoF measures the fraction of revenue sacrificed by adopting a fair allocation. Using the known CTRs from Table W10 in Web Appendix D, expected clicks are calculated by multiplying 500,000 arrivals by the segment-weighted CTR under each policy, $\sum_k \pi_k \sum_a p_{ka}\psi_a$. The static blind policy yields 27,540 expected clicks by exclusively featuring article 109692, compared with 25,858 under both static utilitarian and proportional policies featuring article 109708, and 26,029 under the static egalitarian allocating equal probabilities to articles 109703 and 109708. These allocations correspond to static PoF of 6.1% for utilitarian and proportional policies and 5.5% for egalitarian.

The realized dynamic PoFs are slightly lower than their static counterparts because, for each ethical TS, the learning cost $L_{j,T}$ is below $(1 - \text{PoF}_j)\, L_{B,T}$ (Item 2, Proposition 2). Blind TS achieves 26,784 expected clicks, compared with 25,389 for utilitarian TS, 25,339 for proportional TS, and 25,377 for egalitarian TS, corresponding to dynamic PoFs of 5.2%, 5.4%, and 5.3%, respectively. Applying Proposition 2, Table W14 in Web Appendix H shows that blind TS incurs a learning cost of 756 clicks, compared with 469–652 under ethical TS. These differences reduce the realized efficiency losses relative to blind TS to 1,395–1,445 clicks, below the static fairness costs of 1,511–1,683 clicks. These results highlight how sequential learning affects the realized efficiency cost of fairness relative to its static benchmark.

Fairness Gain for Minority With known CTRs, the static blind policy yields a minority utility of 0.520, compared with 1.000 under both static utilitarian and proportional policies and 0.832 under static egalitarian. These differences reflect the policies' distinct allocation objectives. The blind policy favors the majority's preferred article, leaving the minority substantially

[4]The cross-day results in Web Appendix C clarify how the scope for improvement depends on preference alignment. On Yahoo Day 4, for example, all segments prefer the same article, and blind TS delivers minority utility of 0.993, leaving little room for further gains. On Days 5 and 7, when preferences diverge, the ethical policies raise minority utility from 0.592 to 0.779–0.821 and from 0.556 to 0.819–0.933, respectively.

underserved. In contrast, utilitarian and proportional policies both feature the minority's preferred article, maximizing its utility to the highest while reducing majority utility to 0.736. The egalitarian instead distributes impressions equally between two articles, equalizing majority and minority utilities at 0.832. These full information allocations follow from the utilities reported in Table W10.

Such fairness gains become smaller under learning. Over the bandit horizon, minority utility is 0.556 under blind TS, compared with 0.924, 0.933, and 0.819 under utilitarian, proportional, and egalitarian bandits, respectively. These improvements represent 77%, 79%, and 84% of the corresponding static fairness gains, e.g., $(0.924 - 0.556)/(1.000 - 0.520) = 0.77$ for utilitarian. That is, all three ethical bandits achieve consistently *lower* dynamic PoF than their static counterparts, but also deliver *smaller* fairness gains. Section 6.3 will examine how these costs and benefits evolve over the learning process and vary across ethical principles.

Overall, the Yahoo application illustrates how the separation between demographic-aware learning and demographic-blind featuring can alter the allocation of a common set of products across heterogeneous consumers, and why the choice of ethical objective matters for the resulting tradeoff between firm reward and consumer equity.

6.2 Application II: ZOZO Ecommerce

The second application employs the Open Bandit Dataset from a large Japanese fashion ecommerce platform ZOZOTOWN. The dataset was collected during a seven-day experiment in November 2019, in which fashion items were recommended under either a random policy or Bernoulli Thompson Sampling. We use observations from the random policy for offline evaluation. Each observation records the recommended item, its display position, click outcome, recommendation propensity, and hashed user features. We use the whole campaign data with a selection of 80 candidate items (Saito et al. 2021).

For illustrative purposes, we use `user_feature_0` as the sensitive variable to define user segments. This feature consists of four hashed categorical labels whose substantive meanings are not disclosed in the data documentation. Because these categories have no natural ordering, the range-based 1/3 and 2/3 cutoffs used for the Yahoo data are not applicable. Instead, we rank the four categories by population size, retain the two largest as majority and intermediate, and combine the two smallest into minority. The resulting segments account for 53.04%, 35.36%, and 11.59% of observations, with mean CTRs of 0.32%, 0.38%, and 0.34%, respectively.

These CTRs are substantially lower and more uniform across segments than in the Yahoo

data, where segment CTRs range from 3% to 6%. In ZOZO, items are selected uniformly from 80 fashion products, making clicks relatively rare. Even the Bernoulli TS evaluated by Saito et al. (2021) achieves a CTR of only about 0.50%. In contrast, Yahoo features news headlines, which attract clicks more frequently. The ZOZO application provides a *complementary* setting for examining whether our algorithms generate meaningful fairness gains when average CTRs are similar across segments but preferences over individual items differ.

Table 2: Ethical vs. Benchmark Thompson Sampling Policies in ZOZO Application

	Consumer Utility				Firm Reward			
MAB Model	Majority	Interm.	Minority	RUD	Majority	Interm.	Minority	Total
Personalized	0.865	0.742	0.624	0.214	7,148	2,456	1,229	10,833
	[0.858, 0.872]	[0.736, 0.747]	[0.614, 0.634]	[0.203, 0.225]	[7,092, 7,205]	[2,439, 2,473]	[1,209, 1,249]	[10,767, 10,900]
Utilitarian	0.584	0.761	0.305	0.504	4,828	2,522	601	7,951
	[0.574, 0.595]	[0.748, 0.775]	[0.297, 0.312]	[0.489, 0.519]	[4,741, 4,915]	[2,479, 2,566]	[586, 615]	[7,868, 8,033]
Proportional	0.501	0.725	0.383	0.323	4,138	2,402	756	7,295
	[0.494, 0.508]	[0.712, 0.738]	[0.375, 0.392]	[0.310, 0.338]	[4,083, 4,193]	[2,358, 2,446]	[738, 773]	[7,250, 7,341]
Egalitarian	0.479	0.681	0.407	0.249	3,955	2,254	802	7,011
	[0.474, 0.484]	[0.673, 0.689]	[0.400, 0.413]	[0.240, 0.258]	[3,915, 3,996]	[2,228, 2,280]	[789, 814]	[6,973, 7,049]
Blind	0.800	0.658	0.294	0.574	6,606	2,180	580	9,365
	[0.787, 0.812]	[0.637, 0.679]	[0.290, 0.299]	[0.569, 0.581]	[6,502, 6,710]	[2,111, 2,248]	[571, 588]	[9,319, 9,412]

Notes: "Personalized" is demographic-aware. "Blind" is demographic-blind.

Table 2 reports the performance of the five TS policies for the ZOZO data, averaged over 1,374,327 arrivals and 30 independent runs per bandit, with 95% confidence intervals shown beneath each estimate. These results illustrate how different ethical principles shape utility distribution across segments. Under blind TS, minority utility is only 0.294, compared with 0.658 for intermediate and 0.800 for majority. The utilitarian, proportional, and egalitarian TS increase minority utility to 0.305, 0.383, and 0.407, respectively, while reducing majority utility to 0.584, 0.501, and 0.479. Among the ethical policies, the utilitarian yields the highest RUD of 0.504, whereas the egalitarian achieves the lowest at 0.249. The reductions in relative utility disparity appear less pronounced here than in the Yahoo application due to ZOZO's much rarer clicks and more sparse rewards. Learning item preferences and approaching fairness targets require hundreds of thousands rather than tens of thousands of arrivals.

Using the known CTRs for ZOZO reported in Web Appendix Table W11, the static PoF is 20.0% for utilitarian, 19.0% for proportional, and 23.3% for egalitarian allocation. The corresponding dynamic PoFs are 15.1%, 22.1%, and 25.1%, respectively. Thus, proportional and egalitarian TS lose a larger share of clicks than their full information counterparts, yet deliver only 36% and 41% of their respective static minority utility gains. Item 2 of Proposition 2 explains why this higher percentage need not imply more lost clicks because learning by blind

TS reduces the denominator. Table W14 reports that, for example, egalitarian TS loses 2,354 clicks relative to blind TS, below its static fairness cost of 2,514 clicks. Utilitarian TS incurs a lower click loss but achieves just 4% of its static minority utility gain. Comparisons across the Yahoo and ZOZO applications reveal that realized click costs and delivered fairness can diverge: lower costs can come with incomplete fairness gains, and higher costs with unmet fairness targets. In this sparse feedback environment, substantial efficiency losses coexist with incomplete learning of the allocations needed to deliver the target benefits. We next examine how these costs and benefits evolve over the learning horizon.

6.3 Learning Dynamics

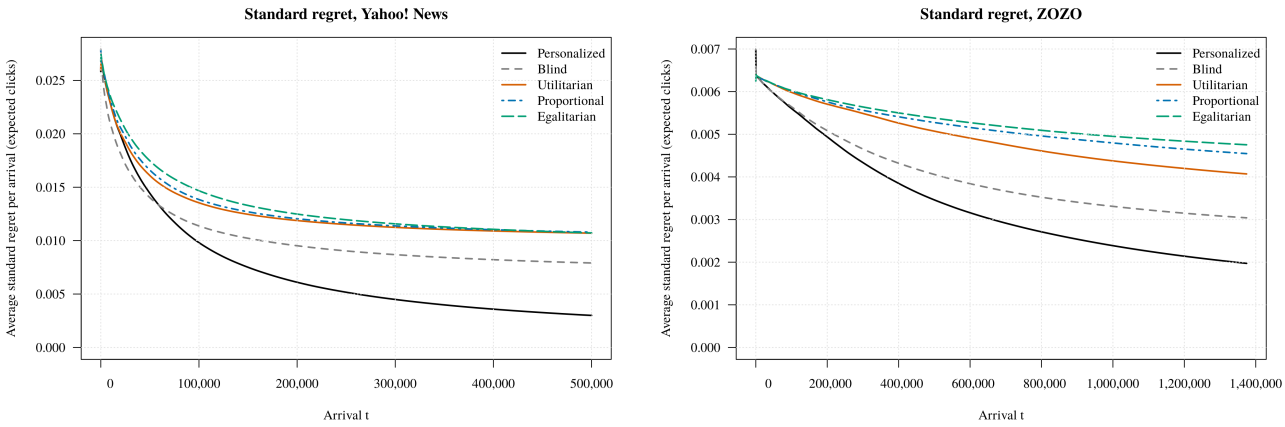
The sequential learning process in the proposed bandits shapes both standard regret, measured by forgone clicks, and ethical regret, which captures the shortfall relative to the selected ethical objective. Moreover, fairness improvements may emerge gradually as segment-specific preferences are learned, even as firms incur the associated efficiency costs. We examine these two aspects of learning dynamics below.

Standard vs. Ethical Regrets We report two metrics of learning performance over a finite horizon. Standard regret measures the clicks lost relative to an oracle that knows the true CTRs and always features the preferred item of each segment. Ethical regret, defined in (30), instead captures the gap between each ethical policy's actual performance and what it would have achieved with known CTRs. At each arrival, both metrics are averaged over the visitors served so far and across the 30 runs reported in Tables 1 and 2.

Figure 3 compares standard regret declines across policies as learning progresses. The ethical policies consistently incur greater click losses than blind TS, reflecting their emphasis on balancing segment utilities rather than maximizing total clicks. By the end of the Yahoo simulation, cumulative click losses per 1,000 visitors relative to the oracle are 3.0 for personalized TS, 7.9 for blind TS, and 10.7, 10.8, and 10.7 for utilitarian, proportional, and egalitarian TS, respectively. The corresponding figures for ZOZO are 2.0, 3.0, 4.07, 4.55, and 4.75. The smaller absolute losses in ZOZO primarily reflect its substantially lower baseline CTRs. Notably, the three ethical policies incur nearly identical click losses in Yahoo but differ more substantially in ZOZO, reflecting both the greater convergence toward full information allocations in Yahoo and the more distinct optimal product mixes in ZOZO, where learning remains incomplete.

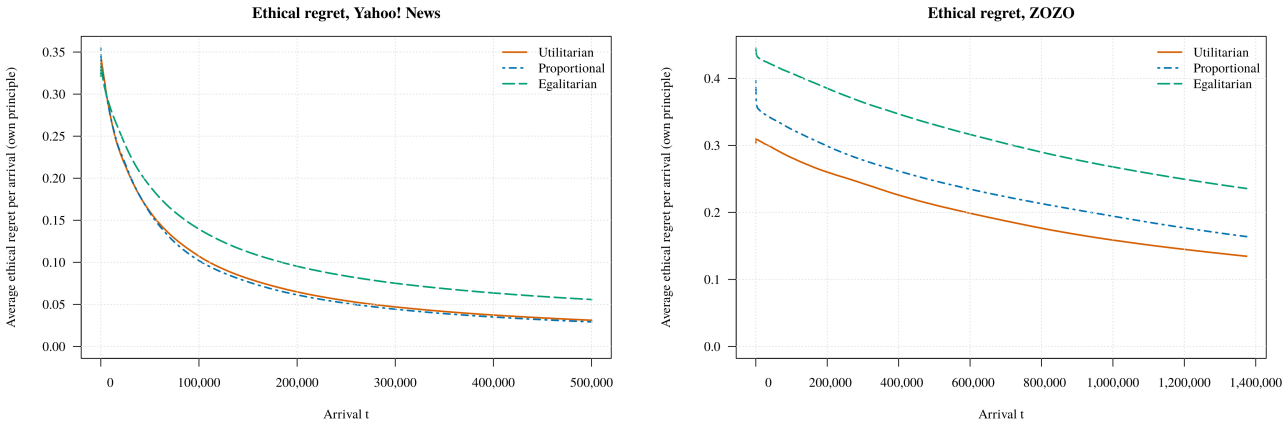
Figure 4 reveals differences in learning performance that standard regret alone does not

capture. In Yahoo, ethical regret falls to 48%–58% of its initial value after 50,000 visitors. Utilitarian and proportional TS follow similar trajectories, whereas egalitarian TS reduces its regret more slowly relative to its initial level. Despite incurring nearly identical standard regrets, these policies approach their respective ethical targets at different speeds.



(a) Yahoo **(b)** ZOZO

Figure 3: Standard Regrets for Benchmark and Ethical Policies

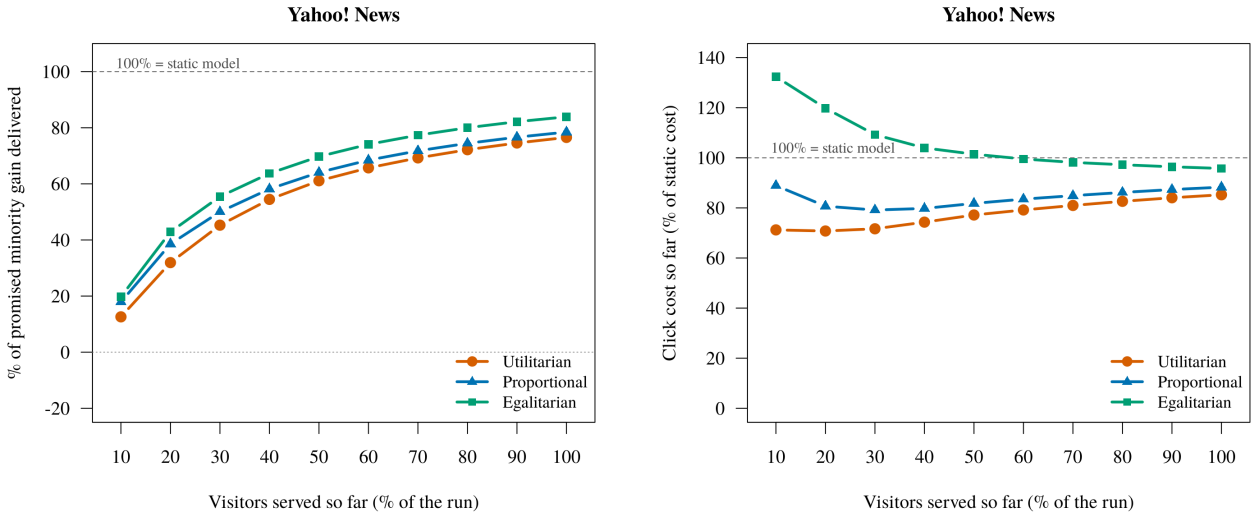


(a) Yahoo **(b)** ZOZO

Figure 4: Ethical Regrets for Ethical Policies

Ethical learning is substantially slower in ZOZO, where clicks are approximately ten times rarer. After 50,000 visitors, ethical regret remains at 93%–96% of its initial value and declines to only 43%–54% after all 1.37 million visitors. Also, Figure 4b shows the persistence of egalitarian regret as it remains substantially elevated even after more than one million visitors, and its average ethical regret over the final 10% of arrivals is still 0.139 in ZOZO, compared with 0.024 in Yahoo (Web Appendices I and J). This empirical pattern is consistent with the greater information demands of egalitarian allocation (Bistritz et al. 2020; Sarkar et al. 2026).

Timing of Fairness Improvements Welfare improvements over blind TS can lag behind progress toward an ethical target. Web Appendices I and J evaluate all policies under each ethical criterion, allowing comparisons on a common welfare objective. In Yahoo, the ethical policies approach their full information targets relatively quickly and achieve substantially higher welfare than blind TS under their corresponding criteria. In contrast, welfare improvements emerge more slowly in ZOZO, where proportional TS outperforms blind TS on the proportional criterion only after approximately 62% of arrivals, while utilitarian TS does so on the utilitarian criterion only during the final 3% of the run. Egalitarian TS, however, outperforms blind TS on its own criterion throughout, despite remaining short of its static target. These findings distinguish the difficulty of reaching an ethical target from the ability to improve upon blind allocation. When feedback is sparse, improvement in the corresponding welfare criterion can take much of the horizon to appear. In fact, a policy designed to promote fairness may spend much of the learning horizon performing worse on its own objective than a policy that ignores fairness altogether.



(a) Minority Benefit: Utility Improvement **(b)** Firm Cost: Clicks Lost to Blind

Figure 5: Dynamics of Minority Benefits and Firm Cost in Yahoo

Beyond these welfare comparisons, we examine how quickly ethical policies deliver improvements in minority utility and the associated costs to firms. Table 1 reports the average outcomes over an entire learning horizon, but these averages obscure when fairness benefits emerge. Ethical bandits must first learn consumer preferences across segments, and learning about the minority relies on observations from these visitors. In Yahoo, the minority accounts for only 12% of the traffic. To examine the timing, we evaluate the Yahoo run at successive stopping points, after 10%, 20%, and so on up to 100% of visitors have arrived. At each stopping point,

we ask how much of the improvement in minority utility promised by the static ethical model has actually been delivered. For example, 100% indicates that the minority receives the full fairness gain of the static model, consistent with the convergence result in Theorem 1, whereas 0% indicates no improvement relative to the blind.

Figure 5 illustrates the dynamics of minority benefits and firm costs under the ethical bandits throughout the sequence of consumer arrivals in the Yahoo data. The estimates are averaged over the same 30 runs reported in Table 1, with the final points corresponding to the full horizon results in that table. Each point in Figure 5a represents the difference in minority utility between an ethical policy and blind TS, divided by the corresponding difference when CTRs are known a priori. This figure shows that, while the price of fairness is paid upfront, its benefits materialize gradually. After the first 10% of visitors (50,000 individuals), the three ethical bandits have delivered only 13%–20% of the minority utility gains promised by the corresponding static models. This proportion increases to 61%–70% halfway through the sequence and reaches 77%–84% by the end. These delays reflect that a firm pursuing an ethical objective cannot effectively serve minority consumers until it learns their preferences, a process requiring substantial data collection given that minority consumers account for only 12% of the visitors in the Yahoo data. By contrast, the static model abstracts away from this learning process and assumes that the full benefits of fairness are realized from the outset.

The cost of fairness to the firm, however, is already paid early in the run, as Figure 5b shows. After the first 10% of visitors, utilitarian and proportional TS have reached 71% and 89% of their static PoFs, respectively, while egalitarian TS reaches 132%. By the end of the run, all three ethical bandits reach 85%–96% of their static PoFs. In Yahoo, delayed minority benefits coexist with click losses over the full horizon below the static fairness cost (Web Appendix H). Web Appendix F provides the estimates underlying Figure 5b. The corresponding ZOZO analysis in Web Appendix E shows a more pronounced gap between costs and benefits. Over 1.37 million visits, the ethical policies deliver only 4%–41% of their static minority utility gains, while proportional and egalitarian TS incur dynamic PoFs above their static benchmarks.

These dynamics bear on which ethical rule suits which business context. For firms prioritizing the worst off segment, the substantial early costs of egalitarian TS make a long horizon and frequent feedback particularly important. Its costs can remain substantial even when the full information utilitarian and proportional policies select the same article as the blind benchmark. On Yahoo Days 1, 4, and 6, egalitarian TS loses 8.7%, 2.5%, and 6.4% of clicks relative to blind TS, compared with 0.2%–1.8% for utilitarian and proportional TS (Web Appendix C). These

lower costs make utilitarian and proportional rules potentially more suitable for products with short life cycles, such as news articles or seasonal fashion items, when their full information allocations coincide with the blind benchmark. When feedback is sparse, firms should assess how much of the intended fairness gain can be achieved within the available horizon. Pooling information across segments with similar preferences could accelerate learning, a possibility we leave for future research.

6.4 Robustness Checks

We conduct two robustness checks to test the sensitivity of our empirical findings to key assumptions underlying the analyses. First, we relax the assumption regarding the number of customer segments and examine how varying this number affects the performance of ethical TS. Second, we vary the relative size of minority segment to assess its impact on bandit learning. All other settings remain largely unchanged from the main analyses.

Number of Segments For Yahoo, we partition the range of the selected sensitive user feature into 2, 3, 4, or 5 segments of equal width. For ZOZO, where the sensitive feature has four categories, we consider segmentations of 2, 3, or 4 groups, assigning each of the largest categories to a separate segment and combining the remaining categories into the final segment. For each segmentation, we recalculate segment sizes and CTRs. Table 3 reports utility of the smallest segment and click loss of each ethical policy relative to the blind. Similar to Tables 1 and 2, results are averaged over all arrivals and across 30 runs for Yahoo and 10 runs for ZOZO.

In Yahoo, the finer the segmentation, the worse all policies generally serve the smallest segment, although ethical TS consistently delivers higher utility than blind TS. Across the tested segmentations, utility remains around 0.90–0.94 under utilitarian and proportional policies and 0.78–0.84 under egalitarian. The lower utility under egalitarian is consistent with its objective of protecting the worst off segment, which need not be the smallest by segment size. Meanwhile, egalitarian's click loss rises from 3.8% with two segments to 7.2% with five, whereas utilitarian and proportional losses remain around 5%. Finer segmentation can reveal additional preference conflicts and leave fewer observations for each segment-specific posterior, potentially increasing the click cost of protecting the worst off.

In ZOZO, smallest-segment utility also declines under every policy as segmentation becomes finer, but egalitarian TS has the highest point estimate in all three configurations. This ranking is consistent with the three-segment results in Table 2, where the minority remains the worst

Table 3: Robustness Check: Varying Number of Segments

Number	Segment Size (%)	Smallest Segment Utility				Clicks Lost to Blind (%)		
Segments	Largest to Smallest	Blind	Utilitarian	Proportional	Egalitarian	Utilitarian	Proportional	Egalitarian
Panel A: Yahoo Data								
2	80.64/19.36	0.629	0.942	0.942	0.840	5.1	5.2	3.8
		[0.620, 0.638]	[0.937, 0.947]	[0.937, 0.948]	[0.834, 0.847]	[4.9, 5.3]	[5.0, 5.4]	[3.6, 4.0]
3	70.21/17.79/12.00	0.556	0.924	0.933	0.819	5.2	5.4	5.3
		[0.548, 0.564]	[0.917, 0.931]	[0.925, 0.941]	[0.812, 0.825]	[5.0, 5.4]	[5.2, 5.6]	[5.0, 5.5]
4	57.85/22.79/10.68/8.68	0.511	0.902	0.928	0.789	5.0	5.2	6.6
		[0.503, 0.520]	[0.890, 0.914]	[0.921, 0.935]	[0.785, 0.794]	[4.8, 5.2]	[5.0, 5.4]	[6.4, 6.8]
5	51.27/23.59/10.37/8.09/6.69	0.503	0.909	0.927	0.779	5.1	5.1	7.2
		[0.476, 0.531]	[0.896, 0.922]	[0.922, 0.932]	[0.773, 0.785]	[4.7, 5.4]	[4.7, 5.4]	[6.8, 7.6]
Panel B: ZOZO Data								
2	53.04/46.96	0.645	0.666	0.740	0.750	3.1	8.7	11.2
		[0.630, 0.660]	[0.645, 0.687]	[0.720, 0.759]	[0.737, 0.763]	[2.2, 4.0]	[7.9, 9.6]	[10.0, 12.4]
3	53.04/35.36/11.59	0.295	0.301	0.387	0.411	15.1	21.8	25.0
		[0.290, 0.300]	[0.286, 0.316]	[0.373, 0.400]	[0.397, 0.425]	[13.5, 16.8]	[20.8, 22.7]	[24.3, 25.7]
4	53.04/35.36/11.11/0.49	0.017	0.053	0.058	0.067	32.8	36.2	36.4
		[0.013, 0.020]	[0.047, 0.060]	[0.051, 0.064]	[0.057, 0.077]	[31.6, 34.0]	[35.4, 37.0]	[35.8, 37.1]

Notes: 95% confidence intervals for click loss are based on 10,000 bootstrap resamples per policy.

off under all ethical policies. The click loss relative to blind TS increases substantially across the ethical rules, from 3.1%–11.2% with two segments to 32.8%–36.4% with four. With four segments, the smallest group accounts for merely 0.49% of arrivals and records only 12 clicks in the data. Its utility falls below 0.07 under every policy, highlighting the difficulty of maintaining fairness when the minority segment becomes extremely small. These comparisons also change segment composition, CTRs, and utility scaling, so the observed patterns reflect changes in the groups being evaluated as well as the conditions for learning.

Size of Minority Segment In Web Appendix G, we vary the minority's share of arrivals from 1% to equal shares across segments, conditional on each segment's CTRs at their empirical values. Both applications show a similar pattern. When the minority is very small, ethical policies offer limited benefits. At 1% of arrivals, they increase its utility from 0.53 under blind to 0.61–0.76 in Yahoo, but provide no improvement in ZOZO, as too few minority visitors make bandit learning difficult. As its share grows, the gains become substantial. At 10% in Yahoo, utility rises from 0.55 under blind to 0.82–0.93 under ethical policies; at 20% in ZOZO, it rises from 0.31 to 0.42 under proportional and to 0.44 under egalitarian. However, when the minority accounts for 1/3 of arrivals, the blind bandit begins serving it well because its preferences carry sufficient weight in maximizing overall clicks. In Yahoo, its utility under the blind reaches 0.97, exceeding that under any ethical policy. Thus, fairness principles matter most when the minority is small enough to be overlooked by the blind bandit but large enough

for ethical policies to learn its preferences in a finite time horizon.

The decomposition in Web Appendix H complements these utility results. At a 1% minority share, egalitarian TS in Yahoo loses 5,925 clicks relative to blind TS, compared with a static fairness cost of 3,868; proportional TS in ZOZO loses 3,196 clicks, compared with 2,811. In both cases, the ethical policy incurs a higher learning cost than blind TS, so limited minority benefits coexist with click losses above the static fairness cost. This additional cost is not confined to small minorities: at larger shares, differences in learning costs can generate click gains or losses even when the static fairness cost is zero. Thus, segment composition and the ethical rule jointly shape both the benefits delivered to the minority and whether learning amplifies or offsets the static cost of fairness.

7 Conclusion

This paper develops an ethical bandit framework that integrates utilitarian, proportional, and egalitarian principles into Thompson Sampling for allocating limited marketing resources among heterogeneous consumers. Through information separation, the framework combines demographic-aware posterior updating within a restricted learning module with demographic-blind resource allocation, allowing firms to accommodate differences in consumer preferences without conditioning individual decisions on the arriving consumer's privacy information. The model requires sensitive customer attributes for the restricted preference learning, which reflects the information needs to identify utility disparities across demographic groups and learn how to mitigate them. The practices at LinkedIn, Airbnb, and Meta discussed in Section 4.1 illustrate how demographic information can support disparity assessment and mitigation while limiting downstream access to individual sensitive attributes. These organizational examples suggest the feasibility of decoupling sensitive data access from marketing decisions.

Our study extends the static price of fairness concept to a dynamic learning environment. Using randomized exposure data from Yahoo and ZOZO in two empirical applications, we show that learning shapes both the magnitude and timing of the efficiency-equity tradeoff. Click costs emerge immediately after learning starts, whereas fairness benefits develop gradually and remain below their full information benchmarks over the evaluated horizons. In Yahoo, ethical policies deliver substantial minority improvements with relatively modest click losses, but their lower dynamic PoFs coexist with incomplete fairness gains. In the sparser ZOZO environment, proportional and egalitarian TS incur dynamic PoFs above their static counterparts while

delivering smaller fractions of the anticipated minority benefits. Some ethical policies also spend much of the learning horizon performing worse than blind TS under the welfare criterion they seek to optimize. These findings distinguish the cost of adopting an ethical objective from the fairness actually delivered. A lower dynamic PoF can coexist with smaller fairness gains, and higher costs can coexist with limited welfare improvements.Outcomes further depend on preference heterogeneity, segment composition, and the ethical principle, with extremely small segments particularly difficult to serve.

Methodologically, we establish statistical guarantees connecting preference learning, ethical allocation, and realized decision outcomes. Under standard regularity conditions, we prove posterior preference consistency, convergence of learned allocations to the set of full information ethical optima, as well as convergence of ethical welfare to its optimal value. We introduce ethical regret to measure welfare losses during learning, establish vanishing average ethical regret, and derive finite sample bounds in terms of utility estimation error. Concentration bounds further show that fluctuations from categorical product sampling have a vanishing effect on average realized utilities. Finally, we decompose dynamic PoF into the static fairness cost and the difference in learning costs between the two policies and establish conditions under which dynamic PoF converges to its static counterpart.

Our research suggests three practical considerations for deploying ethical bandits, both in product featuring and in other resource-constrained marketing settings, such as promotional communications under customer contact limits. First, examining item-level preferences and utility disparities under blind TS helps identify which segments are underserved and the scope for improvement. This assessment should consider both segment definitions and sizes, since similar aggregate response rates may conceal different distributional outcomes. Second, assessing whether sufficient informative feedback can arrive within the operating horizon is particularly valuable for smaller segments. With sparse responses, an ethical policy can incur significant click losses before outperforming blind TS on its own welfare criterion. Third, the choice of welfare objective should reflect both managerial priorities and the learning horizon. The egalitarian rule prioritizes the segment with the lowest utility, which need not be the smallest segment, and may pay substantial costs early in learning. Long horizons and frequent feedback are particularly important when protecting the worst off is the firm's priority. Utilitarian and proportional rules may be more suitable for products with short life cycles when their lower click costs are accompanied by meaningful welfare gains within the available horizon. Considering these dimensions collectively may help firms match ethical objectives to preference structure,

learning opportunities, and welfare priorities.[5]

Algorithmic audits are typically conducted periodically, often annually, with resulting policy changes taking effect months after data collection. Our method enables real time auditing and policy adjustment. By continuously updating segment-level preference estimates, the ethical bandit allows firms to monitor each segment's estimated utility under the current allocation, disparities across segments, and deviations from the chosen ethical target. Because the ethical objective is embedded in each allocation decision, detected disparities can be addressed immediately rather than at the next review. However, this model-based monitoring depends on the accuracy of preference estimates, particularly limited early in learning and for small segments, and covers only predefined segments and ethical objectives. It complements periodic external audits, which can examine other groups and forms of harm.

It is also important to acknowledge limitations of our model and highlight opportunities for future research. First, the ethical bandit requires sensitive consumer attributes for restricted preference learning. Its applicability depends on whether these attributes can be collected and processed for that purpose in relevant organizational settings. Future research could examine whether alternative learning architectures, including federated computation, can reduce access to sensitive data and what additional privacy guarantees they can establish. Second, our theoretical analysis assumes stationary consumer preferences, independent arrivals, and sufficient exploration. Extending the algorithm to nonstationary cases, evolving consumer segments, and limited bandit feedback would broaden its applicability. Third, our analysis measures consumer benefit using click rates, which capture engagement but may not fully reflect consumer satisfaction or long term welfare. Future research could incorporate alternative measures such as post consumption satisfaction to better capture consumer welfare beyond immediate engagement. Fourth, although the empirical analyses consider two distinct data environments, the actual benefits of ethical allocation are still context dependent. Field experiments could further evaluate the effectiveness of an ethical bandit in operation. Moreover, we consider fixed ethical objectives within TS. Future research could investigate mixed strategies that switch ethical principles in response to evolving managerial priorities and fairness requirements. Finally, it would be interesting to extend our approach to other MAB algorithms such as Upper Confidence Bound, Gittins index, Whittle index, and others.

[5]We have developed a teaching case, "Private Data: Friend or Foe?", with an interactive dashboard illustrating these allocation tradeoffs. See Web Appendix K for more details.

References

Airbnb (2024), 'How we're using data to make travel more open for all', Airbnb Newsroom. December 12.
URL: *https://news.airbnb.com/2024-project-lighthouse-update*

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

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.

Badrinarayanan, Saikrishna, Osonde Osoba, Miao Cheng, Ryan Rogers, Sakshi Jain, Rahul Tandra and Natesh S. Pillai (2024), 'Privacy-preserving race/ethnicity estimation for algorithmic bias measurement in the U.S.', *arXiv preprint arXiv:2409.04652* .

Baek, Jackie and Vivek Farias (2021), 'Fair exploration via axiomatic bargaining', *Advances in Neural Information Processing Systems* **34**, 22034–22045.

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* .

Bentham, Jeremy (2024 [1789]), From an introduction to the principles of morals and legislation, *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.

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

Bertsimas, Dimitris, Vivek F Farias and Nikolaos Trichakis (2012), 'On the efficiency-fairness trade-off', *Management Science* **58**(12), 2234–2250.

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.

Bistritz, Ilai, Tavor Baharav, Amir Leshem and Nicholas Bambos (2020), My fair bandit: Distributed learning of max-min fairness with multi-player bandits, *in* 'International Conference on Machine Learning', PMLR, pp. 930–940.

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.

Caiata, Joshua, Carter Blair and Kate Larson (2026), 'Procedural fairness in multi-agent bandits', *arXiv preprint arXiv:2601.10600* .

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.

Chohlas-Wood, Alex, Madison Coots, Henry Zhu, Emma Brunskill and Sharad Goel (2026), 'Learning to be fair: A consequentialist approach to equitable decision making', *Management Science* **72**(1), 456–473.

Chu, Wei, Seung-Taek Park, Todd Beaupre, Nitin Motgi, Amit Phadke, Seinjuti Chakraborty and Joe Zachariah (2009), A case study of behavior-driven conjoint analysis on yahoo! front page today module, *in* 'Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining', pp. 1097–1104.

Coenen, Anna (2019), 'How the new york times is experimenting with recommendation algorithms', *NYT Open, Medium* .

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', *Foundations and trends® in theoretical computer science* **9**(3-4), 211–487.

Dwork, Cynthia, Nitin Kohli and Deirdre Mulligan (2019), 'Differential privacy in practice: Expose your epsilons!', *Journal of Privacy and Confidentiality* **9**(2).

Fleurbaey, Marc, S. M. 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.

Gillen, Stephen, Christopher Jung, Michael Kearns and Aaron Roth (2018), Online learning with an unknown fairness metric, *in* 'Advances in Neural Information Processing Systems', Vol. 31.

Grazzi, Riccardo, Arya Akhavan, John Isak Texas Falk, Leonardo Cella and Massimiliano Pontil (2022), Group meritocratic fairness in linear contextual bandits, *in* 'Advances in Neural Information Processing Systems', Vol. 35, pp. 24392–24404.

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

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.

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.

Hossain, Safwan, Evi Micha and Nisarg Shah (2021), 'Fair algorithms for multi-agent multi-armed bandits', *Advances in Neural Information Processing Systems* **34**, 24005–24017.

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**.

Kaneko, Mamoru and Kenjiro Nakamura (1979), 'The nash social welfare function', *Econometrica: Journal of the Econometric Society* pp. 423–435.

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.

Korganbekova, Malika and Cole Zuber (2023), 'Balancing user privacy and personalization', *Work in progress* **6**.

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 Szepesvári (2020), *Bandit Algorithms*, Cambridge University Press.

Li, Fengjiao, Jia Liu and Bo Ji (2019), 'Combinatorial sleeping bandits with fairness constraints', *IEEE Transactions on Network Science and Engineering* **7**(3), 1799–1813.

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.

Liu, Yang, Goran Radanovic, Christos Dimitrakakis, Debmalya Mandal and David C. Parkes (2017), 'Calibrated fairness in bandits', *arXiv preprint arXiv:1707.01875* .

Meta (2023), 'A new system to help ensure ads are delivered fairly to different demographic groups', Meta AI Blog. January 9.
URL: *https://ai.meta.com/blog/advertising-fairness-variance-reduction-system-vrs/*

Patil, Vishakha, Ganesh Ghalme, Vineet Nair and Yadati Narahari (2021), 'Achieving fairness in the stochastic multi-armed bandit problem', *Journal of Machine Learning Research* **22**(174), 1–31.

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* **41**(3), 529–546.

Rafieian, Omid and Hema Yoganarasimhan (2021), 'Targeting and privacy in mobile advertising', *Marketing Science* **40**(2), 193–218.

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.

Saito, Yuta, Shunsuke Aihara, Megumi Matsutani and Yusuke Narita (2021), Open bandit dataset and pipeline: Towards realistic and reproducible off-policy evaluation, *in* 'Advances in Neural Information Processing Systems'.

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.

Sarkar, Dhruv, Soumyadeep Dutta and Sayak Ray Chowdhury (2026), 'Price of fairness in bandits: A tight minimax characterization', *arXiv preprint arXiv:2607.13402* .

Schumann, Candice, Zhi Lang, Nicholas Mattei and John P. Dickerson (2019), 'Group fairness in bandit arm selection', *arXiv preprint arXiv:1912.03802* .

Schwartz, Eric M., Eric T. Bradlow and Peter S. Fader (2017), 'Customer acquisition via display advertising using multi-armed bandit experiments', *Marketing Science* **36**(4), 500–522.

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

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, William 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 21st ACM International Conference on Knowledge Discovery and Data Mining', pp. 1195–1204.

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.

Wang, Zilong, Zhiyao Zhang and Shuai Li (2025), Optimal algorithm for max-min fair bandit, *in* 'Proceedings of the 42nd International Conference on Machine Learning', Vol. 267 of *Proceedings of Machine Learning Research*, PMLR, pp. 65480–65495.

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

Yahoo Research (2009), 'Machine learning'.
URL: *https://www.yahooinc.com/research/research-areas/machine-learning*

Web Appendices:
Ethical Thompson Sampling for Marketing Resource Allocation

The following supplementary materials are provided by the authors to facilitate understanding of the paper.

Web Appendix A: Proof of Theoretical Results 2

Web Appendix B: Ethical Weight Optimization Solvers 8

Web Appendix C: Sensitivity to Other Days in Yahoo 10

Web Appendix D: CTR and Utility Results 13

Web Appendix E: Cost and Benefit of Fairness over Time in ZOZO 17

Web Appendix F: Cost of Fairness over Time in Yahoo 18

Web Appendix G: Sensitivity to Size of Minority Segment 19

Web Appendix H: Decomposition of Price of Fairness for Data Application . . 20

Web Appendix I: Bandit Performance across Ethical Objectives in Yahoo 24

Web Appendix J: Bandit Performance across Ethical Objectives in ZOZO . . . 26

Web Appendix K: Teaching Case for Ethical Bandits 28

Web Appendix A: Proof of Theoretical Results

A.1 Special Cases of Generalized Power Mean

For strictly positive segment utilities $U_1, \ldots, U_K$, the expression in (6) satisfies

$$\lim_{\rho \to 1} W_\rho = \frac{1}{K} \sum_{k=1}^{K} U_k \qquad \text{(Arithmetic Mean)}, \tag{39}$$

$$\lim_{\rho \to 0} W_\rho = \left(\prod_{k=1}^{K} U_k \right)^{1/K} \qquad \text{(Geometric Mean)}, \tag{40}$$

$$\lim_{\rho \to -1} W_\rho = \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{U_k} \right)^{-1} \qquad \text{(Harmonic Mean)}, \tag{41}$$

$$\lim_{\rho \to \infty} W_\rho = \max_{1 \le k \le K} U_k \qquad \text{(Maximum)}, \tag{42}$$

$$\lim_{\rho \to -\infty} W_\rho = \min_{1 \le k \le K} U_k \qquad \text{(Minimum)}. \tag{43}$$

Proof. For different values of ρ,

Case 1: Arithmetic Mean ($\rho = 1$) Substituting $\rho = 1$ gives the result.

Case 2: Geometric Mean ($\rho \to 0$) Take logarithms, $\ln W_\rho = (1/\rho) \ln \left((1/K) \sum_{k=1}^{K} U_k^\rho \right)$. Since $U_k^\rho \to 1$, the numerator tends to $\ln 1 = 0$. Applying L'Hôpital's rule,

$$\lim_{\rho \to 0} \ln W_\rho = \lim_{\rho \to 0} \frac{\sum_{k=1}^{K} U_k^\rho \ln U_k}{\sum_{k=1}^{K} U_k^\rho} = \frac{1}{K} \sum_{k=1}^{K} \ln U_k.$$

Exponentiating,

$$\lim_{\rho \to 0} W_\rho = \exp \left(\frac{1}{K} \sum_{k=1}^{K} \ln U_k \right) = \left(\prod_{k=1}^{K} U_k \right)^{1/K}.$$

Case 3: Harmonic Mean ($\rho = -1$) Substituting $\rho = -1$ gives the result.

Case 4: Maximum ($\rho \to \infty$) Let $U_{\max} = \max U_k$. Then $U_k^\rho \le U_{\max}^\rho$. Hence,

$$\frac{1}{K} U_{\max}^\rho \le \frac{1}{K} \sum_{k=1}^{K} U_k^\rho \le U_{\max}^\rho.$$

Taking the $1/\rho$ power,

$$K^{-1/\rho} U_{\max} \le W_\rho \le U_{\max}.$$

Since $K^{-1/\rho} \to 1$, then

$$\lim_{\rho \to \infty} W_\rho = U_{\max}.$$

Case 5: Minimum ($\rho \to -\infty$) Let $U_{\min} = \min U_k$. Write $r = -\rho$ and $r \to \infty$, then

$$W_\rho = \left(\frac{1}{K} \sum_{k=1}^{K} U_k^{-r} \right)^{-1/r}.$$

Since $U_{\min}^{-r} = \max_k U_k^{-r}$, we obtain

$$\frac{1}{K} U_{\min}^{-r} \leq \frac{1}{K} \sum_{k=1}^{K} U_k^{-r} \leq U_{\min}^{-r}.$$

Taking the $-1/r$ power reverses the inequalities, $U_{\min} \leq W_\rho \leq K^{1/r} U_{\min}$. As $K^{1/r} \to 1$,

$$\lim_{\rho \to -\infty} W_\rho = U_{\min}.$$

□

A.2 Proof of Lemma 1

Proof. Fix (k, a). Conditional on the number of observations $N_{ka}(t) = n$, let $S_{ka}(t)$ denote the corresponding number of successes. With a Beta$(\alpha_{ka,0}, \beta_{ka,0})$ prior, the posterior is

$$p_{ka} \mid \mathcal{H}_t \sim \text{Beta}\left(\alpha_{ka,0} + S_{ka}(t), \beta_{ka,0} + n - S_{ka}(t)\right). \tag{44}$$

The posterior mean is

$$m_{ka,t} = \frac{\alpha_{ka,0} + S_{ka}(t)}{\alpha_{ka,0} + \beta_{ka,0} + n}. \tag{45}$$

By Assumption 2, $n \to \infty$ almost surely. Conditional on the occasions when pair (k, a) is observed, the outcomes are independent Bernoulli draws with mean p_{ka}. Thus,

$$\frac{S_{ka}(t)}{N_{ka}(t)} \longrightarrow p_{ka} \qquad \text{almost surely.} \tag{46}$$

It follows that

$$m_{ka,t} \longrightarrow p_{ka} \qquad \text{almost surely.} \tag{47}$$

The posterior variance is

$$s_{ka,t}^2 = \frac{(\alpha_{ka,0} + S_{ka}(t))(\beta_{ka,0} + N_{ka}(t) - S_{ka}(t))}{\big(\alpha_{ka,0} + \beta_{ka,0} + N_{ka}(t)\big)^2 \big(\alpha_{ka,0} + \beta_{ka,0} + N_{ka}(t) + 1\big)}. \tag{48}$$

Using $xy \leq (x+y)^2/4$ gives

$$s_{ka,t}^2 \leq \frac{1}{4\big(\alpha_{ka,0} + \beta_{ka,0} + N_{ka}(t) + 1\big)} \longrightarrow 0. \tag{49}$$

For any $\varepsilon > 0$, by the triangle inequality,

$$|\widetilde{p}_{ka,t} - p_{ka}| \leq |\widetilde{p}_{ka,t} - m_{ka,t}| + |m_{ka,t} - p_{ka}|.$$

Therefore,

$$\{|\widetilde{p}_{ka,t} - p_{ka}| > \varepsilon\} \subseteq \left\{|\widetilde{p}_{ka,t} - m_{ka,t}| > \frac{\varepsilon}{2}\right\} \cup \left\{|m_{ka,t} - p_{ka}| > \frac{\varepsilon}{2}\right\}.$$

Applying the union bound gives

$$\begin{aligned} \Pr\left(|\widetilde{p}_{ka,t} - p_{ka}| > \varepsilon\right) \leq \Pr\left(|\widetilde{p}_{ka,t} - m_{ka,t}| > \frac{\varepsilon}{2}\right) \\ + \Pr\left(|m_{ka,t} - p_{ka}| > \frac{\varepsilon}{2}\right). \end{aligned}$$

The first term converges to zero because the posterior variance converges to zero. The second term converges to zero because the posterior mean is consistent by the strong law of large numbers. Thus,

$$\widetilde{p}_{ka,t} \longrightarrow p_{ka} \qquad \text{in probability.}$$

The joint convergence in (24) follows from a union bound over the finite set of KA entries.

□

A.3 Proof of Lemma 2

Proof. Fix a segment k and define

$$M_P = \max_a p_{ka}, \qquad M_Q = \max_a q_{ka}.$$

The maximum operator is 1-Lipschitz under supremum norm, so

$$|M_Q - M_P| \le \|Q - P\|_\infty.$$

Since $M_P \ge M_{\min} > 0$ and $0 < q_{ka}/M_Q \le 1$,

$$\begin{aligned}
|\mathcal{N}_k(Q)_a - \mathcal{N}_k(P)_a| &= \left| \frac{q_{ka}}{M_Q} - \frac{p_{ka}}{M_P} \right| \\
&= \left| \frac{q_{ka} - p_{ka}}{M_P} + \frac{q_{ka}}{M_Q} \frac{M_P - M_Q}{M_P} \right| \\
&\le \frac{|q_{ka} - p_{ka}| + |M_P - M_Q|}{M_P} \\
&\le \frac{2}{M_{\min}} \|Q - P\|_\infty.
\end{aligned}$$

Taking the maximum over segments k and products a establishes (25). The Beta posterior draws lie in $(0, 1)$ almost surely. Applying (25) to $Q = \widetilde{P}_t$ gives

$$\|\widetilde{V}_t - V\|_\infty \le \frac{2}{M_{\min}} \|\widetilde{P}_t - P\|_\infty.$$

By Lemma 1, the right hand side converges to zero in probability, proving (26).

□

A.4 Proof of Lemma 3

Proof. For any $\psi \in \Delta_A$ and segment k,

$$\begin{aligned}
\left| U_k(\psi; \widetilde{V}_t) - U_k(\psi; V) \right| &= \left| \sum_{a=1}^{A} \psi_a (\widetilde{v}_{ka,t} - v_{ka}) \right| \\
&\le \sum_{a=1}^{A} \psi_a |\widetilde{v}_{ka,t} - v_{ka}| \\
&\le \|\widetilde{V}_t - V\|_\infty. \qquad (50)
\end{aligned}$$

For the utilitarian objective,

$$\left| W_{\mathfrak{U}}(\psi; \widetilde{V}_t) - W_{\mathfrak{U}}(\psi; V) \right| \le K \|\widetilde{V}_t - V\|_\infty. \qquad (51)$$

For the egalitarian objective, the inequality $|\min_k x_k - \min_k y_k| \le \max_k |x_k - y_k|$ implies

$$\left| W_{\mathfrak{E}}(\psi; \widetilde{V}_t) - W_{\mathfrak{E}}(\psi; V) \right| \le \|\widetilde{V}_t - V\|_\infty. \qquad (52)$$

For the proportional objective, use the fact

$$\prod_{k=1}^{K} x_k - \prod_{k=1}^{K} y_k = \sum_{k=1}^{K} (x_k - y_k) \left(\prod_{j<k} x_j \right) \left(\prod_{j>k} y_j \right),$$

then

$$\left|\prod_{k=1}^{K} x_k - \prod_{k=1}^{K} y_k\right| \leq \sum_{k=1}^{K} |x_k - y_k|$$

when $x_k, y_k \in [0,1]$. Since both scaled utilities and their convex combinations lie in $[0,1]$, we have

$$\left|W_{\mathfrak{P}}(\psi; \widetilde{V}_t) - W_{\mathfrak{P}}(\psi; V)\right| \leq \sum_{k=1}^{K} \left|U_k(\psi; \widetilde{V}_t) - U_k(\psi; V)\right| \leq K\|\widetilde{V}_t - V\|_\infty.$$

Thus, the three ethical objectives are Lipschitz continuous with respect to the scaled utility matrix. All bounds are uniform in ψ. The result follows from Lemma 2.

□

A.5 Proof of Theorem 1

Proof. Fix an ethical rule $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$. Since Δ_A is compact and $W_j(\psi; V)$ is continuous in ψ, the set Ψ_j^* defined in each of (11)–(13) is nonempty and compact. For any $\psi_j^* \in \Psi_j^*$, we apply the set stability argument underlying Berge's maximum theorem (Berge 1963). For any open set $\mathcal{O}$ containing Ψ_j^*, if $\Delta_A \setminus \mathcal{O}$ is nonempty, compactness and continuity imply the difference $\gamma_{\mathcal{O}} = W_j(\psi_j^*; V) - \max_{\psi \in \Delta_A \setminus \mathcal{O}} W_j(\psi; V) > 0$. Whenever $\sup_{\psi \in \Delta_A} \left|W_j(\psi; \widetilde{V}_t) - W_j(\psi; V)\right| < \gamma_{\mathcal{O}}/2$, every allocation outside $\mathcal{O}$ has a strictly lower sampled objective value than ψ_j^*. Thus, a sampled objective maximizer $\widetilde{\psi}_{j,t}$ must lie in $\mathcal{O}$. By Lemma 3, $\Pr(\widetilde{\psi}_{j,t} \in \mathcal{O}) \longrightarrow 1$. The conclusion is immediate if $\Delta_A \setminus \mathcal{O}$ is empty. Equivalently,

$$\inf_{\phi \in \Psi_j^*} \|\widetilde{\psi}_{j,t} - \phi\| \longrightarrow 0 \qquad \text{in probability.}$$

By definition, $\widetilde{\psi}_{j,t}$ maximizes $W_j(\psi; \widetilde{V}_t)$ over Δ_A, and hence $W_j(\psi_j^*; \widetilde{V}_t) - W_j(\widetilde{\psi}_{j,t}; \widetilde{V}_t) \leq 0$. Moreover, ψ_j^* maximizes the full information objective, so $W_j(\psi_j^*; V) \geq W_j(\widetilde{\psi}_{j,t}; V)$. Adding and subtracting the sampled objectives gives

$$\begin{aligned}
0 &\leq W_j(\psi_j^*; V) - W_j(\widetilde{\psi}_{j,t}; V) \\
&= \left[W_j(\psi_j^*; V) - W_j(\psi_j^*; \widetilde{V}_t)\right] \\
&\quad + \left[W_j(\psi_j^*; \widetilde{V}_t) - W_j(\widetilde{\psi}_{j,t}; \widetilde{V}_t)\right] \\
&\quad + \left[W_j(\widetilde{\psi}_{j,t}; \widetilde{V}_t) - W_j(\widetilde{\psi}_{j,t}; V)\right] \\
&\leq \left[W_j(\psi_j^*; V) - W_j(\psi_j^*; \widetilde{V}_t)\right] + \left[W_j(\widetilde{\psi}_{j,t}; \widetilde{V}_t) - W_j(\widetilde{\psi}_{j,t}; V)\right] \\
&\leq 2 \sup_{\psi \in \Delta_A} \left|W_j(\psi; \widetilde{V}_t) - W_j(\psi; V)\right|. \qquad (53)
\end{aligned}$$

By Lemma 3, the right hand side of (53) converges to zero in probability. Consequently,

$$W_j(\widetilde{\psi}_{j,t}; V) \longrightarrow W_j(\psi_j^*; V) \qquad \text{in probability.}$$

Since all $\psi_j^* \in \Psi_j^*$ yield the same optimal welfare value, this conclusion does not depend on the uniqueness of ψ_j^*. □

A.6 Proof of Corollary 1

Proof. Define

$$Z_{j,t} = W_j(\psi_j^*; V) - W_j(\widetilde{\psi}_{j,t}; V).$$

Theorem 1 implies $Z_{j,t} \to 0$ in probability. Because the one-period regret converges to zero while remaining uniformly bounded (Lemma 3), its time average also converges to zero. That is,

$$\frac{1}{T}\sum_{t=1}^{T} Z_{j,t} \longrightarrow 0 \qquad \text{in probability.}$$

□

A.7 Proof of Theorem 2

Proof. Let L_j denote a uniform Lipschitz constant for W_j with respect to the supremum norm error in V. The proof of Lemma 3 gives

$$L_{\mathfrak{U}} = K, \qquad L_{\mathfrak{E}} = 1, \qquad L_{\mathfrak{P}} = K. \tag{54}$$

Following the derivation in (53),

$$\begin{aligned} W_j(\psi_j^*; V) - W_j(\widetilde{\psi}_{j,t}; V) &= \Big[W_j(\psi_j^*; V) - W_j(\psi_j^*; \widetilde{V}_t)\Big] \\ &\quad + \Big[W_j(\psi_j^*; \widetilde{V}_t) - W_j(\widetilde{\psi}_{j,t}; \widetilde{V}_t)\Big] \\ &\quad + \Big[W_j(\widetilde{\psi}_{j,t}; \widetilde{V}_t) - W_j(\widetilde{\psi}_{j,t}; V)\Big] \\ &\le 2L_j \|\widetilde{V}_t - V\|_\infty. \end{aligned} \tag{55}$$

Taking expectations and summing over t yields $\mathbb{E}[\mathcal{R}_{j,T}] \le 2L_j \sum_{t=1}^{T} b_t$. Substitution of (54) gives (33)–(35). □

A.8 Proof of Proposition 1

Proof. Let $\mathcal{F}_t$ denote the history of arrivals, featured products, and clicks before t, which determines $\mathcal{H}_t$. Conditional on $\mathcal{F}_t$ and the ethical weight vector $\widetilde{\psi}_t$, the featured product is sampled from $A_t \sim \text{Categorical}(\widetilde{\psi}_t)$. Hence, $E[\xi_{k,t} \mid \mathcal{F}_t] = 0$, so that

$$\{\xi_{k,t}\}$$

forms a martingale difference sequence. Also, $0 \le v_{ka} \le 1$, which implies $|\xi_{k,t}| \le 1$. Applying the Azuma-Hoeffding inequality to the martingale $\sum_{t=1}^{T} \mathbb{I}\{k\}\xi_{k,t}$ yields

$$\Pr\left(\left|\sum_{t=1}^{T} \mathbb{I}\{k\}\xi_{k,t}\right| > \sqrt{2N_k(T)\log(2/\lambda)}\right) \le \lambda.$$

Dividing both sides by $N_k(T)$ gives the desired inequality. Finally, as $N_k(T)/T \to \pi_k$ almost surely, the bound for the average execution error is of order $1/\sqrt{\pi_k T}$, which converges to zero. □

A.9 Proof of Proposition 2

Proof. Below are the proofs of the three items.

Item 1 By (21), $C_{B,T} = TC_B - L_{B,T}$ and $C_{j,T} = TC(\psi_j^*) - L_{j,T}$. Subtracting yields the stated identity.

Item 2 Using Item 1, static PoF in (19), and dynamic PoF in (20), we obtain

$$\mathrm{PoF}_j(T) - \mathrm{PoF}_j = \frac{L_{j,T} - (1 - \mathrm{PoF}_j)L_{B,T}}{T \cdot C_B - L_{B,T}}. \tag{56}$$

Since the denominator equals $C_{B,T} > 0$, the difference is positive if and only if $L_{j,T} > (1 - \mathrm{PoF}_j)L_{B,T}$.

Item 3 Let $c_a = \sum_{k=1}^{K} \pi_k p_{ka}$, so that $C_B = \max_a c_a > 0$ and $C(\psi) = \sum_{a=1}^{A} c_a \psi_a$. For the blind TS, under Assumption 1, the arriving consumer's segment is independent of the past, and blind TS does not condition its action on that segment. Consequently, featuring product a yields a Bernoulli click with mean c_a, so blind TS faces a standard Bernoulli bandit. By iterated expectations, $\mathbb{E}[L_{B,T}]$ equals its expected pseudo-regret, which is sublinear (Agrawal and Goyal 2012). Since $L_{B,T} \geq 0$, Markov's inequality gives $L_{B,T}/T \xrightarrow{p} 0$. For the ethical TS, because $C(\psi)$ is constant on Ψ_j^* and $0 < c_a < 1$,

$$\left| C(\widetilde{\psi}_{j,t}) - C(\psi_j^*) \right| \leq \inf_{\phi \in \Psi_j^*} \left\| \widetilde{\psi}_{j,t} - \phi \right\|_1 \xrightarrow{p} 0, \tag{57}$$

where convergence follows from Theorem 1 and the equivalence of norms on $\mathbb{R}^A$. The click difference is bounded by 1, so it also converges to zero in $\mathcal{L}^1$. Hence, by the triangle inequality and Cesàro averaging,

$$\mathbb{E}\left| \frac{L_{j,T}}{T} \right| \leq \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left| C(\psi_j^*) - C(\widetilde{\psi}_{j,t}) \right| \longrightarrow 0. \tag{58}$$

It follows that $L_{j,T}/T \xrightarrow{p} 0$. Finally, dividing the numerator and denominator of the identity in Item 2 by T gives

$$\mathrm{PoF}_j(T) - \mathrm{PoF}_j = \frac{L_{j,T}/T - (1 - \mathrm{PoF}_j)L_{B,T}/T}{C_B - L_{B,T}/T} \xrightarrow{p} 0, \tag{59}$$

by the continuous mapping theorem, since $C_B > 0$.

□

Web Appendix B: Ethical Weight Optimization Solvers

At each allocation update, the ethical bandit draws CTRs from the segment-product posteriors and scales them to obtain the sampled utility matrix $\widetilde{V}_t$. Each segment's utility is linear in ψ and strictly positive on Δ_A. The utilitarian allocation follows directly from (37). The procedures below compute the proportional and egalitarian allocations. Replacing $\widetilde{V}_t$ with the true utility matrix V gives the corresponding full-information allocation problems.

Proportional Because the logarithm is strictly increasing, maximizing proportional welfare is equivalent to maximizing its logarithm. The resulting log objective is concave in ψ, and every local maximum over the simplex is global.

We first screen out products that are Pareto dominated under the current sampled utility matrix. Product a is *Pareto dominated* if another product a' satisfies $\widetilde{v}_{ka',t} \geq \widetilde{v}_{ka,t}$ for every segment k, with strict inequality for at least one segment. Shifting probability mass from a to a' cannot lower any segment's utility and cannot reduce proportional welfare, i.e., screening preserves the optimal objective value. It applies only to the current draw: a screened product retains its posterior and may reenter the consideration set after a subsequent posterior draw.

Let $\mathcal{A}_t$ denote the retained products. Starting from equal positive weights on $\mathcal{A}_t$ and zero weights on screened products, we apply the multiplicative update

$$\widetilde{\psi}_{a,t}^{(r+1)} = \widetilde{\psi}_{a,t}^{(r)} \frac{1}{K} \sum_{k=1}^{K} \frac{\widetilde{v}_{ka,t}}{U_k(\psi^{(r)})}, \qquad a \in \mathcal{A}_t, \tag{60}$$

where r indexes optimization iterations. The updates preserve the simplex because, for every k, $\sum_a \widetilde{\psi}_{a,t}\tilde{v}_{ka,t}/U_k(\psi) = 1$. Each step does not decrease the log objective, and the fixed points satisfy the optimality conditions of this concave problem. This is the classical algorithm for maximizing a sum of logarithms of linear functions (Cover 1984). We stop when the total absolute change in ψ falls below 10^{-11} or after 3,000 iterations.

Egalitarian The egalitarian problem maximizes the minimum segment utility. Introducing an auxiliary variable z gives the equivalent linear program

$$\begin{aligned} \max_{\widetilde{\psi}_t,\, z} \quad & z \\ \text{s.t.} \quad & \sum_{a=1}^{A} \widetilde{\psi}_{a,t}\, \widetilde{v}_{ka,t} \geq z, \qquad k = 1, \ldots, K, \\ & \sum_{a=1}^{A} \widetilde{\psi}_{a,t} = 1, \\ & \widetilde{\psi}_{a,t} \geq 0, \qquad a = 1, \ldots, A. \end{aligned} \tag{61}$$

In short, the firm's problem and the worst off segment's problem are two sides of the same linear program: we solve the segment's side, and the firm's optimal product weights are obtained from its solution as the shadow prices of the product constraints. Formally, (61) is the value of a zero-sum game in which the firm chooses $\widetilde{\psi}_t$ and an adversary selects the worst off segment (Dantzig 1951). The implementation adds 1 to every utility. This raises every segment's utility, and hence the optimal value, by exactly 1 and leaves the optimal allocations unchanged, but it makes the value $w = z + 1$ strictly positive. With the shifted utilities, the change of variables $y_a = \widetilde{\psi}_{a,t}/w$ turns (61) into

$$\min_{y \geq 0} \sum_{a=1}^{A} y_a \quad \text{s.t.} \quad \sum_{a=1}^{A} (\widetilde{v}_{ka,t} + 1)\, y_a \geq 1, \qquad k = 1, \ldots, K, \tag{62}$$

whose optimal value is $1/w^*$. Its linear programming dual is

$$\max_{x \geq 0} \sum_{k=1}^{K} x_k \quad \text{s.t.} \quad \sum_{k=1}^{K} (\widetilde{v}_{ka,t} + 1)\, x_k \leq 1, \qquad a = 1, \ldots, A. \tag{63}$$

We solve the dual with the simplex method (Dantzig 1963). The entering variable is chosen by Dantzig's rule (the most negative reduced cost), the leaving variable by the minimum ratio test with ties broken by the lowest index, and the solver stops at optimality or after 500 pivots. The optimal y equals the vector of dual values (shadow prices) of the product constraints in the dual problem, and normalizing it gives the egalitarian allocation, $\widetilde{\psi}_{a,t} = y_a / \sum_b y_b$. Dantzig's rule does not rule out cycling in degenerate problems; the pivot limit guards against this, and the accuracy check below shows that the solutions match an independent solver.

Updating and Sampling Posteriors are updated after every arrival. In the simulations, new CTRs are drawn and the allocation is reoptimized every two arrivals. Between allocation updates, the most recently computed distribution remains in use. Each arriving consumer's product is independently sampled from this distribution.

Numerical Accuracy The solvers were evaluated on $1,200$ randomly generated sampled utility matrices with 2 to 5 segments and either 20 or 80 products. For the egalitarian problem, the objective values matched those from the HiGHS linear programming solver (Huangfu and Hall 2018) as implemented in SciPy (Virtanen et al. 2020) to within 10^{-14} in every case. For the proportional problem, optimality conditions for the concave log objective provide an upper bound on the shortfall from the optimum. Expressed using the equivalent geometric mean $\left(\prod_{k=1}^{K} U_k(\psi)\right)^{1/K}$, the relative shortfall was within 10^{-6} in $1,162$ of the $1,200$ cases and within 0.02% in all cases.

Web Appendix C: Sensitivity to Other Days in Yahoo

Whereas the ZOZO analysis pools observations across the entire campaign, the Yahoo! Front Page data exhibit nonstationarity across days because news articles are short-lived and the candidate pool refreshes continuously. Consequently, our main analysis in Section 6.1 focuses on a single batch of 500,000 consecutive visits (Batch 9 of Day 7). In this appendix, we replicate the analysis across all other days in the dataset that contain at least nine batches of 500,000 consecutive visits—namely, Batch 9 of Days 1, 4, 5, and 6 (Days 2, 3, 8, 9, and 10 record fewer total visits and do not reach a ninth batch). To ensure reliable ground-truth CTR estimates for the offline simulations, we retain candidate articles with at least 30 impressions in every segment; this screening rule excludes a single article (109611) on Day 5 (which received 25 and 22 impressions in the intermediate and minority segments, respectively) and no articles on the remaining days. All other settings mirror Section 6.1: consumers are partitioned into three equal-width segments based on feature 5, utilities are scaled according to (1), and estimates are averaged over all 500,000 arrivals across 30 independent runs per policy. Table W1 summarizes minority utility and the price of fairness across all five days. Tables W2, W4, W6, and W8 report the segment distributions for Days 1, 4, 5, and 6, while Tables W3, W5, W7, and W9 report the full policy evaluation results in the format of Table 1.

Table W1: Minority Utility and Price of Fairness across Days in Yahoo

			Minority Utility					Static PoF (%)			Dynamic PoF (%)		
Day	Segment Shares (%)	Mean CTR (%)	Blind	Util.	Prop.	Egal.	Blind RUD	Util.	Prop.	Egal.	Util.	Prop.	Egal.
1	73.3 / 16.6 / 10.1	3.0 / 4.6 / 6.4	0.820	0.819	0.819	0.831	0.148	0.0	0.0	7.5	1.5	1.8	8.7
4	70.7 / 17.8 / 11.5	3.6 / 5.4 / 7.8	0.993	0.991	0.991	0.973	0.001	0.0	0.0	0.0	0.2	0.3	2.5
5	70.6 / 17.8 / 11.5	3.0 / 4.5 / 6.5	0.592	0.785	0.821	0.779	0.348	9.7	6.8	5.7	5.0	5.9	6.3
6	69.8 / 18.1 / 12.1	2.7 / 4.0 / 6.3	0.936	0.933	0.933	0.921	0.043	0.0	0.0	2.7	0.6	0.9	6.4
7 (Table 1)	70.2 / 17.8 / 12.0	2.7 / 4.0 / 5.8	0.556	0.924	0.933	0.819	0.389	6.1	6.1	5.5	5.2	5.4	5.3

Notes: All results are evaluated on Batch 9 of each day. Segment shares and mean CTRs are ordered as majority / intermediate / minority. Minority utility and the blind RUD are averaged over 500,000 arrivals and 30 runs. Static PoF measures the percentage click loss relative to the full-information blind benchmark; Dynamic PoF measures the realized percentage click loss relative to blind TS over the simulation horizon. A static PoF of 0.0 indicates that the full-information ethical policy selects the same article as the blind benchmark.

Key Findings. Because the pool of candidate articles changes daily, the degree of preference alignment across segments varies across days, yielding four main insights. First, *when segment preferences diverge, blind TS underserves the minority segment.* On Days 5 and 7, the segments prefer different articles. Under blind TS, minority utility falls to 0.592 and 0.556 (with RUD of 0.348 and 0.389), whereas the ethical TS policies raise minority utility to between 0.779 and 0.933 at a dynamic PoF of 5.0% to 6.3%. Conversely, on Day 4, a single article (109612) is the most preferred choice for every segment, allowing blind TS to achieve a minority utility of 0.993. On Days 1 and 6, although the minority's favorite article differs from the blind benchmark's selection, the blind-selected article already provides high utility to the minority (0.820 and 0.936).

Second, *when segment preferences align, utilitarian and proportional TS coincide with blind TS in the static limit.* On Days 1, 4, and 6, the full-information utilitarian and proportional policies select the same article as the blind benchmark, resulting in a static PoF of 0.0%. Over the finite learning horizon, however, utilitarian and proportional TS still incur a dynamic PoF of 0.2% to 1.8% relative to blind TS because they estimate segment-specific posteriors rather than pooling observations across the population.

Third, *egalitarian TS can incur substantial click losses even when realized minority gains are modest.* On Day 1, the full-information egalitarian policy mixes two articles to raise static minority utility from 0.828 to 0.890 at a static PoF of 7.5%; under sequential learning, egalitarian TS loses 8.7% of clicks relative to blind TS while increasing realized minority utility by

only 0.011 (from 0.820 to 0.831). Even on Day 4, where its static PoF is 0.0%, egalitarian TS incurs a dynamic PoF of 2.5% due to posterior uncertainty over which segment is worst off.

Fourth, *personalized TS does not always achieve the highest total clicks over a finite learning horizon.* On Days 4 and 6, personalized TS earns fewer total clicks than blind TS (56,822 vs. 57,554 on Day 4; 35,204 vs. 35,652 on Day 6). Because personalized TS learns each segment's preferences exclusively from arrivals within that segment, whereas blind TS pools arrivals across all segments, blind TS converges faster whenever all segments share the same optimal article. Personalized allocation is guaranteed to upper-bound aggregate clicks only under full information.

Table W2: Segment Summary Statistics in Yahoo (Day 1, Batch 9)

Segment	Feature 5 Range	Visits	Share (%)	Mean CTR (%)
Majority	[0.0000, 0.3317]	366,739	73.35	2.98
Intermediate	[0.3317, 0.6634]	82,992	16.60	4.59
Minority	[0.6634, 0.9951]	50,269	10.05	6.44

Notes: Segments correspond to three equal-width intervals of feature 5, ordered by population share. All 23 candidate articles exceeded the minimum threshold of 30 impressions per segment.

Table W3: Ethical vs. Benchmark Policies in Yahoo (Day 1, Batch 9)

	Consumer Utility				Firm Reward			
MAB Model	Majority	Interm.	Minority	RUD	Majority	Interm.	Minority	Total
Personalized	0.972	0.928	0.922	0.039	18,567	5,794	4,990	29,352
	[0.971, 0.973]	[0.925, 0.932]	[0.918, 0.926]	[0.035, 0.044]	[18,549, 18,586]	[5,772, 5,817]	[4,967, 5,013]	[29,322, 29,382]
Utilitarian	0.960	0.969	0.819	0.135	18,325	6,058	4,427	28,809
	[0.957, 0.962]	[0.968, 0.971]	[0.818, 0.820]	[0.133, 0.137]	[18,281, 18,368]	[6,050, 6,066]	[4,422, 4,432]	[28,760, 28,859]
Proportional	0.956	0.967	0.819	0.132	18,249	6,042	4,430	28,721
	[0.953, 0.958]	[0.965, 0.969]	[0.818, 0.820]	[0.130, 0.134]	[18,196, 18,302]	[6,030, 6,053]	[4,424, 4,436]	[28,657, 28,785]
Egalitarian	0.864	0.911	0.831	0.043	16,500	5,693	4,496	26,689
	[0.858, 0.870]	[0.908, 0.914]	[0.827, 0.836]	[0.033, 0.052]	[16,385, 16,615]	[5,672, 5,714]	[4,469, 4,522]	[26,574, 26,804]
Blind	0.978	0.981	0.820	0.148	18,676	6,132	4,435	29,243
	[0.977, 0.979]	[0.980, 0.982]	[0.820, 0.821]	[0.147, 0.149]	[18,654, 18,698]	[6,126, 6,138]	[4,432, 4,438]	[29,214, 29,272]

Notes: Estimates are averaged over 500,000 arrivals and 30 independent runs per policy on Batch 9 of Day 1, with 95% confidence intervals reported in brackets (analytic intervals across runs for utility and firm reward; bootstrap intervals for RUD).

Table W4: Segment Summary Statistics in Yahoo (Day 4, Batch 9)

Segment	Feature 5 Range	Visits	Share (%)	Mean CTR (%)
Majority	[0.0000, 0.3326]	353,739	70.75	3.58
Intermediate	[0.3326, 0.6652]	88,933	17.79	5.42
Minority	[0.6652, 0.9978]	57,328	11.47	7.83

Notes: Segments correspond to three equal-width intervals of feature 5, ordered by population share. All 27 candidate articles exceeded the minimum threshold of 30 impressions per segment.

Table W5: Ethical vs. Benchmark Policies in Yahoo (Day 4, Batch 9)

	Consumer Utility				Firm Reward			
MAB Model	Majority	Interm.	Minority	RUD	Majority	Interm.	Minority	Total
Personalized	0.987	0.975	0.958	0.024	33,574	13,957	9,291	56,822
	[0.987, 0.987]	[0.975, 0.976]	[0.956, 0.960]	[0.023, 0.025]	[33,556, 33,592]	[13,941, 13,973]	[9,270, 9,312]	[56,801, 56,844]
Utilitarian	0.990	0.989	0.991	0.002	33,667	14,154	9,612	57,432
	[0.989, 0.990]	[0.989, 0.989]	[0.991, 0.991]	[0.001, 0.002]	[33,658, 33,676]	[14,150, 14,158]	[9,610, 9,614]	[57,417, 57,447]
Proportional	0.989	0.988	0.991	0.002	33,645	14,144	9,607	57,395
	[0.989, 0.989]	[0.988, 0.989]	[0.990, 0.991]	[0.002, 0.002]	[33,634, 33,656]	[14,139, 14,149]	[9,604, 9,609]	[57,377, 57,413]
Egalitarian	0.966	0.965	0.973	0.006	32,876	13,806	9,432	56,114
	[0.965, 0.968]	[0.963, 0.966]	[0.972, 0.974]	[0.006, 0.006]	[32,831, 32,922]	[13,787, 13,826]	[9,422, 9,441]	[56,041, 56,188]
Blind	0.992	0.991	0.993	0.001	33,739	14,185	9,629	57,554
	[0.991, 0.992]	[0.991, 0.992]	[0.993, 0.993]	[0.001, 0.001]	[33,730, 33,748]	[14,181, 14,190]	[9,627, 9,631]	[57,539, 57,569]

Notes: Estimates are averaged over 500,000 arrivals and 30 independent runs per policy on Batch 9 of Day 4, with 95% confidence intervals reported in brackets (analytic intervals across runs for utility and firm reward; bootstrap intervals for RUD).

Table W6: Segment Summary Statistics in Yahoo (Day 5, Batch 9)

Segment	Feature 5 Range	Visits	Share (%)	Mean CTR (%)
Majority	[0.0000, 0.3321]	353,175	70.64	2.99
Intermediate	[0.3321, 0.6641]	89,139	17.83	4.51
Minority	[0.6641, 0.9962]	57,686	11.54	6.46

Notes: Segments correspond to three equal-width intervals of feature 5, ordered by population share. Of the 27 candidate articles, 26 were retained and 1 (article 109611) was excluded by the 30-impression screening rule.

Table W7: Ethical vs. Benchmark Policies in Yahoo (Day 5, Batch 9)

	Consumer Utility				Firm Reward			
MAB Model	Majority	Interm.	Minority	RUD	Majority	Interm.	Minority	Total
Personalized	0.963	0.919	0.941	0.012	17,538	6,390	6,930	30,857
	[0.960, 0.966]	[0.913, 0.925]	[0.939, 0.944]	[0.008, 0.015]	[17,482, 17,593]	[6,348, 6,431]	[6,909, 6,951]	[30,785, 30,929]
Utilitarian	0.816	0.892	0.785	0.049	14,858	6,203	5,780	26,841
	[0.793, 0.838]	[0.882, 0.902]	[0.751, 0.820]	[0.004, 0.103]	[14,445, 15,270]	[6,135, 6,272]	[5,527, 6,032]	[26,611, 27,071]
Proportional	0.792	0.882	0.821	0.012	14,422	6,132	6,043	26,596
	[0.780, 0.803]	[0.876, 0.887]	[0.802, 0.840]	[0.001, 0.042]	[14,209, 14,634]	[6,092, 6,171]	[5,903, 6,183]	[26,478, 26,715]
Egalitarian	0.808	0.867	0.779	0.044	14,721	6,029	5,734	26,483
	[0.805, 0.811]	[0.862, 0.871]	[0.774, 0.784]	[0.037, 0.052]	[14,663, 14,778]	[5,998, 6,059]	[5,695, 5,773]	[26,428, 26,539]
Blind	0.951	0.946	0.592	0.348	17,326	6,580	4,359	28,265
	[0.948, 0.954]	[0.941, 0.951]	[0.589, 0.596]	[0.343, 0.353]	[17,271, 17,381]	[6,545, 6,615]	[4,334, 4,384]	[28,197, 28,332]

Notes: Estimates are averaged over 500,000 arrivals and 30 independent runs per policy on Batch 9 of Day 5, with 95% confidence intervals reported in brackets (analytic intervals across runs for utility and firm reward; bootstrap intervals for RUD).

Table W8: Segment Summary Statistics in Yahoo (Day 6, Batch 9)

Segment	Feature 5 Range	Visits	Share (%)	Mean CTR (%)
Majority	[0.0000, 0.3324]	349,156	69.83	2.73
Intermediate	[0.3324, 0.6647]	90,303	18.06	3.97
Minority	[0.6647, 0.9971]	60,541	12.11	6.32

Notes: Segments correspond to three equal-width intervals of feature 5, ordered by population share. All 27 candidate articles exceeded the minimum threshold of 30 impressions per segment.

Table W9: Ethical vs. Benchmark Policies in Yahoo (Day 6, Batch 9)

	Consumer Utility				Firm Reward			
MAB Model	Majority	Interm.	Minority	RUD	Majority	Interm.	Minority	Total
Personalized	0.979	0.943	0.931	0.037	21,683	7,038	6,484	35,204
	[0.978, 0.979]	[0.940, 0.947]	[0.927, 0.935]	[0.034, 0.040]	[21,663, 21,702]	[7,011, 7,065]	[6,454, 6,514]	[35,164, 35,245]
Utilitarian	0.976	0.980	0.933	0.040	21,628	7,306	6,496	35,429
	[0.975, 0.977]	[0.979, 0.980]	[0.932, 0.933]	[0.039, 0.041]	[21,605, 21,651]	[7,301, 7,310]	[6,492, 6,499]	[35,402, 35,457]
Proportional	0.972	0.978	0.933	0.036	21,541	7,291	6,498	35,330
	[0.970, 0.974]	[0.977, 0.979]	[0.932, 0.934]	[0.035, 0.038]	[21,500, 21,583]	[7,285, 7,297]	[6,493, 6,503]	[35,285, 35,375]
Egalitarian	0.903	0.932	0.921	0.012	20,004	6,949	6,414	33,367
	[0.896, 0.909]	[0.930, 0.934]	[0.918, 0.924]	[0.005, 0.020]	[19,860, 20,147]	[6,932, 6,966]	[6,393, 6,436]	[33,226, 33,509]
Blind	0.983	0.985	0.936	0.043	21,786	7,349	6,518	35,652
	[0.982, 0.984]	[0.985, 0.986]	[0.936, 0.936]	[0.042, 0.043]	[21,767, 21,805]	[7,344, 7,353]	[6,515, 6,521]	[35,628, 35,677]

Notes: Estimates are averaged over 500,000 arrivals and 30 independent runs per policy on Batch 9 of Day 6, with 95% confidence intervals reported in brackets (analytic intervals across runs for utility and firm reward; bootstrap intervals for RUD).

Web Appendix D: CTR and Utility Results

To aid in interpreting the empirical results in Tables 1 and 2, Tables W10 and W11 report the segment-level impressions, empirical CTRs, and scaled utilities v_{ka} for all candidate articles in the Yahoo application (Day 7, Batch 9) and all candidate fashion items in the ZOZO application, respectively.

Table W10: Empirical Click Through Rates and Scaled Utilities in Yahoo

	Majority (70.21%)			Intermediate (17.79%)			Minority (12.00%)		
Article	Impr.	CTR	v	Impr.	CTR	v	Impr.	CTR	v
109635	15,496	3.60	0.709	3,910	4.45	0.634	2,602	6.69	0.603
109651	15,396	1.91	0.376	3,903	3.23	0.460	2,578	5.16	0.465
109660	15,171	0.76	0.149	3,862	1.29	0.184	2,603	1.65	0.149
109667	15,296	1.50	0.295	3,913	2.02	0.288	2,615	3.29	0.297
109673	15,196	3.22	0.635	3,808	3.91	0.557	2,551	6.23	0.562
109675	15,591	3.62	0.713	3,904	5.17	0.737	2,649	6.68	0.602
109683	15,718	2.70	0.532	3,834	3.99	0.568	2,632	7.26	0.654
109689	15,235	2.99	0.588	3,936	4.24	0.604	2,604	4.84	0.436
109690	15,331	2.46	0.484	3,952	4.10	0.584	2,744	7.14	0.644
109692	4,664	**5.08**	**1.000**	1,168	**7.02**	**1.000**	833	5.76	0.520
109693	2,526	2.45	0.483	696	2.87	0.409	437	5.49	0.495
109696	15,143	2.68	0.528	4,012	4.01	0.572	2,658	5.76	0.519
109697	15,120	1.80	0.354	3,910	2.48	0.353	2,612	4.17	0.376
109698	15,512	1.99	0.391	3,920	3.24	0.461	2,579	5.54	0.500
109699	15,294	3.18	0.625	3,832	5.79	0.825	2,610	7.28	0.656
109700	15,253	2.98	0.586	3,859	4.38	0.624	2,633	6.87	0.620
109701	15,117	1.72	0.338	3,796	3.48	0.495	2,562	5.07	0.457
109702	15,304	1.04	0.204	3,884	2.37	0.337	2,597	4.24	0.382
109703	15,507	4.72	0.929	3,854	5.86	0.835	2,718	7.36	0.663
109704	15,424	3.96	0.780	3,926	5.58	0.795	2,636	7.78	0.701
109705	11,703	2.98	0.587	2,988	3.78	0.539	2,014	4.92	0.443
109706	15,446	3.19	0.627	3,881	3.09	0.440	2,636	3.26	0.294

(continued on next page)

(continued)

	Majority (70.21%)			Intermediate (17.79%)			Minority (12.00%)		
Article	Impr.	CTR	v	Impr.	CTR	v	Impr.	CTR	v
109707	13,491	2.01	0.395	3,450	2.90	0.413	2,270	5.15	0.465
109708	12,589	3.74	0.736	3,151	6.82	0.972	2,227	**11.09**	**1.000**
109709	11,047	3.06	0.602	2,721	5.22	0.743	1,836	7.24	0.653
109710	3,498	3.23	0.636	859	4.77	0.680	567	6.17	0.557

Notes: Impr. denotes total impressions in Batch 9 of Day 7; CTR denotes clicks per impression (in %), which serves as the ground-truth click probability p_{ka} in the simulations; v denotes scaled utility v_{ka} from (1). Bold values indicate each segment's most preferred article ($v_{ka} = 1.000$). All 26 articles exceeded the 30-impression screening threshold. The full information allocations in Section 6.1 follow directly from this table: the blind benchmark selects article 109692, which maximizes the population-weighted CTR ($0.7021\,\mathrm{CTR}_{\mathrm{Maj}} + 0.1779\,\mathrm{CTR}_{\mathrm{Int}} + 0.1200\,\mathrm{CTR}_{\mathrm{Min}} = 5.51\%$, vs. 5.24% for the next highest article, 109703). Both the utilitarian and proportional policies select article 109708, which maximizes average scaled utility across segments ($(0.736+0.972+1.000)/3 = 0.903$, vs. 0.840 for 109692) as well as the product of utilities ($0.736 \times 0.972 \times 1.000 = 0.716$, with no mixture of articles achieving a higher product). The egalitarian policy assigns equal weight (0.5) to articles 109703 and 109708, equalizing majority and minority utilities at $0.5 \times 0.929 + 0.5 \times 0.736 = 0.5 \times 0.663 + 0.5 \times 1.000 = 0.832$.

Table W11: Empirical Click Through Rates and Scaled Utilities in ZOZO

	Majority (53.04%)			Intermediate (35.36%)			Minority (11.59%)		
Item	Impr.	CTR	v	Impr.	CTR	v	Impr.	CTR	v
0	9,404	0.13	0.113	6,236	0.42	0.612	2,017	0.50	0.401
1	9,208	0.13	0.115	6,251	0.27	0.399	2,157	0.46	0.375
2	9,300	0.05	0.047	6,093	0.21	0.313	1,906	0.21	0.170
3	8,804	0.18	0.160	6,018	0.27	0.390	1,907	0.58	0.466
4	9,318	0.04	0.038	6,131	0.24	0.359	1,979	0.05	0.041
5	8,797	0.23	0.201	5,895	0.29	0.423	1,902	0.21	0.170
6	9,365	0.53	0.471	6,143	0.49	0.717	2,047	0.00	0.000
7	9,349	0.64	0.566	6,141	0.47	0.693	2,026	0.15	0.120
8	8,968	0.51	0.453	6,069	0.41	0.604	2,025	0.20	0.160
9	9,318	0.33	0.294	6,288	0.21	0.303	2,045	0.05	0.040
10	9,084	0.02	0.019	5,929	0.24	0.346	1,937	0.52	0.417
11	8,688	0.14	0.122	6,042	0.50	0.729	2,000	0.75	0.606
12	9,377	0.04	0.038	6,227	0.31	0.448	2,028	0.39	0.319
13	9,108	0.10	0.087	5,916	0.30	0.446	1,904	0.05	0.042
14	9,114	0.05	0.048	6,180	0.36	0.522	1,945	0.15	0.125
15	9,352	0.04	0.038	6,104	0.39	0.577	2,026	0.44	0.359
16	9,337	0.11	0.095	6,104	0.29	0.433	2,054	1.12	0.905
17	8,932	0.10	0.089	5,806	0.19	0.278	1,964	0.20	0.165
18	9,585	0.32	0.285	6,422	0.37	0.548	2,177	0.87	0.706
19	9,045	0.08	0.068	6,080	0.30	0.434	1,988	0.75	0.610
20	9,430	0.06	0.056	6,133	0.21	0.311	2,001	0.30	0.242
21	9,110	0.04	0.039	6,078	0.33	0.483	2,007	1.00	0.806
22	8,942	0.10	0.089	6,023	0.38	0.560	2,021	**1.24**	**1.000**
23	8,882	0.11	0.099	5,987	0.25	0.368	1,967	0.20	0.164
24	9,146	0.04	0.039	6,104	0.28	0.409	2,035	1.08	0.874
25	9,120	0.04	0.039	6,071	0.23	0.338	1,944	0.77	0.624
26	9,158	0.16	0.145	6,101	0.36	0.529	1,946	0.15	0.125
27	8,992	0.11	0.098	6,056	0.20	0.291	1,991	0.15	0.122
28	8,742	0.02	0.020	5,758	0.42	0.612	1,895	0.53	0.427
29	9,269	0.04	0.038	6,039	0.46	0.680	1,916	0.42	0.338
30	9,354	0.36	0.321	6,173	0.42	0.618	2,011	0.15	0.121
31	9,442	0.40	0.355	6,305	0.41	0.605	2,034	0.15	0.119
32	9,099	0.26	0.233	6,077	0.49	0.724	1,982	0.20	0.163
33	9,482	0.31	0.270	6,208	0.27	0.402	2,043	0.00	0.000
34	9,028	0.62	0.547	6,168	0.65	0.952	1,989	0.35	0.285
35	9,192	0.88	0.778	6,111	0.67	0.984	2,032	0.34	0.278
36	8,675	0.47	0.417	5,965	0.35	0.517	1,911	0.05	0.042
37	9,443	0.43	0.383	6,214	0.35	0.520	2,131	0.89	0.721
38	9,024	0.58	0.509	6,082	0.61	0.893	1,872	0.11	0.086
39	9,253	0.91	0.801	6,114	0.62	0.912	2,029	0.05	0.040
40	9,706	0.15	0.136	6,340	0.19	0.278	1,970	0.10	0.082
41	9,203	0.62	0.547	5,967	0.57	0.836	2,037	0.15	0.119
42	9,234	0.55	0.487	6,040	0.60	0.875	2,013	0.00	0.000
43	9,175	0.56	0.491	6,163	**0.68**	**1.000**	2,059	0.15	0.118
44	8,866	0.32	0.279	6,036	0.38	0.559	2,006	0.50	0.403
45	9,193	0.52	0.461	6,161	0.50	0.738	1,985	0.10	0.081
46	9,159	0.50	0.443	6,120	0.34	0.504	2,017	0.05	0.040

(continued on next page)

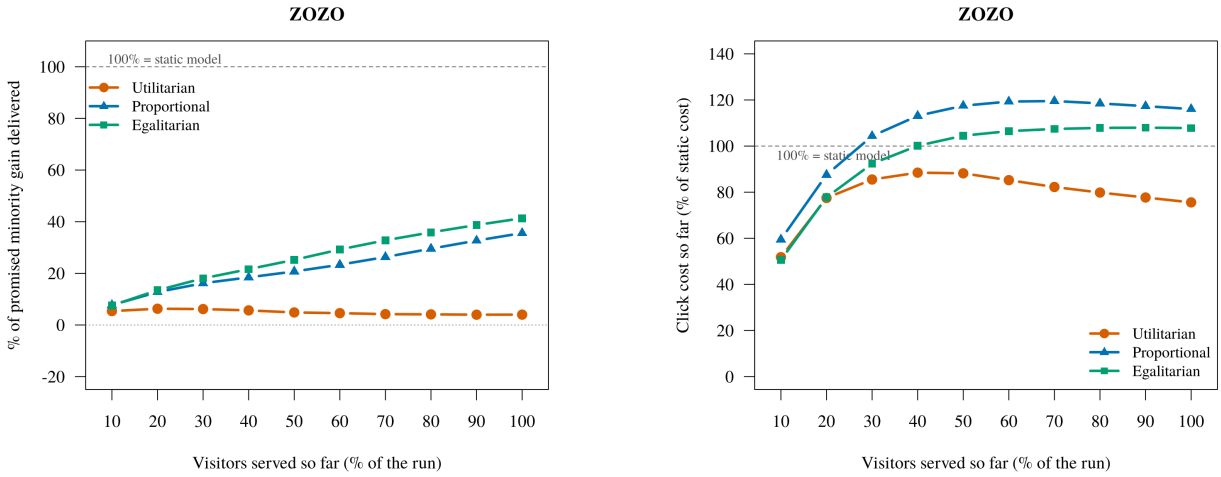
(continued)

	Majority (53.04%)			Intermediate (35.36%)			Minority (11.59%)		
Item	Impr.	CTR	v	Impr.	CTR	v	Impr.	CTR	v
47	9,150	0.71	0.627	6,125	0.38	0.551	1,884	0.42	0.343
48	8,802	0.60	0.531	5,893	0.63	0.921	1,881	0.74	0.602
49	9,047	0.64	0.566	6,166	0.55	0.809	1,960	0.00	0.000
50	8,776	0.52	0.463	5,887	0.56	0.823	1,976	0.30	0.245
51	9,103	0.45	0.397	6,021	0.60	0.877	1,978	0.35	0.286
52	9,062	0.44	0.390	5,990	0.37	0.539	2,073	0.19	0.156
53	8,810	0.31	0.270	6,019	0.40	0.585	1,904	0.21	0.170
54	8,667	0.10	0.092	5,720	0.19	0.282	1,943	0.00	0.000
55	8,904	0.33	0.287	5,669	0.26	0.388	1,979	0.61	0.490
56	9,318	0.18	0.161	6,240	0.14	0.212	2,142	0.00	0.000
57	9,431	0.59	0.524	6,105	0.43	0.625	1,962	0.10	0.082
58	9,014	0.80	0.705	5,888	0.63	0.922	2,011	0.20	0.161
59	8,775	0.52	0.463	5,898	0.49	0.721	1,920	0.10	0.084
60	8,648	**1.13**	**1.000**	5,765	0.38	0.560	1,912	0.42	0.338
61	9,161	0.56	0.491	6,130	0.60	0.886	2,046	0.10	0.079
62	9,056	0.67	0.594	6,063	0.38	0.557	1,995	0.05	0.041
63	9,446	0.74	0.654	6,217	0.42	0.614	2,090	0.19	0.155
64	9,039	0.29	0.254	6,018	0.32	0.463	1,919	0.31	0.253
65	9,237	0.42	0.373	6,175	0.36	0.523	1,977	0.10	0.082
66	9,181	0.37	0.327	6,343	0.22	0.324	2,077	0.05	0.039
67	9,098	0.02	0.019	5,994	0.30	0.441	1,951	0.92	0.746
68	9,180	0.10	0.087	6,117	0.34	0.504	2,083	0.53	0.427
69	9,163	0.25	0.222	6,133	0.42	0.622	2,098	0.62	0.501
70	9,267	0.02	0.019	6,317	0.25	0.372	1,980	0.30	0.245
71	9,252	0.14	0.124	6,039	0.36	0.535	1,965	0.20	0.165
72	9,250	0.16	0.143	6,174	0.44	0.642	2,022	0.45	0.360
73	8,937	0.09	0.079	5,869	0.20	0.300	1,906	1.00	0.806
74	9,057	0.09	0.078	6,083	0.12	0.169	2,007	0.35	0.282
75	8,681	0.12	0.102	5,963	0.30	0.443	1,950	0.36	0.290
76	8,760	0.26	0.232	5,973	0.33	0.491	1,925	0.31	0.252
77	9,138	0.31	0.270	6,404	0.59	0.871	2,055	0.24	0.197
78	8,701	0.52	0.456	5,827	0.46	0.680	1,964	0.36	0.288
79	9,102	0.54	0.475	6,099	0.43	0.626	1,936	0.15	0.125

Notes: Impr. denotes total impressions in the random-policy data of the ALL campaign (1,374,327 total arrivals across 80 items); CTR denotes clicks per impression (in %); v denotes scaled utility v_{ka} from (1). Bold values indicate each segment's most preferred item ($v_{ka} = 1.000$; items 60, 43, and 22 for the majority, intermediate, and minority segments, respectively). All 80 items exceeded the 30-impression screening threshold. Under full information (Section 6.2), the blind benchmark selects item 60, the utilitarian policy selects item 48, the proportional policy allocates 95% weight to item 48 and 5% to item 60, and the egalitarian policy allocates 36%, 14%, and 50% weight to items 22, 48, and 60, respectively.

Web Appendix E: Cost and Benefit of Fairness over Time in ZOZO

Section 6.3 demonstrates in the Yahoo application that the revenue cost of fairness is incurred immediately while minority utility gains materialize gradually as preferences are learned (Figure 5). This appendix replicates the dynamic analysis for the ZOZO application. Although the minority segment share in ZOZO (11.59%) is comparable to that in Yahoo (12.00%), click responses are roughly an order of magnitude sparser (approximately 3.5 clicks per 1,000 visits, compared with 27 to 58 clicks per 1,000 visits in Yahoo). Following the procedure in Section 6.3, we evaluate cumulative performance at decile stopping points ($10\%, 20\%, \ldots, 100\%$ of total arrivals). Figure W1a plots the realized minority utility improvement over blind TS as a percentage of the full-information static improvement predicted by Bertsimas et al. (2011), where 100% indicates full delivery of the static fairness gain and 0% indicates no improvement over blind TS. Figure W1b plots the cumulative click loss relative to blind TS as a percentage of the static PoF (20.0% for utilitarian, 19.0% for proportional, and 23.3% for egalitarian).



(a) Minority Benefit: Share of Utility Gain

(b) Firm Cost: Clicks Lost to Blind

Figure W1: Dynamics of Minority Benefits and Firm Cost in ZOZO

In the sparser ZOZO environment, the divergence between realized costs and realized benefits is substantially wider than in Yahoo. After the first 10% of arrivals (approximately 137,000 visits), the ethical TS policies have already incurred 51% to 60% of their static click costs while delivering only 5% to 8% of the static minority utility improvement. Beyond the first 30% of arrivals, the percentage click losses of proportional and egalitarian TS relative to blind TS exceed their static PoFs (peaking at 119% and 108% of static PoF, respectively), while the policies deliver only 36% and 41% of the static minority utility gain by the end of the 1.37-million-visit horizon. Neither trajectory plateaus because, under sparse click feedback, preference learning remains incomplete even after 1.37 million visits: over the final 10% of arrivals, proportional and egalitarian TS deliver 62% and 65% of their static fairness gains.

Utilitarian TS exhibits a distinct pattern due to near-ties in aggregate segment utilities. Under full information, the utilitarian target (item 48, with mean segment utility 0.685) is nearly tied with item 35 (mean segment utility 0.680). Under sparse click feedback, utilitarian TS frequently concentrates on item 35, which generates a high overall CTR among the majority and intermediate segments but provides low utility to the minority segment (0.278 vs. 0.602 under item 48; see Table W11). Consequently, its realized minority benefit remains between 4% and 6% of the static gain while its click loss declines below the static PoF.

Web Appendix F: Cost of Fairness over Time in Yahoo

Table W12 reports the numerical estimates underlying Figure 5b in Section 6.3. At each decile of the 500,000-visit horizon, the table presents the cumulative percentage of expected clicks lost by each ethical TS policy relative to blind TS, alongside the ratio of this cumulative click loss to the corresponding full-information static PoF (6.1% for utilitarian and proportional, and 5.5% for egalitarian).

Table W12: Cumulative Click Loss Relative to Blind TS over Time in Yahoo

	Clicks Lost to Blind (%)			Proportion of Static PoF (%)		
Visitors Served	Utilitarian	Proportional	Egalitarian	Utilitarian	Proportional	Egalitarian
First 10%	4.4	5.4	7.3	71	89	132
First 20%	4.3	4.9	6.6	71	81	120
First 30%	4.4	4.8	6.0	72	79	109
First 40%	4.5	4.9	5.7	74	80	104
First 50%	4.7	5.0	5.6	77	82	101
First 60%	4.8	5.1	5.5	79	84	100
First 70%	5.0	5.2	5.4	81	85	98
First 80%	5.1	5.3	5.3	83	86	97
First 90%	5.1	5.3	5.3	84	87	96
First 100%	5.2	5.4	5.3	85	88	96
Static Benchmark	6.1	6.1	5.5	100	100	100

Notes: Clicks lost are computed as the percentage reduction in cumulative expected clicks relative to blind TS over the first X% of arrivals, averaged across the 30 runs in Table 1. Static Benchmark reports the corresponding full-information PoF when click rates are known a priori. The First 100% row matches the full horizon dynamic PoF values reported in Section 6.1.

Web Appendix G: Sensitivity to Size of Minority Segment

To examine how segment composition affects sequential learning and allocation outcomes (Section 6.4), we vary the arrival probability of the minority segment π_{Min} from 1% up to an equal one-third share while holding the empirical segment-item CTR matrix P fixed at its observed values. As the minority share varies, the arrival probabilities of the majority and intermediate segments are scaled proportionally to preserve their observed relative ratio (except in the final row of each panel, where all three segments are assigned an equal 1/3 share). Table W13 reports the resulting minority utility under blind TS and the three ethical TS policies, averaged over the full simulation horizon.

Table W13: Minority Segment Utility across Varying Minority Arrival Shares

Minority Share	Blind	Utilitarian	Proportional	Egalitarian
Panel A: Yahoo Data (30 runs)				
1%	0.531	0.610	0.760	0.700
	[0.529, 0.532]	[0.590, 0.629]	[0.732, 0.789]	[0.682, 0.718]
2.5%	0.532	0.694	0.863	0.787
	[0.529, 0.534]	[0.660, 0.729]	[0.850, 0.876]	[0.777, 0.798]
10%	0.546	0.906	0.928	0.817
	[0.541, 0.551]	[0.892, 0.920]	[0.920, 0.936]	[0.812, 0.822]
12.0% (observed)	0.556	0.924	0.933	0.819
	[0.548, 0.564]	[0.917, 0.931]	[0.925, 0.941]	[0.812, 0.825]
20%	0.874	0.939	0.938	0.810
	[0.847, 0.901]	[0.932, 0.947]	[0.932, 0.943]	[0.805, 0.815]
33%	0.968	0.945	0.939	0.806
	[0.966, 0.970]	[0.941, 0.950]	[0.933, 0.945]	[0.802, 0.809]
Equal (1/3 each)	0.975	0.956	0.946	0.814
	[0.974, 0.977]	[0.950, 0.961]	[0.941, 0.951]	[0.810, 0.818]
Panel B: ZOZO Data (10 runs)				
1%	0.263	0.232	0.256	0.268
	[0.236, 0.291]	[0.220, 0.244]	[0.249, 0.263]	[0.265, 0.272]
2.5%	0.272	0.256	0.279	0.302
	[0.249, 0.295]	[0.247, 0.264]	[0.267, 0.290]	[0.294, 0.310]
4.6%	0.285	0.258	0.320	0.324
	[0.275, 0.295]	[0.247, 0.270]	[0.308, 0.332]	[0.312, 0.336]
10%	0.278	0.308	0.365	0.385
	[0.264, 0.293]	[0.285, 0.330]	[0.353, 0.378]	[0.371, 0.398]
11.6% (observed)	0.295	0.301	0.387	0.411
	[0.290, 0.300]	[0.286, 0.316]	[0.373, 0.400]	[0.397, 0.425]
20%	0.307	0.335	0.420	0.443
	[0.299, 0.315]	[0.315, 0.354]	[0.411, 0.430]	[0.427, 0.458]
33%	0.357	0.348	0.428	0.471
	[0.339, 0.374]	[0.324, 0.372]	[0.413, 0.444]	[0.462, 0.481]
Equal (1/3 each)	0.421	0.341	0.443	0.471
	[0.394, 0.448]	[0.320, 0.361]	[0.426, 0.460]	[0.459, 0.483]

Notes: Estimates represent scaled minority utility from (1) averaged over the full simulation horizon, with 95% confidence intervals across runs reported in brackets. Panel A averages over 30 runs of 500,000 arrivals per configuration (the 12.0% observed-share row matches Table 1). Panel B averages over 10 runs of 1,374,327 arrivals per configuration (matching the 10-run sample in Table 3, Panel B). In the 33% row, the minority share is 33.0% and the remaining 67.0% is split proportionally between the majority and intermediate segments; in the Equal (1/3 each) row, all three segments arrive with probability 1/3.

Web Appendix H: Decomposition of Price of Fairness for Data Application

This appendix applies Proposition 2 to the two data applications. All click quantities are expected clicks over the full simulation horizon: 500,000 arrivals for Yahoo and 1,374,327 for ZOZO. They are evaluated using the empirical CTR matrices held fixed in the simulations. Full information allocations are computed using the procedures in Web Appendix B and are fixed across runs within each configuration.

The static fairness cost is $T[C_B - C(\psi_j^*)]$. The learning costs, $L_{B,T}$ and $L_{j,T}$, and the realized click loss, $C_{B,T} - C_{j,T}$, follow (21) and Proposition 2. Denoting averages across runs by bars, the reported quantities satisfy

$$\overline{C}_{B,T} - \overline{C}_{j,T} = T[C_B - C(\psi_j^*)] - \overline{L}_{B,T} + \overline{L}_{j,T}.$$

Thus, the learning contribution to the realized click loss is $\overline{L}_{j,T} - \overline{L}_{B,T}$. An ethical learning cost can be negative because its full information benchmark maximizes the ethical objective rather than clicks.

The analysis reuses the simulation runs underlying Tables 1, 2, and W13. Table W14 uses 30 runs per policy in each application. The minority share results use 30 runs per configuration for Yahoo and 10 for ZOZO. The observed share ZOZO row uses the first 10 runs of Table 2, so its estimates differ slightly from those based on all 30 runs. Brackets report 95% confidence intervals for the reported estimates, conditional on the fixed empirical CTR matrices and arrival shares. Learning-cost intervals use mean $\pm\, t_{0.975,M-1}\mathrm{SD}/\sqrt{M}$, where $M = 30$ for Yahoo and $M = 10$ for ZOZO. Realized loss intervals are for differences between independent policy means, reflecting the separate simulation seeds. Dynamic PoF intervals use 10,000 bootstrap resamples of each policy's runs. Static baselines are fixed within each configuration and have no simulation interval.

Table W14: Decomposition of Realized Click Loss Relative to Blind TS

MAB Model	Static Fairness Cost	Blind TS Learning Cost	Ethical TS Learning Cost	Realized Loss	Static PoF (%)	Dynamic PoF (%)
Panel A: Yahoo Data						
Utilitarian	1,683	756	469	1,395	6.1	5.2
		[696, 817]	[440, 497]	[1,329, 1,461]		[5.0, 5.4]
Proportional	1,683	756	518	1,445	6.1	5.4
		[696, 817]	[488, 549]	[1,378, 1,511]		[5.2, 5.6]
Egalitarian	1,511	756	652	1,407	5.5	5.3
		[696, 817]	[612, 693]	[1,336, 1,479]		[5.0, 5.5]
Panel B: ZOZO Data						
Utilitarian	2,155	1,417	676	1,415	20.0	15.1
		[1,370, 1,463]	[594, 759]	[1,322, 1,508]		[14.2, 16.0]
Proportional	2,054	1,417	1,433	2,070	19.0	22.1
		[1,370, 1,463]	[1,388, 1,479]	[2,006, 2,134]		[21.5, 22.7]
Egalitarian	2,514	1,417	1,257	2,354	23.3	25.1
		[1,370, 1,463]	[1,219, 1,295]	[2,296, 2,413]		[24.6, 25.7]

Notes: Static fairness cost and static PoF do not depend on the runs and have no interval.

In Yahoo, ethical learning costs range from 469 to 652 clicks, below the 756 clicks lost by blind TS while learning (Table W14). Consequently, the realized losses of 1,395–1,445 clicks are below the static fairness costs of 1,511–1,683 clicks. In ZOZO, utilitarian and egalitarian

learning costs are also below that of blind TS. The proportional learning cost is slightly higher, i.e., 1,433 versus 1,417 clicks. Its realized loss of 2,070 clicks is close to its static fairness cost of 2,054, which lies within the reported confidence interval for the realized loss.

ZOZO also illustrates why absolute and percentage comparisons differ. Egalitarian TS loses 2,354 clicks relative to blind TS, fewer than its static fairness cost of 2,514, yet its dynamic PoF is 25.1%, above the static value of 23.3%. Proportional TS also has a higher dynamic PoF, i.e., 22.1% versus 19.0%. Item 2 of Proposition 2 explains this pattern. Dynamic PoF exceeds static PoF when $L_{j,T} > (1 - \text{PoF}_j)L_{B,T}$, because learning by blind TS reduces the denominator. For egalitarian TS, the reported learning cost of 1,257 clicks exceeds the corresponding threshold of approximately 1,087, even though it is below the blind learning cost of 1,417. For proportional TS, 1,433 exceeds the threshold of approximately 1,148. In Yahoo, the condition fails for all three rules and their dynamic PoFs are below their static counterparts.

Tables W15 and W16 apply the decomposition to the minority shares in Table W13. Large positive learning contributions occur in some rare-minority configurations. At a 1% minority share in Yahoo, egalitarian TS loses 5,925 clicks relative to blind TS, compared with a static fairness cost of 3,868. Its learning cost of 2,699 clicks substantially exceeds the 643 clicks lost by blind TS. At a 2.5% share, its realized loss is 4,567 clicks against a static fairness cost of 3,547. In ZOZO, proportional TS also loses more than its static fairness cost at every minority shares except the observed 11.6% and 20%.

Negative ethical learning costs require a different interpretation. For Yahoo utilitarian TS at a 1% minority share, $\overline{L}_{\mathfrak{U},T} = -2{,}437$, i.e., the learning policy earns more clicks than its full information ethical target. This does not establish better attainment of the ethical objective. Indeed, minority utility is 0.610 in Table W13, compared with the full information utilitarian target of 1.000.

Table W16: Decomposition by Minority Share in ZOZO

Minority Share	Ethical TS	Static Fairness Cost	Blind TS Learning Cost	Ethical TS Learning Cost	Realized Loss	Static PoF (%)	Dynamic PoF (%)
1%	U	2,950	1,406	631	2,175	25.9	21.8
			[1,307, 1,504]	[529, 732]	[2,044, 2,306]		[20.7, 22.8]
	P	2,811	1,406	1,790	3,196	24.7	32.0
			[1,307, 1,504]	[1,719, 1,861]	[3,082, 3,309]		[31.2, 32.8]
	E	3,370	1,406	1,311	3,276	29.6	32.8
			[1,307, 1,504]	[1,247, 1,375]	[3,166, 3,386]		[32.1, 33.6]
2.5%	U	2,838	1,380	761	2,218	25.1	22.4
			[1,259, 1,501]	[530, 992]	[1,971, 2,466]		[20.3, 24.4]
	P	2,704	1,380	1,684	3,007	23.9	30.3
			[1,259, 1,501]	[1,610, 1,758]	[2,874, 3,141]		[29.3, 31.2]
	E	3,249	1,380	1,333	3,202	28.8	32.3
			[1,259, 1,501]	[1,256, 1,409]	[3,067, 3,336]		[31.3, 33.2]
4.6%	U	2,680	1,387	524	1,817	24.0	18.6
			[1,310, 1,464]	[374, 674]	[1,656, 1,978]		[17.2, 19.9]
	P	2,554	1,387	1,636	2,802	22.8	28.6
			[1,310, 1,464]	[1,518, 1,754]	[2,670, 2,935]		[27.5, 29.7]
	E	3,079	1,387	1,301	2,993	27.5	30.6
			[1,310, 1,464]	[1,194, 1,407]	[2,870, 3,116]		[29.6, 31.6]

Continued on next page

Table W16 – Continued from previous page

Minority Share	Ethical TS	Static Fairness Cost	Blind TS Learning Cost	Ethical TS Learning Cost	Realized Loss	Static PoF (%)	Dynamic PoF (%)
10%	U	2,275	1,509	788	1,554	20.9	16.6
			[1,380, 1,639]	[656, 921]	[1,382, 1,726]		[15.1, 18.1]
	P	2,168	1,509	1,548	2,207	19.9	23.6
			[1,380, 1,639]	[1,464, 1,633]	[2,061, 2,352]		[22.4, 24.7]
	E	2,643	1,509	1,322	2,456	24.3	26.2
			[1,380, 1,639]	[1,191, 1,453]	[2,285, 2,627]		[24.9, 27.6]
11.6% (obs.)	U	2,155	1,445	697	1,407	20.0	15.1
			[1,378, 1,513]	[513, 881]	[1,217, 1,597]		[13.5, 16.9]
	P	2,054	1,445	1,425	2,033	19.0	21.8
			[1,378, 1,513]	[1,331, 1,519]	[1,925, 2,141]		[20.8, 22.7]
	E	2,514	1,445	1,265	2,334	23.3	25.0
			[1,378, 1,513]	[1,202, 1,328]	[2,249, 2,420]		[24.2, 25.7]
20%	U	1,525	1,423	998	1,100	14.8	12.4
			[1,325, 1,520]	[833, 1,163]	[919, 1,281]		[10.7, 14.1]
	P	1,453	1,423	1,420	1,450	14.1	16.3
			[1,325, 1,520]	[1,345, 1,494]	[1,335, 1,564]		[15.3, 17.3]
	E	1,835	1,423	1,296	1,709	17.8	19.2
			[1,325, 1,520]	[1,202, 1,390]	[1,582, 1,835]		[18.1, 20.3]
33%	U	549	1,509	1,512	552	5.7	6.9
			[1,405, 1,614]	[1,418, 1,606]	[422, 683]		[5.5, 8.2]
	P	523	1,509	1,630	644	5.5	8.0
			[1,405, 1,614]	[1,510, 1,750]	[496, 792]		[6.5, 9.6]
	E	785	1,509	1,348	623	8.2	7.7
			[1,405, 1,614]	[1,245, 1,450]	[488, 759]		[6.3, 9.2]
Equal	U	0	1,489	1,760	272	0.0	3.6
			[1,370, 1,607]	[1,682, 1,838]	[139, 405]		[2.1, 5.1]
	P	9	1,489	1,594	114	0.1	1.5
			[1,370, 1,607]	[1,490, 1,697]	[−32, 260]		[−0.2, 3.2]
	E	510	1,489	1,194	215	5.6	2.9
			[1,370, 1,607]	[1,040, 1,347]	[35, 396]		[0.9, 5.0]

Positive learning contributions also occur at larger minority shares. In Yahoo, utilitarian and proportional allocations have zero static fairness cost at shares of 20% and above, meaning that their full information click yields equal the blind optimum. Their realized click differences arise entirely from differences in learning costs as they outperform blind TS by 123 and 104 clicks at 20%, but lose 370 and 460 clicks at 33%, and 370 and 560 clicks under equal shares. Learning can either offset or add to the static fairness cost, depending on the policy and segment composition.

These full horizon results complement Figure 5 and Web Appendices E and F, which trace how click costs and minority benefits develop over time. In particular, the Yahoo results show that delayed minority benefits can coexist with realized click losses below the static fairness cost at the observed shares.

Table W15: Decomposition by Minority Share in Yahoo

Minority Share	Ethical TS	Static Fairness Cost	Blind TS Learning Cost	Ethical TS Learning Cost	Realized Loss	Static PoF (%)	Dynamic PoF (%)
1%	U	5,224	643 [615, 670]	−2,437 [−2,690, −2,184]	2,144 [1,890, 2,398]	19.1	8.0 [7.2, 9.0]
	P	5,224	643 [615, 670]	425 [256, 594]	5,006 [4,835, 5,177]	19.1	18.7 [18.1, 19.3]
	E	3,868	643 [615, 670]	2,699 [2,564, 2,834]	5,925 [5,788, 6,062]	14.1	22.2 [21.7, 22.6]
2.5%	U	4,741	654 [608, 700]	−1,669 [−1,996, −1,341]	2,418 [2,088, 2,749]	17.3	9.0 [7.9, 10.2]
	P	4,741	654 [608, 700]	352 [259, 446]	4,439 [4,337, 4,542]	17.3	16.6 [16.2, 16.9]
	E	3,547	654 [608, 700]	1,674 [1,577, 1,771]	4,567 [4,461, 4,673]	12.9	17.1 [16.7, 17.4]
10%	U	2,327	736 [681, 790]	336 [272, 401]	1,927 [1,845, 2,010]	8.5	7.2 [6.9, 7.5]
	P	2,327	736 [681, 790]	434 [396, 471]	2,025 [1,960, 2,090]	8.5	7.6 [7.3, 7.8]
	E	1,940	736 [681, 790]	698 [660, 736]	1,902 [1,837, 1,967]	7.1	7.1 [6.9, 7.3]
12.0% (obs.)	U	1,683	756 [696, 817]	469 [440, 497]	1,395 [1,329, 1,461]	6.1	5.2 [5.0, 5.4]
	P	1,683	756 [696, 817]	518 [488, 549]	1,445 [1,378, 1,511]	6.1	5.4 [5.2, 5.6]
	E	1,511	756 [696, 817]	652 [612, 693]	1,407 [1,336, 1,479]	5.5	5.3 [5.0, 5.5]
20%	U	0	841 [780, 901]	718 [681, 755]	−123 [−192, −53]	0.0	−0.4 [−0.7, −0.2]
	P	0	841 [780, 901]	736 [709, 763]	−104 [−169, −39]	0.0	−0.4 [−0.6, −0.2]
	E	689	841 [780, 901]	668 [623, 713]	517 [443, 590]	2.4	1.9 [1.6, 2.1]
33%	U	0	704 [673, 734]	1,073 [1,014, 1,133]	370 [304, 435]	0.0	1.1 [1.0, 1.3]
	P	0	704 [673, 734]	1,164 [1,090, 1,238]	460 [381, 540]	0.0	1.4 [1.2, 1.7]
	E	2,089	704 [673, 734]	860 [820, 899]	2,244 [2,195, 2,294]	6.3	7.0 [6.8, 7.1]
Equal	U	0	667 [639, 695]	1,037 [962, 1,112]	370 [291, 450]	0.0	1.0 [0.8, 1.3]
	P	0	667 [639, 695]	1,227 [1,157, 1,298]	560 [485, 635]	0.0	1.6 [1.4, 1.8]
	E	3,084	667 [639, 695]	493 [424, 562]	2,909 [2,836, 2,983]	8.5	8.2 [8.0, 8.4]

Web Appendix I: Bandit Performance across Ethical Objectives in Yahoo

Whereas Figure 4 in Section 6.3 evaluates each ethical TS policy exclusively under its own welfare criterion, this appendix evaluates all four demographic-blind policies, including utilitarian TS, proportional TS, egalitarian TS, and blind TS, across all three ethical welfare criteria $j \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}\}$ defined in (11)–(13). For bandit policy $i \in \{\mathfrak{U}, \mathfrak{P}, \mathfrak{E}, B\}$ evaluated under ethical objective j, the per-period cross-objective regret at arrival t is

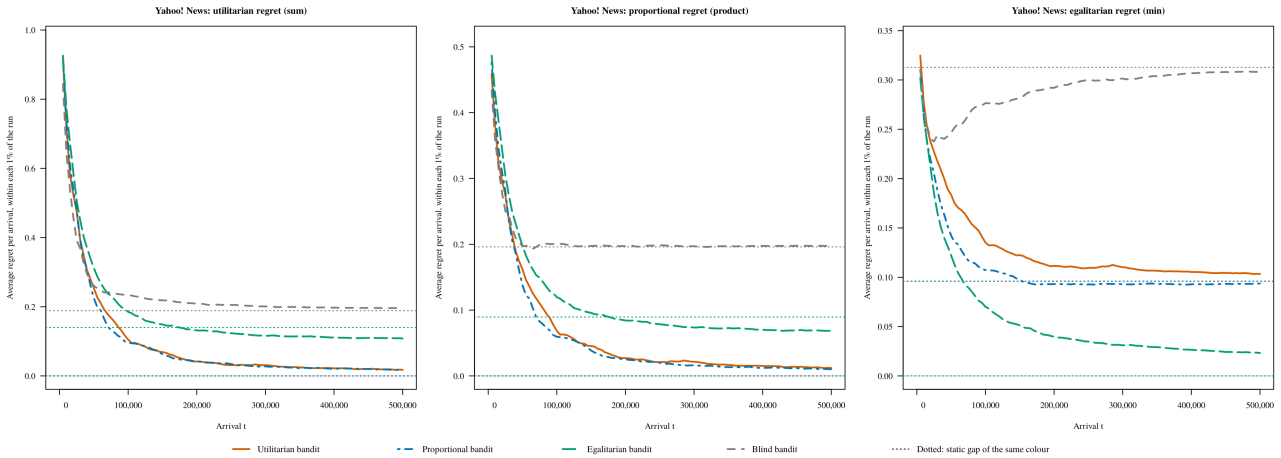
$$r_{ij,t} = W_j(\psi_j^*; V) - W_j(\widetilde{\psi}_{i,t}; V), \tag{64}$$

where ψ_j^* is the full information allocation maximizing criterion j, and $\widetilde{\psi}_{i,t}$ is the allocation distribution selected by bandit i at arrival t (for blind TS, the one-hot vector corresponding to the displayed article a_t). When $i = j$, $r_{jj,t}$ equals the per-period ethical regret in Definition 1. Here, $W_{\mathfrak{U}}$, $W_{\mathfrak{P}}$, and $W_{\mathfrak{E}}$ denote the sum, product, and minimum of the three segment utilities, respectively. All estimates are evaluated across the 30 runs of Table 1 using identical random seeds.

Table W17: Static Welfare Gaps and Terminal Cross-Objective Regret in Yahoo

	Utilitarian		Proportional		Egalitarian	
MAB Model	Static Gap	Final 10%	Static Gap	Final 10%	Static Gap	Final 10%
Utilitarian	0.000	0.019	0.000	0.013	0.096	0.104
Proportional	0.000	0.018	0.000	0.011	0.096	0.093
Egalitarian	0.140	0.109	0.090	0.069	0.000	0.024
Blind	0.189	0.196	0.196	0.198	0.313	0.308

Notes: Under full information, the utilitarian and proportional policies select article 109708, the egalitarian policy splits weight equally between articles 109703 and 109708, and the blind policy selects article 109692 (see Table W10 notes in Web Appendix D). Static Gap reports $W_j(\psi_j^*; V) - W_j(\psi_i^*; V)$. Final 10% reports average per-period regret $r_{ij,t}$ over the last 50,000 arrivals across the 30 runs in Table 1. Diagonal entries correspond to each ethical bandit's own ethical regret.



Notes: Curves plot per-period regret $r_{ij,t}$ averaged within each 1% bin of the run across the 30 runs in Table 1. Horizontal dotted lines indicate the full-information static welfare gap for the policy of the corresponding color.

Figure W2: Cross-Objective Ethical Regret Trajectories in Yahoo

For off-diagonal evaluations ($i \neq j$), the asymptotic benchmark is the full information *static welfare gap*, $W_j(\psi_j^*; V) - W_j(\psi_i^*; V)$, which measures how far policy i's full-information target

allocation ψ_i^* falls short of the maximum attainable welfare under criterion j. If bandit i converges to its own target allocation, its regret under criterion j approaches this static welfare gap rather than zero.

Two primary patterns emerge from Table W17 and Figure W2. First, when evaluated on its own objective ($i = j$), each ethical TS policy exhibits vanishing per-period regret, consistent with Corollary 1. Second, when evaluated on an alternative ethical criterion ($i \neq j$), regret converges toward the corresponding static welfare gap rather than zero. Because the full information utilitarian and proportional policies select the same article in the Yahoo data, their trajectories coincide across all three panels, settling approximately 0.10 below the egalitarian optimum on the minimum utility criterion. Blind TS similarly converges to its static gap across all three criteria. Egalitarian TS converges more gradually, i.e., retaining an average own-objective regret of 0.024 over the final 10% of arrivals, while remaining below its static gap on the utilitarian and proportional objectives. Thus, the residual cross-objective gap reflects the structural tradeoff among ethical principles rather than estimation error: a bandit that converges on its own welfare objective need not perform well under alternative normative criteria.

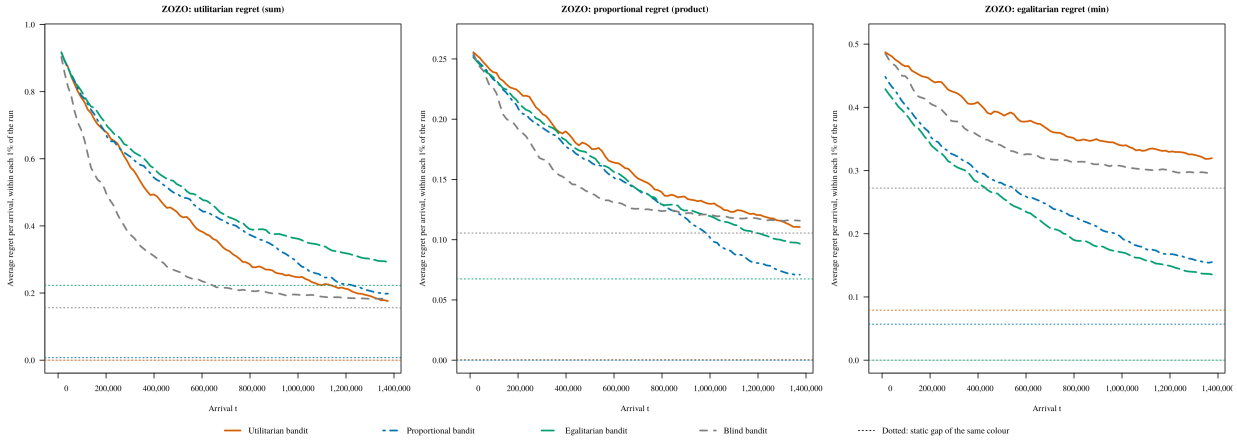
Web Appendix J: Bandit Performance across Ethical Objectives in ZOZO

This appendix replicates the cross-objective regret analysis from Web Appendix I for the ZOZO application. Table W18 reports the full information static welfare gaps and average cross-objective regrets over the final 10% of the entire horizon of the 1,374,327 visits, and Figure W3 plots the corresponding learning trajectories across each 1% bin of arrivals.

Table W18: Static Welfare Gaps and Terminal Cross-Objective Regret in ZOZO

	Utilitarian		Proportional		Egalitarian	
MAB Model	Static Gap	Final 10%	Static Gap	Final 10%	Static Gap	Final 10%
Utilitarian	0.000	0.188	0.000	0.114	0.079	0.323
Proportional	0.007	0.206	0.000	0.073	0.057	0.158
Egalitarian	0.223	0.300	0.067	0.099	0.000	0.139
Blind	0.156	0.183	0.106	0.116	0.272	0.297

Notes: Under full information, the utilitarian policy selects item 48, the proportional policy mixes items 48 (95%) and 60 (5%), the egalitarian policy mixes items 22 (36%), 48 (14%), and 60 (50%), and the blind policy selects item 60 (see Table W11 in Web Appendix D). Static Gap reports $W_j(\psi_j^*; V) - W_j(\psi_i^*; V)$. Final 10% reports average per-period regret $r_{ij,t}$ over the final 10% of arrivals across the 30 runs in Table 2. Diagonal entries correspond to each ethical bandit's own ethical regret.



Notes: Curves plot per-period regret $r_{ij,t}$ averaged within each 1% bin of the run across the 30 runs in Table 2. Horizontal dotted lines indicate the full information static welfare gap for the policy of the corresponding color.

Figure W3: Cross-Objective Ethical Regret Trajectories in ZOZO

In ZOZO, where clicks are approximately ten times rarer than in Yahoo, none of the ethical TS policies has fully converged by the end of the horizon (Table W18 and Figure W3). Each ethical bandit's own-objective regret ($i = j$) continues to decline at the end of the run but remains well above zero (0.188, 0.073, and 0.139 over the final 10% of arrivals for utilitarian, proportional, and egalitarian TS, respectively), while off-diagonal regrets remain above their static welfare gaps (consistent with the gradual learning documented in Web Appendix E). Over the full horizon, blind TS achieves lower cumulative utilitarian regret than utilitarian TS and performs comparably to proportional TS on the proportional criterion. Within the run, proportional TS overtakes blind TS on its own criterion only after approximately 62% of arrivals, and utilitarian TS overtakes blind TS on its own criterion only during the final 3% of arrivals. By contrast, egalitarian TS outperforms blind TS on the egalitarian criterion throughout the entire horizon. Thus, when feedback is sparse, an ethical bandit can spend much of a finite learning horizon performing worse on its own welfare objective than a policy

that ignores fairness altogether.

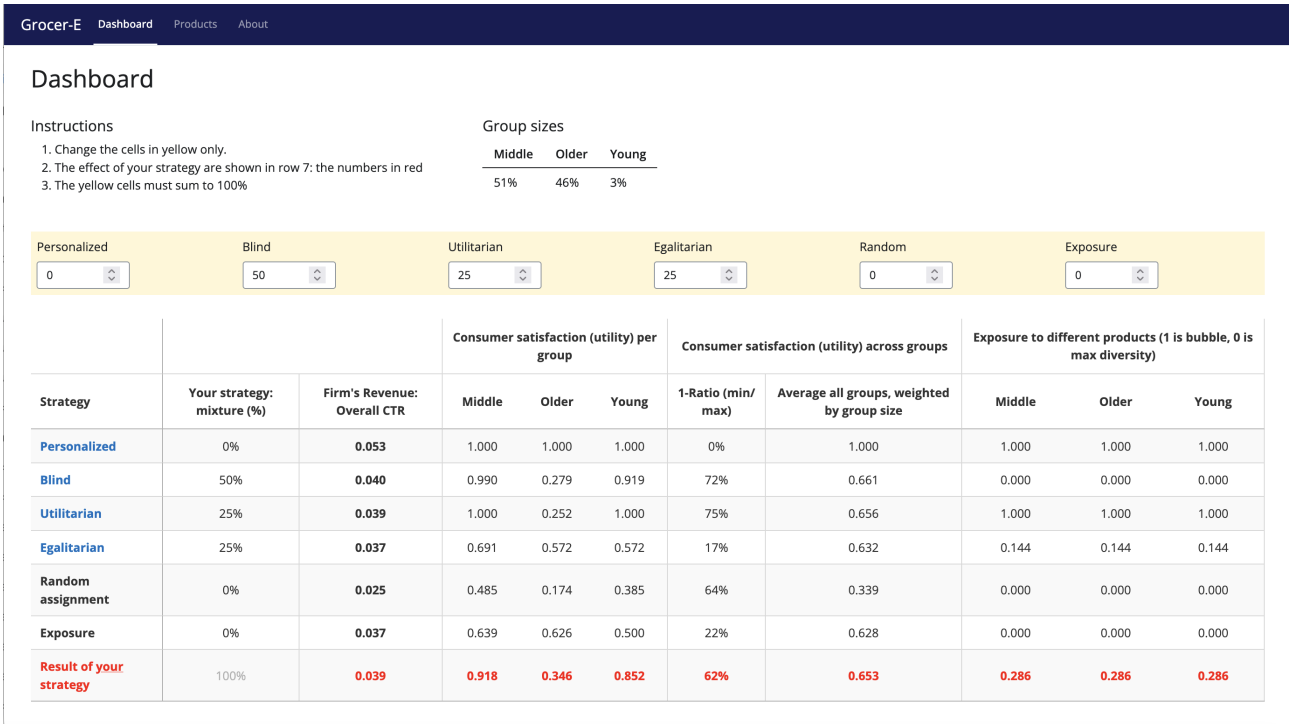
Web Appendix K: Teaching Case for Ethical Bandits

The teaching case *Private Data: Friend or Foe?* provides an educational application of the ethical resource allocation framework developed in this paper. The case features a fictional online grocer, Grocer-E, that must select a common landing page for customers belonging to different demographic segments without using sensitive personal information to personalize individual allocation decisions. Students assume the role of a marketing manager tasked with recommending an allocation strategy to the company's board. In doing so, they must consider the tradeoffs among firm revenue, customer welfare, and data privacy.

The case is accompanied by an interactive dashboard that allows students to explore and compare alternative allocation strategies. In particular, students can examine the implications of the utilitarian, proportional, and egalitarian objectives considered in this paper, as well as mixed strategies that combine different allocation approaches. By adjusting the allocation weights, students can observe how prioritizing different customer segments affects overall business performance and the distribution of benefits across groups. Figure W4 illustrates the Grocer-E dashboard.

The teaching case complements the theoretical and empirical analyses by making the efficiency–equity tradeoffs of demographic-blind allocation accessible through an interactive managerial decision exercise. It illustrates how a common allocation policy can produce different outcomes across heterogeneous consumer segments and how alternative ethical objectives can reshape these outcomes. The dashboard focuses on comparing static allocation strategies rather than replicating the sequential learning process of the ethical Thompson Sampling algorithms developed in the main paper.

The teaching case and accompanying dashboard are available at `https://placeholderURL/r-apps/grocer-e/`.



			Consumer satisfaction (utility) per group			Consumer satisfaction (utility) across groups		Exposure to different products (1 is bubble, 0 is max diversity)		
Strategy	Your strategy: mixture (%)	Firm's Revenue: Overall CTR	Middle	Older	Young	1-Ratio (min/max)	Average all groups, weighted by group size	Middle	Older	Young
Personalized	0%	**0.053**	1.000	1.000	1.000	0%	1.000	1.000	1.000	1.000
Blind	50%	**0.040**	0.990	0.279	0.919	72%	0.661	0.000	0.000	0.000
Utilitarian	25%	**0.039**	1.000	0.252	1.000	75%	0.656	1.000	1.000	1.000
Egalitarian	25%	**0.037**	0.691	0.572	0.572	17%	0.632	0.144	0.144	0.144
Random assignment	0%	**0.025**	0.485	0.174	0.385	64%	0.339	0.000	0.000	0.000
Exposure	0%	**0.037**	0.639	0.626	0.500	22%	0.628	0.000	0.000	0.000
Result of your strategy	100%	**0.039**	**0.918**	**0.346**	**0.852**	**62%**	**0.653**	**0.286**	**0.286**	**0.286**

In the dashboard interface of the teaching case, consumers are segmented into Middle, Older, and Young age groups.

Figure W4: Grocer-E Teaching Dashboard

References

Agrawal, Shipra and Navin Goyal (2012), Analysis of thompson sampling for the multi-armed bandit problem, *in* 'Proceedings of the 25th Annual Conference on Learning Theory', Vol. 23, pp. 39.1–39.26.

Berge, Claude (1963), *Topological Spaces*, Macmillan, New York.

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

Cover, Thomas M. (1984), 'An algorithm for maximizing expected log investment return', *IEEE Transactions on Information Theory* **30**(2), 369–373.

Dantzig, George B. (1951), A proof of the equivalence of the programming problem and the game problem, *in* T. C.Koopmans, ed., 'Activity Analysis of Production and Allocation', Wiley, New York, pp. 330–335.

Dantzig, George B. (1963), *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ.

Huangfu, Qi and J. A. Julian Hall (2018), 'Parallelizing the dual revised simplex method', *Mathematical Programming Computation* **10**(1), 119–142.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant et al. (2020), 'SciPy 1.0: Fundamental algorithms for scientific computing in Python', *Nature Methods* **17**(3), 261–272.