# Mitigating Targeting Bias in Content Recommendation with Causal Bandits*

YAN ZHAO, MITCHELL GOODMAN, SAMEER KANASE, SHENGHE XU, YANNICK KIMMEL, BRENT PAYNE, SAAD KHAN, and PATRICIA GRAO, Amazon.com,Inc, USA

Recommendations systems play a central role in improving customer experience on the Amazon retail website. Commonly, Learning-to-Rank (LTR) methods are employed to rank content, however these methods are subject to bias inherent in the observational data that they use for training. This paper studies a domain-specific self-selection bias, called Content Targeting Bias, introduced when content is generated for specific targeted customers. When content specifically targets classes of customers who are more or less likely to take actions associated with traditional recommendations algorithms (clicks, purchases), the resulting observations reflect a biased relationship between the content and feedback. These observations do not account for the counterfactual condition, or what would have happened if the customer had not received a recommendation. In many cases, customers will have a high propensity of generating rewards, independent of the recommendations shown on the website. In this work we incorporate causal uplift modeling with contextual bandits in order to consider the heterogeneous treatment effect as an adjusted objective for top-k content selection. We demonstrate the performance and impact of the framework through both offline model evaluations and multiple live A/B experiments.

CCS Concepts: • **Computing methodologies** → **Sequential decision making**; **Batch learning**; *Learning from implicit feedback*; Causal reasoning and diagnostics; **Learning to rank**; *Supervised learning by regression*; • **Applied computing** → *Online shopping*; • **Information systems** → **Content ranking**; **Personalization**; *Top-k retrieval in databases*; *Recommender systems*; • **Mathematics of computing** → Bayesian computation.

Additional Key Words and Phrases: Personalization, Recommender system, Content optimization, Content ranking, Selection bias, Causal bandit, Contextual bandit, Uplift, View-through attribution, Fairness, Counterfactual learning

## 1 INTRODUCTION

In many e-commerce applications, customers rely on recommender systems to help sort through large corpuses of content in order to discover the small fraction of content that they would be interested in. Amazon's content optimization/ranking system is designed as a self-service tool which enables teams across the company to build content recommendation strategies once and run anywhere. Such a content optimization system is challenging mostly due to following two reasons: (1) continuous learning: surfacing the right content to the right user at the right time requires a ranking system to continually adapt to users' shifting interested along with newly introduced content; (2) content bias reduction: learning the unbiased incremental value of each piece of content given the context is typically unachievable given the limits of partial observations.

To continuously learn new content and adapt to changing customer behaviors, exploration/exploitation trade-off in the context of Reinforcement learning, is currently an active area of research. In literature, numerous techniques have emerged which are competitive and have shown promising results. These include epsilon-greedy ([26]), adding random noise to parameters[8], bootstrap sampling[19], and Thompson Sampling[6].

Content bias is introduced by the process of making recommendations online, which influences the way users interact with the system and how the data collected from users is fed back into the system. This leads to several types of biases such as popularity bias[5], human decision bias[7], position bias[14] or selection bias[13]. Traditional learning-to-rank approaches must contend with the these biases, and most approaches are focusing on position[14] or selection bias[30].

*Copyright 2022 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
Presented at the MORS workshop held in conjunction with the 16th ACM Conference on Recommender Systems (RecSys), 2022, in Seattle, USA.



CEUR-WS.org/Vol-3268/paper11.pdf

In this paper, we identified a new type of bias called Content Targeting Bias, introduced in a recommendations systems at industry scale due to content targeting criteria. Here, content targeting criteria, defined by content recommendation strategy owners, target only certain populations of customer or types of page contexts. These content owners then participate in ranking competitions only when targeting criteria is met. Content targeting criteria can be as simple as "all customers and context" or can be specific to a small portion of customers who have taken particular actions over the past month(s). For example, Content Targeting Bias is introduced when content owners target only signed-in customers, which are known to spend more on average than the population as a whole regardless of the recommendations provided. Another way this targeting bias can happen is when content owners target recommendations to some negative-profit item pages. Not accounting for such biases in ranking means we end up over or under estimating a content incremental performance, thus unfair ranking and degraded customer experiences. For practical reasons it is nearly impossible for a ranking system to have awareness of all targeting criteria that each content owner uses, along with an awareness of the detailed context and customer information at the level of each content owner. Thus, there needs to be a feature-agnostic way to mitigate Content Targeting Bias and thereby improve ranking.

On top of this, we further proposed quantified measurement for Content Targeting Bias within recommendation systems, and proposed solutions for reducing such biases. Our solution to reduce Content Targeting Bias was intuited by causal bandits work[25] [24]. In this work, we incorporate uplift model [27][20][9], using meta learning approaches (e.g. x-learner, r-learner)[17] [32], into contextualized bandits[18][3]. This approach was designed to consider the heterogeneous treatment effect between when content eligible to show but not actually observed (not treated) v.s. observed (treated), with the goal of improving equal opportunity of showing content without targeting criteria impacts, thus maximizing customer experience. To the best of our knowledge, our work is the first to identify content targeting as a unique bias and incorporate uplift modeling into bandit approaches including Bayesian Linear Regression Model (BLIR [10]), to reduce such biases for the content ranking problem. During experimentation in a Amazon commercial system[15], this work achieved significant online improvements across multiple pages within Amazon e-commerce website.

This paper is organized as follows: In section 2, we describe problem definitions. Section 3 describes the proposed solution. Section 4 covers offline model evaluation and online live experiments results with learnings. Finally, Section 5 details conclusions and future work.

## 2 PROBLEM DESCRIPTION

### 2.1 Formalizing problems

We define widget group as a region of experience on Amazon's e-commerce website which can be populated with recommended content (a.k.a. widget) provided by different teams. Note here the number of content rendered on widget group is much less than the number of all possible candidate content that is generated. Eligibility of rendering widget $w_i$ is typically determined by a combined factor of both the widget's targeting criteria and ranking system valuation.

The metric for measuring reward $R$ is determined by $MOI$ , short for 'metric of interest'. In our setting, $MOI$ takes into account the short-term as well as long-term impact to the customer's shopping experience, and helps us to fairly balance multiple and differing objectives of various stakeholders. Many Learning-to-Rank systems in the literature ([2][11][29]) optimize for Click-through Rate (CTR), while we are more interested in site-wise $MOI$. Towards this end, we have adopted an attribution modeling using view-through attribution (VTA)[16][31], which credits widgets for all rewards following an impression within an attribution window (e.g. 100 minutes). For example, if a customer view

some content recommendation $A$ for longer than 1 second (defined as an impression) and then make a purchase in the subsequent 100 minutes, the reward $R$ generated by this purchase along with all other high value actions taken in these 100 minutes will all be attributed back to the impressed content $A$.

The drawback of this methodology is that it loosens the connection between the content and the associated reward, which makes ranking system more vulnerable to Content Targeting Bias. Specifically, by considering feedback as "response" without considering counterfactual cases (i.e. what customer would have behaved if he/she had not received a recommendation) we end up with a biased estimate. Cases where customers have a high probability of generating down session rewards independent of the recommendations shown, can lead the system to over-estimate the value of widgets shown. This misspecification of value due to Content Targeting Bias disrupts fairness guarantees provided by the ranking system and ultimately leads to a suboptimal customer experience.

### 2.2 Formalizing Content Targeting Bias and ranking fairness

To formalize Content Targeting Bias, we adopt a recently introduced idea of opportunity bias[33] , a formula designed to evaluate whether different types of content receive clicks (or other engagement metrics) proportional to their true targeted population sizes (i.e. do content with different targeting criteria receive similar true positive rates?). This method assumes that the content recommended by content owners are all relatively in good quality. We believe this formalization of Content Targeting Bias is directly aligned with user satisfaction and economic gains of content owners.

To quantify the impact of Content Targeting Bias to recommendation systems, we need to first calculate the true positive rate for each content. Using show rate as an example, suppose content $i$ has been exposed to customers $E$ times in total, the true positive rate for $i$ is $TPR_i = E_i/A_i$, where $A_i$ is the total times of content $i$ is generated based on content owners' targeting criteria. Then, we can use the Gini Coefficient[4][28] to measure the inequality in true positive rates corresponding to content generation

$$Gini = \frac{\sum (2 * i - M - 1) * TPR_i}{(M * \sum TPR_i)} \tag{1}$$

where contents are indexed from 1 to $M$ in targeted audience size, non-descending. We use $-1 \leq Gini \leq 1$ to quantify the Content Targeting Bias in recommendation system: a close to 0 $Gini$ indicates a low bias; $Gini > 0$ represents that true positive rate is positively correlated to content targeted audience size; and $Gini < 0$ represents that the true positive rate is negatively correlated to audience size.

## 3 METHODOLOGY

To address problems above, we propose a framework employing uplift techniques with contextual bandits on top of VTA, to de-bias observations for the ranking system. In more detail, we consider adding contextual features in an uplift model in order to estimate Conditional Average Treatment Effect (CATE [1]) between exposure v.s. non-exposure of a recommendation to customers, using both Randomized Controlled Trial (RCT) or observational data. Under this framework, we also propose a modeling architecture incorporating Bayesian Linear Regression (BLIR) with Thompson Sampling to achieve online exploration-exploitation trade-offs.

### 3.1 Assumptions and definitions

We divided the causal impacts of showing a widget to customers on reward $R$ into two parts, (1) request-level incrementality, or the incremental value of showing top $K$ widgets to customer within request. This is to remove confounding

factors which impact the overall down-session reward $R$ independent of the recommendations received. (2) widget-level attribution: out of the top $K$ widgets, each widget's contribution to request-level incrementality. Ideally, we want to have a single causal model that can solve request-level incrementality and widget-level attribution at the same time, but for scope of this paper, we focus on the first problem, which is more related to the Content Targeting Bias issue, as removing customer/context intrinsic value from observed reward $R$.

Let $T$ be a dummy variable indicating treatment status, with $T = 1$ if a customer receives recommended contents (treatment) for a given request and $T = 0$ otherwise. The observed reward is defined as $Y \equiv T \cdot Y(1) + (1 - T) \cdot Y(0)$, where $Y(1)$ and $Y(0)$ are the potential outcomes when people receive the treatment or not. Under the unconfoundedness assumption[23], $Y(0), Y(1) \perp T$, we can approximate $Y$ using $Y(1)$ if treatment is imposed , or $Y(0)$ otherwise. In practice, instead of the simple unconfoundedness assumption, we make a conditional unconfoundedness assumption due to the non-random treatment assignment, which is also known as "strongly ignorable treatment assignment"[22], $Y(0), Y(1) \perp T|X$. In other words, given all the covariates X, the treatment assignment will be independent of the potential outcomes. Given $X = x$, the CATE $\tau(x)$ is then defined as $\tau(x) \equiv E[Y(1) - Y(0)|X = x]$

In this work, treatment group is defined as request-level exposure of top $K$ widgets while control group is defined as non-exposure of the entire widget group. This definition is to account for exogenous factors in the top-k ranking problem, where widgets on top ranks may have an impact on lower ranked widget's exposures.

### 3.2 Features

Another key point related to the unconfoundedness assumption is the set of covariates $X$ which can support it. In the real world, finding all confounders can be intractable, however in this work we simplify the problem by limiting to two confounding factors which impact user propensities: (1) content targeting where content is only generated to subgroup of customers; (2) model targeting where the model returns content non-uniformly given different page context. The conditional unconfoundedness assumption then becomes true as long as we can capture page context and a customer's intrinsic values in $X$. In the Content Targeting Bias problem, a customer's intrinsic values can be related only to the candidate widgets generated for a given request. Thus, with feature representations of candidate widgets as well as other page context and customer information, we can appropriately counteract the confounding problem.

### 3.3 Two-model framework estimating pseudo-effect

We further propose a two-model framework composed of a baseline (control) model and an uplift model using Linear Regression, to reduce Content Targeting Bias. Note we use linear model in this paper, but this approach can be applied to any types of Machine Learning models.

*3.3.1 Model 1 - Baseline Model.* The baseline model measures the observation in the control group, which is to estimate the expected down-session rewards of various contents being eligible to show but not actually observed by customers, defined at request level as $\hat{\mu}_0 = \beta^T X + \epsilon$, where $\beta$ is feature weights, and during training we find which best explains data. $\epsilon$ is a noise term to capture unobserved variables. $X$ represents the features used in the baseline model as explained in above sections.

*3.3.2 Model 2 - Uplift Model.* For each request with $K$ individual observation (widget) in the treatment groups, which are returned and actually observed by customers, we define 2

$$D_i^{(1)} = Y_i - \hat{\mu}_0(x) = \tau^{(1)} + \epsilon \tag{2}$$

where $Y_i$ is observed down-session reward for a customer at request $i$, and $D_i^{(1)}$ is the imputed treatment effects for request $i$ in the treated group, based on the baseline outcome estimator. $\tau^{(1)}$ is imputed treatment effects estimation for a given request. This "pseudo-effect" is an adjusted objective with Content Targeting Bias reduction, and it is then used as the outcome in a secondary machine learning model to obtain the response functions with treatment effects estimated. To achieve online exploration-exploitation trade-off, we utilize Bayesian Linear Regression Model (BLIR) [10] approach, $\hat{\tau}_A^{(1)} \sim \mathcal{N}(\theta^T X_A, \sigma^2 I)$, where $\hat{\tau}_A^{(1)}$ is the estimated score for widget $A$, $\theta$ is the coefficients of features, $X_A$ represents the features used in the uplift model. Note that $X_A$ is different from $X$ in that $X$ contains all candidate widgets information and is only used in offline, while $X_A$ only contains features for focal widget $A$. This uplift model estimates "pseudo-treatment-effects" for the observations in the treatment group, and can help reduce Content Targeting Bias, since we remove the counterfactual effect using the baseline (control) model. Finally, we do point-wise ranking online, by estimating a score for every candidate widget, sorting and returning top-K to customers. The exploration-exploitation trade-off is achieved by sampling model parameters from their posterior distributions through Thompson Sampling.

In addition to bias reduction with causal effect estimations, another fact of this two-model framework is that the baseline model is only used offline, generating uplift objectives for the second model. This simplification empowers us to include as many features as we can in $X$ without concern for latency or other requirements for online systems, such as representations of all candidate widgets features, while keeping online feature set $X_A$ relatively simple but achieve similar effect on bias reductions.

### 3.4 Log-tricks on objectives

One trick we performed is to transform reward $R$ and uplift estimations into log-scale. Transforming reward into log-scale is a widely used trick for removing outliers, thus we first introduced it on top of baseline model estimations. Here instead of using $log$, we used $signedLog1p(v) = signed(v) * log1p(|v|)$ to achieve symmetry and valid values at zeros (will still denote using $log$ in following context). Due to Jensen's inequality, transforming log-scaled baseline estimations $log\_\hat{\mu}_0$ back is biased, thus, we directly perform treatment effects estimations by 3

$$log\_D^{(1)} = log\_Y_i - log\_\hat{\mu}_0(x) = log\_\tau^{(1)} + \epsilon \tag{3}$$

where $log\_\hat{\mu}_0(x)$ can be treated as geometric mean of baseline values given covariates $X$, like $log((\Pi_m \hat{\mu}_0)^{1/M})|X$, while treatment effects $log\_\tau^{(1)}$ becomes multiplicative, like $log(\frac{D^{(1)}}{geoMean\_\hat{\mu}_0})$. Although this definition differs from additive uplift in above section, the signs still has meaning, in that positive value indicate positive incrementality whereas negative value indicate negative incrementality. In addition, defining this relationship as multiplicative also lends an intuitive semantic meaning. For example, say customers who are not signed-in spend \$10 on average while signed-in customers spend \$100, when representing uplift for some content $A$, instead of an additive uplift of \$10 (thus \$20 for unsigned-in customer and \$110 for signed-in customer), a multiplicative lift ratio of 1.1 makes more sense in terms of e-commerce context (thus \$11 for unsigned-in customer and \$110 for signed-in customers). Results from A/B testing show log scaled uplift (multiplicative) performs better than additive uplift.

### 3.5 Data collection

Randomly hiding data, as in **RCT**, could provide us with a more unbiased estimate of CATE, however it is costly to proactively hide content from customers. In RCT setup, we randomly punt (do not display) the entire widget group

a small percentage $\zeta$ of the time, and train baseline model using only this punted traffic. In this way, using RCT, we are able to remove the confounder effect resulting from customer selection bias, e.g. customer's propensities to scroll down the page and browse content. **Observational data**, in contrast to RCT, can provide sufficient data with minimal cost, but at the expense of increased data bias. In practice, when top K widgets are returned, they are actually not always all shown on viewport to customers. Our system is able to capture this client-side impression behaviors, and our proposed work is able to train baseline&uplift models using this observational data. Results show limited biases using observational data compared to RCT.

## 4 MODEL EVALUATION AND EXPERIMENTS

### 4.1 Ranking fairness estimation

In offline model evaluation of ranking fairness, we compared *Gini* scores defined in section 2.2 by ranking content using production model (non-uplift linear bandits directly regressed on VTA) v.s. two-model uplift bandit approach. Also, in order to measure fairness in multiple dimensions, we defined $TPR$ in 2 ways

$$TPR_i = \begin{cases} \frac{\sum \text{content exposure}}{\sum \text{content generated}} & \text{TPR of show rate} \\ \frac{\sum \text{observed content reward}}{\sum \text{content exposure}} & \text{TPR of performance} \end{cases} \tag{4}$$

From **??**, uplift approach reduces bias in terms of both content show rate and content average performance, especially the latter, which indicates that our approach improves fairness based more on content's performance than their show rate coverage, which better aligns with business goals.

Table 1. uplift Gini metrics

| | production model | uplift model |
|---|---|---|
| Gini of content show rate | 0.818 | 0.815 |
| Gini of content performance | 0.55 | 0.48 |

### 4.2 Effectiveness of heterogeneous treatment effect estimation

Through analysis on real online data, we identified several widgets targeting customers with high intrinsic rewards, compare their observed score with predicted uplift values, and validated that uplift is able to reduce those biases. From Table 2, widget C and D are widgets targeting at high-valued customer only, while widget A, B, E, F are common widgets targeting at all customers. The scores are observed or estimated rewards as defined in section 2.1, and are represented in tuples formatting as ***(average across all customers, average across high-valued customers only)***. Evaluating using proposed framework, we intentionally exclude all customer related features, so model won't depend on customer profile. We can see that although widget C has the highest average observed score across all customers, the actual uplift prediction is not as high as widget B after reducing Content Targeting Bias (0.15 v.s. 0.24), this aligns with our observations of these widgets through online experimentation. A similar pattern can be found on widget D.

### 4.3 RCT and observational data

We also evaluate the baseline model in two-model uplift approach, trained with either RCT or observation data. Through offline analysis, we see that the observational uplift approach is able to achieve similar results compared to RCT

Table 2. Effectiveness of uplift estimation (average across all customers, average across high-valued customers only)

| | widget A | widget B | widget C | widget D | widget E | widget F |
|---|---|---|---|---|---|---|
| average observed reward | (5.42,6.73) | (6.14,7.49) | (7.70,7.70) | (3.74,3.74) | (3.48,4.52) | (3.58,4.82) |
| Control estimation | (4.88,6.13) | (5.20,6.45) | (6.37,6.37) | (3.84,3.84) | (3.97,5.02) | (4.79,6.34) |
| Treatment estimation | (4.97,6.15) | (5.44,6.63) | (6.52,6.52) | (3.76,3.76) | (3.86,4.86) | (4.83,6.35) |
| average uplift estimation | (0.08,0.01) | (0.24,0.18) | (0.15,0.15) | (-0.09,-0.09) | (-0.11,-0.16) | (0.03,0.02) |

model, without significant difference in rankings. However, RCT baseline model gives a general lower estimation of counterfactuals (5% $\sim$ 10%). This gap can be interpreted by customers' selection biases, that when customers intentionally don't view content (observational data), they might be attracted by other content on the page or already have a clear shopping mission. This in turn leads to higher estimates in observational baseline models vs RCT. Estimating this gap is important, since this can be used to adjust observational modeling, and improve model interpretability while avoiding the high cost imposed by RCT, e.g. showing sub-optimal results on a certain percentage of the populations.

### 4.4 Online experiments

Content Targeting Bias might appear in different formats across different pages. For example, on homepage, Content Targeting Bias is mostly introduced by different targeting criteria on customer populations; while on product detail pages, biases are mostly introduced by targeting criteria towards context information. To gain a thorough understanding about this proposed work, we have completed five online randomized A/B experiments[12] [21].

Table 3. Aggregated table for all experiments results

| Experiment Iteration | A/B Treatment | Annualized Impact (% improvement) | Confidence interval | p-value |
|---|---|---|---|---|
| EXP-1 | Observational Uplift | +0.13% | +0.02% ~ +0.23% | 0.020 |
| EXP-1 | RCT Uplift | +0.05% | -0.06% ~ +0.15% | 0.380 |
| EXP-2 | Observational Uplift | +0.11% | -0.09% ~ +0.32% | 0.280 |
| EXP-3 | Observational Uplift | +0.07% | +0.00% ~ +0.13% | 0.039 |
| EXP-4 | Observational Uplift | +0.06% | -0.02% ~ +0.14% | 0.170 |
| EXP-4 | Observational Uplift with log trick | +0.09% | +0.01% ~ +0.17% | 0.024 |
| EXP-5 | Observational Uplift with log trick | +0.13% | +0.09% ~ +0.18% | 0.000 |

*4.4.1 Online experiment setup.* In our online experimentation setting, observational units (or shopping sessions) are randomly exposed to either the baseline control policy or the alternative treatment policies. Here, we track the impact to our metric of interest *MOI* , which is a measure of improved site-wide customer shopping experience. In our results, we include the causal effect w.r.t percentage improvement in this metric at Amazon's scale. The experiments are conducted across all of Amazon's world-wide marketplaces and product categories. Level of significance $\alpha$ for these experiments was determined by Amazon's business objectives and was set to 0.10. Duration for these experiments was estimated from statistical power analysis. We allocated equal traffic to both the control and treatment groups. During the course of the experiment, the models were incrementally trained using their own set of logged feedback.

Through all experiments, A/B test Control group is a non-uplift linear contextual bandit regressed on VTA. In **Experiment 1**, we ran on slots located at the bottom of different pages (e.g. detail page, homepage page etc.). The following treatments are performed, (1) observational uplift: two-model uplift with observational data;(2) RCT uplift: two-model uplift with RCT data. In **Experiment 2**, we ran on slots located at top of product detail pages, with

treatment group as observational uplift. In **Experiment 3**, we ran on cart pages, using uplift with observational data. In **Experiment 4**, we ran on desktop product detail page, the following treatments are performed, (1) observational uplift; (2) observational uplift with log scale tricks. In **Experiment 5** ran on mobile app product detail page, using observational uplift with log scale tricks as treatment group.

*4.4.2 Online experiment results.* We observed consistent improvements across experiments. Table 3 shows details on *MOI* results with confidence intervals. Through these experiments, we proved (1) Improvements using heterogeneous treatment effect estimation on top of bandits approach, on different pages including homepage, detail page, cart pages, across Amazon e-commerce websites. Out of these, experiment 1, 3, 4, 5 achieved statistically significant improvements with p-value less than 0.05, with experiment 5 having p-value close to 0.000. (2) the proposed method using observational data achieved significant improvements (with p-value 0.02) while RCT only outperformed production model with low confidence (with p-value 0.38). This demonstrated that observational uplift modeling can achieve similar results as using RCT, by successfully minimizing potential bias in training examples. Conversely, RCT depends on random hiding contents which is guaranteed to be suboptimal some percentage of the time, thus the overall RCT performance is hurt; (3) the proposed uplift model with log tricks outperforms additive uplift, which can be observed directly from experiment 4 where uplift with log tricks achieved significant improvements (with p-value 0.024) while additive uplift improvements was not significant with p-value as 0.17. This further demonstrates our hypothesis on log tricks for better managing outliers and more reasonable semantic meanings from multiplicative uplift.

## 5 CONCLUSION

In this paper, we studied a new type of bias in Learning-to-Rank systems, called Content Targeting Bias. We defined such bias, proposed quantified measurement and further proposed an online ranking approach using BLIR considering contextual features into uplift modeling to reduce such bias for top-K content selection. Through this work, we introduced log-tricks for treatment effect estimations between exposure v.s. non-exposure of a recommendation and compared baseline models trained using both RCT and observational data. This work demonstrates significant bias reduction as well as significant *MOI* improvements both offline and online. In future work, building on top of current framework, we will improve uplift estimation by applying propensity-weighting based meta learner approach e.g. double ML (R-learner) to improve current uplift modeling, to further reduce display biases in content rankings.

## REFERENCES

[1] Jason Abrevaya, Yu-Chin Hsu, and Robert P Lieli. Estimating conditional average treatment effects. *Journal of Business & Economic Statistics*, 33(4): 485–505, 2015.

[2] Aman Agarwal, Kenta Takatsu, Ivan Zaitsev, and Thorsten Joachims. A general framework for counterfactual learning-to-rank. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 5–14, 2019.

[3] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural network. In *International Conference on Machine Learning*, pages 1613–1622. PMLR, 2015.

[4] Malcolm C Brown. Using gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on alberta data. *Social science & medicine*, 38(9):1243–1256, 1994.

[5] Òscar Celma and Pedro Cano. From hits to niches? or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, pages 1–8, 2008.

[6] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, 24:2249–2257, 2011.

[7] Li Chen, Marco De Gemmis, Alexander Felfernig, Pasquale Lops, Francesco Ricci, and Giovanni Semeraro. Human decision making and recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(3):1–7, 2013.

[8] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[9] Graton Gathright, Ranjan Roopesh, Vasudev Rahul, Marshall Yan, and Fan Zhang. Cross-channel attribution of consumer marketing. In *Amazon Machine Learning Conference*, 2017.

[10] Thore Graepel, Joaquin Quinonero Candela, Thomas Borchert, and Ralf Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *ICML*, 2010.

[11] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: a factorization-machine based neural network for ctr prediction. *arXiv preprint arXiv:1703.04247*, 2017.

[12] Somit Gupta, Ronny Kohavi, Diane Tang, Ya Xu, Reid Andersen, Eytan Bakshy, Niall Cardin, Sumita Chandran, Nanyu Chen, Dominic Coey, et al. Top challenges from the first practical online controlled experiments summit. *ACM SIGKDD Explorations Newsletter*, 21(1):20–35, 2019.

[13] James J Heckman. *Sample selection bias as a specification error with an application to the estimation of labor supply functions*. Princeton University Press, 2014.

[14] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. Unbiased learning-to-rank with biased feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 781–789, 2017.

[15] Sameer Kanase, Yan Zhao, Shenghe Xu, Mitchell Goodman, Manohar Mandalapu, Benjamyn Ward, Chan Jeon, Shreya Kamath, Ben Cohen, Vlad Suslikov, Yujia Liu, Hengjia Zhang, Yannick Kimmel, Saad Khan, Brent Payne, and Patricia Grao. An application of causal bandit to content optimization. In *Proceedings of the 5th Workshop on Online Recommender Systems and User Modeling (ORSUM 2022), in conjunction with the 16th ACM Conference on Recommender Systems (RecSys 2022), Seattle, WA, USA*, 2022.

[16] Pavel Kireyev, Koen Pauwels, and Sunil Gupta. Do display ads influence search? attribution and dynamics in online advertising. *International Journal of Research in Marketing*, 33(3):475–490, 2016.

[17] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

[18] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

[19] Ian Osband, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. Deep exploration via bootstrapped dqn. *Advances in neural information processing systems*, 29:4026–4034, 2016.

[20] Roopesh Ranjan, Narayanan Sadagopan, and Guido Imbens. A propensity matching approach to multi touch attribution. In *Amazon Machine Learning Conference*, 2016.

[21] Thomas S Richardson, Yu Liu, James McQueen, and Doug Hains. A bayesian model for online activity sample sizes. In *International Conference on Artificial Intelligence and Statistics*, pages 1775–1785. PMLR, 2022.

[22] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[23] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[24] Neela Sawant, Chii Babu Namballa, Narayanan Sadagopan, and Houssam Nassif. Multi-armed bandit framework for causal effect optimization. In *Amazon Machine Learning Conference*, 2017.

[25] Neela Sawant, Chitti Babu Namballa, Narayanan Sadagopan, and Houssam Nassif. Contextual multi-armed bandits for causal marketing. *arXiv preprint arXiv:1810.01859*, 2018.

[26] Bradly C Stadie, Sergey Levine, and Pieter Abbeel. Incentivizing exploration in reinforcement learning with deep predictive models. *arXiv preprint arXiv:1507.00814*, 2015.

[27] Bo Tan, Pramod Muralidharan, Naveen Nair, Wenduo Wang, Shaurya Gupta, Jimmy Issac, Vignesh Kannappan, Prakash Bulusu, and Phil Leslie. Attribution of prime member signups to prime benefits. In *Amazon Machine Learning Conference*, 2016.

[28] Adam Wagstaff, Pierella Paci, and Eddy Van Doorslaer. On the measurement of inequalities in health. *Social science & medicine*, 33(5):545–557, 1991.

[29] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1235–1244, 2015.

[30] Xuanhui Wang, Michael Bendersky, Donald Metzler, and Marc Najork. Learning to rank with selection bias in personal search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 115–124, 2016.

[31] Shenghe Xu, Yan Zhao, Sameer Kanase, Mitchell Goodman, Saad Khan, Brent Payne, and Patricia Grao. Machine learning attribution: Inferring item-level impact from slate recommendation in e-commerce. In *KDD 2022 Workshop on First Content Understanding and Generation for e-Commerce*, 2022. URL https://www.amazon.science/publications/machine-learning-attribution-inferring-item-level-impact-from-slate-recommendation-in-e-commerce.

[32] Zhenyu Zhao and Totte Harinen. Uplift modeling for multiple treatments with cost optimization. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 422–431. IEEE, 2019.

[33] Ziwei Zhu, Yun He, Xing Zhao, and James Caverlee. Popularity bias in dynamic recommendation. 2021.