

Reaching the End of Unbiasedness: Uncovering Implicit Limitations of Click-Based Learning to Rank

Harrie Oosterhuis
Radboud University
Nijmegen, The Netherlands
harrie.oosterhuis@ru.nl

ABSTRACT

Click-based learning to rank (LTR) tackles the mismatch between click frequencies on items and their actual relevance. The approach of previous work has been to assume a model of click behavior and to subsequently introduce a method for unbiasedly estimating preferences under that assumed model. The success of this approach is evident in that unbiased methods have been found for an increasing number of behavior models and types of bias.

This work aims to uncover the implicit limitations of the high-level prevalent approach in the counterfactual LTR field. Thus, in contrast with limitations that follow from explicit assumptions, our aim is to recognize limitations that the field is currently unaware of. We do this by inverting the existing approach: we start by capturing existing methods in generic terms, and subsequently, from these generic descriptions we derive the click behavior for which these methods can be unbiased. Our inverted approach reveals that there are indeed implicit limitations to the counterfactual LTR approach: we find counterfactual estimation can only produce unbiased methods for click behavior based on affine transformations. In addition, we also recognize previously undiscussed limitations of click-modelling and pairwise approaches to click-based LTR. Our findings reveal that it is impossible for existing approaches to provide unbiasedness guarantees for all plausible click behavior models.

CCS CONCEPTS

• Information systems → Learning to rank.

KEYWORDS

Counterfactual Learning to Rank; Unbiased Estimation; Clicks

ACM Reference Format:

Harrie Oosterhuis. 2022. Reaching the End of Unbiasedness: Uncovering Implicit Limitations of Click-Based Learning to Rank. In *Proceedings of the 2022 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '22)*, July 11–12, 2022, Madrid, Spain. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3539813.3545137>

1 INTRODUCTION

Search and recommendation systems make large collections easily accessible and navigable for users [12, 17, 20, 23, 52, 61]. A vital component of such systems are ranking models that - based on the

features of the items - create short ranked lists of items to present to the user [11, 36, 42, 70]. The experience of users depends on these rankings providing them the items that are relevant to their current needs or match their preferences [30, 38, 39, 58, 59]. Ranking systems are applied to an enormous variety of settings, e.g. web-search engines [12, 20, 52], e-commerce [61], video platforms [17, 23], online advertising [34, 35], etc. As a result of this large variety, what is relevant or preferred also heavily varies over applications and contexts. Consequently, methods for optimizing ranking systems based on user click behavior have become increasingly popular, since they have the potential to automatically adapt to the exact setting in which they are applied [25, 31, 41].

However, click behavior is not an unbiased representation of user preferences or relevance [32, 54, 59], since there are many other factors that affect whether an item is clicked or not: i.e. items that are not displayed cannot be clicked [48], and furthermore, it is well known that the position at which an item is displayed heavily affects the number of clicks it receives [18, 50]. Accordingly, the unbiased learning to rank (LTR) field has introduced several techniques for dealing with the mismatch between clicks and actual preferences [1, 6, 33, 41, 68]. Typically, previous work in this field first identifies certain forms of bias, e.g. position-bias [1, 26, 33, 68], item-selection-bias [44, 48], trust-bias [2, 4, 50, 66], etc., then it assumes a mathematical model of click behavior that describes the identified bias, and subsequently, introduces a method that can provenly optimize a ranking model unbiasedly. Thus far, this approach has been very successful as methods have been found for an exceeding number of behavior models [33, 41, 44, 66, 68]. The large number of recent publications make it evident that the success of this approach is still ongoing and that unbiased LTR remains a very active field [1, 2, 5, 6, 27, 41, 44, 46, 48, 65, 66, 75].

In this work, we aim to investigate the implicit limitations of the prevalent approach in unbiased LTR. In other words, our goal is to better understand what problems we can and cannot expect unbiased LTR to solve in the future. In particular, our focus is on limitations that are implicit, i.e. limitations that follow from implicit assumptions in the existing approach that the field is thus currently unaware of. We believe this is important for two main reasons: Firstly, by understanding how the current approach is lacking, the field could be better equipped to find alternative approaches in the future. Secondly, due to its success, the expectation could exist that every click-based LTR method should follow the prevalent unbiased LTR approach. However, such an expectation could - unknowingly - systematically exclude problem settings where this approach can never be successful. For the field, it is thus important to understand when unbiasedness is impossible and change its expectations for future work accordingly.



This work is licensed under a Creative Commons Attribution International 4.0 License.

ICTIR '22, July 11–12, 2022, Madrid, Spain
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9412-3/22/07.
<https://doi.org/10.1145/3539813.3545137>

Our method of finding the limitations of the prevalent approach is to invert it: we start with generic formulations that describe the main branches of existing methods in the unbiased LTR field: counterfactual estimation [33, 41, 68] and click modelling [15, 65, 75]. Subsequently, from these generic descriptions, we derive the conditions under which they can be unbiased. Our process reveals that conditions exist where counterfactual estimation can *never* produce an unbiased method. Furthermore, we find that click modelling does not seem to provide robust unbiasedness guarantees. Lastly, we prove that the starting assumptions of the *Unbiased LambdaMART* method [27] are problematic since they implicitly assume clicks are unbiased indicators of relevance. These findings have serious implications for the unbiased LTR field: most importantly, they reveal that either very different novel approaches are needed for unbiased click-based LTR in certain settings, or that it may be the case that unbiasedness will never be feasible in certain circumstances.

2 PRELIMINARIES: THE LTR GOAL

Before we start our discussion of click-based LTR, this section will explicitly state the assumptions underlying the LTR task. Firstly, LTR work generally assumes that each item has a certain merit [7, 11, 21, 36, 40], often referred to as its *relevance* (in search contexts) or *user preference* (in recommendation contexts). For the sake of simplicity, we use the *relevance* terminology for the remainder of this work. In formal terms, relevance is often interpreted as a scalar value; without loss of generality, we define relevance as the probability that a user prefers this item conditioned on the context q , i.e. for an item d :

$$P(R = 1 \mid d, q) = R_{d|q} \in [0, 1]. \quad (1)$$

In a web-search setting, the context q is generally referred to as the *query*, that represents the user's search request so that $R_{d|q}$ can indicate how well a document and the user's information need matches. Alternatively, in a recommendation setting q may contain *personalized information* about the user from which their preferences could be inferred. In addition, we allow q to contain any other contextual information that could affect relevance or preference, e.g. the time of day/year, preferred language by the user, etc. Importantly, our definition of q excludes information specific to any item d , i.e. for each individual ranking y , q is the same for all items. Our intentionally generic definition of the context q allows our assumptions to cover an enormous variety of ranking settings. For simplicity, this work will use the *query* terminology.

The goal of the ranking system is to provide a ranking y of items to the user [11, 36, 42, 70, 72], y is simply an ordered list of items:

$$y = [d_1, d_2, \dots]. \quad (2)$$

The quality of a ranking for a query q is determined by a weighted sum of associated the $R_{d|q}$ values, where the weight per item is determined by its rank in y [30, 38, 39]. Let the $\text{rank}(d \mid y) \in \mathbb{N}_{>0}$ function indicate the rank of d in y and $\lambda : \mathbb{N}_{>0} \rightarrow \mathbb{R}$ the weight function, the quality of ranking is then:

$$\mathcal{R}_q(y) = \sum_{d \in y} \lambda(\text{rank}(d \mid y)) R_{d|q}. \quad (3)$$

In general, ranking models predict a score for each item conditioned on the query and a ranking is created by sorting items according to their predicted scores [11, 36, 42, 70, 72]. In formal terms, the quality of a ranking model s is the expected quality of its rankings over the natural distribution of the queries:

$$\mathcal{R}(s) = \mathbb{E}_q[\mathcal{R}_q(y_{q,s})]. \quad (4)$$

In practice this expectation is often estimated based on a dataset that contains a large sample of queries, i.e. web-search datasets that contain many queries logged from users [12, 20, 52].

This concludes our very generic description of the LTR goal and its explicit assumptions. We note that even this generic description brings several limitations, for instance, its main assumption is that ranking quality can be decomposed into rank-weights and item relevances that only depend on query-item combinations (Eq. 3). Consequently, this explicitly excludes any notion of diversity in a ranking, e.g. it assumes users do not mind if all items in a ranking are very similar or identical [16, 28, 73]. We will not further discuss this particular limitation since it follows from explicit assumptions, and it is well-known and discussed in previous work [19, 53, 56].

3 EXISTING CLICK-BASED LTR APPROACH

This section will describe the two main families of methods in click-based LTR: counterfactual estimation and click modelling.

3.1 Counterfactual Estimation

Counterfactual estimation appears to be the most common approach in click-based LTR literature [1, 2, 5, 27, 33, 41, 44, 46, 66, 68, 71]. This approach starts by assuming a probabilistic model of click behavior, and subsequently, derives an estimation method for estimating the relevance of an item based on clicks. Mathematically, the method can usually be proven to provide unbiased estimates of the relevances. To illustrate this approach, we will briefly describe earlier work by Vardasbi et al. [66] for affine click models [2].

Let k denote a display position where an item can be displayed, per position there are the parameters $\alpha_k \in [0, 1]$ and $\beta_k \in [0, 1]$ so that $\alpha_k + \beta_k \in [0, 1]$. The affine model assumes that the probability of a click $C \in \{0, 1\}$ at position k is an affine transformation of the item's relevance:

$$P(C = 1 \mid d, k, q) = \alpha_k R_{d|q} + \beta_k. \quad (5)$$

This model was originally proposed by Agarwal et al. [2] to capture both position-bias and trust-bias. In the earliest unbiased LTR work a linear transformation was assumed that only describes position-bias [1, 5, 33, 68], in the affine model this scenario can be captured with: $\forall k, \beta_k = 0$. In the linear case, α_k is called a propensity, when $\beta_k \neq 0$, α_k can be interpreted as the correlation between relevance and clicks at position k . Non-zero β_k allow the model to capture trust-bias: an increase in clicks regardless of relevance at higher ranks due to the trust users have in the ranking system [2, 4, 32, 46, 50, 65, 66].

Clicks under this affine model are biased in that the expected click-through-rates (CTRs) are not equal to relevances. Let \mathcal{D}_q indicate a set of observed interactions with query q , where $c_i(d) \in \{0, 1\}$ indicates whether item d was clicked at interaction i and $k_i(d)$ the position at which d was displayed at interaction i . We can

then express the bias of CTRs formally:

$$\mathbb{E}_{c,k} \left[\frac{1}{|\mathcal{D}_q|} \sum_{i \in \mathcal{D}_q} (c_i(d) - R_{d|q}) \right] = \mathbb{E}_k [(\alpha_k(d) - 1)R_{d|q} + \beta_k(d)] \neq 0.$$

Depending on the α and β parameters and the positions where d was displayed, its CTR will over or underestimate its relevance. Naively ranking items according to their CTRs will give the items displayed at higher ranks an unfair advantage due to position-bias and trust-bias [2, 18, 32, 33]. In other words, high CTRs are often more indicative of how an item was displayed than its relevance.

Based on an inversion of Eq. 5, Vardasbi et al. [66] introduced an affine estimator that uses the estimated $\hat{\alpha}$ and $\hat{\beta}$ parameters:

$$\hat{R}_{d|q} = \frac{1}{|\mathcal{D}_q|} \sum_{i \in \mathcal{D}_q} \frac{c_i(d) - \hat{\beta}_{k_i(d)}}{\hat{\alpha}_{k_i(d)}}, \quad (6)$$

that can easily be used to estimate the quality of a ranking:

$$\hat{\mathcal{R}}_q(y) = \sum_{d \in y} \lambda(\text{rank}(d | y)) \hat{R}_{d|q}. \quad (7)$$

The affine estimator has the following bias:

$$\mathbb{E}_{c,k} [\hat{R}_{d|q} - R_{d|q}] = \mathbb{E}_k \left[\frac{(\alpha_k(d) - \hat{\alpha}_k(d))R_{d|q} + \beta_k(d) - \hat{\beta}_k(d)}{\hat{\alpha}_k(d)} \right].$$

It is easy to see that this estimator is unbiased when the estimated bias parameters are correct [46, 65, 66]. Moreover, it can then also be used for an unbiased estimate of ranking quality:

$$\begin{aligned} & \left(\forall d \in y, \forall i \in \mathcal{D}_q, \hat{\alpha}_{k_i(d)} = \alpha_{k_i(d)} \wedge \hat{\beta}_{k_i(d)} = \beta_{k_i(d)} \right) \\ & \longrightarrow \forall d \in y, \mathbb{E}_{c,k} [\hat{R}_{d|q}] = R_{d|q} \longrightarrow \mathbb{E}_{c,k} [\hat{\mathcal{R}}_q(y)] = \mathcal{R}_q(y). \end{aligned} \quad (8)$$

Trivially, this can provide an unbiased estimate of model performance $\mathcal{R}(s)$ which in turn can be optimized with standard LTR techniques [1, 33, 44, 70]. Therefore, Vardasbi et al. [66] provide an unbiased LTR method for their assumed affine click model.

The above example illustrates how Vardasbi et al. [66] applied the common counterfactual approach to unbiased LTR, other work has applied it to many different models of click behavior: For example, early work applied this approach to position-bias only [1, 5, 33, 68]; Fang et al. [22] to position-bias that varies per query; Oosterhuis and de Rijke [44] to position-bias and item-selection-bias; Vardasbi et al. [64] to cascading examination behavior; Wu et al. [71] to a bias from surrounding items; and Oosterhuis and de Rijke [46] address position-bias, trust-bias and item-selection-bias all at once.

3.2 Click Modelling

The other main approach in click-based LTR is click modelling, this approach is both used to make relevance estimates [8–10, 13–15, 75] as well as estimating the bias parameters values for counterfactual estimation [45, 66, 68, 69]. The approach is to fit a generative model of click behavior on the observed click data, the hope is that the resulting fitted model captures both the bias and relevances by modelling both simultaneously [15]. Similar to the counterfactual estimation, the click modelling approach starts by assuming a probabilistic model of click behavior. For example, we can again assume

the affine click model in Eq. 5 [2, 66], the predictive model we want to fit would thus be:

$$\hat{P}(C = 1 | d, k, q) = \hat{\alpha}_k \hat{R}_{d|q} + \hat{\beta}_k. \quad (9)$$

Let \hat{A} and \hat{B} indicate the set of estimated $\hat{\alpha}$ and $\hat{\beta}$ parameters and \hat{R}_q a vector of relevance estimates of all items $\hat{R}_{d|q}$ for query q . The model can be fitted to data using the negative log-likelihood loss:

$$\begin{aligned} \mathcal{L}(\hat{R}_q, \hat{A}, \hat{B}, \mathcal{D}_q) &= \frac{-1}{|\mathcal{D}_q| |\hat{R}_q|} \sum_{i \in \mathcal{D}_q} \sum_{d \in \hat{R}_q} c_i(d) \log(\hat{\alpha}_{k_i(d)} \hat{R}_{d|q} \\ &+ \hat{\beta}_{k_i(d)}) + (1 - c_i(d)) \log(1 - (\hat{\alpha}_{k_i(d)} \hat{R}_{d|q} + \hat{\beta}_{k_i(d)})). \end{aligned} \quad (10)$$

An important difference with the counterfactual estimation approach is that the relevance estimates \hat{R}_q are not the output of an estimator function but parameters to be optimized. If we consider the expected value of the loss, we see that in expectation it is equal to a negative log-likelihood loss between the predictive click model and the actual click probabilities:

$$\begin{aligned} & \mathbb{E}_{c,k} [\mathcal{L}(\hat{R}_q, \hat{A}, \hat{B}, \mathcal{D}_q)] \\ &= \mathbb{E}_k \left[\frac{-1}{|\hat{R}_q|} \sum_{d \in \hat{R}_q} (\alpha_k(d) R_{d|q} + \beta_k(d)) \log(\hat{\alpha}_k(d) \hat{R}_{d|q} + \hat{\beta}_k(d)) \right. \\ &\quad \left. + (1 - (\alpha_k(d) R_{d|q} + \beta_k(d))) \log(1 - (\hat{\alpha}_k(d) \hat{R}_{d|q} + \hat{\beta}_k(d))) \right] \\ &= \frac{-1}{|\hat{R}_q|} \sum_{d \in \hat{R}_q} \mathbb{E}_k [P(C = 1 | d, k, q) \log(\hat{P}(C = 1 | d, k, q)) \\ &\quad + (1 - P(C = 1 | d, k, q)) \log(1 - \hat{P}(C = 1 | d, k, q))]. \end{aligned} \quad (11)$$

The hope is that by minimizing the loss, the model learns the relevances and bias parameters correctly, since correct parameters would minimize the loss in expectation:

$$\begin{aligned} & \left((\forall d \in \hat{R}_q, \hat{R}_{d|q} = R_{d|q}) \wedge (\forall k, \hat{\alpha}_k = \alpha_k \wedge \hat{\beta}_k = \beta_k) \right) \\ & \longrightarrow \mathbb{E}_{c,k} [\mathcal{L}(\hat{R}_q, \hat{A}, \hat{B}, \mathcal{D}_q)] = \min_{\hat{R}'_q, \hat{A}', \hat{B}'} \mathbb{E}_{c,k} [\mathcal{L}(\hat{R}'_q, \hat{A}', \hat{B}', \mathcal{D}_q)]. \end{aligned} \quad (12)$$

However, it is important to realize the direction of the implication which reveals that this is not a strong guarantee: minimizing the expected loss does not imply that the correct parameters have been found. Section 6 will take a critical look at the unbiasedness guarantees this approach can provide.

Click modelling was initially - and could still be - seen as its own field within information retrieval: early methods used intuitive graphical models [8, 13, 15], later work relies on neural models that are hard to interpret but have better predictive performance [9, 10, 14]. With the inception of unbiased LTR, click models became a common choice for estimating bias parameters for counterfactual estimation instead of relevance estimation [45, 66, 68, 69]. However, recently novel click models for relevance estimation have been introduced again: For example, Vardasbi et al. [65] proposed a click model that uses a prior distribution over relevances; and Zhuang et al. [75] used a click model with a state-of-the-art deep learning architecture for relevance estimation in a grid layout. They found their click model to be more effective than existing counterfactual estimation techniques in a real-world recommendation setting.

3.3 Shortcomings of the Existing Approach

The unbiased LTR field has been very successful both in terms of theoretical results [33, 40, 41, 46] as in performance improvements for real-world search and recommendation systems [2, 33, 68, 75]. However, there are also some well-known shortcomings with the existing approach that should be discussed:

First, while in theory counterfactual estimation can provide provenly unbiased estimators, in practice the application of these estimators is rarely actually unbiased. There are three main reasons for this mismatch: (i) the assumed behavior model is incorrect - real-world user behavior is rarely captured perfectly by mathematical models; (ii) the propensities are inaccurate - perfect propensity estimation is infeasible in practice [3, 69]; and most importantly (iii) it is standard to apply clipping in practice - propensities are clipped by a threshold $\tau \in [0, 1]$: $\hat{\alpha}_{k(d)}^{\text{clip}} = \max(\hat{\alpha}_{k(d)}, \tau)$, this introduces some bias but can greatly reduce variance [33, 62]. The first two points are understandable: models and parameters are often inaccurate and small inaccuracies are often inconsequential. However, clipping generally has a large effect and purposefully introduces notable bias to the estimation process. It is at least surprising that most work in the unbiased LTR field puts much emphasis on the unbiased estimators they introduce, while simultaneously applying biased versions of the estimators in their experiments [2, 33, 44, 46, 66].

Second, while click modelling has seen real-world success [75], there are very little guarantees regarding its performance. In stark contrast with counterfactual estimation - and to the best of our knowledge - it is currently unclear whether there are broad theoretical guarantees for when click models can provide correct relevance estimates. Even when the structure of the assumed behavior model is correct, that does not guarantee that a fitted click model will have unbiased relevance estimates. From a theoretical perspective, it is thus unclear in what circumstances click models are reliable or not.

Lastly, while there have been several high-level looks at the unbiased LTR field, they have mainly compared the subgroups of online and counterfactual LTR methods [6, 29, 47]. To the best of our knowledge, previous work has not looked at the properties of the high-level approach of the Unbiased LTR field, instead of its individual methods. It is thus currently not known what the limitations of the prevalent approach are, nor what they may entail.

4 METHODOLOGY

Section 3 described the prevalent approaches in the unbiased LTR field and noted that the limitations of the high-level approach are currently unclear. In this work, we will investigate whether there are implicit assumptions in the two main families of debiasing approaches. In particular, we want to find out whether they impose limits on which circumstances estimation methods can or cannot be unbiased or consistent. This section will explain our methodology to investigating such implicit limitations, we will start by defining the two theoretical properties that we are interested in.

We begin by formally defining unbiasedness:

Definition 4.1. Unbiasedness. A click-based relevance estimation method is *unbiased*, if the expected values of all the resulting relevance estimates are equal to the true relevances:

$$\forall(d, q), \mathbb{E}[\hat{R}_{d|q}] = R_{d|q}. \quad (13)$$

Our definition is slightly different than the standard definition in the previous literature: where existing work has focussed on the estimated performance of a ranking system (Eq. 4) [1, 33, 41], we focus on bias of the individual relevance estimates. We motivate this difference with three arguments: (i) Our definition based on individual relevance estimates practically guarantees an unbiased performance estimate (c.f. Eq. 3); (ii) To the best of our knowledge, there is no existing work that provides an unbiased performance estimate without also providing unbiased relevance estimates [1, 5, 27, 33, 41, 44, 46, 64, 66, 68, 71]. In other words, all existing unbiased LTR methods are also unbiased under our definition. (iii) Our definition is simpler and easier to work with. To summarize, we define unbiasedness based on individual relevance estimates because it is easier to work with and is interchangeable with the unbiased definition based on system performance in previous work.

While unbiasedness has been the main focus of most click-based LTR work, it is not the only property that is important to the field. As Section 3.3 noted, in practice counterfactual estimation methods are often deployed in a biased manner. In particular, propensity clipping introduces some bias but reduces variance greatly and is widely applied in counterfactual LTR [2, 33, 44, 46, 66]. Similarly, it is conceptually hard to apply the unbiasedness property to click modelling methods [65]. For these reasons, we will also consider *consistency* [45] with the following formal definition:

Definition 4.2. Consistency. A click-based relevance estimation method is *consistent*, if the values of all the resulting relevance estimates are equal to the true relevances in the limit of an infinite number of interactions:

$$\lim_{|\mathcal{D}_q| \rightarrow \infty} \hat{R}_{d|q} = R_{d|q}. \quad (14)$$

Consistency is a desirable property since it guarantees accurate convergence as interaction data continues to increase. Furthermore, variance is generally negligible when the number of interaction is extremely large [33], therefore unlike unbiasedness, there is often not a trade-off between variance and consistency.

Our goal is thus to identify under which conditions the unbiased LTR approach can provide methods that are unbiased and consistent according to our definitions. Our method to identifying these conditions inverts the prevalent approach that starts with behavior assumptions and derives an unbiased method. In contrast, we first describe the two main families: counterfactual estimation and click modelling, in the most generic terms. Subsequently, we derive click behavior conditions from these generic methods, thereby revealing the assumptions that are implicitly present in the existing approach. The following two sections will do this for counterfactual estimation and click modelling, in addition, Section 7 will discuss other methods that fall outside the former two categories.

5 THE LIMITATIONS OF CLICK-BASED COUNTERFACTUAL ESTIMATORS

As discussed in Section 3.1, counterfactual estimation represents the largest branch of the unbiased LTR field [33, 41, 68]. The underlying assumption of these methods is that click probabilities are decomposable into relevance and display factors, e.g. the estimator in Eq. 6 assumes click probabilities are determined by relevance and display-position according to Eq. 5 [66]. Click-based counterfactual

estimators aim to convert a click signal into an unbiased relevance signal by correcting for the display factors. In order to investigate the entire family of click-based counterfactual estimators, we first define a display context $x(d)$ in the most generic and broad terms:

Definition 5.1. Display Context. The display context $x_i(d)$ contains all information about how item d is displayed that could affect the click probability of d at interaction i . It does not contain any information about the relevance d : $R_{d|q}$. In other words, one should be able to determine $x_i(d)$ without any knowledge of $R_{d|q}$.

For example, for the estimator in Eq. 6 the display context is the display position: $x_i(d) = k_i(d)$ [2]. Many alternatives are possible, e.g. $x_i(d)$ can also represent a probability distribution over positions as in the policy-aware estimator [44, 46]. With this broad definition of the display context, we propose a generic description of a click-based counterfactual estimator:

Definition 5.2. Counterfactual Relevance Estimate. A click-based counterfactual relevance estimate is an average over independently sampled interactions where each click or non-click is transformed by a function f such that:

$$\hat{R}_{d|q} = \frac{1}{N_q} \sum_{i \in \mathcal{D}_q} f(c_i(d), x_i(d)). \quad (15)$$

Accordingly, f only has two relevant values per context $x(d)$ for the counterfactual estimate: $f(1, x(d))$ when a click takes place on d and $f(0, x(d))$ when no click takes place.

To the best of our knowledge, Definition 5.2 covers all existing click-based counterfactual estimators [1, 2, 5, 33, 41, 44, 46, 64, 66, 68, 71] with the exception of three counterfactual pairwise estimators [27, 57, 67] that will be discussed in Section 7. For example, we see that the estimator in Eq. 6 is a specific instance where f is chosen so that $f(1, x_i(d)) = \frac{1 - \beta_{k_i(d)}}{\alpha_{k_i(d)}}$ and $f(0, x_i(d)) = \frac{-\beta_{k_i(d)}}{\alpha_{k_i(d)}}$. The aim of our definition is to cover all existing estimators and as many future estimators as possible. Accordingly, - to the best of our knowledge - different choices of f can cover almost all existing counterfactual estimators, and moreover, they can also cover many more possible estimators that have yet to be introduced to the field.

With these definitions, we can now start to derive the conditions for which a click-based counterfactual estimator is unbiased or consistent. First, we note that the expected value of a relevance estimate $\hat{R}_{d|q}$ is simply the expected value of f conditioned on q :

LEMMA 5.3. *In expectation a counterfactual relevance estimate is:*

$$\mathbb{E}_{c,x}[\hat{R}_{d|q}] = \mathbb{E}_{c,x}[f(c(d), x(d)) | q]. \quad (16)$$

PROOF. Using Definition 5.2:

$$\begin{aligned} \mathbb{E}_{c,x}[\hat{R}_{d|q}] &= \mathbb{E}_{c,x}\left[\frac{1}{N_q} \sum_{i \in \mathcal{D}_q} f(c_i(d), x_i(d))\right] \\ &= \mathbb{E}_{c,x}[f(c_i(d), x_i(d)) | i \in \mathcal{D}_q] = \mathbb{E}_{c,x}[f(c(d), x(d)) | q]. \quad \square \end{aligned} \quad (17)$$

With this Lemma, we can prove that an unbiased estimator is always consistent and vice-versa:

THEOREM 5.4. *A counterfactual estimator is consistent if and only if it is unbiased:*

$$\mathbb{E}_{c,x}[\hat{R}_{d|q}] = R_{d|q} \iff \lim_{|\mathcal{D}_q| \rightarrow \infty} \hat{R}_{d|q} = R_{d|q}. \quad (18)$$

PROOF. With the use of Definition 5.2 and Lemma 5.3:

$$\lim_{|\mathcal{D}_q| \rightarrow \infty} \hat{R}_{d|q} = \mathbb{E}_{c,x}[f(c_i(d), x_i(d)) | q] = \mathbb{E}_{c,x}[\hat{R}_{d|q}]. \quad (19)$$

Therefore, it appears that the focus of previous work to prove unbiasedness was not incorrectly placed, since implicitly this has also proved consistency. Nevertheless, we think it is important that we have a theoretical motivation for this focus now. Finally, we can derive the following conditions for the unbiasedness - and therefore also the consistency - of click-based counterfactual estimators:

THEOREM 5.5. *A counterfactual estimator can only be unbiased if click probabilities follow an affine transformation of relevance s.t:*

$$R_{d|q} = \mathbb{E}_x\left[\frac{P(C = 1 | d, x, q) - \beta_{x(d)}}{\alpha_{x(d)}} | q\right]. \quad (20)$$

PROOF. From Lemma 5.3 it follows that the unbiasedness criteria for a counterfactual relevance estimate can be reformulated as:

$$\mathbb{E}_{c,x}[\hat{R}_{d|q}] = R_{d|q} \iff \mathbb{E}_{c,x}[f(c(d), x(d)) | q] = R_{d|q}. \quad (21)$$

The expected value can be rewritten to:

$$\begin{aligned} \mathbb{E}_{c,x}[f(c(d), x(d)) | q] &= \mathbb{E}_x[P(C = 1 | d, x, q)f(1, x(d)) \\ &\quad + (1 - P(C = 1 | d, x, q))f(0, x(d)) | q] \\ &= \mathbb{E}_x[P(C = 1 | d, x, q)(f(1, x(d)) - f(0, x(d))) + f(0, x(d)) | q]. \end{aligned} \quad (22)$$

Combining Eq. 21 and 22 reveals that the unbiasedness criteria also has implications on the click probability:

$$\mathbb{E}_{c,x}[\hat{R}_{d|q}] = R_{d|q} \iff \mathbb{E}_x[P(C = 1 | d, x, q)(f(1, x(d)) - f(0, x(d))) + f(0, x(d)) | q] = R_{d|q}. \quad (23)$$

To prove Theorem 5.5, we derive the following $\alpha_{x(d)}$ and $\beta_{x(d)}$ from Eq. 23 which show that click probability is an affine transformation of the form stated in Eq. 20:

$$\begin{aligned} \alpha_{x(d)} &= (f(1, x(d)) - f(0, x(d)))^{-1}, \\ \beta_{x(d)} &= -f(0, x(d))(f(1, x(d)) - f(0, x(d)))^{-1}. \quad \square \end{aligned} \quad (24)$$

Theorem 5.5 can be difficult to interpret since the expectation over the display context in Eq. 20 depends on the choice of logging policy. Nonetheless, Theorem 5.5 clearly shows that - in spite of being broad and generic - Definition 5.2 implicitly limits the click models counterfactual relevance estimation can be unbiased or consistent for. It appears that this is a result of the estimate being a mean over individually transformed clicks (Eq. 15) which makes its expected value a linear interpolation between the possible $f(0, x(d))$ and $f(1, x(d))$ values. Consequently, an counterfactual estimator following Definition 5.2 can only be unbiased or consistent if clicks follow a transformation that can be corrected by such an interpolation, Theorem 5.5 proves that such a transformation has to match the affine form of Eq. 20.

To better understand what Theorem 5.5 entails, we consider what it would mean for a deterministic logging policy:

COROLLARY 5.6. *If an item is displayed without randomization then a counterfactual estimator can only be unbiased if the item's click probability is an affine transformation of relevance in the form:*

$$P(x(d) | q) = 1 \rightarrow P(C = 1 | d, x, q) = \alpha_{x(d)}R_{d|q} + \beta_{x(d)}. \quad (25)$$

PROOF. This follows directly from Theorem 5.5. \square

Furthermore, often practitioners have no control over the exact logging policy that gathers data. In such a scenario, unbiasedness guarantees cannot rely on a specific choice of logging policy as the deployment of that policy may not be possible:

COROLLARY 5.7. *Without control over the logging policy, unbiasedness guarantees for a counterfactual estimator are only possible if click probabilities are affine transformations of relevance in the form:*

$$\forall (d, x, q), P(C = 1 | d, x, q) = \alpha_{x(d)} R_{d|q} + \beta_{x(d)}. \quad (26)$$

PROOF. We cannot rule out the possibility that for each possible display context an item exists which the logging policy will only display in that context, thus, Cor. 5.7 follows from Cor. 5.6. \square

While the individual steps of our theoretical derivation are very straightforward, the result is quite significant: we have proven that the form of click-based counterfactual estimators entail implicit assumptions about the click behavior they are able to unbias. In other words, because click-based counterfactual estimators transform individual click or non-click signals and then average the result, they can only correct for click probabilities that are affine transformations of relevances. Conversely, this proves that no unbiased counterfactual relevance estimator (Def. 5.2) can exist for click behavior that does not match the affine transformation of Eq. 20.

Furthermore, there is a less obvious implicit requirement that $\alpha_{x(d)}$ and $\beta_{x(d)}$ should only depend on the display context and not on the item relevance:

COROLLARY 5.8. *If two items are always displayed in the same single display context, their $\alpha_{x(d)}$ and $\beta_{x(d)}$ values should be equal:*

$$\begin{aligned} (\exists x', P(x(d) = x' | q) = 1 \wedge P(x(d') = x' | q) = 1) \\ \longrightarrow P(\alpha_{x(d)} = \alpha_{x(d')} \wedge \beta_{x(d)} = \beta_{x(d')} | q) = 1. \end{aligned} \quad (27)$$

PROOF. According to Definition 5.1, $x(d)$ is not determined by the relevance of d . Therefore, two items with the same display context $x(d)$ should also have the same values from f :

$$\begin{aligned} (x(d) = x(d')) \leftrightarrow (f(0, x(d)) = f(0, x(d')) \\ \wedge f(1, x(d)) = f(1, x(d'))). \end{aligned} \quad (28)$$

Cor. 5.6 and Eq. 24 show $\alpha_{x(d)}$ and $\beta_{x(d)}$ must then be equal. \square

This requirement becomes more obvious when one considers the implications of its negation: a setting where the value of $\alpha_{x(d)}$ and $\beta_{x(d)}$ are determined by $R_{d|q}$. In such a setting the value of $f(0, x(d))$ and $f(1, x(d))$ are also determined by $R_{d|q}$, and therefore, one should first know the relevance of an item before being able to unbiasedly estimate it. In other words, in such a setting one can only unbiasedly estimate relevances if one already knows those relevances and thus has no need for estimation. Importantly, we are not arguing that such settings do not exist but merely that unbiased counterfactual estimation is infeasible in such cases.

In addition to the requirement stated in Theorem 5.5 and Corollaries 5.6, 5.7 and 5.8, there could be more requirements that have yet to be proven. In particular, our analysis has not considered under what conditions the values of f or α and β can be determined. Previous work on position-bias estimation [3, 15, 18, 69] shows that bias estimation is far from a trivial task. Therefore, it seems

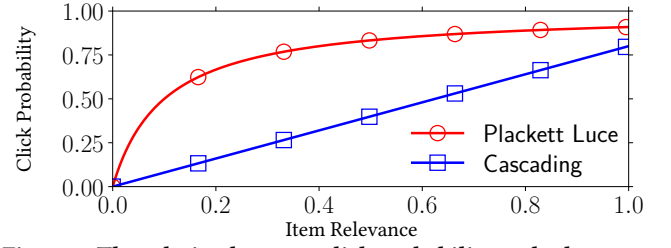


Figure 1: The relation between click probability and relevance in the Plackett-Luce and cascading scenarios of Section 5.1.

reasonable that it is also necessary that the bias parameters α and β can be accurately inferred for unbiased counterfactual estimation. However, we currently lack a theoretical approach to prove broad conditions about bias estimation in general (Section 6.2 will discuss bias estimation with click modelling).

To summarize our theoretical findings on counterfactual estimation for click-based LTR: We have provided a generic definition of click-based counterfactual estimators that captures almost all methods in the field [1, 2, 5, 33, 41, 44, 46, 64, 66, 68, 71]. From this definition, we have shown that estimators of this form can only provide unbiased and consistent estimates when click probabilities are affine transformations of relevances. We have thus identified a significant limitation of the high-level approach for click-based counterfactual estimation: it can never find unbiased or consistent methods for non-affine click behavior.

The remainder of this section will provide some example scenarios that illustrate this limitation and discuss some edge-cases related to our definitions.

5.1 Example Scenarios

To better understand what the implicit limitations of counterfactual estimation regarding affine click-behavior entail, we will briefly discuss two models of click behavior that illustrate scenarios where unbiased estimation is possible and impossible.

To start, we will consider a model of behavior where a user first considers all items in a ranking and then chooses according to a Plackett-Luce decision model:

$$P_{\text{plackett-luce}}(c(d) = 1 | y, q) = \frac{\mathbb{1}[d \in y] \cdot R_{d|q}}{\sum_{d' \in y} R_{d'|q}}. \quad (29)$$

Thus the probability of a user choosing an item in a ranking is equal to its relevance divided by the sum of the relevances of all items in the ranking. This Plackett-Luce click model is based on well-established decision models from the economic field [37, 51]. To keep our example simple, we have not added any form of position-bias or trust-bias in the model, although such extensions are certainly possible.

At first glance the Plackett-Luce click model may incorrectly seem to fall within the limitations of counterfactual estimation with the following parameters: $\alpha_{x(d)} = \mathbb{1}[d \in y] / \sum_{d' \in y} R_{d'|q}$ and $\beta_{x(d)} = 0$ (cf. Eq. 25). However, the denominator of this $\alpha_{x(d)}$ is directly determined by the relevances of *all* items in the ranking, thus also by the item for which relevance is estimated itself. It therefore does not match the definition of an affine click-model. This can be clearly seen when plotting how the click probability changes with an item's relevance; Figure 1 displays this for a scenario where

the other relevances that sum up to 0.01: $\sum_{d' \in y/d} R_{d'|q} = 0.01$. Clearly, there is not an affine transformation from relevance to click probabilities, and consequently, it is impossible for a click-based counterfactual estimator according to Definition 5.2 to provide an unbiased or consistent estimate in this scenario.

This does not mean that unbiased counterfactual estimation is impossible for all click models where click probabilities depend on the relevances of other items. For example, let us consider a classic cascading click model where a user considers one item at a time until one is clicked [15, 18]. The click probability is thus:

$$P_{\text{casc}}(c(d) = 1 \mid d, y, q) \\ = P_{\text{casc}}(\forall d' \in y, k(d') < k(d) \rightarrow c(d') = 0 \mid y, q) \cdot R_{d|q}. \quad (30)$$

Therefore, this scenario does provide an affine transformation with $\alpha_{x(d)} = P_{\text{casc}}(\forall d' \in y, k(d') < k(d) \rightarrow c(d') = 0 \mid y, q)$ – the probability that no earlier item was clicked – and $\beta_{x(d)} = 0$ (cf. Eq. 25). Figure 1 also shows how the click probability changes with relevance in a scenario where $\alpha_{x(d)} = 0.7$, which clearly reveals an affine transformation. The crucial difference in this scenario is that the value of $\alpha_{x(d)}$ only depends on other items and not on d itself.

These examples illustrate an intuitive way to think of Theorem 5.5: unbiasedness and consistency guarantees for counterfactual estimation are only possible if there is a linear relation between click probability and item relevance.

5.2 Adaptive Normalization and Clipping

While our definition of counterfactual relevance estimates (Definition 5.2) covers almost all click-based counterfactual estimators, it does not fit perfectly with adaptive self-normalization [63] and clipping strategies that change with $|\mathcal{D}_q|$ [44, 46]. For instance, a popular clipping strategy is one where $\lim_{|\mathcal{D}_q| \rightarrow \infty} \tau = 0$, e.g. with $\tau \propto 1/\sqrt{|\mathcal{D}_q|}$ [44, 46]. To incorporate these exceptions one could use the function: $f(c_i(d), x_i(d), |\mathcal{D}_q|)$ in Definition 5.2. This change would mean unbiasedness now has to be proven for all possible $|\mathcal{D}_q|$ values, and consistency only requires unbiasedness in the limit, i.e. $\lim_{|\mathcal{D}_q| \rightarrow \infty} \mathbb{E}[\hat{R}_{d|q}] = R_{d|q}$. In spite of these differences, Theorem 5.5 would only need minor changes and the main point would still hold: unbiasedness and consistency are only possible under affine click behavior. We argue that the simplicity of our theoretical analysis with Definition 5.2 and the lack of significant implications of including adaptive strategies, justify our choice of excluding them.

6 THE LIMITATIONS OF CLICK MODELLING

The second branch of click-based LTR is click-modelling [9, 13, 15, 75]; as described in Section 3.2 click modelling methods fit a predictive click model to a dataset of observed user interactions. Part of this fitted model parameters represent relevance, thus the click model relevance estimates \hat{R}_q are variables that are optimized, in contrast with counterfactual estimation where they are the output of the estimator function (see Section 5). We will define a click modelling method by the loss that they optimize:

Definition 6.1. A click model loss is a function \mathcal{L} that takes as input the vector of relevance estimates \hat{R}_q , the estimated latent variables \hat{Z} and the observed data \mathcal{D}_q : $\mathcal{L}(\hat{R}_q, \hat{Z}, \mathcal{D}_q)$. The value of

$\mathcal{L}(\hat{R}_q, \hat{Z}, \mathcal{D}_q)$ indicates the quality of both the relevance and latent estimates where minimizing indicates improvement.

Our definition is intentionally as generic as possible and – to the best of our knowledge – covers virtually all click modelling methods used for click-based LTR [8–10, 13–15, 69, 75]. Usually, there is a model of user behavior based around the latent variables \hat{Z} and relevances \hat{R}_q and the loss represents the predictive power of that model to describe the data \mathcal{D}_q . In traditional click modelling work, this model is generally a graphical model with a clear interpretation [8, 13], e.g. the rank-based position-biased model has a single latent variable per rank representing user examination [15, 18]. More recent work has introduced neural networks for click modelling where the latent variables are the parameters of the networks, these models provide no intuitive interpretation but they have enormous predictive power [9, 10, 14, 75]. Our definition covers both traditional and neural click modelling methods by avoiding assumptions about the underlying predictive model.

While in the context of click models, relevance estimates are parameters to be optimized, a click modelling method still has to produce relevance estimates to be used for LTR. We define the output of a click modelling method as the optimal relevance estimates:

Definition 6.2. The optimal relevance estimate of a click model are any relevance estimates that minimize the click model loss:

$$(\hat{R}_q^*, \hat{Z}^*) = \arg \min_{\hat{R}_q, \hat{Z}} \mathcal{L}(\hat{R}_q, \hat{Z}, \mathcal{D}_q). \quad (31)$$

We note that this definition is intentionally ambiguous in that it allows for the possibility that multiple relevance estimates could minimize the loss. Depending on the exact loss and optimization method, it is both possible that there is a single unique vector of optimal estimate values or that multiple values are optimal.

With these definitions, we can now analyse the consistency and unbiasedness of click modelling methods. To start, we prove the following conditions for consistency:

THEOREM 6.3. A click model estimator is consistent if and only if the only relevance estimates that minimize its loss are the true relevances as $|\mathcal{D}_q|$ tends to infinity:

$$\lim_{|\mathcal{D}_q| \rightarrow \infty} (R_q = \hat{R}_q \leftrightarrow \min_{\hat{Z}} \mathcal{L}(\hat{R}_q, \hat{Z}, \mathcal{D}_q) = \min_{\hat{R}'_q, \hat{Z}} \mathcal{L}(\hat{R}'_q, \hat{Z}, \mathcal{D}_q)). \quad (32)$$

PROOF. From Definition 4.2 and 6.2, we see that the true relevance estimates need to minimize the click loss in order to be a possible optimal estimate:

$$\lim_{|\mathcal{D}_q| \rightarrow \infty} (R_q = \hat{R}_q \rightarrow \min_{\hat{Z}} \mathcal{L}(\hat{R}_q, \hat{Z}, \mathcal{D}_q) = \min_{\hat{R}'_q, \hat{Z}} \mathcal{L}(\hat{R}'_q, \hat{Z}, \mathcal{D}_q)). \quad (33)$$

However, Definition 6.2 also reveals that multiple relevance estimates may be valid. Thus to guarantee that the correct values are estimated, we also have to exclude other possible values that minimize the loss:

$$\lim_{|\mathcal{D}_q| \rightarrow \infty} (R_q = \hat{R}_q \leftarrow \min_{\hat{Z}} \mathcal{L}(\hat{R}_q, \hat{Z}, \mathcal{D}_q) = \min_{\hat{R}'_q, \hat{Z}} \mathcal{L}(\hat{R}'_q, \hat{Z}, \mathcal{D}_q)). \quad (34)$$

Combining Equation 33 and 34 directly proves Theorem 6.3. \square

In other words, Theorem 6.3 proves that a click modelling method is consistent if – in the limit of infinite data – its loss is only minimized by the true relevance values. Moreover, it also proves that

these are the only conditions under which it can be consistent. Importantly, these conditions relate to both the click modelling method, i.e. the loss and the underlying predictive model, and the data collection procedure, since it is both the loss function and the data that determine where minimum values of the loss are. Thus, while Theorem 6.3 provides a limitation that is conceptually straightforward: the true relevance values should provide the only loss minimum; it appears difficult to understand when this condition is met in practice.

Finally, we prove the following unbiasedness condition:

THEOREM 6.4. *A click modelling method is unbiased if and only if the expected value of its optimal relevance estimates are equal to the true relevances:*

$$\mathbb{E}_{\mathcal{D}_q} [\hat{R}_q^*] = R_q \iff \forall d, \mathbb{E}_{\mathcal{D}_q} [\hat{R}_{d|q}^*] = R_{d|q}. \quad (35)$$

PROOF. Follows from the Definition 4.1 and 6.2. \square

Admittedly, the condition proved in Theorem 6.4 is extremely straightforward: the expected value of the estimates that minimize the loss should be equal to the true relevance values. Similar to the consistency conditions of Theorem 6.3, this condition relates to both the loss function of the click model and to the process that gathers the data \mathcal{D}_q . Furthermore, while the unbiasedness condition is also conceptually straightforward, it is again hard to understand in practical terms.

To summarize our theoretical findings on click modelling for click-based LTR: We have provided a very general definition of click modelling methods that covers virtually all existing methods in the field [8–10, 13–15, 69, 75]. From this definition, we derived the condition for consistency: correct relevance estimates need to provide the only minimum of the loss in the limit of infinite data; and the condition for unbiasedness: the expected value of the estimates that minimize the loss need to be equal to the correct relevance estimates. While these conditions are conceptually straightforward, it heavily depends on the exact model and data collection procedure whether they are met. As a result, it seems there is no straightforward general way to determine whether a click modelling method is consistent or unbiased in practice.

On a higher level, it thus appears that click modelling cannot provide robust guarantees on unbiasedness or consistency, but simultaneously, we were unable to find clear and substantial limitations regarding the click behavior they can debias, as we have found for counterfactual estimation. Moreover, our theoretical findings contrast with the recent empirical success of this approach that showed it to be an effective approach for click-based LTR. We thus conclude that the main limitation of click modelling methods is that we currently lack robust theoretical guarantees regarding their unbiasedness or consistency.

6.1 Example Scenario

We found that the conditions we derived for the unbiasedness and consistency of click modelling methods (Theorem 6.4 and 6.3) are conceptually straightforward, but they also heavily depend on the exact model and data collection procedure. To illustrate this dependency, we will consider the following example scenario that consists of two rankings with the following click probabilities:

Items	Ranking 1				Ranking 2			
	A	B	C	D	B	A	D	C
Click Prob.	0.90	0.64	0.40	0.05	0.8	0.72	0.20	0.10

Given all the click probabilities, one could fit the commonly used rank-based position bias model [15, 18] to this scenario: $P(c(d) | d, k, q) = \alpha_k R_{d|q}$ with $\alpha_1 = 1$. Clearly, the result would follow $\alpha_2 = 0.8$, $\alpha_3 = 4 \cdot \alpha_4$, $R_A = 0.9$, $R_B = 0.8$ and $R_C = 2 \cdot R_D$. Thus, while the values for R_A and R_B are determined exactly, R_C and R_D are merely constrained and a multitude of values would fit the data equally well. We can thus conclude that this click model does not provide consistent estimates for the relevances in this scenario.

Our example scenario shows the difficulty in determining whether a click model is consistent. In this scenario, even when all click probabilities and one bias parameter are known, we are unable to determine all document relevances. Importantly, if we had an additional ranking that placed item C or D at one of the first two ranks this would be possible, illustrating consistency depends both on the model and how data is gathered. We have not considered unbiasedness because it requires an extensive amount of assumptions and derivation: including a probabilistic model of what rankings are displayed and how many interactions are logged. Finally, we note that the click model of our example scenario is extremely simple compared to recent neural click models [9, 10, 14, 75], we expect that analysing unbiasedness or consistency conditions for such complex models is even more infeasible.

6.2 Bias Estimation With Click Modelling

While our discussion has focused on click modelling for relevance estimation, it is also often applied to estimate bias parameters for counterfactual estimation [45, 66, 68, 69]. While a full discussion of this application falls outside the scope of this work, we note that our methodology in Section 6 can be applied analogously to bias estimation. This would lead to similar a conclusion: for unbiased and consistent estimation of the bias parameters, only the correct vector of values should minimize the loss in expectation and in the limit. Correspondingly, in spite of its empirical success, it thus seems that click modelling currently lacks any robust theoretical guarantees for bias estimation.

7 OTHER CLICK-BASED LTR METHODS

We have attempted to cover as many existing click-based LTR methods in our discussion as possible. Nonetheless, we will now briefly discuss several methods that fall outside of our definitions of counterfactual estimation and click modelling.

Firstly, we have not considered the multitude of online LTR methods [25, 55, 60, 74], because several earlier works have already determined they are not able to provide theoretical guarantees w.r.t. unbiasedness [6, 41, 43, 46]. Secondly, Ovaisi et al. propose using Heckman corrections [24], commonly used in econometrics, to the click-based LTR problem [48, 49]. Unfortunately, this approach fell outside the scope of this paper due it being so different from the rest of the click-based LTR field. Lastly, there are several click-based pairwise LTR works: some were excluded because they do not aim to be unbiased [31, 67]. A notable exception is the pairwise method by Saito [57] which does not match our LTR problem setting, but does rely on counterfactual relevance estimation, making our analysis in

Section 5 still applicable. Finally, there is the *Unbiased LambdaMart* method by Hu et al. [27] which is unbiased but not due to its method, the next section will address this interesting case separately.

7.1 Unbiased LambdaMart: Trivially Unbiased

Hu et al. [27] propose *Pairwise Debiasing* that aims to estimate a pairwise loss, based on the assumption that per rank there is a static ratio between click probabilities and relevances across queries:

Definition 7.1. Pairwise Debiasing Assumption. Per position k there are two ratios $t_k^+ \in \mathbb{R}_{>0}$ and $t_k^- \in \mathbb{R}_{>0}$ between click and non-click probabilities and relevances when any item is displayed at position k :

$$P_{\text{pair}}(C = 1 \mid d, k, q) = t_k^+ R_{d|q}, \quad (36)$$

$$P_{\text{pair}}(C = 0 \mid d, k, q) = t_k^- (1 - R_{d|q}). \quad (37)$$

Ai et al. [6] pointed out that since Eq. 36 is equivalent to the popular rank-based position-bias model [15, 18, 33, 69], the additional Eq. 37 seems to conflict with that model. We will now prove that the pairwise debiasing assumption actually implicitly assumes that there is no bias at all in non-trivial ranking circumstances:

THEOREM 7.2. *If two items d_1 and d_2 are both displayed at position k for two different queries q_1 and q_2 and they have different relevances s.t. $R_{d_1|q_1} \neq R_{d_2|q_2}$ then under the pairwise debiasing assumption click probabilities at k are equal to the corresponding relevances:*

$$P_{\text{pair}}(C = 1 \mid d, k, q) = R_{d|q}, \quad (38)$$

$$P_{\text{pair}}(C = 0 \mid d, k, q) = 1 - R_{d|q}. \quad (39)$$

PROOF. Because the click probability (Eq. 36) and non-click probability (Eq. 37) have to sum to one, the following must hold:

$$t_k^+ R_{d_1|q_1} = 1 - t_k^- (1 - R_{d_1|q_1}). \quad (40)$$

From this we can express t_k^+ in terms of t_k^- and $R_{d_1|q_1}$, similarly, this can also be done with t_k^- and $R_{d_2|q_2}$, therefore:

$$t_k^+ = \frac{1 - t_k^- (1 - R_{d_1|q_1})}{R_{d_1|q_1}} = \frac{1 - t_k^- (1 - R_{d_2|q_2})}{R_{d_2|q_2}}. \quad (41)$$

From the latter part of that equality, we can derive:

$$t_k^- (R_{d_1|q_1} - R_{d_2|q_2}) = R_{d_1|q_1} - R_{d_2|q_2}. \quad (42)$$

Since $R_{d_1|q_1} \neq R_{d_2|q_2}$: $t_k^- = 1$ proving Eq. 39, in turn, since Eq. 36 and Eq. 37 sum to one: $t_k^+ = 1$ thus also proving Eq. 38. \square

Without discussing the details of Hu et al.'s *Unbiased LambdaMart* method [27], we can already expect it to be unbiased, since its starting assumption already implicitly entails that clicks are unbiased w.r.t. relevance. Therefore, its unbiasedness is completely trivial from a theoretical perspective, i.e. rankings according to CTRs are also unbiased under the pairwise debiasing assumption.

We would like to point out that there may still be practical value in pairwise debiasing, i.e. Hu et al. [27] report promising real-world performance improvements. For our discussion regarding implicit limitations, it provides an illustrative example of why one should be critical when choosing or evaluating assumptions and the importance of investigating implicit limitations: for they can make the subsequent theoretical findings trivial and inconsequential.

8 DISCUSSION AND CONCLUSION: THE FUTURE OF CLICK-BASED LTR

We have critically analyzed the theoretical foundations of the main branches of unbiased LTR for implicit assumptions and the limitations they entail, and came to the following conclusions: (i) Counterfactual estimation can only unbiasedly and consistently estimate relevance when click probabilities follow affine transformations of relevances. (ii) The conditions for click modelling methods to be unbiased and consistent do not translate to robust theoretical guarantees; we currently lack strong guarantees for complex models, i.e. the best performing neural networks [9, 75]. (iii) Assumptions that assert constant ratios between clicks and relevance as well as non-clicks and non-relevance [27] implicitly assume that clicks are unbiased w.r.t. relevance. Our findings have revealed significant limitations of the unbiased LTR approach: (i) There is a clear limit on the click behaviors that the current counterfactual estimation approach could possibly correct for. (ii) The theoretical assumptions of pairwise debiasing [27] are not applicable to biased clicks. (iii) The current unbiased LTR theory lacks the ability to properly analyze the unbiasedness and consistency of click modelling methods. These limitations do not undermine the value of the unbiased LTR field: its usefulness and effectiveness is very evident by a multitude empirical results [2, 33, 68, 69, 75, 75], but could help guide future directions and provide valuable lessons to the field.

The implicit limitations that we have uncovered for the existing approach, reveal that in order to find unbiased counterfactual estimation methods that are unbiased and consistent w.r.t. non-affine click behavior, future research should differentiate from our generic definition (Definition 5.2). Concretely, such methods should not rely on averaging over transformed individual clicks, and as a result, proving unbiasedness would be much more difficult for them. Furthermore, novel theoretical research is needed for finding theoretical guarantees of unbiasedness and consistency for click modelling methods, in particular, those that utilize complex models. Our findings reveal that it is not enough that the underlying click model of a method matches the real-world click behavior; whether the method will produce unbiased or consistent relevance estimates also depends on how data is gathered, and ultimately, where the minima of the resulting loss are. Similarly, theoretical research that explores in what circumstances bias estimation is feasible would be very valuable for further understanding the theoretical guarantees of counterfactual estimation. Lastly, future research that introduces novel assumptions should critically investigate what these assumptions implicitly entail. One should actively avoid assumptions that - intentionally or incidentally - make learning from clicks a trivial problem.

Finally, the findings of our critical analysis also provide some important lessons for the field: Firstly, we should realize that unbiasedness is not always possible. Accordingly, we should thus not invariably expect nor require it from future work, for this could systematically exclude research that tackles novel problem settings (e.g. non-affine click behavior). Secondly, since unbiasedness may not be a realistic long term goal, the field will likely shift to bias mitigation or partial debiasing as feasible future directions. Correspondingly, we recommend replacing the term *unbiased LTR* with the less-demanding *debiased LTR* or the neutral *click-based LTR*.

ACKNOWLEDGMENTS

This work was supported by the Google Research Scholar Program. All content represents the opinion of the author, which is not necessarily shared or endorsed by their employers and/or sponsors.

REFERENCES

- [1] Aman Agarwal, Kenta Takatsu, Ivan Zaitsev, and Thorsten Joachims. 2019. A General Framework for Counterfactual Learning-to-Rank. In *Proceedings of the 42nd International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 5–14.
- [2] Aman Agarwal, Xuanhui Wang, Cheng Li, Michael Bendersky, and Marc Najork. 2019. Addressing Trust Bias for Unbiased Learning-to-Rank. In *The World Wide Web Conference*. ACM, 4–14.
- [3] Aman Agarwal, Ivan Zaitsev, Xuanhui Wang, Cheng Li, Marc Najork, and Thorsten Joachims. 2019. Estimating Position Bias without Intrusive Interventions. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 474–482.
- [4] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. 2006. Learning User Interaction Models for Predicting Web Search Result Preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 3–10.
- [5] Qingyao Ai, Keping Bi, Cheng Luo, Jiafeng Guo, and W Bruce Croft. 2018. Unbiased Learning to Rank with Unbiased Propensity Estimation. In *Proceedings of the 41st International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 385–394.
- [6] Qingyao Ai, Tao Yang, Huazheng Wang, and Jiaxin Mao. 2021. Unbiased Learning to Rank: Online or Offline? *ACM Transactions on Information Systems (TOIS)* 39, 2 (2021), 1–29.
- [7] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 405–414.
- [8] Alexey Borisov, Ilya Markov, Maarten de Rijke, and Pavel Serdyukov. 2016. A Context-aware Time Model for Web Search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 205–214.
- [9] Alexey Borisov, Ilya Markov, Maarten De Rijke, and Pavel Serdyukov. 2016. A Neural Click Model for Web Search. In *Proceedings of the 25th International Conference on World Wide Web*. 531–541.
- [10] Alexey Borisov, Martijn Wardenaar, Ilya Markov, and Maarten de Rijke. 2018. A Click Sequence Model for Web Search. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 45–54.
- [11] Christopher J.C. Burges. 2010. *From RankNet to LambdaRank to LambdaMART: An Overview*. Technical Report MSR-TR-2010-82. Microsoft.
- [12] Olivier Chapelle and Yi Chang. 2011. Yahoo! Learning to Rank Challenge Overview. *Journal of Machine Learning Research* 14 (2011), 1–24.
- [13] Olivier Chapelle and Ya Zhang. 2009. A Dynamic Bayesian Network Click Model for Web Search Ranking. In *Proceedings of the 18th International Conference on World Wide Web*. 1–10.
- [14] Jia Chen, Jiaxin Mao, Yiqun Liu, Min Zhang, and Shaoping Ma. 2020. A Context-Aware Click Model for Web Search. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 88–96.
- [15] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. 2015. *Click Models for Web Search*. Morgan & Claypool Publishers.
- [16] Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and Diversity in Information Retrieval Evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 659–666.
- [17] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM conference on recommender systems*. 191–198.
- [18] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An Experimental Comparison of Click Position-bias Models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 87–94.
- [19] Van Dang and W Bruce Croft. 2012. Diversity by proportionality: an election-based approach to search result diversification. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 65–74.
- [20] Domenico Dato, Claudio Lucchese, Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, Nicola Tonellotto, and Rossano Venturini. 2016. Fast Ranking with Additive Ensembles of Oblivious and Non-Oblivious Regression Trees. *ACM Transactions on Information Systems (TOIS)* 35, 2 (2016), Article 15.
- [21] Fernando Diaz, Bhaskar Mitra, Michael D. Ekstrand, Asia J. Biega, and Ben Carterette. 2020. Evaluating Stochastic Rankings with Expected Exposure. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery, New York, NY, USA, 275–284.
- [22] Zhichong Fang, Aman Agarwal, and Thorsten Joachims. 2019. Intervention Harvesting for Context-Dependent Examination-Bias Estimation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 825–834.
- [23] Carlos A Gomez-Urbe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems (TMIS)* 6, 4 (2015), 1–19.
- [24] James J Heckman. 1979. Sample Selection Bias as a Specification Error. *Econometrica: Journal of the econometric society* (1979), 153–161.
- [25] Katja Hofmann. 2013. *Fast and Reliably Online Learning to Rank for Information Retrieval*. Ph.D. Dissertation. University of Amsterdam.
- [26] Katja Hofmann, Anne Schuth, Alejandro Bellogin, and Maarten de Rijke. 2014. Effects of Position Bias on Click-based Recommender Evaluation. In *European Conference on Information Retrieval*. Springer, 624–630.
- [27] Ziniu Hu, Yang Wang, Qu Peng, and Hang Li. 2019. Unbiased LambdaMART: An Unbiased Pairwise Learning-to-Rank Algorithm. In *The World Wide Web Conference*. ACM, 2830–2836.
- [28] Neil Hurley and Mi Zhang. 2011. Novelty and Diversity in Top-n Recommendation—Analysis and Evaluation. *ACM Transactions on Internet Technology (TOIT)* 10, 4 (2011), 14.
- [29] Rolf Jagerman, Harrie Oosterhuis, and Maarten de Rijke. 2019. To Model or to Intervene: A Comparison of Counterfactual and Online Learning to Rank from User Interactions. In *Proceedings of the 42nd International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 15–24.
- [30] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated Gain-Based Evaluation of IR Techniques. *ACM Transactions on Information Systems (TOIS)* 20, 4 (2002), 422–446.
- [31] Thorsten Joachims. 2002. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 133–142.
- [32] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2017. Accurately Interpreting Clickthrough Data as Implicit Feedback. In *ACM SIGIR Forum*, Vol. 51. Acm New York, NY, USA, 4–11.
- [33] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 781–789.
- [34] Junpei Komiya, Junya Honda, and Hiroshi Nakagawa. 2015. Optimal Regret Analysis of Thompson Sampling in Stochastic Multi-armed Bandit Problem with Multiple Plays. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (Lille, France) (ICML '15)*. JMLR.org, 1152–1161.
- [35] Paul Lagrée, Claire Vernade, and Olivier Cappé. 2016. Multiple-Play Bandits in the Position-Based Model. In *Advances in Neural Information Processing Systems*. 1597–1605.
- [36] Tie-Yan Liu. 2009. Learning to Rank for Information Retrieval. *Foundations and Trends in Information Retrieval* 3, 3 (2009), 225–331.
- [37] R Duncan Luce. 2012. *Individual Choice Behavior: A Theoretical Analysis*. Courier Corporation.
- [38] Alistair Moffat, Peter Bailey, Falk Scholer, and Paul Thomas. 2017. Incorporating User Expectations and Behavior into the Measurement of Search Effectiveness. *ACM Transactions on Information Systems (TOIS)* 35, 3 (2017), 1–38.
- [39] Alistair Moffat, Paul Thomas, and Falk Scholer. 2013. Users Versus Models: What Observation Tells us about Effectiveness Metrics. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 659–668.
- [40] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. *Controlling Fairness and Bias in Dynamic Learning-to-Rank*. ACM, 429–438.
- [41] Harrie Oosterhuis. 2020. *Learning From User Interactions With Rankings: A Unification of the Field*. Ph.D. Dissertation. University of Amsterdam.
- [42] Harrie Oosterhuis. 2021. Computationally Efficient Optimization of Plackett-Luce Ranking Models for Relevance and Fairness. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (Virtual Event, Canada) (SIGIR '21)*. ACM, 1023–1032.
- [43] Harrie Oosterhuis and Maarten de Rijke. 2019. Optimizing Ranking Models in an Online Setting. In *Advances in Information Retrieval*. Springer International Publishing, Cham, 382–396.
- [44] Harrie Oosterhuis and Maarten de Rijke. 2020. Policy-Aware Unbiased Learning to Rank for Top-k Rankings. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 489–498.
- [45] Harrie Oosterhuis and Maarten de Rijke. 2020. Taking the Counterfactual Online: Efficient and Unbiased Online Evaluation for Ranking. In *Proceedings of the 2020 International Conference on The Theory of Information Retrieval*. ACM.
- [46] Harrie Oosterhuis and Maarten de Rijke. 2021. Unifying Online and Counterfactual Learning to Rank. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. ACM.
- [47] Harrie Oosterhuis, Rolf Jagerman, and Maarten de Rijke. 2020. Unbiased Learning to Rank: Counterfactual and Online Approaches. In *Companion Proceedings of the Web Conference 2020*. 299–300.

- [48] Zohreh Ovaisi, Ragib Ahsan, Yifan Zhang, Kathryn Vasilaky, and Elena Zheleva. 2020. Correcting for Selection Bias in Learning-to-Rank Systems. In *Proceedings of The Web Conference 2020*. 1863–1873.
- [49] Zohreh Ovaisi, Kathryn Vasilaky, and Elena Zheleva. 2021. Propensity-Independent Bias Recovery in Offline Learning-to-Rank Systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1763–1767.
- [50] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of computer-mediated communication* 12, 3 (2007), 801–823.
- [51] Robin L Plackett. 1975. The Analysis of Permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 24, 2 (1975), 193–202.
- [52] Tao Qin and Tie-Yan Liu. 2013. Introducing LETOR 4.0 Datasets. *arXiv preprint arXiv:1306.2597* (2013).
- [53] Filip Radlinski and Susan Dumais. 2006. Improving personalized web search using result diversification. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 691–692.
- [54] Filip Radlinski, Madhu Kurup, and Thorsten Joachims. 2008. How Does Click-through Data Reflect Retrieval Quality?. In *CIKM*. ACM, 43–52.
- [55] Karthik Raman, Thorsten Joachims, Pannaga Shivaswamy, and Tobias Schnabel. 2013. Stable Coactive Learning via Perturbation. In *International Conference on Machine Learning*. 837–845.
- [56] LT Rodrygo, Craig Macdonald, and Iadh Ounis. 2015. Search result diversification. *Foundations and Trends in Information Retrieval* 9, 1 (2015), 1–90.
- [57] Yuta Saito. 2020. Unbiased Pairwise Learning from Biased Implicit Feedback. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 5–12.
- [58] Mark Sanderson. 2010. Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval* 4, 4 (2010), 247–375.
- [59] Mark Sanderson, Monica Lestari Paramita, Paul Clough, and Evangelos Kanoulas. 2010. Do User Preferences and Evaluation Measures Line Up?. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. 555–562.
- [60] Anne Schuth, Harrie Oosterhuis, Shimon Whiteson, and Maarten de Rijke. 2016. Multileave Gradient Descent for Fast Online Learning to Rank. In *WSDM*. ACM, 457–466.
- [61] Daria Sorokina and Erick Cantu-Paz. 2016. Amazon Search: The Joy of Ranking Products. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 459–460.
- [62] Alex Strehl, John Langford, Lihong Li, and Sham M Kakade. 2010. Learning from Logged Implicit Exploration Data. In *Advances in Neural Information Processing Systems*. J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta (Eds.), Vol. 23. Curran Associates, Inc.
- [63] Adith Swaminathan and Thorsten Joachims. 2015. The Self-Normalized Estimator for Counterfactual Learning. In *Advances in Neural Information Processing Systems*. 3231–3239.
- [64] Ali Vardasbi, Maarten de Rijke, and Ilya Markov. 2020. Cascade Model-Based Propensity Estimation for Counterfactual Learning to Rank. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2089–2092.
- [65] Ali Vardasbi, Maarten de Rijke, and Ilya Markov. 2021. Mixture-Based Correction for Position and Trust Bias in Counterfactual Learning to Rank. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 1869–1878.
- [66] Ali Vardasbi, Harrie Oosterhuis, and Maarten de Rijke. 2020. When Inverse Propensity Scoring does not Work: Affine Corrections for Unbiased Learning to Rank. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*.
- [67] Nan Wang, Zhen Qin, Xuanhui Wang, and Hongning Wang. 2021. Non-clicks Mean Irrelevant? Propensity Ratio Scoring as a Correction. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 481–489.
- [68] Xuanhui Wang, Michael Bendersky, Donald Metzler, and Marc Najork. 2016. 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*. ACM, 115–124.
- [69] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 2018. Position Bias Estimation for Unbiased Learning to Rank in Personal Search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 610–618.
- [70] Xuanhui Wang, Cheng Li, Nadav Golbandi, Michael Bendersky, and Marc Najork. 2018. The LambdaLoss Framework for Ranking Metric Optimization. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 1313–1322.
- [71] Xinwei Wu, Hechang Chen, Jiashu Zhao, Li He, Dawei Yin, and Yi Chang. 2021. Unbiased Learning to Rank in Feeds Recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 490–498.
- [72] Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. 2008. Listwise Approach to Learning to Rank: Theory and Algorithm. In *Proceedings of the 25th international conference on Machine learning*. 1192–1199.
- [73] Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, Wei Zeng, and Xueqi Cheng. 2017. Adapting Markov Decision Process for Search Result Diversification. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 535–544.
- [74] Yisong Yue and Thorsten Joachims. 2009. Interactively Optimizing Information Retrieval Systems as a Dueling Bandits Problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 1201–1208.
- [75] Honglei Zhuang, Zhen Qin, Xuanhui Wang, Michael Bendersky, Xinyu Qian, Po Hu, and Dan Chary Chen. 2021. Cross-Positional Attention for Debiasing Clicks. In *Proceedings of the Web Conference 2021*. 788–797.