Unidentified and Confounded?

Understanding Two-Tower Models for Unbiased Learning to Rank

Philipp Hager[1], Onno Zoeter[2], and Maarten de Rijke[1]
1. University of Amsterdam, 2. Booking.com

CONSEQUENCES, RecSys 2025 - September 22, 2025

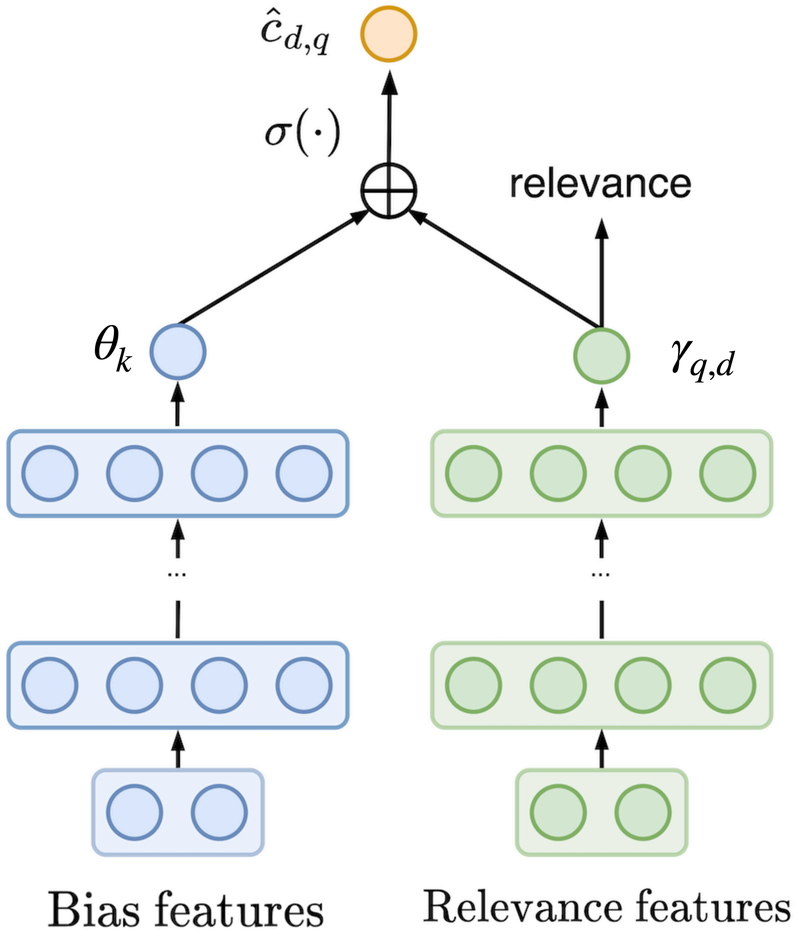
IR Lab AMS

Two-Tower Models for Unbiased LTR

Two-tower models are NNs following the PBM:

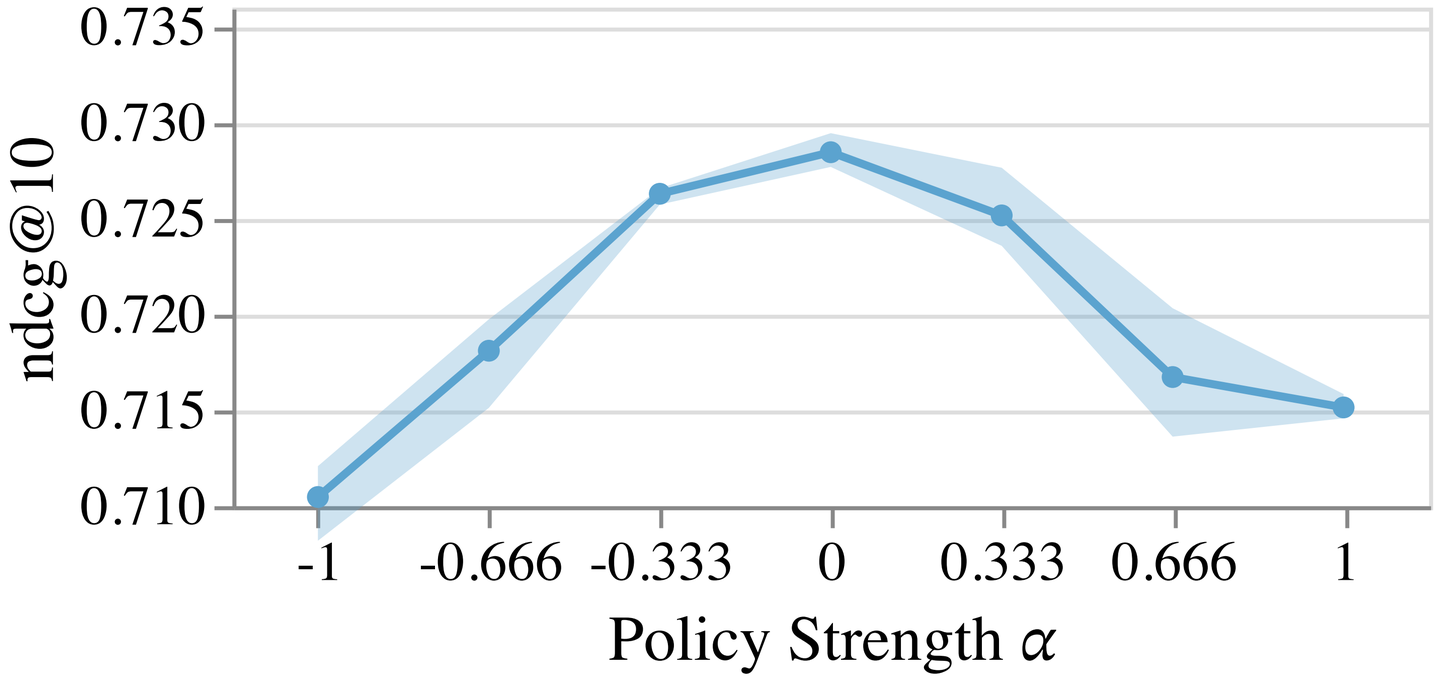
$$P(C = 1 \mid q, d, k) = \sigma(\theta_k + \gamma_{q,d})$$

- **Relevance tower** uses **query-document features:**
  text, bm25, tf-idf, page rank, ...
- **Bias tower** uses **context features:**
  position, device type, item size, dwell time, ...
- Towers are trained to predict click,
  at inference time, only the relevance tower is used.



[1] Le Yan, Zhen Qin, et al. Revisiting two tower models for unbiased learning to rank. In SIGIR 2022.
[2] Xiaoshu Chen et al. Multi-Feature Integration for Perception-Dependent Examination-Bias Estimation. In WSDM 2023.

Trouble with non-random logging policies?



Two-tower models trained on deterministic logging policies of varying strengths (α) on MSLR30K: $\alpha = 1$ sorting by expert annotations, $\alpha = 0$ random sorting, and $\alpha = -1$ ranks from least to most relevant.

[1] Yunan Zhang et al. Towards Disentangling Relevance and Bias in Unbiased Learning to Rank. In KDD 2023.

Hypothesis I

Logging policy performance confounds Two-Tower Models [1]

When bias and relevance are correlated, Two-Tower models struggle to disentangle them.

Intuitive, but theoretically unsatisfying…

Hypothesis II

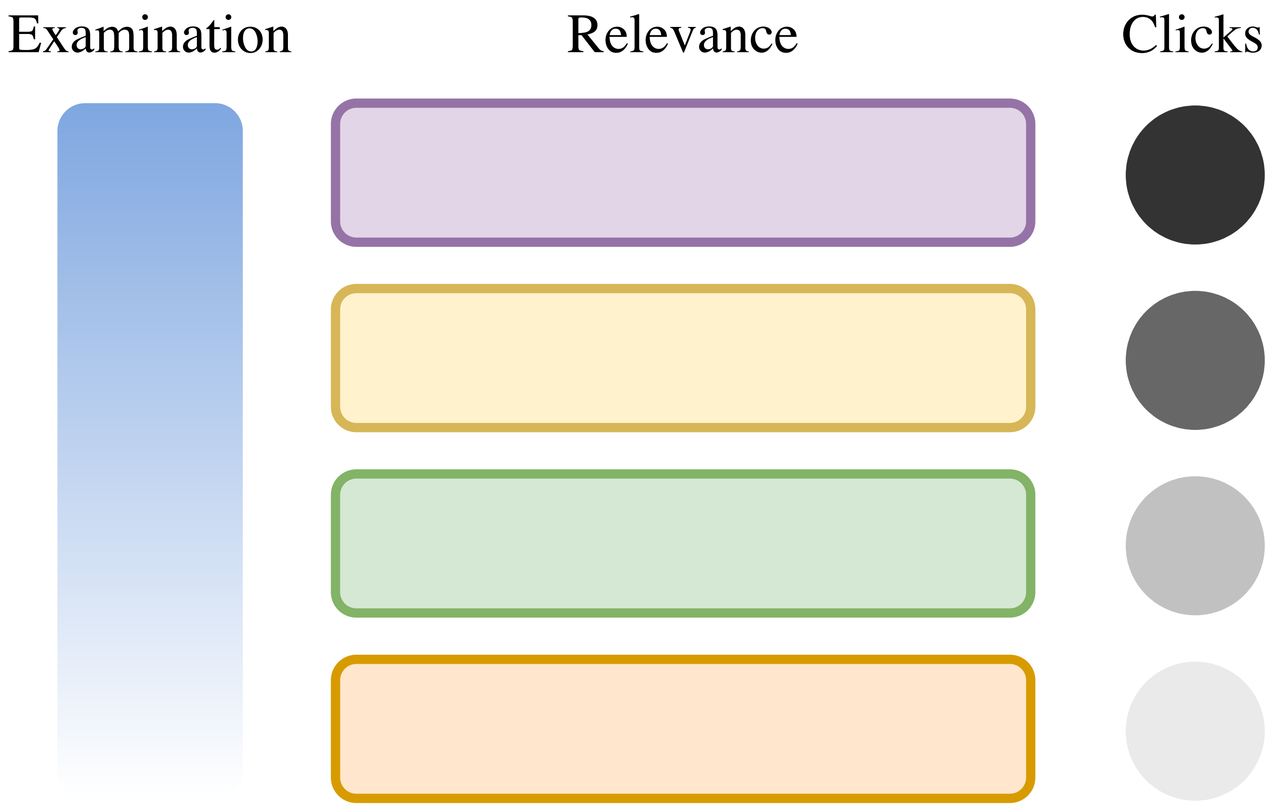
Logging policies impact the identifiability of Two-Tower Models [2]

Logging policies might collect data that is insufficient for recovering the true model parameters.

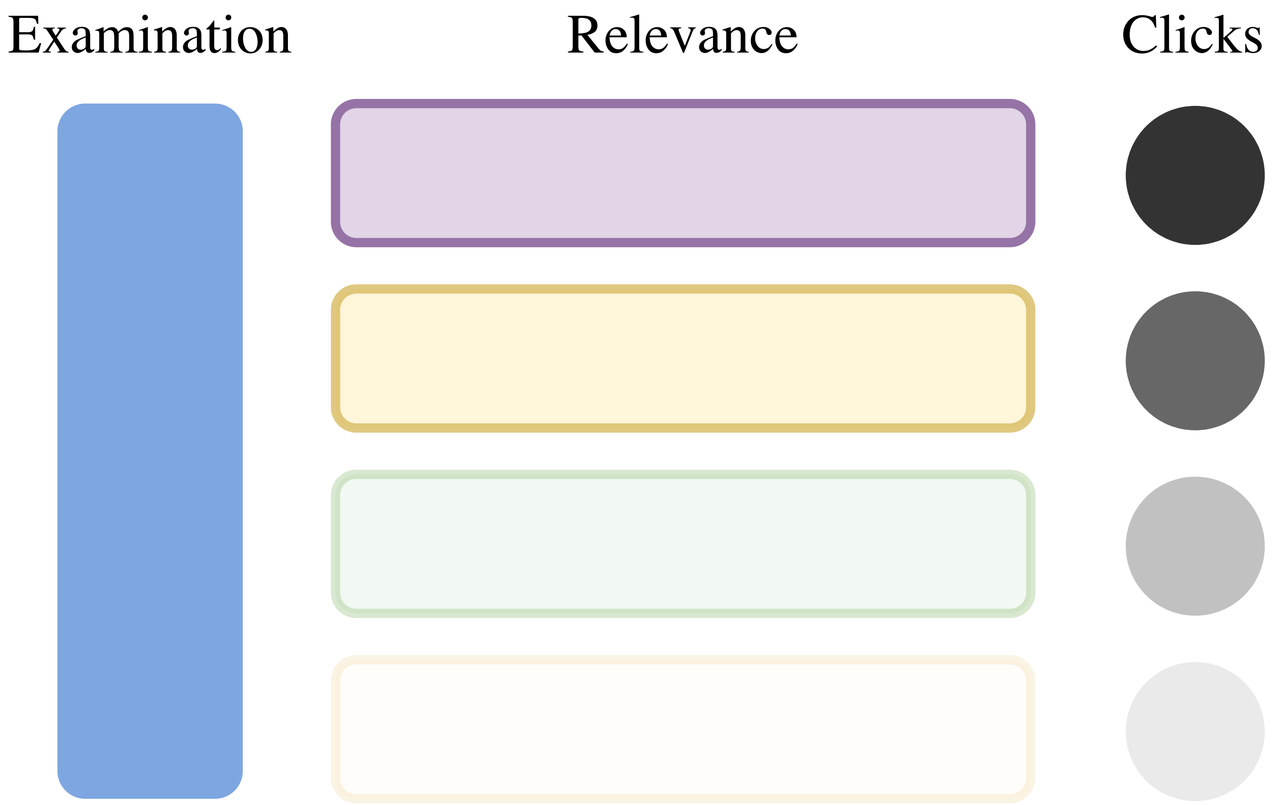
Potential culprit as simulation did not contain any swaps…

[1] Yunan Zhang et al. Towards Disentangling Relevance and Bias in Unbiased Learning to Rank. In KDD 2023.
[2] Mouxiang Chen et al. Identifiability matters: revealing the hidden recoverable condition in unbiased learning to rank. In ICML 2024.

Identification through randomization



Identification through randomization



Identification through overlapping features

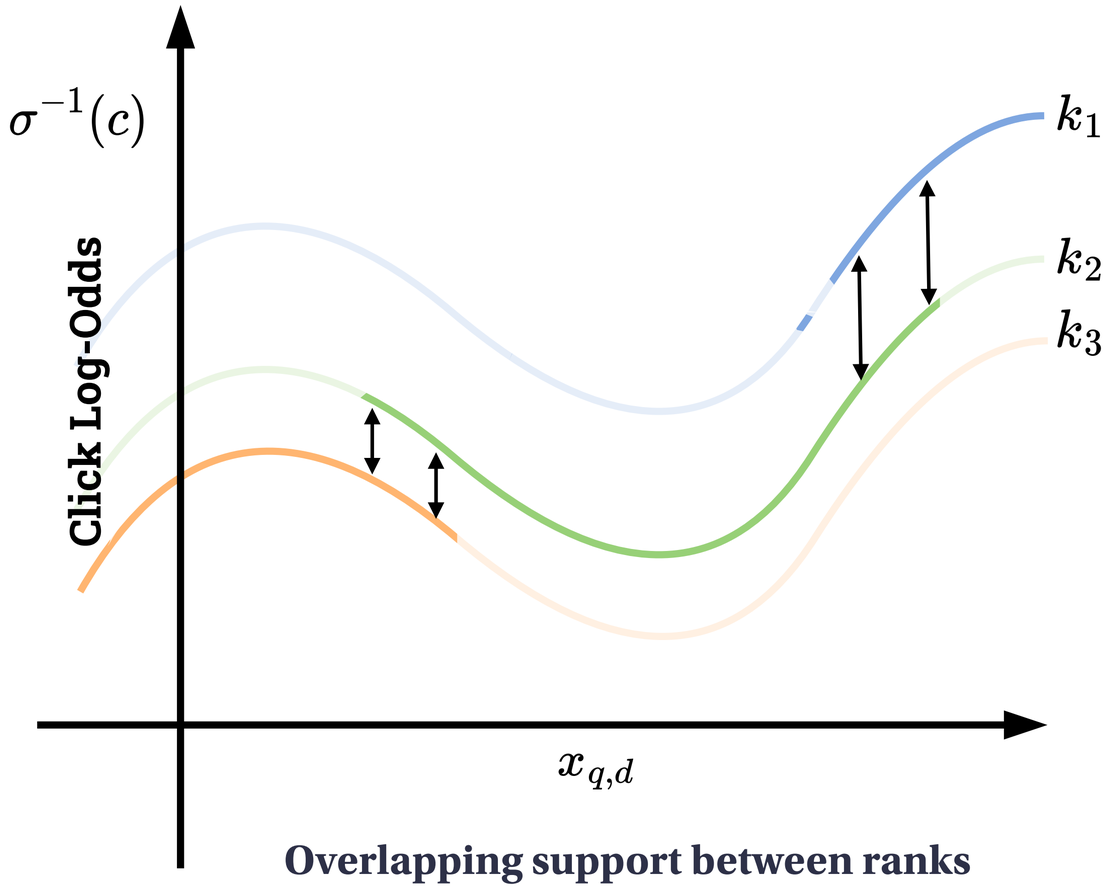
To approx. identify two-tower models:

$$P(C = 1 \mid q, d, k) = \sigma(f(x_{q,d}) + \theta_k),$$

we need **overlapping feature distributions** across ranks*:

$$\text{supp}(P(x \mid k)) \cap \text{supp}(P(x \mid k')) \neq 0,$$

and a **continuous relevance tower.**



Overlapping support between ranks allows identification of bias and relevance

* Such that all ranks are form a connected graph...

Model Misspecification

Let's take a look closer at the derivative w.r.t. a single bias logit:

$$\frac{\partial \mathcal{L}}{\partial \theta_k} = \sum_q P(q) \sum_d \underbrace{\pi(d, k \mid q)}_{\text{Logging Policy}} \Big[\underbrace{P(C=1 \mid q, d, k)}_{\text{Observed CTR}} - \underbrace{\sigma(\theta_k + \gamma_{q,d})}_{\text{Predicted CTR}} \Big] = 0.$$

There is no logging policy influence on an identified and well-specified two-tower model.

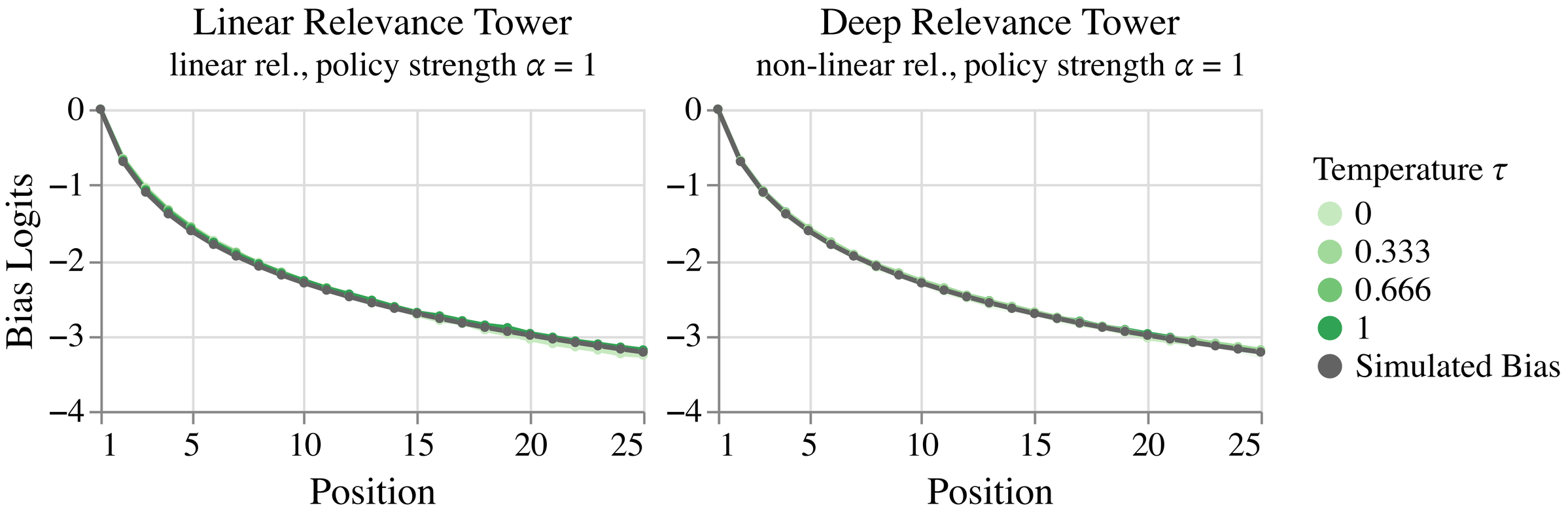
Model Misspecification

Let's take a look closer at the derivative w.r.t. a single bias logit:

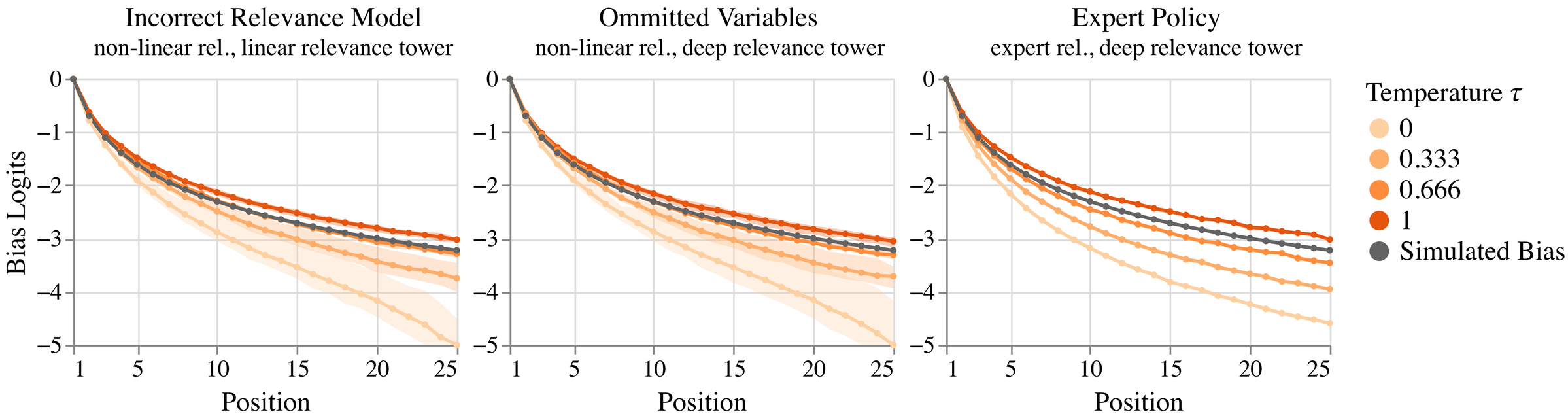
$$\frac{\partial \mathcal{L}}{\partial \theta_k} = \sum_q P(q) \sum_d \underbrace{\pi(d, k \mid q)}_{\text{Logging Policy}} \Big[\underbrace{P(C=1 \mid q, d, k)}_{\text{Observed CTR}} - \underbrace{\sigma(\theta_k + \gamma_{q,d})}_{\text{Predicted CTR}} \Big] = 0.$$

A misspecified model can introduce bias which might be amplified throught the logging policy **when prediction errors correlate with the logging policy.**

Well-specified Models



Model Misspecification



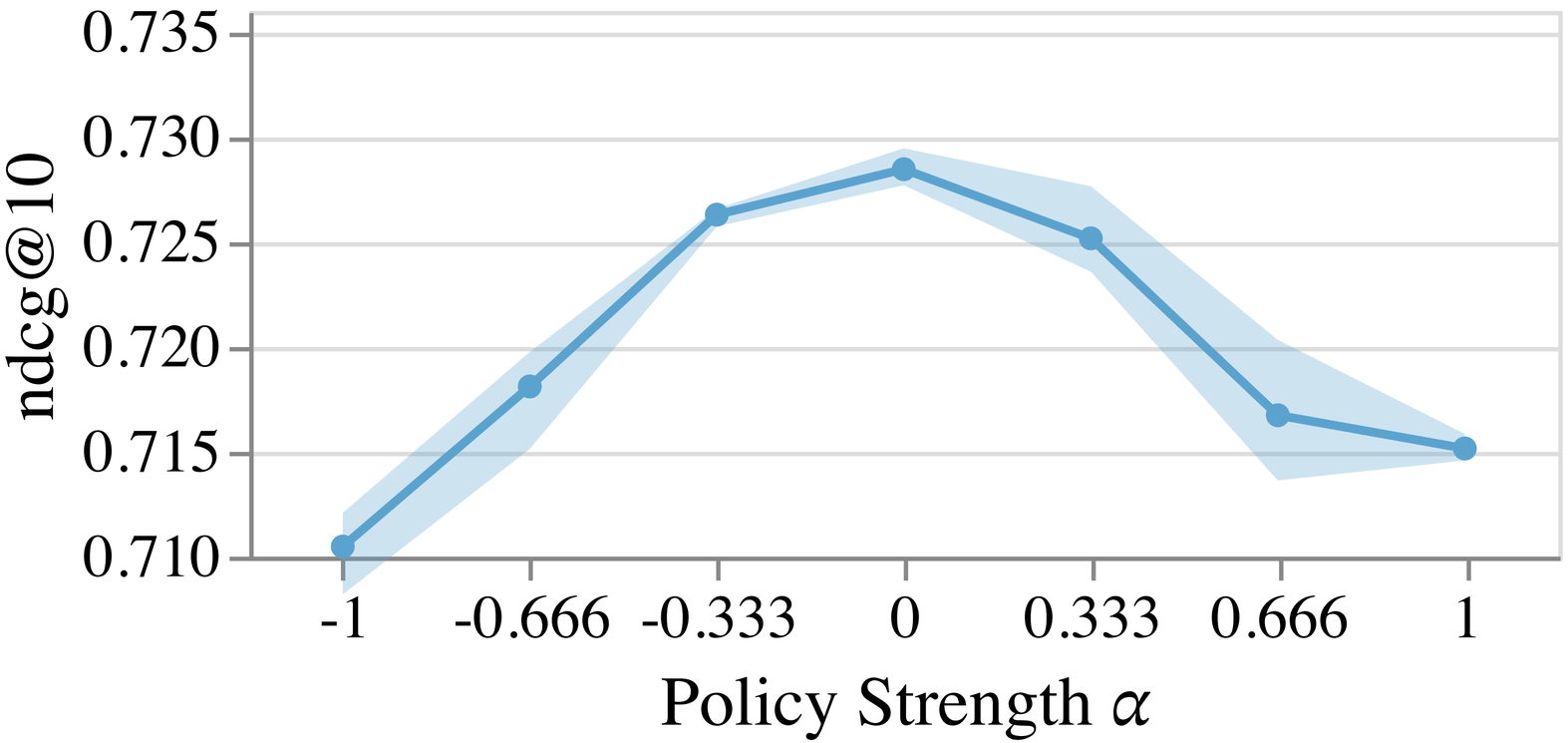
Takeaways

- Two-tower models require **overlapping feature distributions or swaps** across ranks to be identified.
- Logging policies have **no impact on well-specified models** but can amplify bias in misspecified models.
- If model-misspecification cannot be avoided, **consider IPS-based sample weights** (see paper).

For more practical insights,
read the paper and come by our poster!

The observation that two-tower models get worse on non-random logging policies is a simulation artifact…

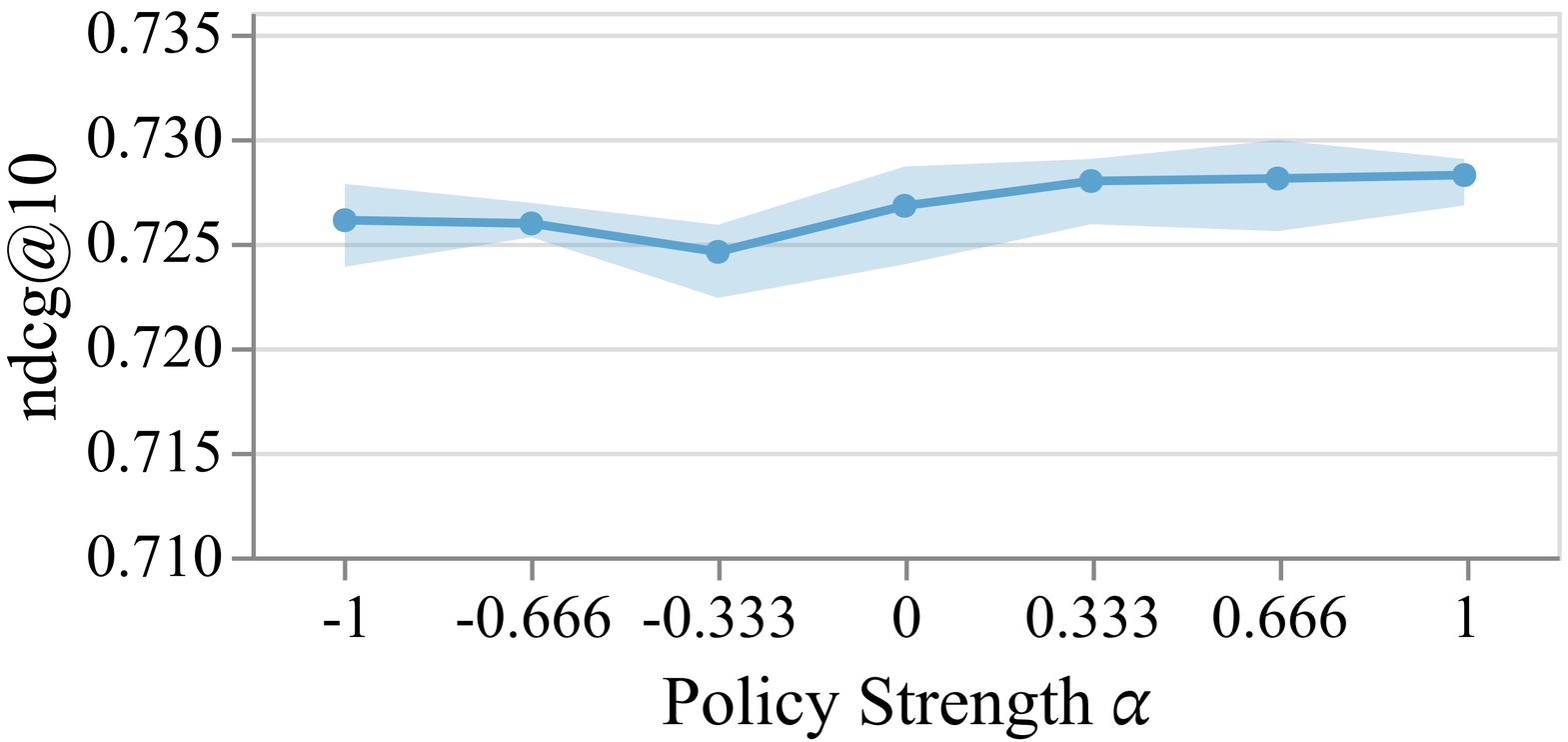
Initial experiments with expert policy



Two-tower models trained on deterministic logging policies of varying strengths (α) on MSLR30K: $\alpha = 1$ sorting by expert annotations, $\alpha = 0$ random sorting, and $\alpha = -1$ ranks from least to most relevant.

[1] Yunan Zhang et al. Towards Disentangling Relevance and Bias in Unbiased Learning to Rank. In KDD 2023.

Initial experiments without expert policy



Two-tower models trained on deterministic logging policies of varying strengths (α) on MSLR30K: $\alpha = 1$ sorting by expert annotations, $\alpha = 0$ random sorting, and $\alpha = -1$ ranks from least to most relevant.