It's Not You, It's Me: The Impact of Choice Models and Ranking Strategies on Gender Imbalance in Music Recommendation

SCAN ME

Andres Ferraro Michael Ekstrand Christine Bauer

Goal

We use simulations to study how **algorithmic strategies** or **user behavior** contribute to improvement (or loss) in **gender fairness** as models are **repeatedly re-trained** on new user feedback data.

Dataset

Lastfm-2b (2013-2020) artist gender information collected from MusicBrainz.org:

- Only 'solo' artists (436,789): female (93,316), male (342,523), and nonbinary (950)
- 15-core of user-artist interactions, retaining 78,021 users and 187,471 artists, of whom **21.675%** represent female artists.

Metrics

- **First-position exposure** (PFA): We average for each user the position of the first female artist in the recommendation ranking, with the first position as 0.
- **AWRF** (Attention-Weighted Rank Fairness). This uses rank-discounting to estimate the exposure value of each rank position, and measuring the fairness of exposure provided to each group.
- **Diversity:** Gini@k to measure how concentrated the recommendations are on a few artists both overall and disaggregated by gender. 0 means all artists are equally recommended.

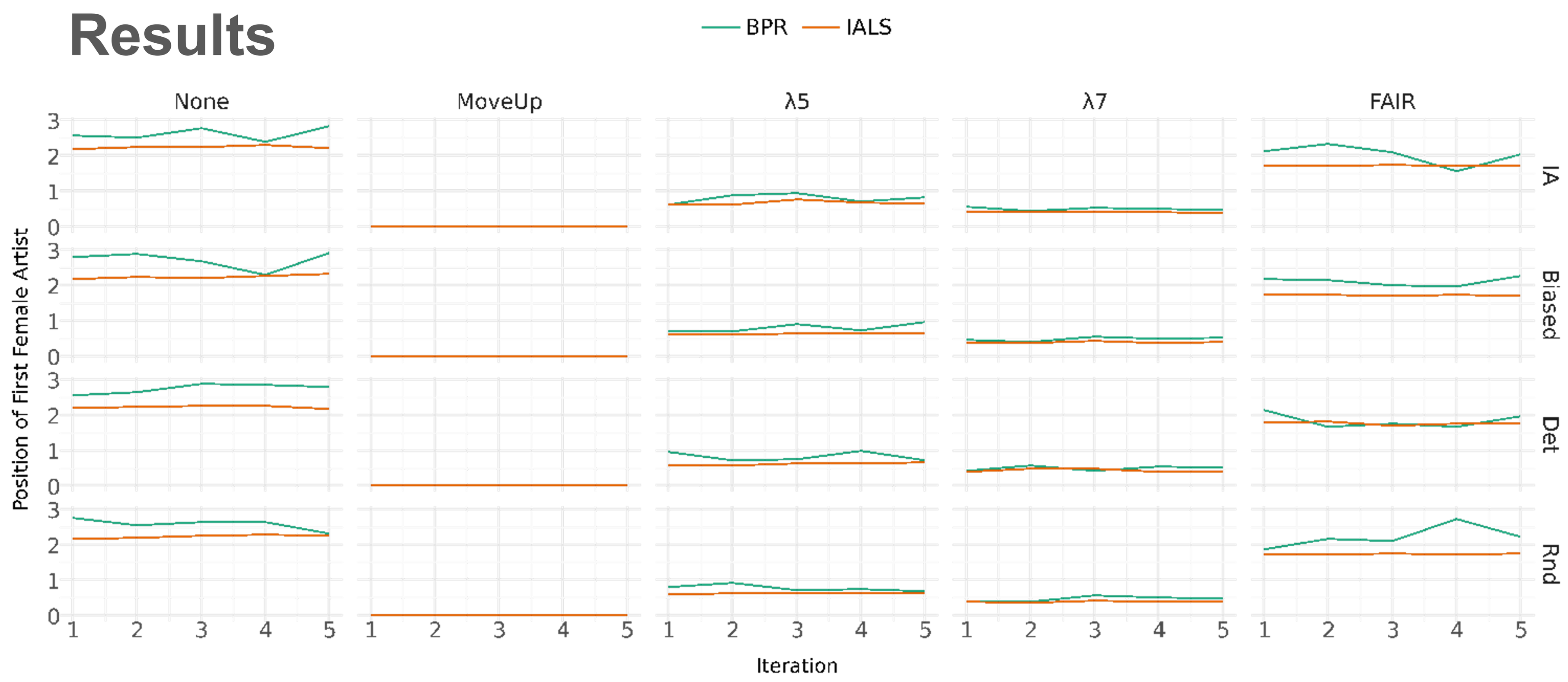
User Choice Models

- **Deterministic (Det):** User listens to all N recommended items.
- **Random (Rnd):** For each item in a user's top N recommended items, the user listens with probability 0.5.
- **InspectionAbandon (IA)** [1]: For each recommended item, with probability 0.5 the user listens to that item. After listening or ignoring the item, they stop listening entirely with probability 0.3; otherwise, they continue to the next item.
- **Biased:** Variant of IA in which users are biased so that they are 10% more likely to listen to male artists than others.

Approaches

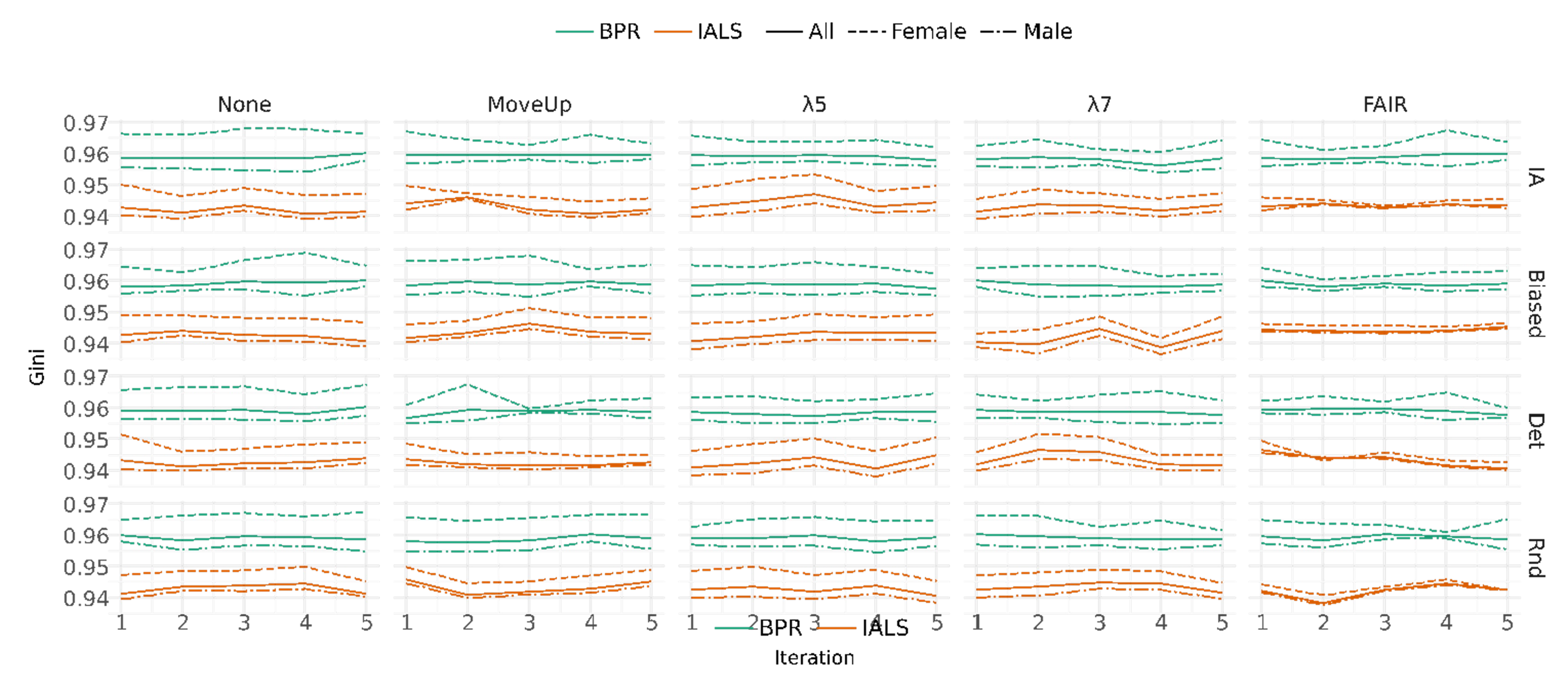
- **MoveUp:** Move the first item by a female artist to the first rank.
- **λ5, λ7:** Penalize items by male artists by moving each of them λ positions downward in the ranking [2]; based on the original paper, we use λ=5 and λ=7.
- **FA*IR** [3]: For each position, select the highest-scored item in the original ranking that will not cause the protected group (non-male artists) to be statistically significantly underrepresented concerning the target proportion (which we set to 50%).

Results



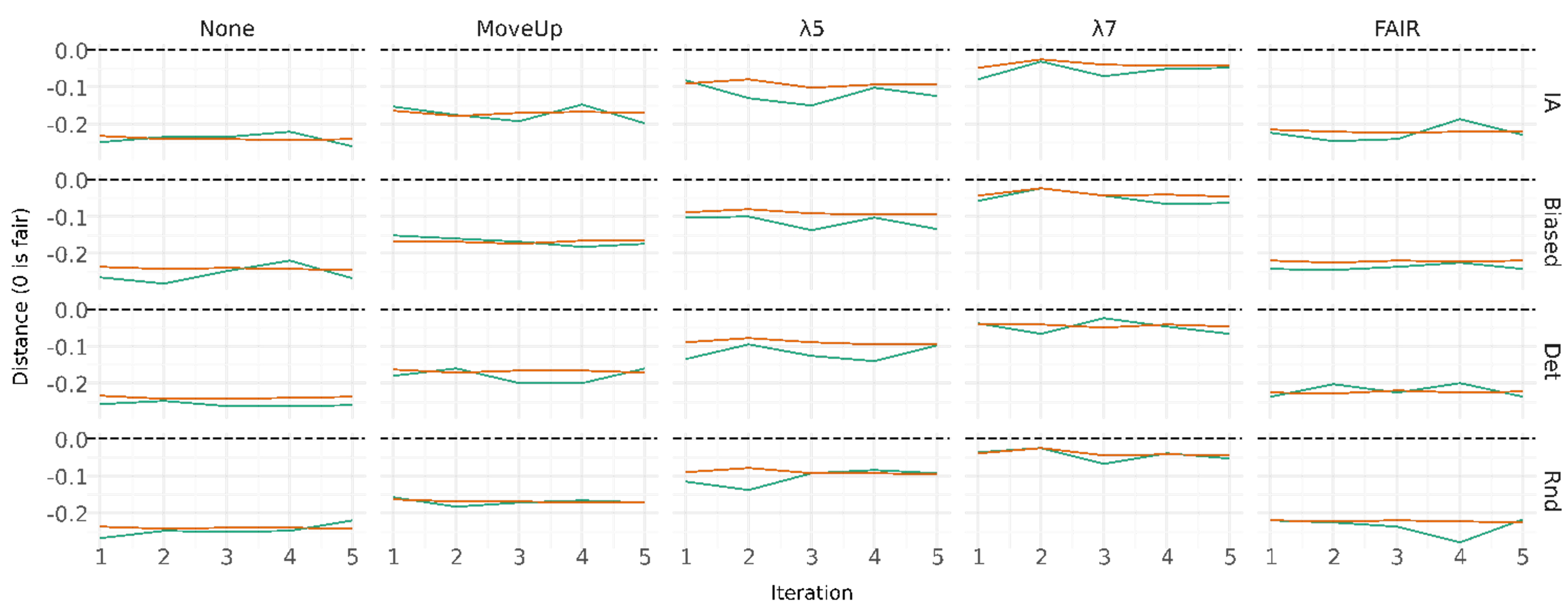
Re-ranking strategy determines the position of the first female artist, whereas the user choice model does not impact the ranking position of the first female artist:

- IALS is more stable than BPR, no variation across iterations or choice models.
- As MoveUp always puts a female artist on the very first position, it optimizes this metric by design. λ7 –followed by λ5– ranks the first female artist only on a slightly lower rank. FAIR ranks the first female artist only slightly higher than the baselines without re-ranking (None).



IALS is more equitable than BPR across all re-ranking strategies and choice models:

- Gini@10_female/male shows more equitable results for male than female artists: not only are female artists under-recommended (see AWRF), the exposure that does go to female artists is more concentrated on a smaller fraction of those artists.
- With FAIR, the gender gap in inequity closes, particularly with the IALS base model.



Re-ranking strategies have more impact than user choice models on AWRF (there is more variation between columns than rows), with little change over time. AWRF is mostly consistent with PFA.

FAIR contributes least to improving AWRF compared to the baselines; moving only the first female artist (MoveUp) is more effective at overall exposure fairness even though it only adjusts the position of a single item. The λ-re-rankers were the most effective, with λ7 improving exposure fairness the most.

As with PFA, IALS-based recommendations had more stable gender exposure balance.

Conclusion

Re-ranking strategies have a greater effect than user choice models on recommendation fairness over time:
This effect is consistent across multiple metrics and underlying recommendation models.

IALS deliver more stable results while BPR showed greater variation in fairness and diversity metrics, although without clear trends.

The model of user choice and response to recommendation had little effect on the fairness metrics we considered.

→ Re-ranking strategies may be a useful tool for intervening in a biased world and addressing 'societal imbalance'[4], as the algorithms have a stronger impact than the users' behavior in choosing items.

→ Re-ranking strategies can break the original feedback loop, IALS seems to anchor a 'new' feedback loop within the 5 iterations observed in our study.

[1] Amifa Raj and Michael Ekstrand. 2023. Unified Browsing Models for Linear and Grid Layouts. CoRR arXiv:2310.12524 (2023),
[2] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the Loop: Gender Imbalance in Music Recommenders. (CHIIR'21).
[3] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA*IR: A Fair Top-k Ranking Algorithm.(CIKM '17).
[4] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. Annual Review of Statistics and Its Application 8.

SiriusXM
Drexel University College of Computing & Informatics
Paris Lodron University Salzburg