

# A More Comprehensive Offline Evaluation of Active Learning in Recommender Systems

Diego Carraro Derek Bridge

## Active Learning

An Active Learning (AL) algorithm interrogates the user to obtain additional training examples that it reasons will be useful for building a better model. Asking the user for preference information has:

- ▶ a cost (the effort it places on the user), and
- ▶ a possible benefit (improvements to the recommendations).

Different AL strategies take different approaches to identifying which items to ask the user about: *we seek the best strategy, which asks the user for as little additional information as possible (reducing the cost) while obtaining the most benefit.*

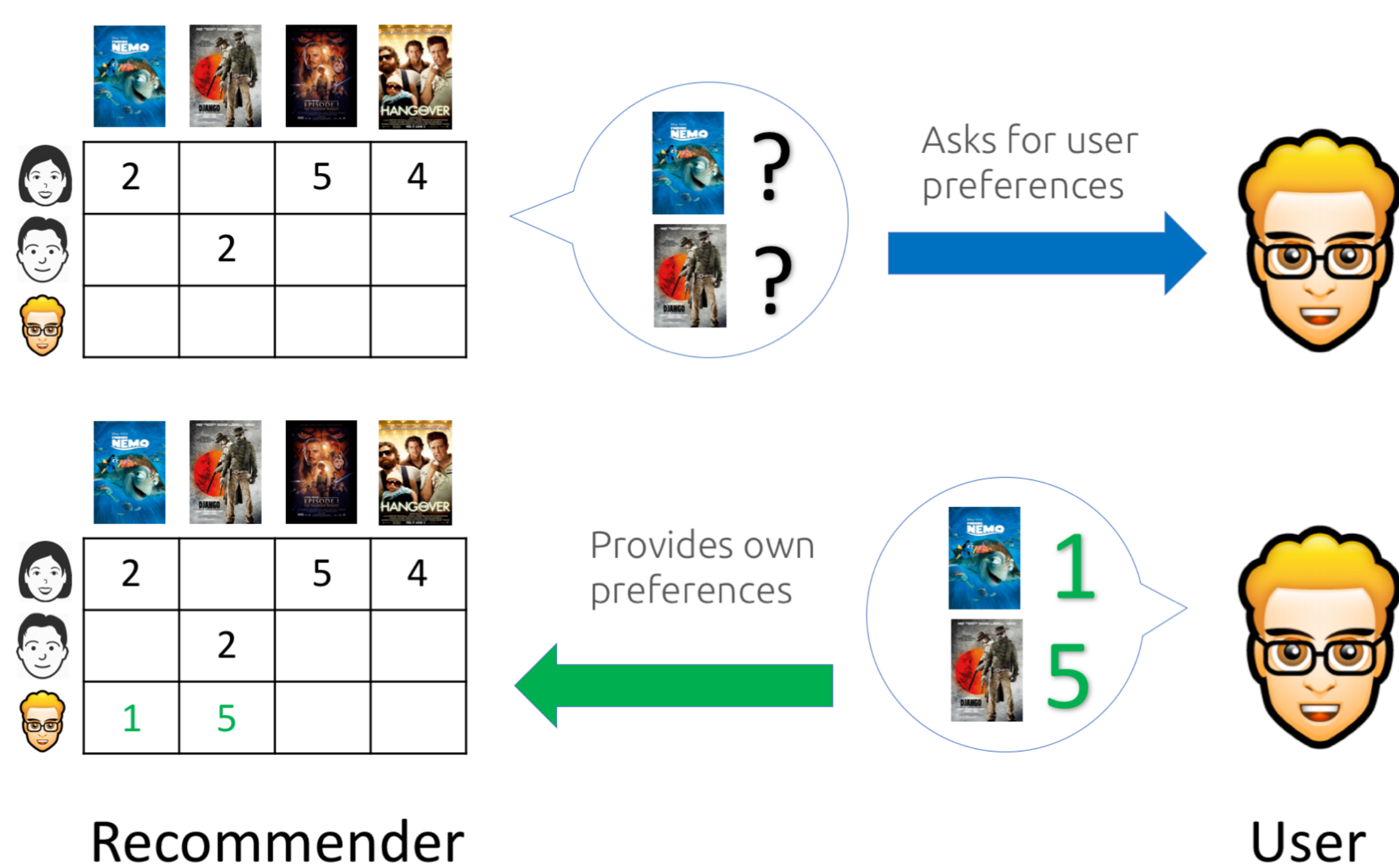
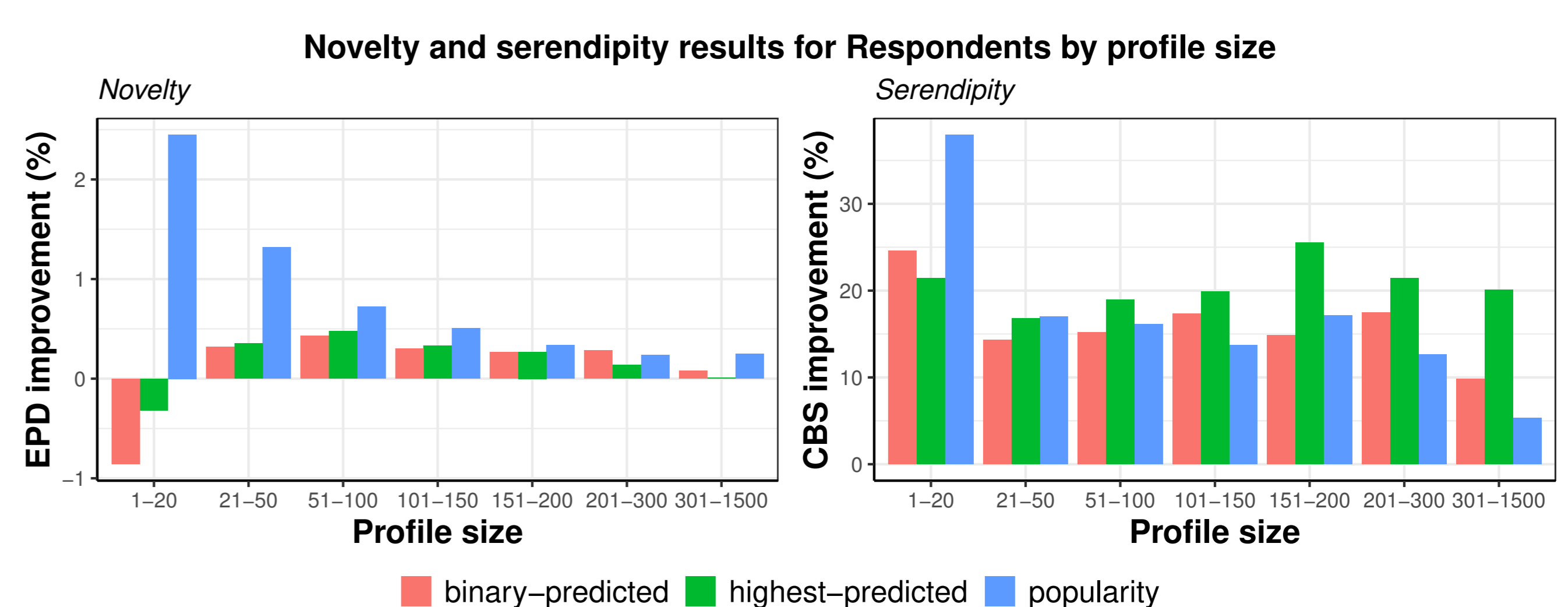
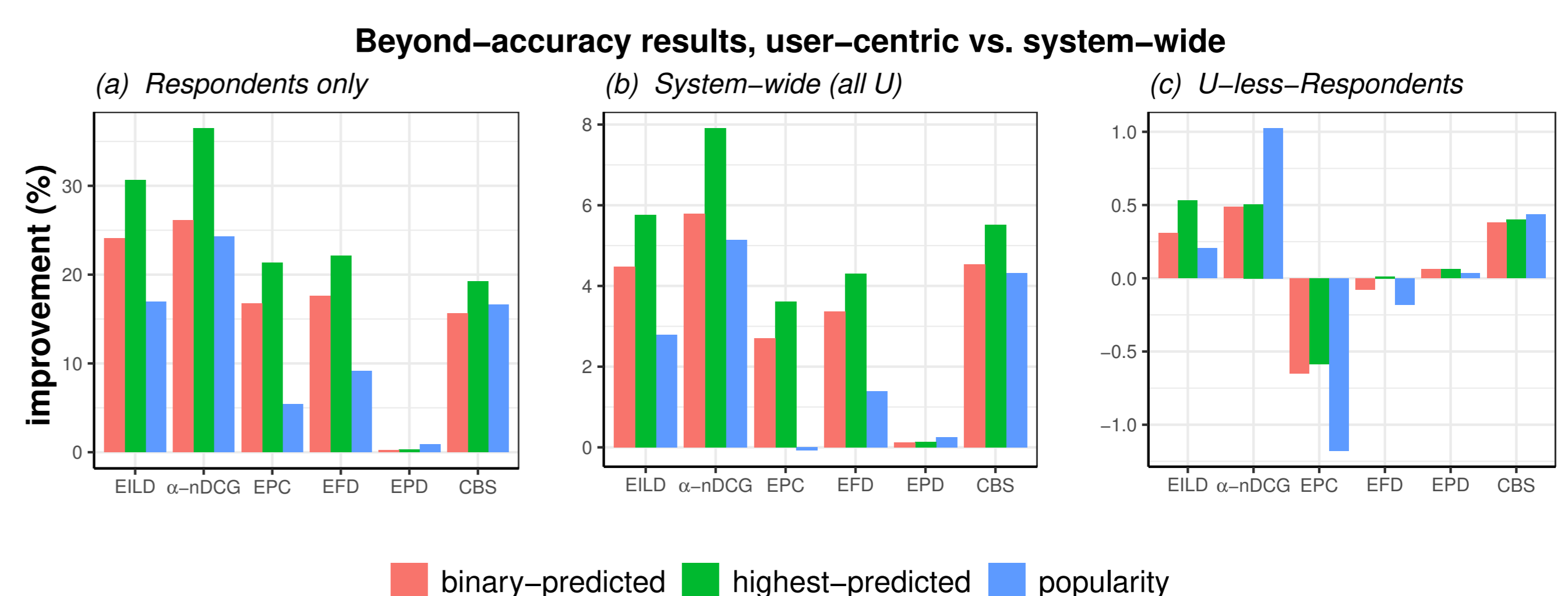
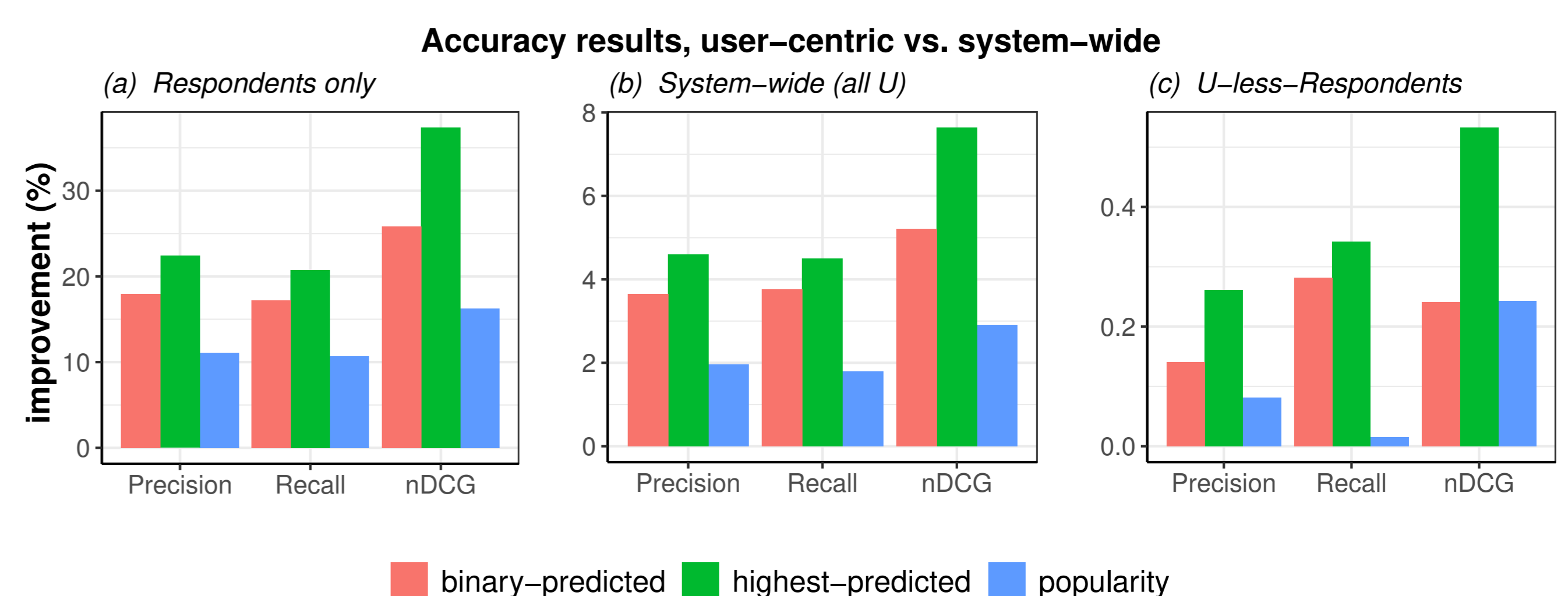


Figure: Example of a movie collaborative recommender eliciting user ratings using Active Learning.

## Preliminary Offline experiments

We applied our methodology to compare 3 AL strategies from the literature:

- ▶ *popularity*: asking for most popular items (non-personalized strategy).
- ▶ *highest-predicted*: asking for items which the recommender thinks the user will like (personalized strategy).
- ▶ *binary-predicted*: asking for items that are likely to be familiar to the user (personalized strategy).



## Research objectives

### HOW we evaluate AL

#### • Accuracy vs. beyond-accuracy metrics

*Accuracy & Ranking*: e.g. Precision, Recall, nDCG

*Beyond-accuracy*: e.g. Diversity, Novelty, Serendipity

#### • User-centric vs. system-wide measures

*Respondents only*: just those users who provide at least one rating

*System-wide*: all user  $U$  in the system

*U-less-Respondents*: users who were asked for ratings but did not provide any

#### • Different kind of users

E.g. grouped by *profile-size*

### WHAT we evaluate

#### • Existing strategies in literature

#### • New strategies

Beyond-accuracy targeted

Personalized for every user

## Future & research directions

- ▶ Extend the analysis of results, so we can see results not just by profile size but also perhaps by rating variance and profile diversity.
- ▶ Employ a similar evaluation method for situations where AL is used to boost the ratings of new items and more mature items (item-perspective).
- ▶ To use our method to help us design new AL strategies targeted to improving beyond-accuracy metrics.
- ▶ To use our method to help us design new AL strategies that are better targeted to the needs of different kinds of users.