

# Practical Evaluation of Repeated Recommendations in Personalized Music Discovery

Brian Manolovitz and Mitsunori Ogiwara  
University of Miami, Department of Computer Science

## Abstract

Studies have shown that repeated exposures to novel songs cause an increase in a person's memory and liking. This paper proposes the "retention rate" as an additional parameter for evaluation, which examines the rate at which the listener revisits the novel items. The authors hypothesize that the retention rate will be proportional to the number of times the discovery engine suggests the pieces to her, *as long as they remain novel*. The authors have tested the hypothesis through a six-week human-subject experiment which simulates a real-world listening environment and a follow-up survey. The researchers conclude that the more times a listener is exposed to a song during the discovery process, the more likely she is to return to the song.

## Introduction

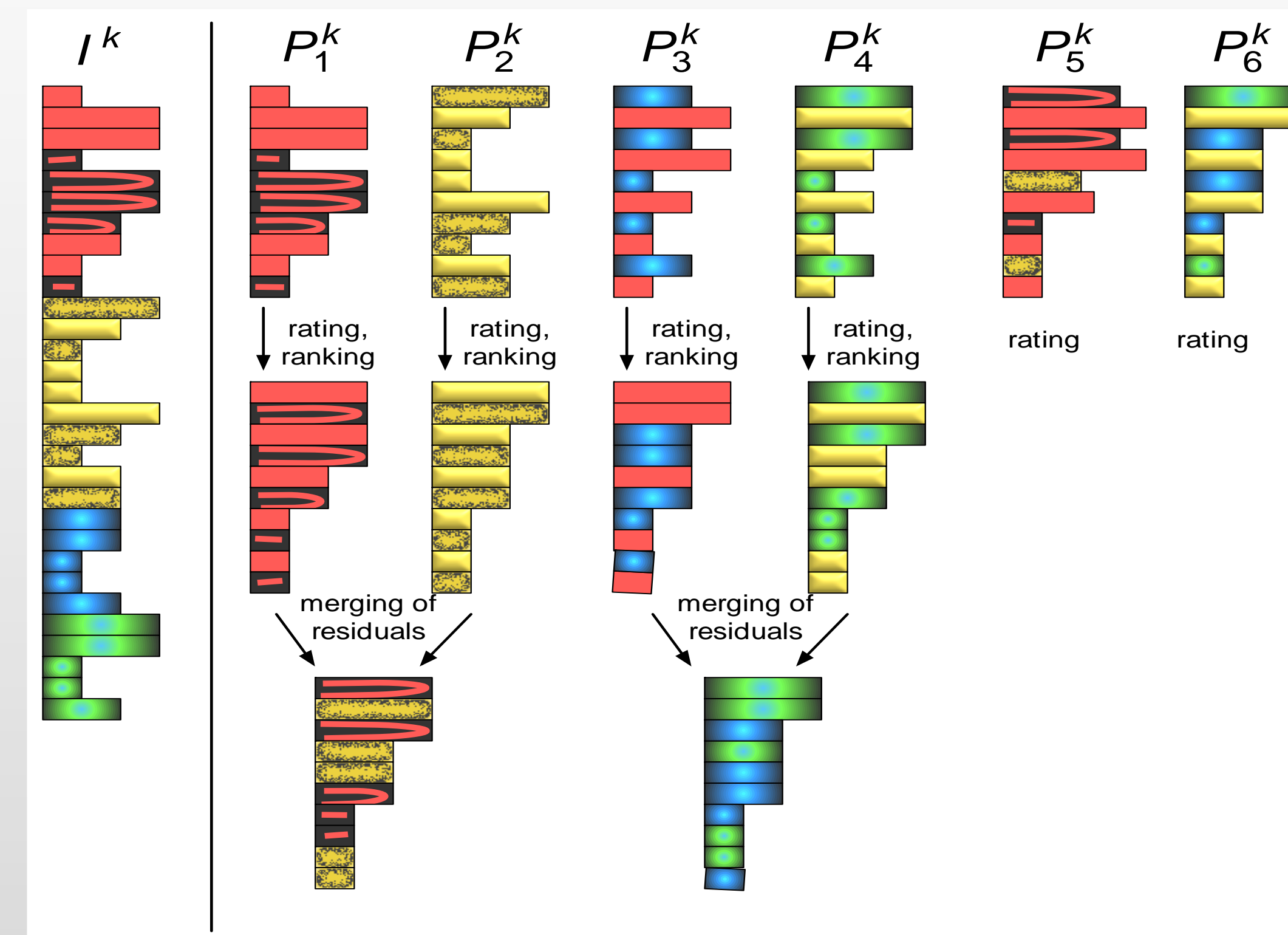
- The long-tail problem = most interactions are with few items, and most items have few interactions
- Items in the long tail are difficult to recommend due to lack of data
- How to recommend long-tail items; songs/artists need repeat consumers in order to have a chance to break out of the long tail**
- Tools like Spotify's Discover Weekly recommend new and interesting music to users on a weekly basis, introducing lesser-known artists
- People can forget about "discovered" songs they liked**
- Also, people may feel neutral toward a song and never revisit, but would have enjoyed it more with additional listens**
- The core foundation = the inverted-U theory:
- This study: a 19-human-subjects listening study using Spotify's Discover Weekly as the recommendation tool, where the unfamiliar songs were played 1, 2, or 3 times during the 6-week experiment
- Novel component = "retention rate": the rate at which the unfamiliar songs were revisited during normal listening behavior
- Research question:** does presenting an unfamiliar song multiple times increase the likelihood for the listener to revisit?

## Materials and Methodology

- Subjects provide 30-song Discover Weekly playlist without listening
- Researchers construct weekly 10-song playlist from those 30 songs
- Subject listens to each playlist and rates preference (1-5), discovery effectiveness (1-5), and liking ratings for each song (Like, Neutral/Not Sure, or Dislike)
- Researchers use ratings to select for repetition: 10 songs once, 10 songs twice, and 10 songs 3 times

## Materials and Methodology, cont.

### Weekly Playlist Construction Process



Initial: 30 songs from Discover Weekly  
Weeks 1 and 2: 10 songs each from Initial  
Weeks 3 and 4: 5 songs from Initial, 5 songs from weeks 1 and 2 respectively: sorted by rated, every other song chosen  
Week 5: Same repeated songs from week 3; choose 5 songs from 10 remaining songs from weeks 1 and 2  
Week 6: Same as week 5 but using the 5 repeated in week 4 and the 10 remaining from weeks 3 and 4

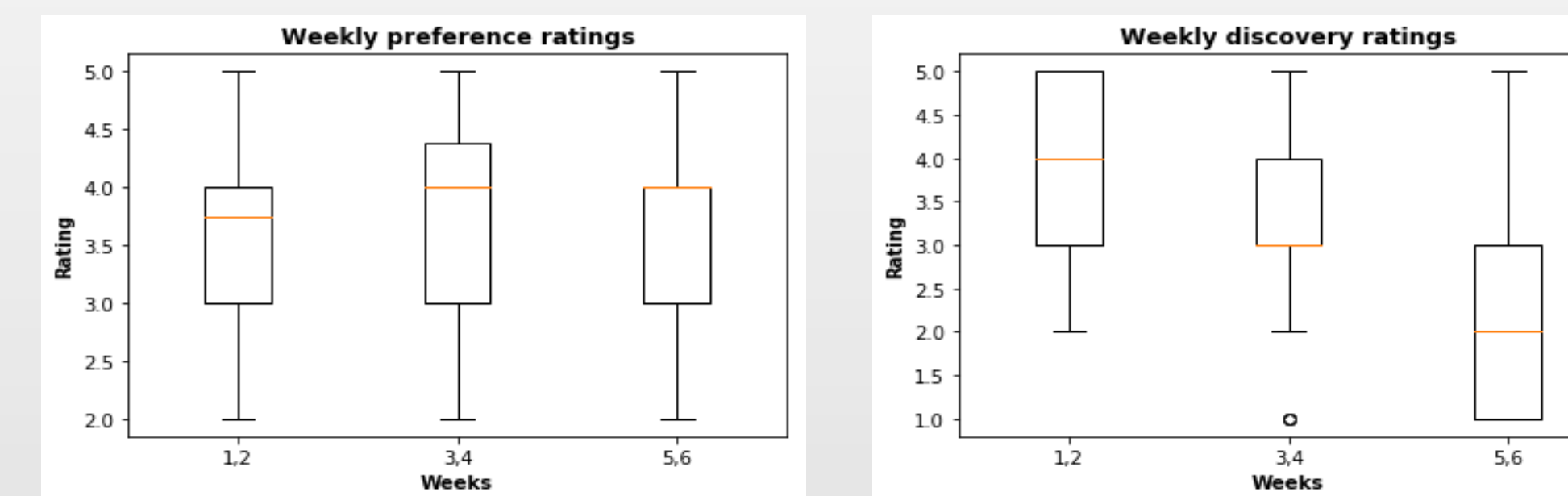
- Even distribution of liking ratings across all 3 play counts
- After listening portion, subjects received a survey asking them their likelihood (1-5 Likert scale) to revisit each of the 30 songs:
  - If they heard the song prior to the experiment, leave blank
  - If they don't remember the song, place an asterisk
  - Forgetting rate** = number of forgotten songs / total # of songs
- A survey 1-month post experience: for each of the 30 songs, they went out of their way to listen in the month since the experiment ended
  - Retention rate** = number of songs retained / total # of songs

## Results

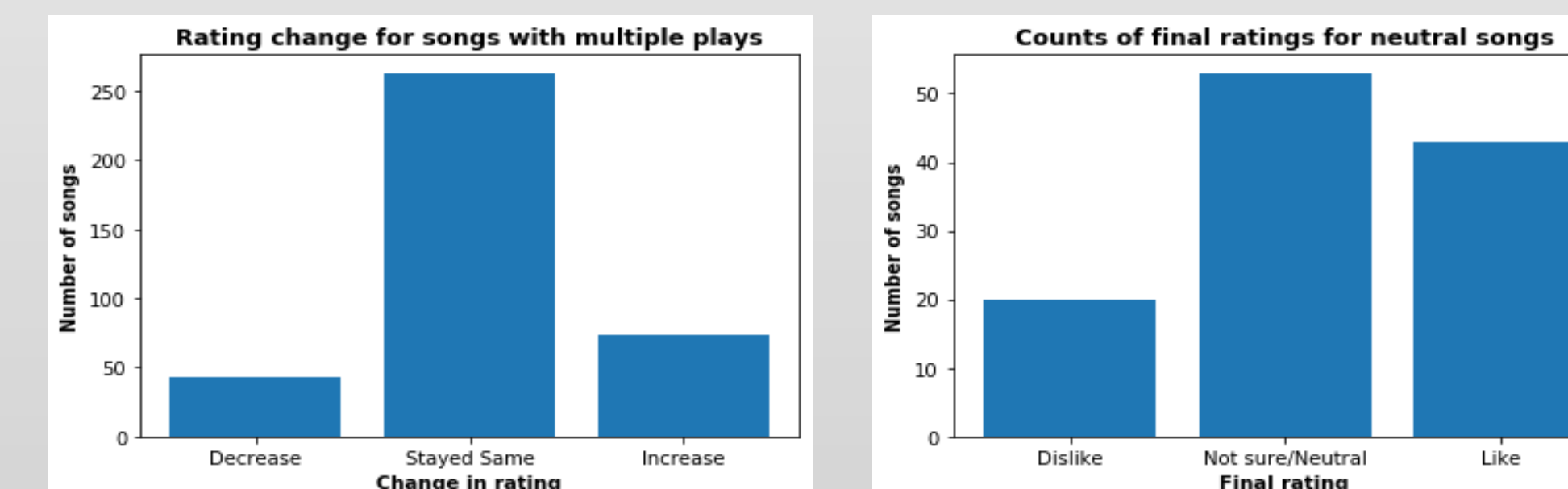
- 26 of 570 songs were heard prior to experiment (< 0.5%)
- Caveat: 19 subjects is not enough to generalize results
- One subject failed to report the retention survey

### Weekly preference/discovery rating boxplots

Grouped by # of repeated songs in playlist (0, 5, 10)

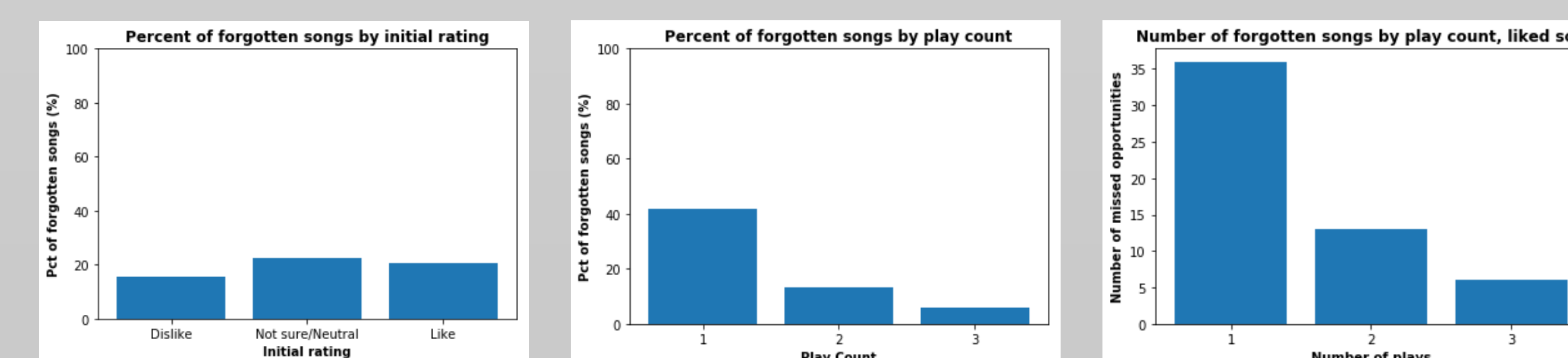


### Rating Changes for Individual Songs

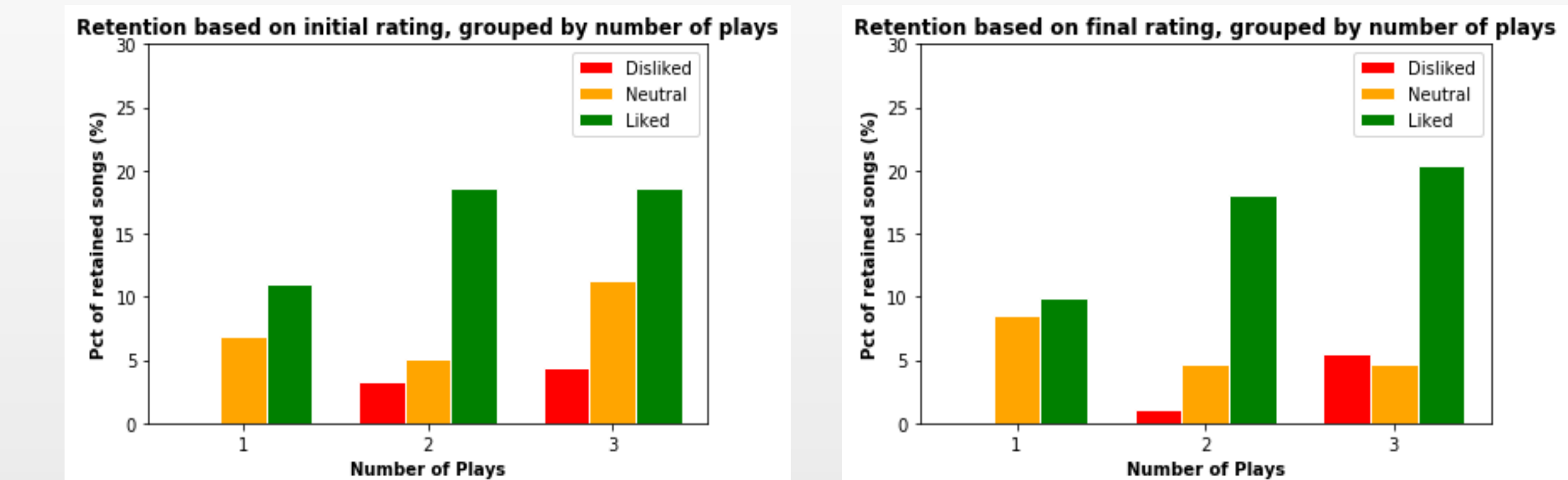


- Songs with initial rating of "Neutral/Not sure" that increased to "Like" would have likely been discarded after 1 listen but grew in favor and were more likely to be revisited

### Forgetting Rate



### Retention Rate



- Dislike and Neutral feelings do not entice listener to revisit (< 5% retention rate)
  - 37% of Neutral songs increased to Like -> difference between graphs
  - 1 play, Neutral and Like similar, seems to indicate curiosity
- Caveat: Could not test same song-subject pair for each play count

### One-tailed paired-samples z-tests

Test	Results
Dislike vs. Neutral	$Z(250) = 3.11, p < 0.001$
Neutral vs. Like	$Z(423) = 5.08, p < 0.00001$
All songs, 1 vs. 2+ plays	$Z(514) = 3.42, p < 0.0005$
Liked songs, 1 vs. 2+ plays	$Z(264) = 3.37, p < 0.0005$
Neutral songs, 1 vs. 2+ plays	$Z(159) = 0.62, p < 0.3$
Disliked songs, 1 vs. 2+ plays	$Z(91) = 1.98, p < 0.025$

- Significant differences between the retention rates w.r.t. preference rating, which is expected
- In practice, songs repeated if user gives feedback that indicates liking

## Conclusion

- Retention rate introduced as metric for evaluating success of music discovery process
  - Uses actual listening behavior, not liking and familiarity ratings
  - Measures effect of repeating recommendations of unfamiliar songs
- One additional play = higher retention rate for liked songs
- Helps unknown songs and artists break out of the long tail
- Also helps listeners broaden their repertoires