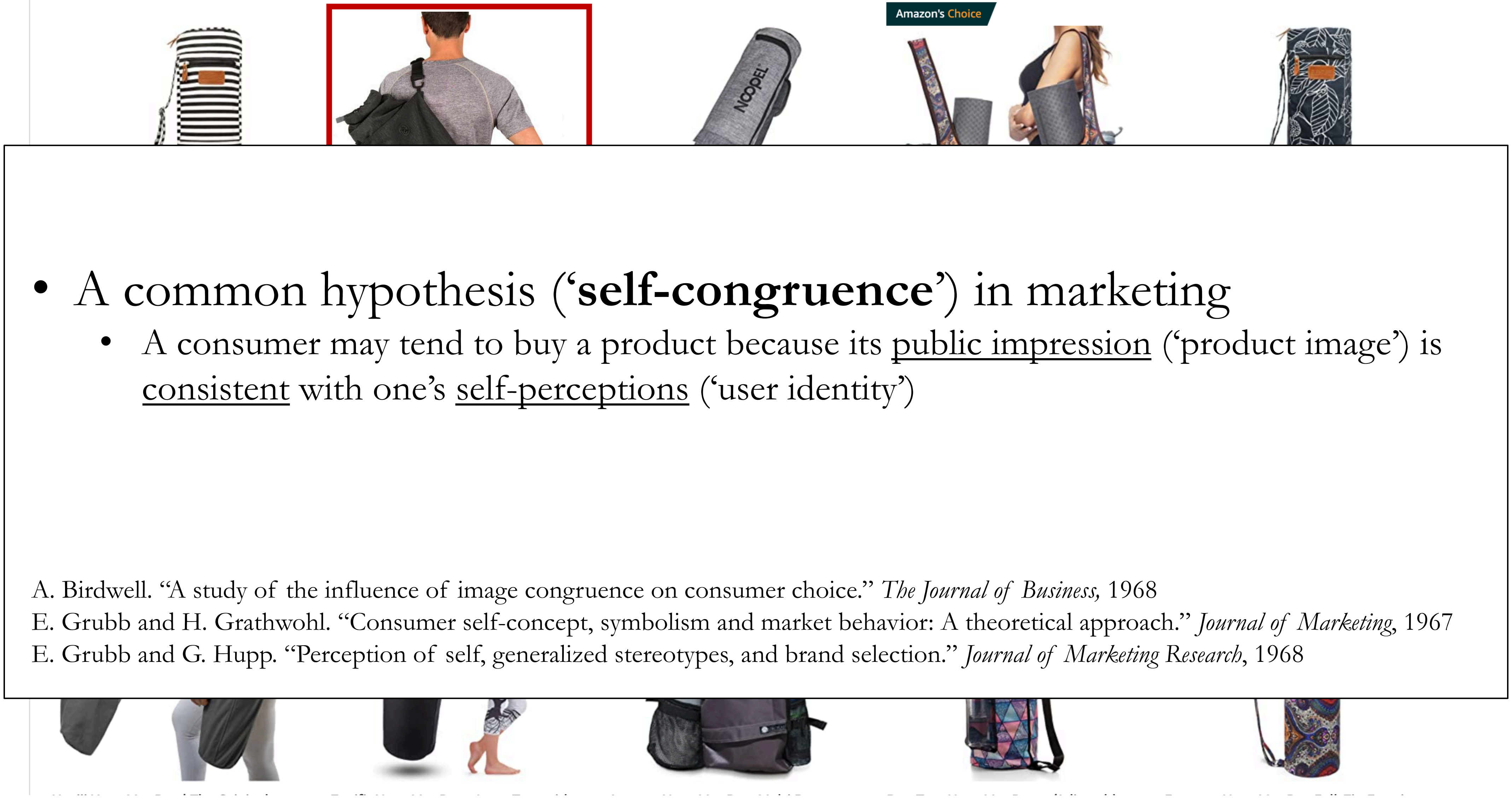
Addressing Marketing Bias in Product Recommendations

Mengting Wan (Airbnb, Inc)*, Jianmo Ni (UC San Diego)
Rishabh Misra (Twitter, Inc), Julian McAuley (UC San Diego)
**Work was done while at UC San Diego*

Marketing Bias

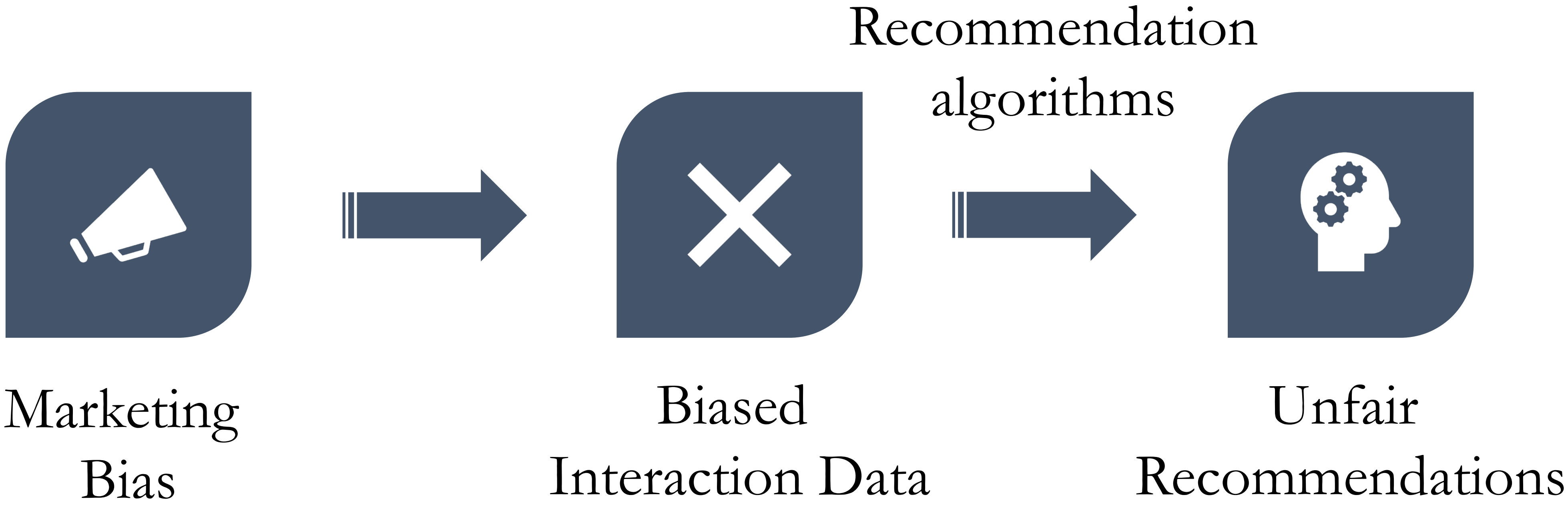
- The same product can be marketed using different human images
  - Body Shapes, Genders, Ages, Ethnicity Groups, etc.
- As indicated in many marketing studies, these strategies could affect consumer behavior

A female user wants to buy a boxing product



- A common hypothesis (‘**self-congruence**’) in marketing
  - A consumer may tend to buy a product because its <u>public impression</u> (‘product image’) is <u>consistent</u> with one’s <u>self-perceptions</u> (‘user identity’)

A. Birdwell. “A study of the influence of image congruence on consumer choice.” *The Journal of Business,* 1968
E. Grubb and H. Grathwohl. “Consumer self-concept, symbolism and market behavior: A theoretical approach.” *Journal of Marketing*, 1967
E. Grubb and G. Hupp. “Perception of self, generalized stereotypes, and brand selection.” *Journal of Marketing Research*, 1968


Recommendation algorithms

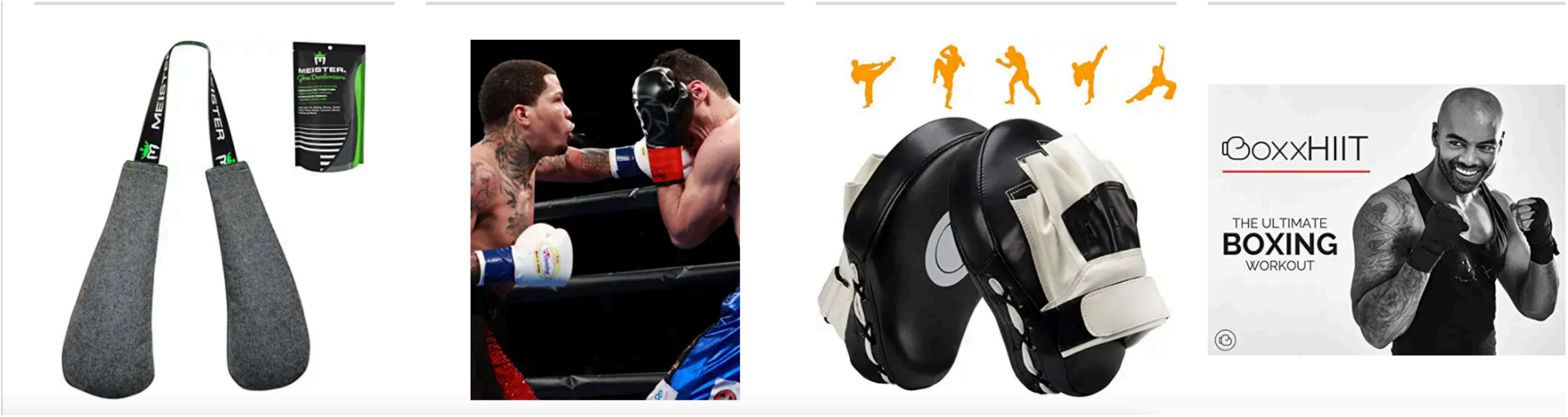
Marketing Bias

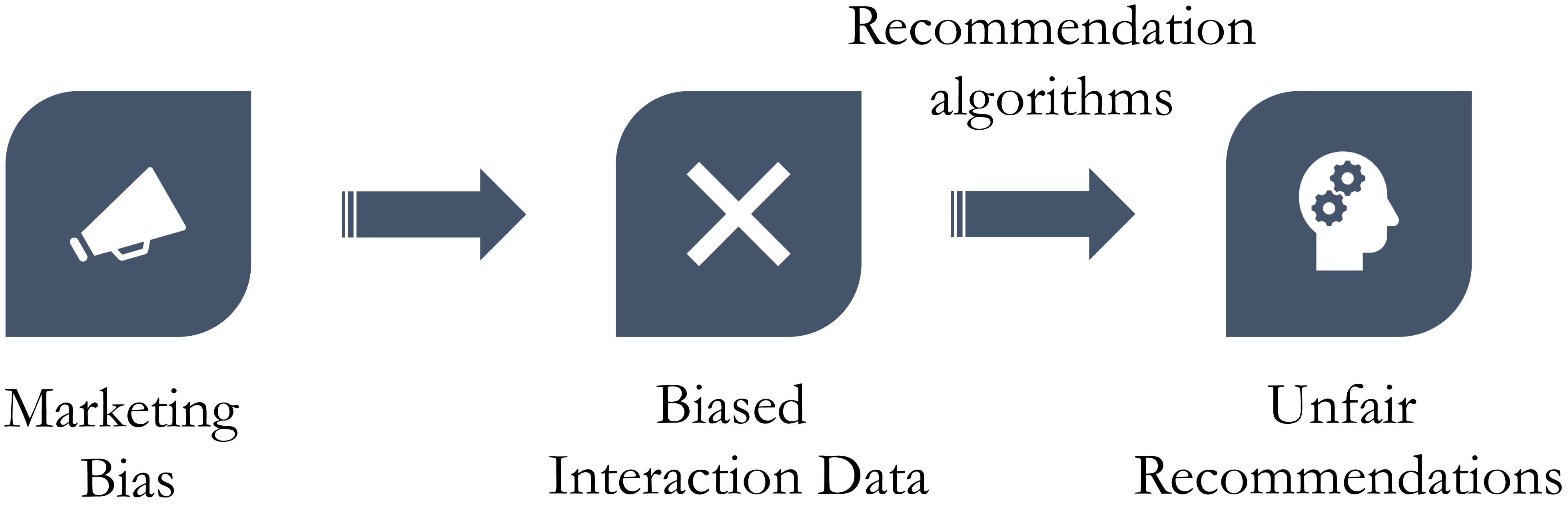
Biased Interaction Data

Unfair Recommendations

- Biased interaction dataset
  - consumers' intrinsic preferences (the target of a RecSys) and marketing preferences (confounding factors) are entangled
- Potential marketing bias could be propagated by ML algorithms
  - Market imbalance can be worsen – even worse recommendation accuracy in the underrepresented market segment

Female user buying a boxing product

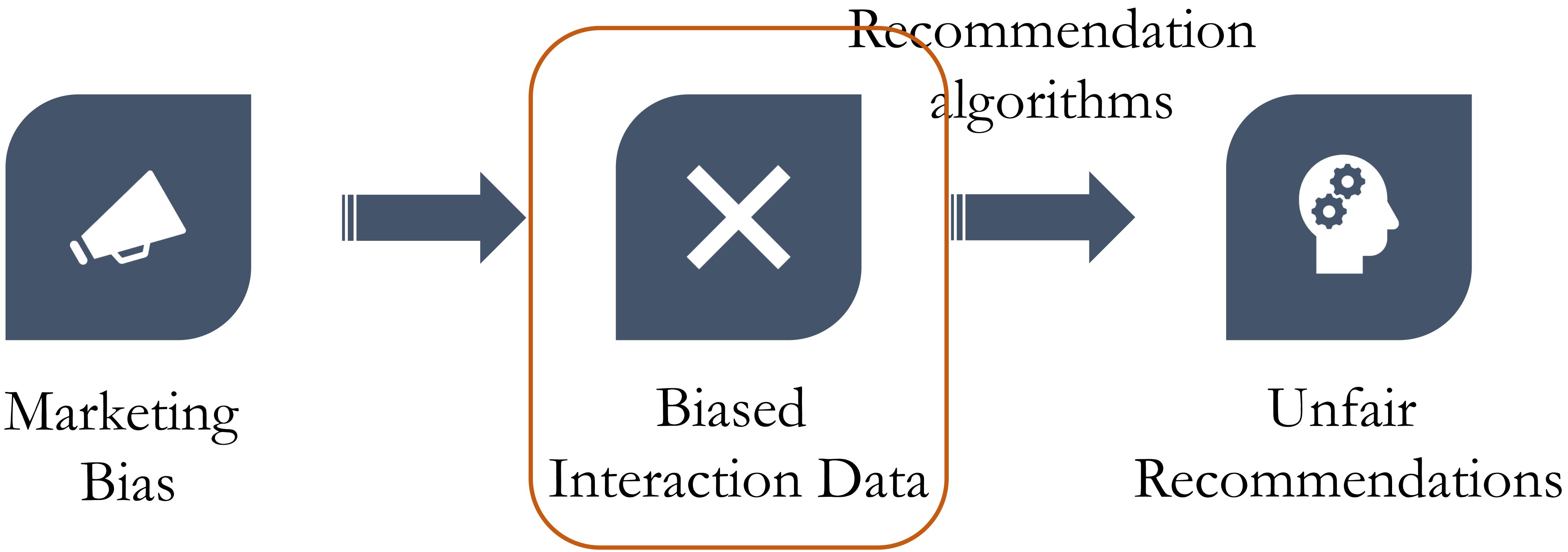




Marketing Bias → Biased Interaction Data → Unfair Recommendations

Recommendation algorithms

- Q1: Does such a marketing bias exist in the input interaction data?
- Q2: How do standard algorithms respond to the biased inputs?
- Q3: How to improve the market fairness of recommendations?



Recommendation algorithms

Marketing Bias

Biased Interaction Data

Unfair Recommendations

- Q1: Does such a marketing bias exist in the input interaction data?
    - Observational analysis on two collected e-commerce datasets (ModCloth & Amazon Electronics)

Data Collection

- **Modcloth**
  - Clothing website
  - Potential marketing bias
    - The **body shape** (small/large) of the human models in product images
- Amazon **Electronics**
  - Electronic products
  - Potential marketing bias
    - The **gender** (male/female) of the human models in product images
- Users' rating scores on product items are available on both websites

Large
(1X, 2X, 3X, 4X)

Small
(XS, S, M, L, XL)

- Product image group
  - Small; Small & Large
- User identity group
  - Small; Large

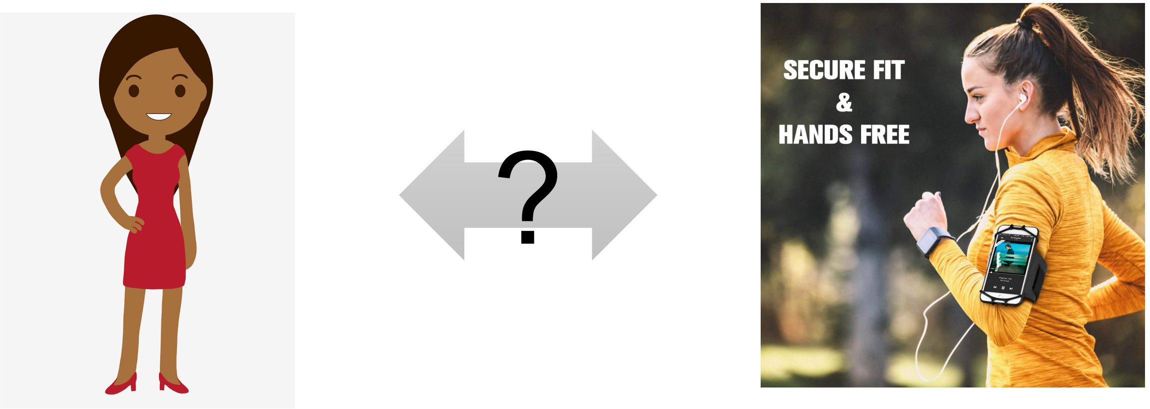
Female

Male

- Product image group
  - Female; Male; Female & Male
- User identity group
  - Female; Male

Q1: Does marketing bias exist in the input data?

Are user identity and product image correlated to each other in the input interactions?



H_0: user identity groups (m) and product image groups (n) are independent in terms of interaction frequency

Q1: Does marketing bias exist in the input data?

H_0: user identity groups (m) and product image groups (n) are <u>independent</u> in terms of interaction frequency

χ^2-test for statistical independence:

$$\boldsymbol{f}_{m,n} - \mathbf{E}\boldsymbol{f}_{m,n}$$

		User Group	
		Small	Large
Product Group	Small	$\boldsymbol{f} - \mathbf{E}\boldsymbol{f}$	
	Small & Large		

Modcloth

		User Group	
		Female	Male
Product Group	Female	$\boldsymbol{f}m, n\ \mathbf{E}\boldsymbol{f}$	
	Female & Male		
	Male		

Electronics

Q1: Does marketing bias exist in the input data?

~~H_0: user identity groups (m) and product image groups (n) are independent in terms of interaction~~

Product image and user identity are correlated with each other

Large χ^2, small p-value, H_0 is rejected

		User Group	
		Small	Large
Product Group	Small	$f - \mathbf{E}f$	
	Small & Large		

Modcloth

$\chi^2 = 158.7, p < 0.001$

		User Group	
		Female	Male
Product Group	Female	$f - \mathbf{E}f$	
	Female & Male		
	Male		

Electronics

$\chi^2 = 581.8, p < 0.001$

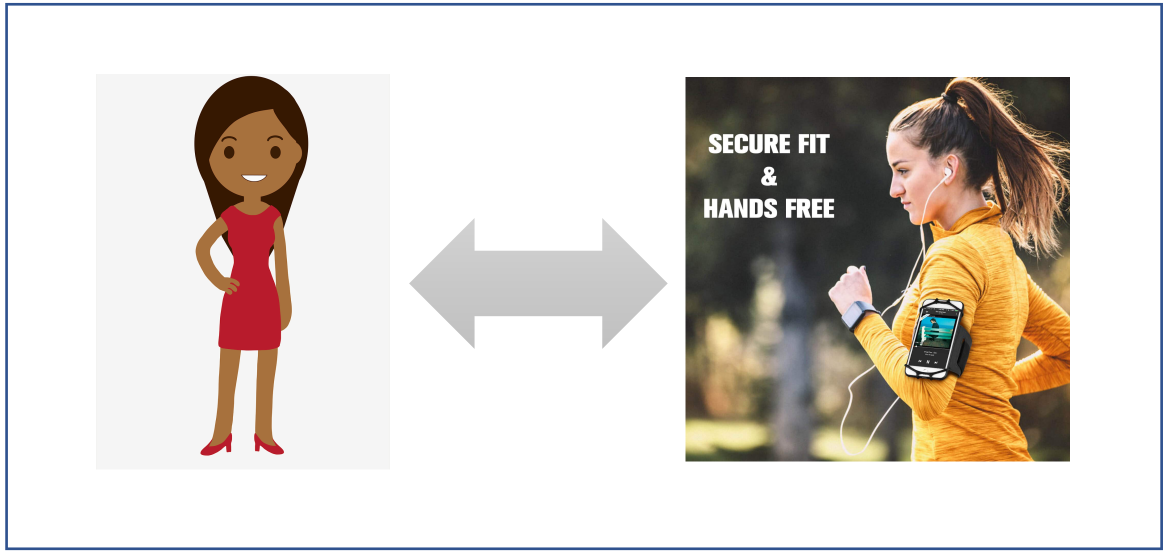
Q1: Does marketing bias exist in the input data?

‘Self-Congruency’ pattern is significant

- People are more likely to consume products represented by someone ‘similar’ to themselves

$$\text{Observed Frequency } (f) - \text{Expected Frequency } (\boldsymbol{Ef}) > 0$$

WHEN “user identity” = “product image”



Product Group	User Group: Female	User Group: Male
Female	+1473	-1473
Female & Male	+881	-881
Male	-2354	+2354

Electronics

$\chi^2 = 581.8, p < 0.001$

Recommendation algorithms

Marketing Bias

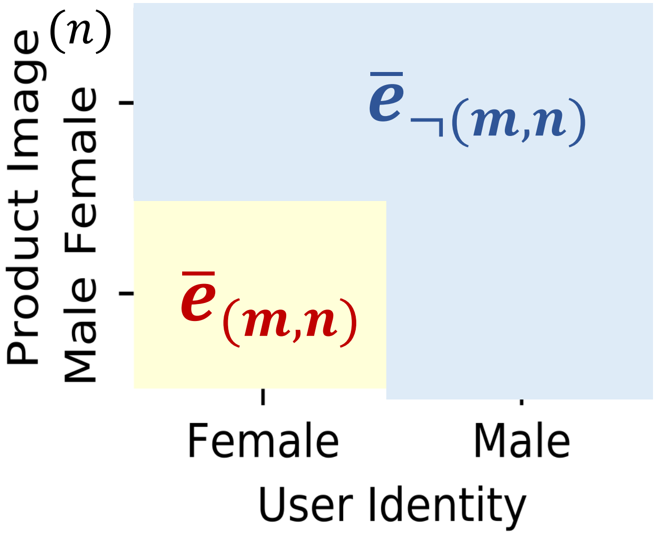
Biased Interaction Data

Unfair Recommendations

- Q2: how do standard recommendation algorithms respond to the biased input data?

Q2: How do standard algorithms respond to the biased inputs?

- Predictive Task
  - Rating Prediction ($s_{u,i} := f(u, i) \rightarrow y_{u,i}$)
- $\text{diff}_{m,n} = \bar{e}_{\neg(m,n)} - \bar{e}_{(m,n)}$
  - > 0: segment (m, n) is favored by the algorithm (smaller prediction error inside the market segment)

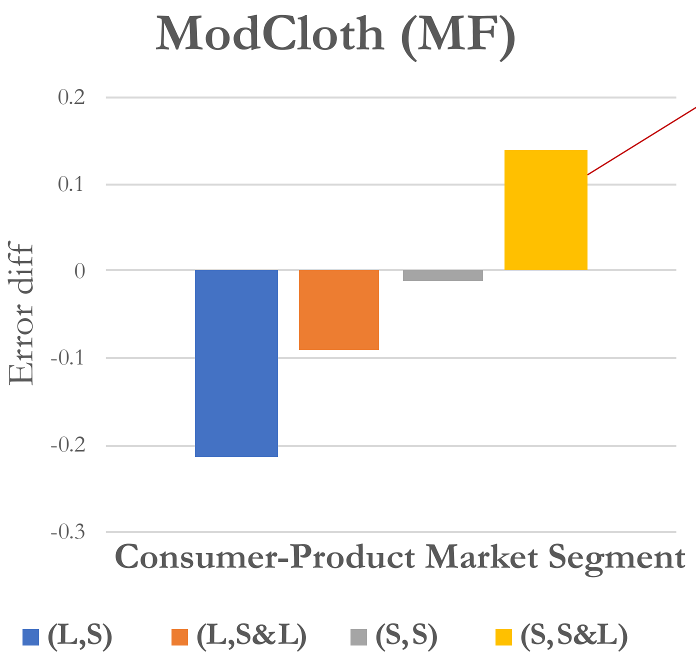


itemCF: B. Sarwar, G. Karypis, J. Konstan, J. Riedl, *et al.* "Item-based collaborative filtering recommendation algorithms." WWW'01.
userCF: J. Herlocker, J. Konstan, A. Borchers, and J. Riedl. "An algorithmic framework for performing collaborative filtering." *SIGIR'99*
MF: K. Yehuda, R. Bell, C. Volinsky. "Matrix factorization techniques for recommender systems." *Computer (2009).*
PoissonMF: P. Gopalan, J. Hofman, and D. Blei. "Scalable Recommendation with Hierarchical Poisson Factorization." UAI'15.

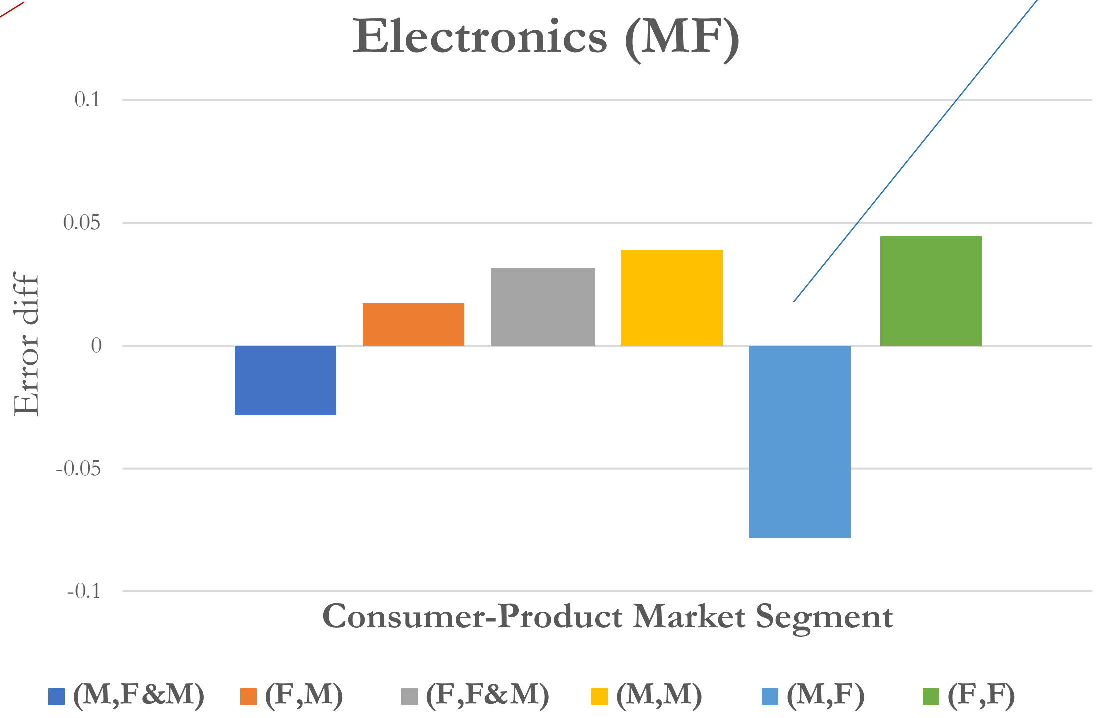
Q2: How do standard algorithms respond to the biased input data?

ML algorithms generally tend to favor dominating market segments

male users buying products w. female model(s)



ModCloth (MF)



Electronics (MF)

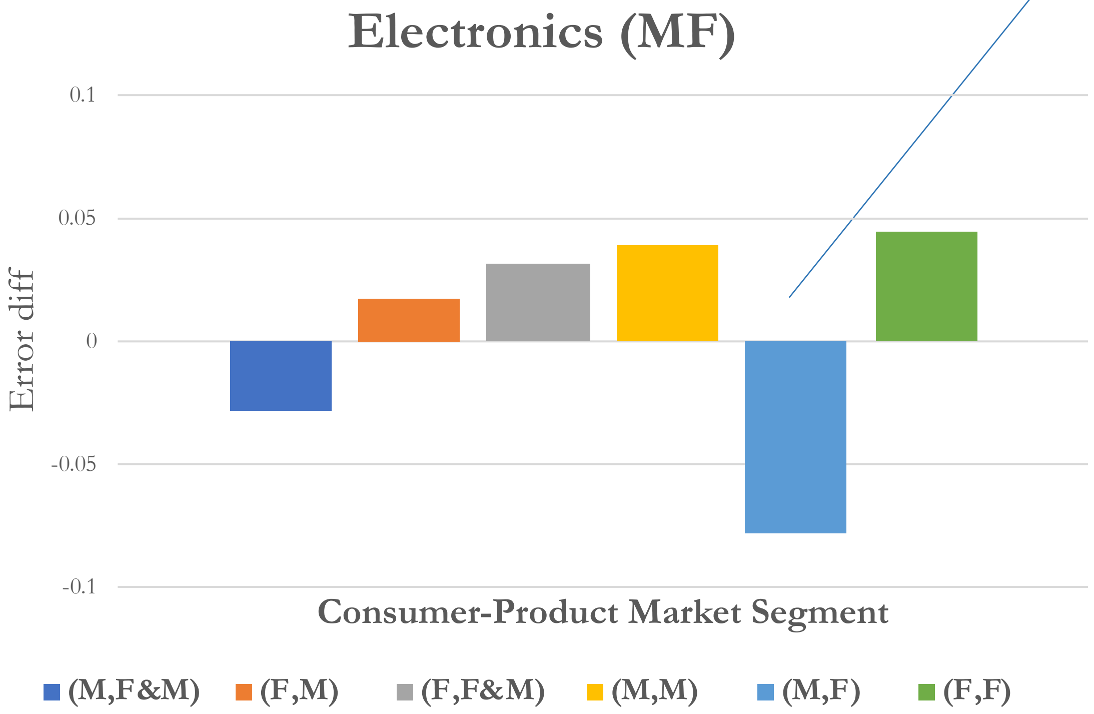
- $\text{diff}_{m,n} = \bar{e}_{\neg(m,n)} - \bar{e}_{(m,n)}$
- Market segments are sorted based on their sizes in training data

Q2: How do standard algorithms respond to the biased input data?

Real market size (f) < Expected market size (Ef)

		User Group	
		Female	Male
Product Group	Female	+1473	-1473
	Female & Male	+881	-881
	Male	-2354	+2354

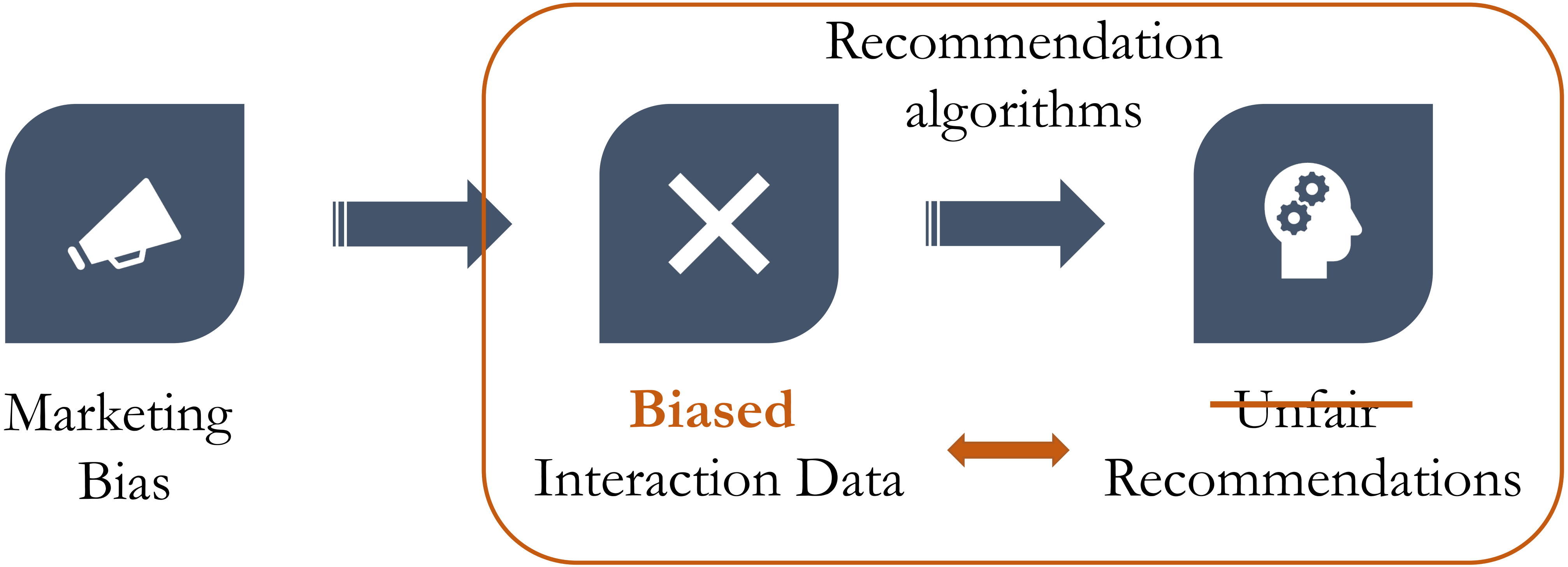
Electronics



Electronics (MF)

male users buying products w. female model(s)

- **Electronics**: the trend correlates to the deviations of the real market size from the expected market size



Recommendation algorithms

Marketing Bias

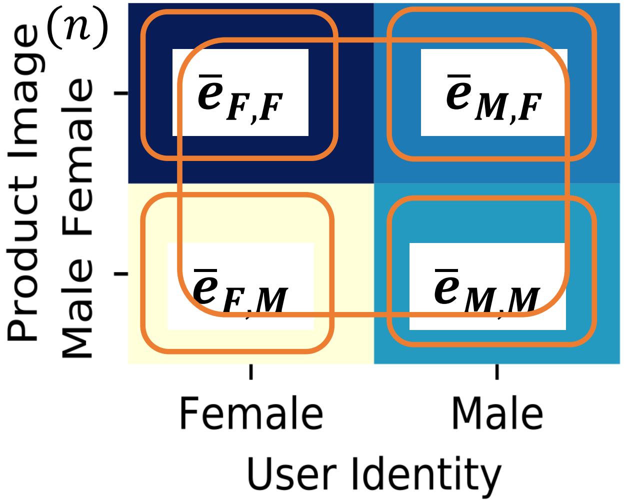
Biased Interaction Data

~~Unfair~~ Recommendations

- Q3: how to mitigate such an algorithmic bias and improve the market fairness of recommendations?

Market Fairness of Rating Predictions

- Rating Prediction Fairness
    - Prediction **errors** across different consumer-product **market segments** (m, n) are expected to be consistent (H_0)



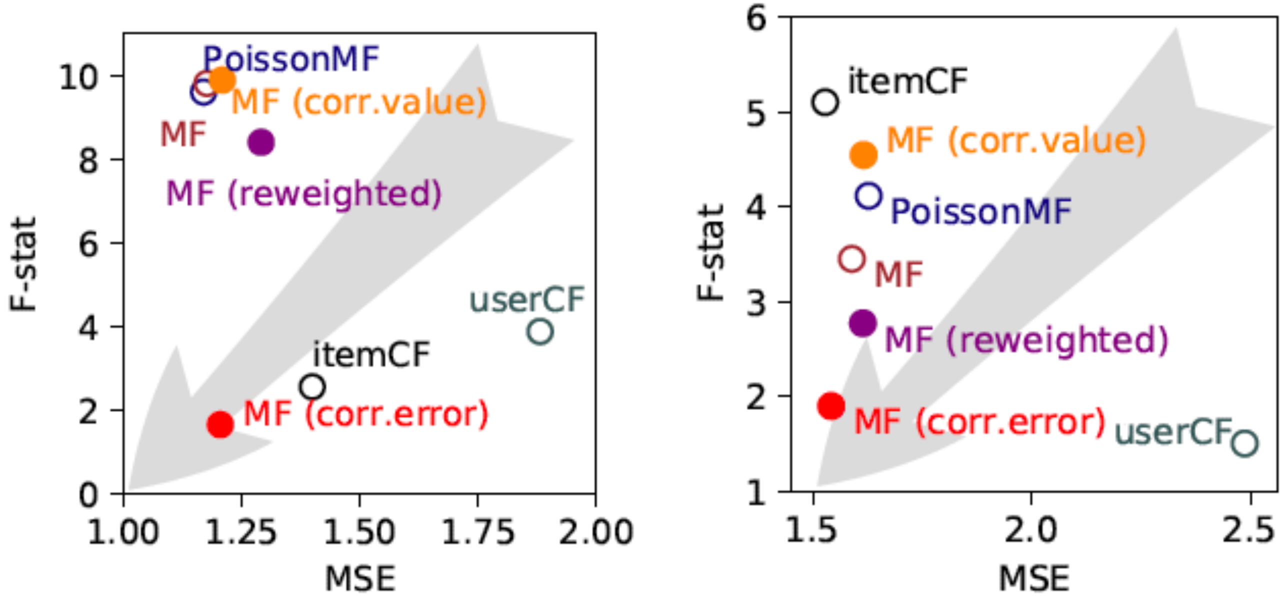
- F-test for statistical independence
    - *Small* **between**-segment error variation (V)
    - *Compared to* **within**-segment error variation (U)
    - *Small* **F-statistic**:
        - $F = \frac{V/(M*N-1)}{U/(|D|-M*N)}$ – deviation of the observed errors from H_0
        - A metric to evaluate the market fairness of rating predictions with a tractable statistical distribution

Q3: How to improve the market fairness of recommendations?

- Matrix Factorization
  - $s_{u,i} := f(u,i) = < \boldsymbol{\gamma_u}, \boldsymbol{\gamma_i} > \rightarrow \boldsymbol{y_{u,i}}$
  - MSE-based loss function: $L = \sum \left(s_{u,i} - \boldsymbol{y_{u,i}}\right)^2$
- Error Correlation Loss
  - $L^* = \sum \left(s_{u,i} - \boldsymbol{y_{u,i}}\right)^2 + \alpha \boldsymbol{L_{corr}}$
  - L_{corr} regularizes the correlation between prediction errors and the distribution of market segments
    - $L_{\mathrm{corr}} = V/U$, where between-segment error variation: V; within-segment error variation: U;
    - Reflecting the previous fairness metric – F-stat

Q3: How to improve the market fairness of recommendations?

Better Fairness



(a) ModCloth (Rating) **(b) Electronics (Rating)**

Better Recommendation

- The proposed framework **MF (corr.error)** provides a superior recommendation fairness without trading-off much recommendation accuracy

Takeaway & Future

- Marketing bias: a resource of bias for recommendation algorithms
  - Possibly due to the 'self-congruence' effect in the training data
- Calibrating prediction errors across different market segments leads to better recommendation fairness
  - Without trading-off much recommendation accuracy
- Encourage RecSys researchers and practitioners to keep investigating this type of bias
  - Better data collection
  - Comprehensive user study
  - Address in algorithms at scale