Quantifying Herding Effects in Crowd Wisdom

Ting Wang Dashun Wang Fei Wang

KDD

IBM

Crowd Wisdom

"One Vote, One Value."

–Francis Galton. *Nature,* 75:414, 1907

(In)dependency of Minds

(In)dependency of Minds

Interaction with each other when forming consensus

Lorenz et al. “How Social Influence can Undermine the Wisdom of Crowd Effect”. *PNAS,* 108(22):9020-9025, 2011

(In)dependency of Minds

Interaction with each other when forming consensus

Lorenz et al. “How Social Influence can Undermine the Wisdom of Crowd Effect”. *PNAS,* 108(22):9020-9025, 2011

Exposure to prior opinions before forming our own

Salganik et al. “Experimental Study of Inequality and Unpredictability in An Artificial Cultural Market”. *Science,* 311(5762):854-856, 2006

Muchnik et al. “Social Influence Bias: A Randomized Experiment”. *Science,* 341(6146):647-651, 2013

Data

Amazon real customer rating data

- Span of 18 years
- 2.4 million products
- 35 million ratings
- One-to-five star rating system

Category	# Products	# Ratings	Average Rating	Average Entropy
Books	929,264	12,886,488	4.271	0.666
Music	556,814	6,396,350	4.410	0.555
Movies & TV	212,836	7,850,072	3.944	0.955
Electronics	82,067	1,241,778	3.791	0.824

Phantom of Herding Effects

Phantom of Herding Effects

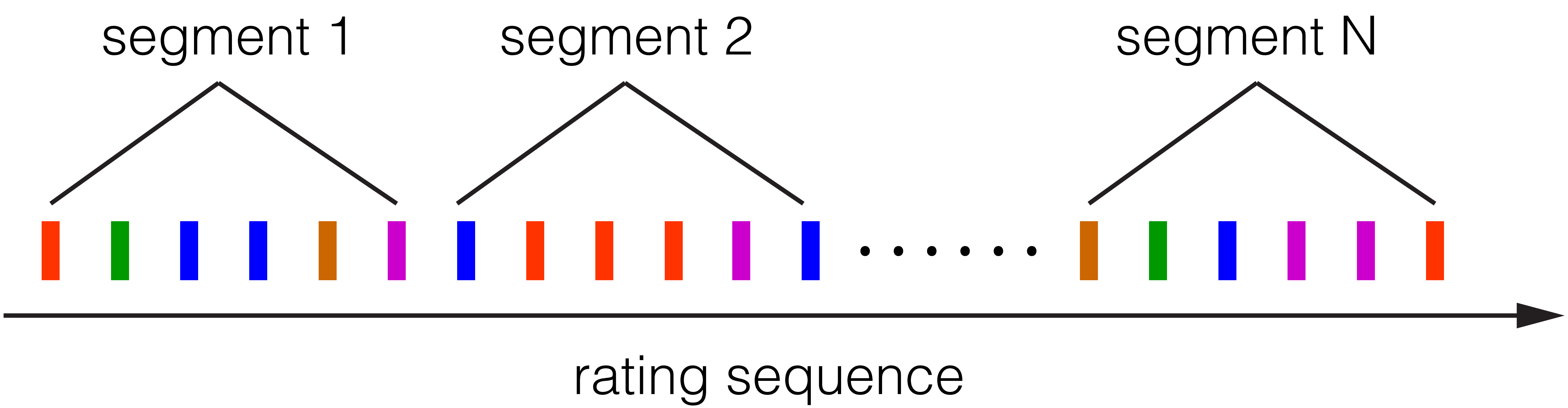
Herding effects-agnostic model

- Ratings drawn from common hidden distribution
- Each rating generated independently
- Segments statistically homogeneous

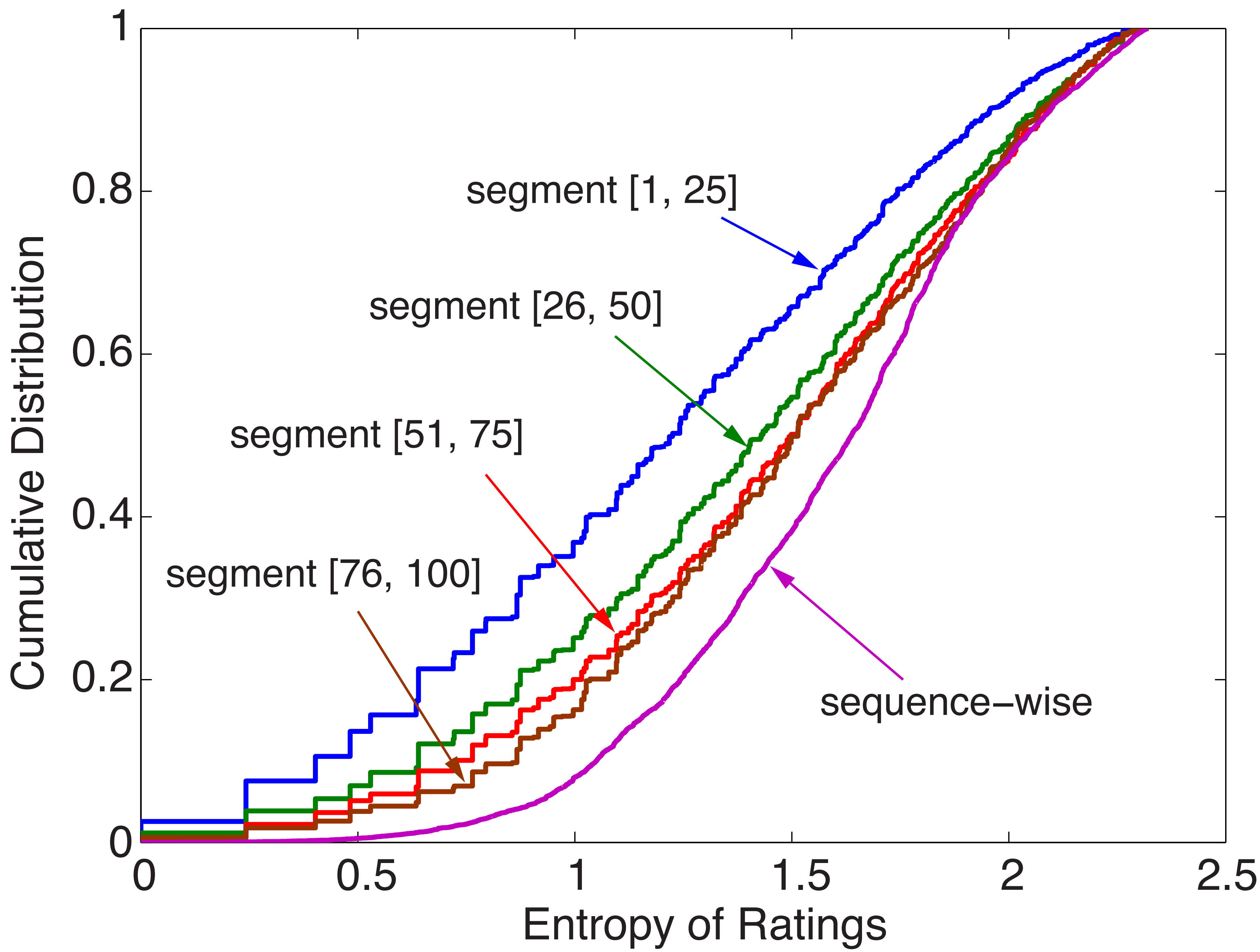
Phantom of Herding Effects

Herding effects-agnostic model

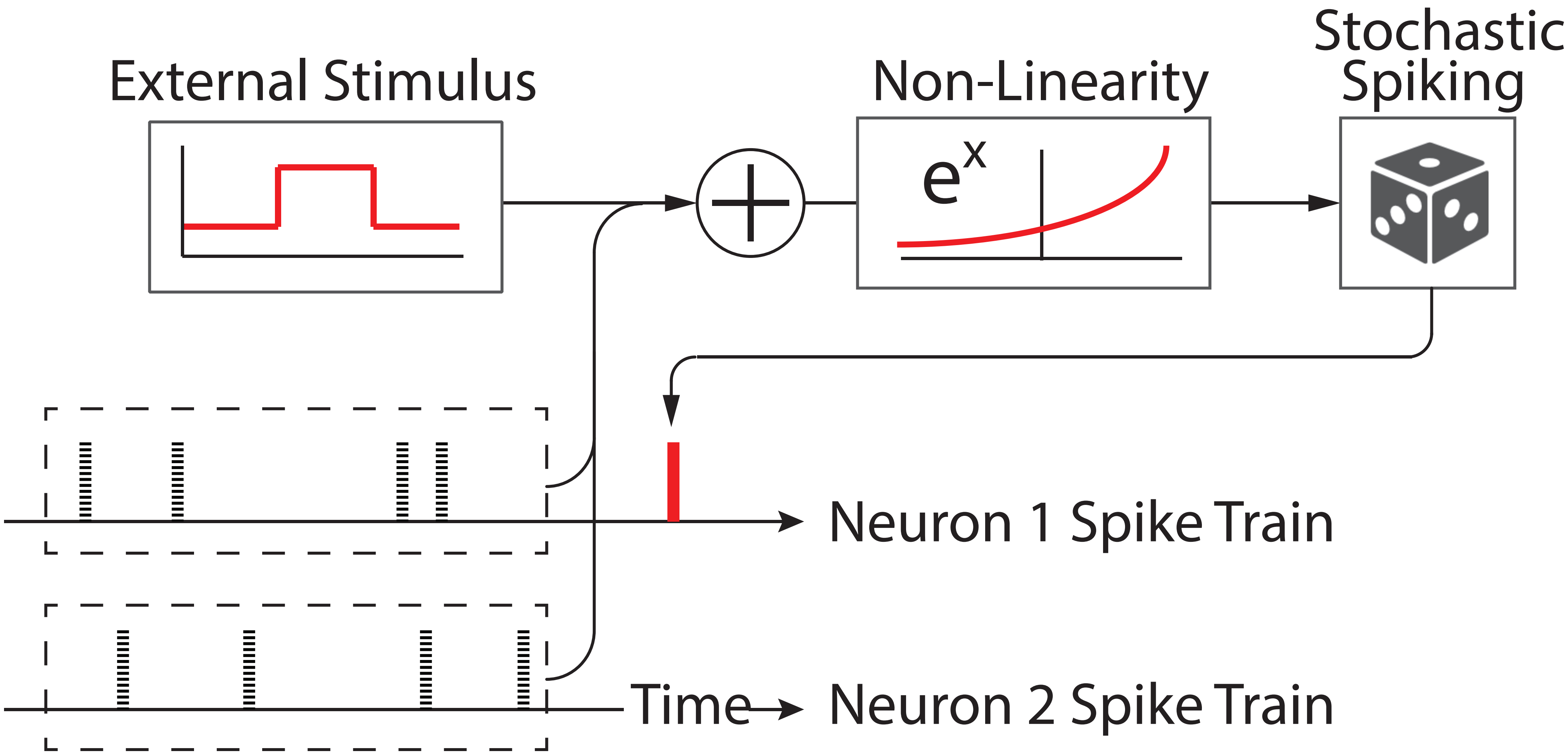
- Ratings drawn from common hidden distribution
- Each rating generated independently
- Segments statistically homogeneous



Phantom of Herding Effects

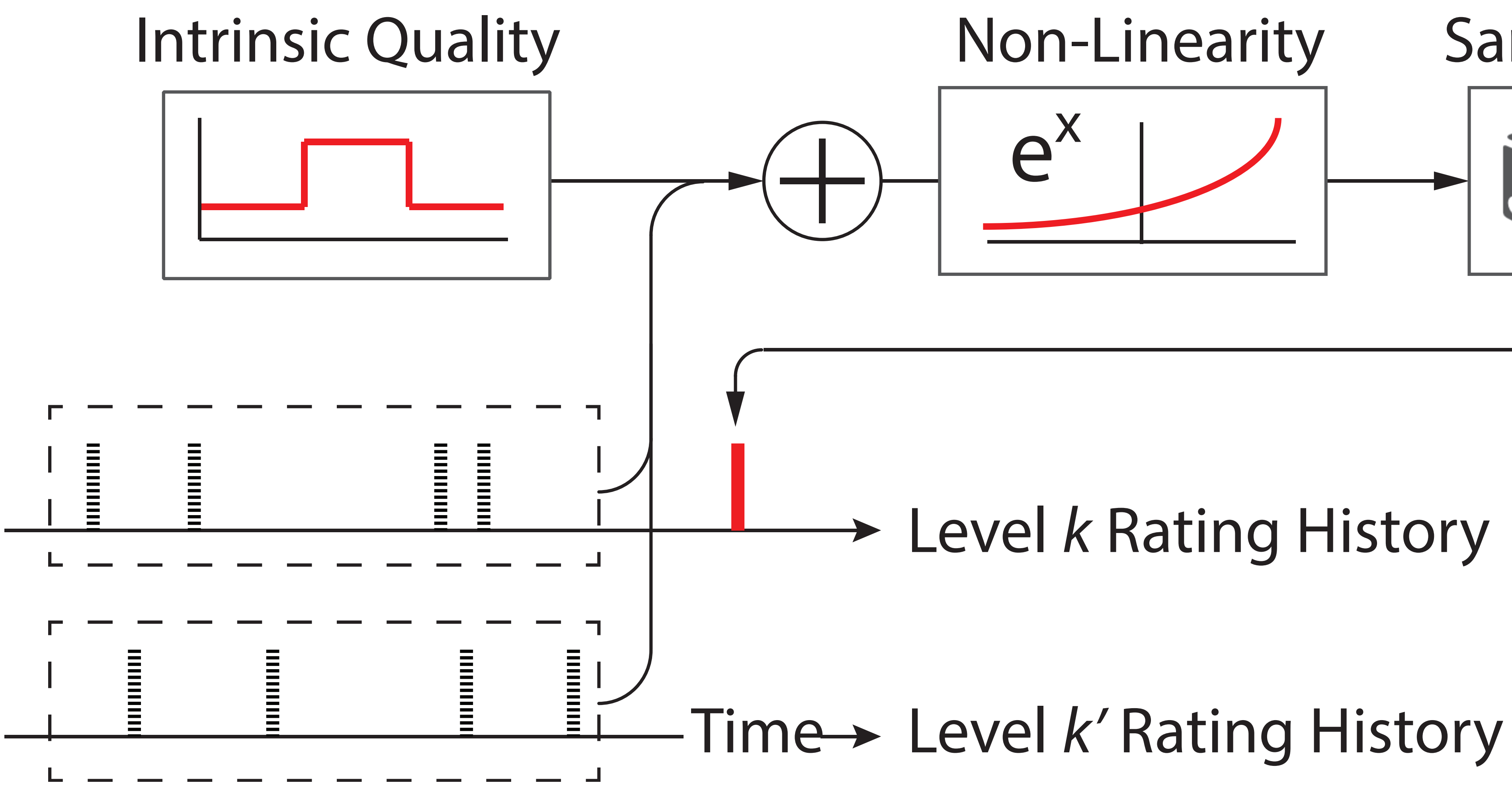


Coupled-Neuron Spiking



J. Pillow et al. “Spatio-Temporal Correlations and Visual Signaling in A Complete Neuronal Population”. *Nature,* 454(7206):995-999, 2008

Dynamics of Rating Growth



Dynamics of Rating Growth

$$Pr(\underbrace{r_i}_{\text{the i-th rating}} = k|x_i) = \frac{1}{Z}\exp(\mu_k + f(i)\theta_k^\top x_i)$$

Dynamics of Rating Growth

$$Pr(r_i = k | x_i) = \frac{1}{Z} \exp(\mu_k + f(i)\theta_k^\top x_i)$$

rating history

Dynamics of Rating Growth

$$Pr(r_i = k | x_i) = \frac{1}{Z} \exp(\mu_k + f(i)\theta_k^\top x_i)$$

$\mu = [\mu_1, \mu_2, \ldots, \mu_K]$: intrinsic quality coefficient

Dynamics of Rating Growth

$$Pr(r_i = k | x_i) = \frac{1}{Z} \exp(\mu_k + f(i) \theta_k^\top x_i)$$

magnitude function : influence of history length

Dynamics of Rating Growth

$$Pr(r_i = k | x_i) = \frac{1}{Z} \exp(\mu_k + f(i)\theta_k^\top x_i)$$

$\theta_k = [\theta_{k1}, \theta_{k2}, \ldots, \theta_{kK}]$: component weights

Objective Function

$$\mathcal{L}_\lambda(\Theta) = -\mathcal{L}(\Theta) + \frac{\lambda}{2}\left(||\Theta||_F^2 + \mathcal{R}(f)\right)$$

$$\begin{cases} \Theta = [\theta_1, \theta_2, \ldots, \theta_K, \mu] \\ \mathcal{L}(\Theta) = \frac{1}{N} \log \prod_{i=1}^{N} Pr\left(r_i | x_i, \Theta\right) \\ \mathcal{R}(f) == \int_0^\infty (f'(t))^2 \mathrm{d}t \end{cases}$$

Model Inference

Surrogate function to decouple parameters

$$\begin{cases} Q(\boldsymbol{\Theta}; \boldsymbol{\Theta}^{(n)}) \geq \mathcal{L}_\lambda(\boldsymbol{\Theta}) & \forall \boldsymbol{\Theta}, \boldsymbol{\Theta}^{(n)} \\ Q(\boldsymbol{\Theta}^{(n)}; \boldsymbol{\Theta}^{(n)}) = \mathcal{L}_\lambda(\boldsymbol{\Theta}^{(n)}) & \forall \boldsymbol{\Theta}^{(n)} \end{cases}$$

Euler-Lagrange equation to fit functional equation

$$\min_{f \in L_1(\mathbb{R})} \sum_i A_i f_i^2 + \sum_i B_i f_i + \frac{\lambda}{2} \int_0^{+\infty} (f'(t))^2 \mathrm{d}t$$

Iteration between updating parameters and updating functional

Predicting Rating Growth

HEARD - Herding Effects-Aware Rating Dynamics Model

IMG - Independent Multinomial Generative Model

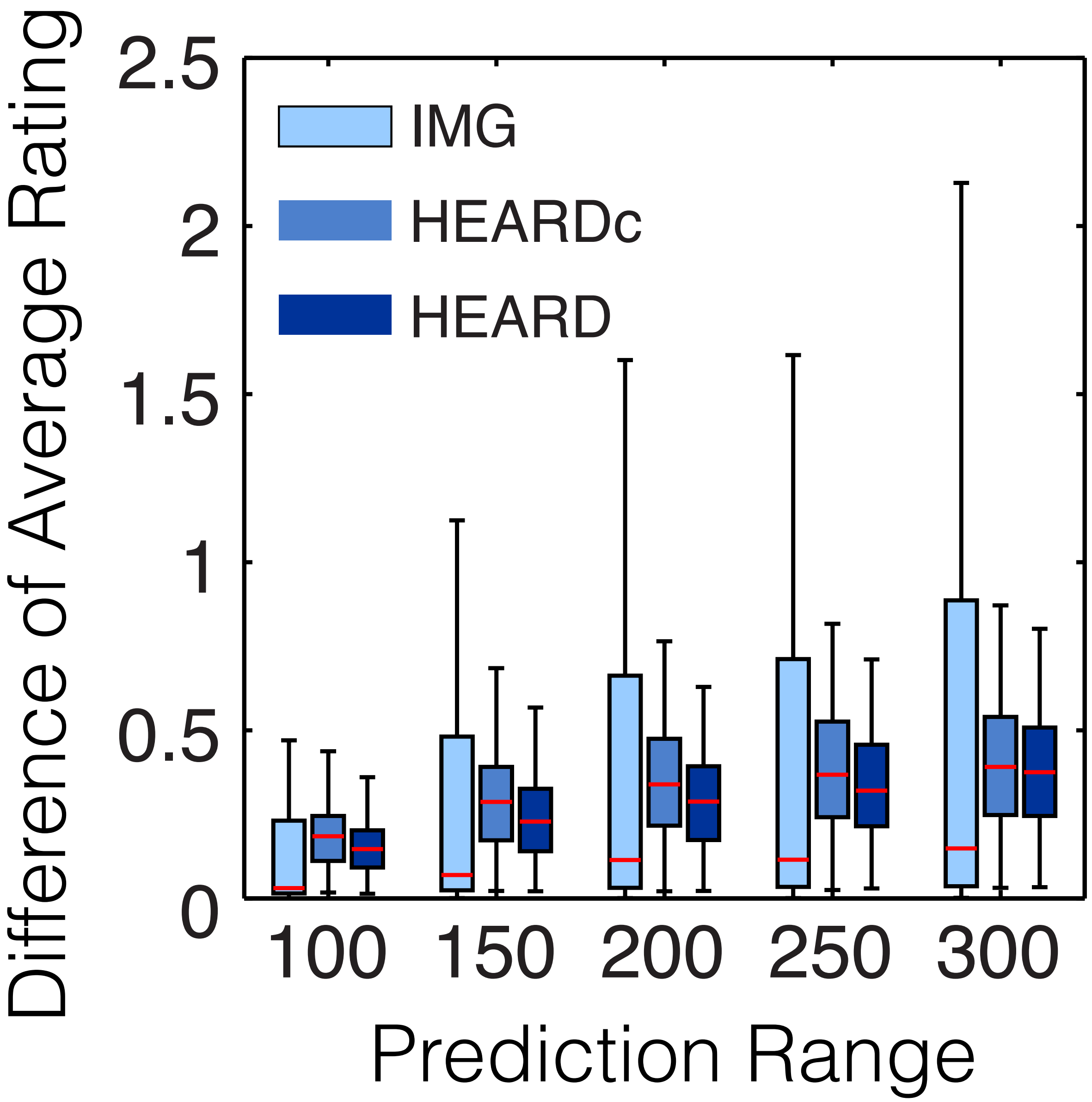
HEARD_c - Constant HEARD Model

Predicting Rating Growth

HEARD - Herding Effects-Aware Rating Dynamics Model

IMG - Independent Multinomial Generative Model

HEARD_c - Constant HEARD Model



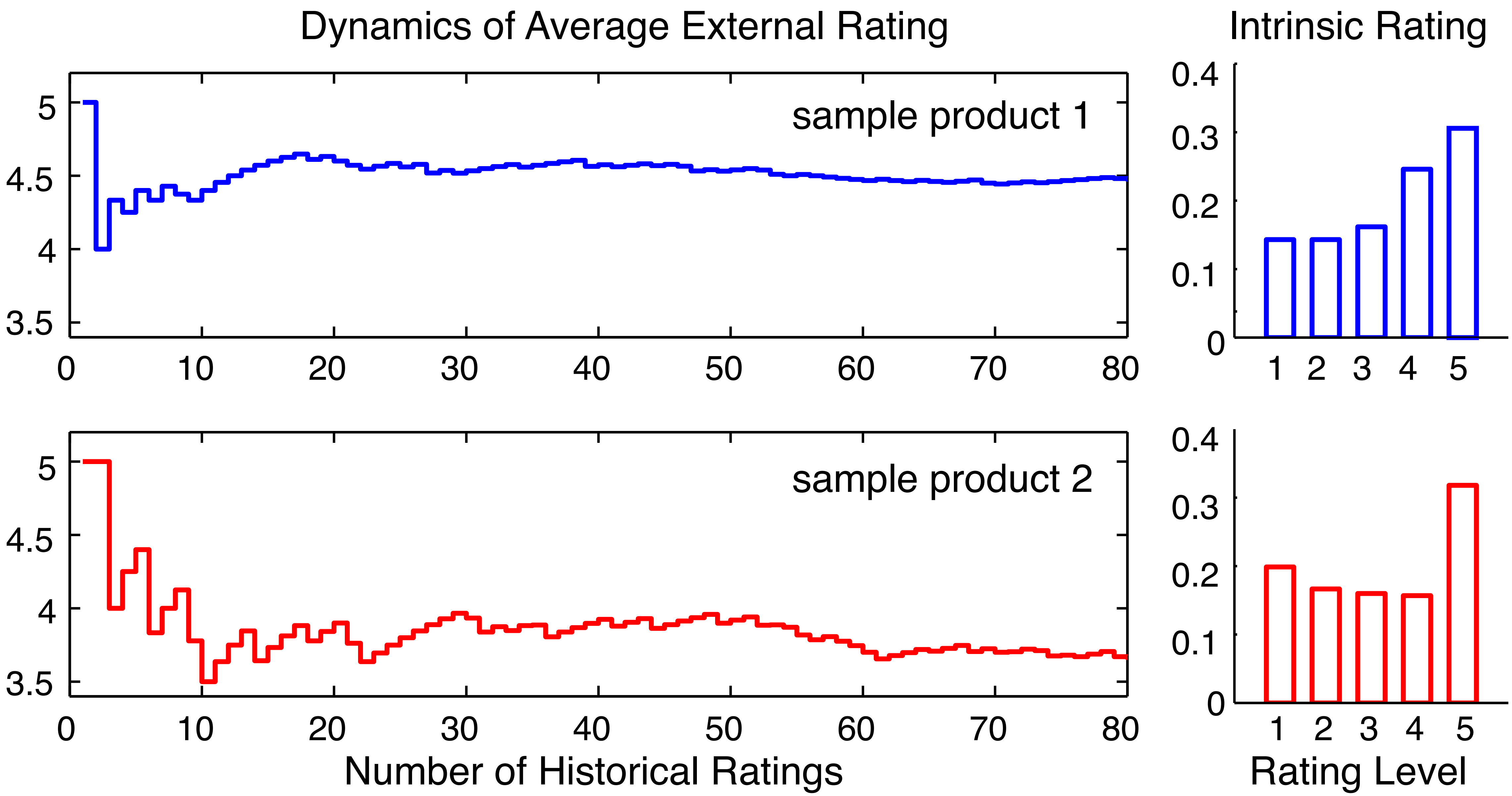
Debiasing Collective Ratings

Debiasing Collective Ratings

$$\boldsymbol{\eta} = \frac{\exp(\boldsymbol{\mu})}{\sum_k \exp(\mu_k)}$$

Debiasing Collective Ratings

$$\boldsymbol{\eta} = \frac{\exp(\boldsymbol{\mu})}{\sum_k \exp(\mu_k)}$$



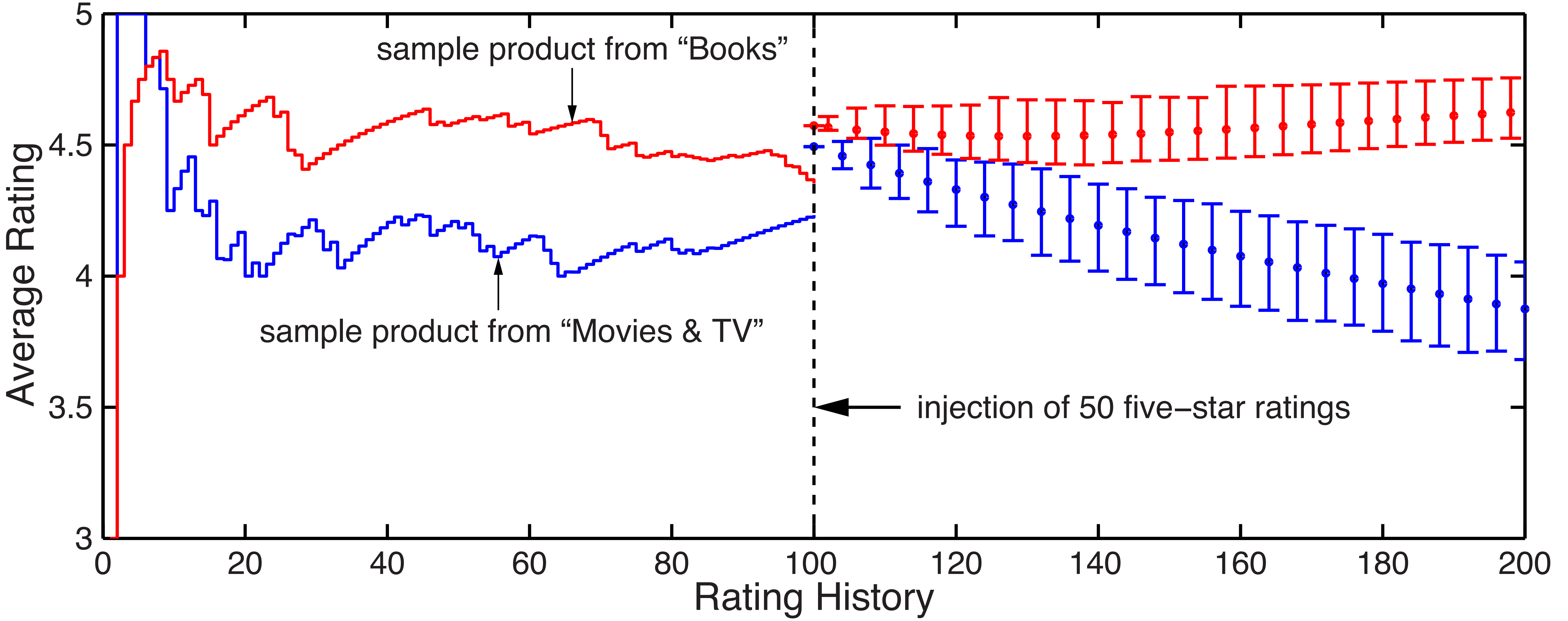
What-If Analysis

What-If Analysis

$$Pr\left(\boldsymbol{x}_{i+1} = \frac{i-1}{i}\boldsymbol{x}_i + \frac{\boldsymbol{e}_k}{i} \,\middle|\, \boldsymbol{x}_i\right)$$

What-If Analysis

$$Pr\left(\boldsymbol{x}_{i+1} = \frac{i-1}{i}\boldsymbol{x}_i + \frac{\boldsymbol{e}_k}{i} \middle| \boldsymbol{x}_i\right)$$



Conclusions

The first mechanistic framework to model the herding effects in crowd wisdom

- De-biasing collective opinions by factoring out herding effects
- Predicting short/long term trajectories of collective opinions
- Performing what-if analysis to untangle manipulations

Future directions

- Text and social aspects of ratings
- Fraudulent and fake ratings
- Temporal dynamics

tingwang@us.ibm.com

Surrogate Function

$$
\begin{aligned}
Q(\Theta;\Theta^{(n)}) = \frac{1}{N}\sum_i\sum_k \left(\phi_{i,k}^2 + \left(\beta_{i,k}^{(n)} - 2\phi_{i,k}^{(n)} - y_{i,k}\right)\phi_{i,k}\right) \\
-\frac{1}{NK}\sum_i \left(\sum_k \phi_{i,k} - 2\sum_k \phi_{i,k}^{(n)}\right)\left(\sum_k \phi_{i,k}\right)
\end{aligned}
$$

$$
\begin{aligned}
\phi_{i,k} &= \mu_k + f_i \theta_k^\top x_i \\
\phi_{i,k}^{(n)} &= \mu_k^{(n)} + f_i^{(n)} \theta_k^{(n)\top} x_i \\
\beta_{i,k}^{(n)} &= \frac{\exp\left(\phi_{i,k}^{(n)}\right)}{\sum_{k'} \exp\left(\phi_{i,k'}^{(n)}\right)} \\
C_i^{(n)} &= \sum_k \left(\phi_{i,k}^{(n)2} - \beta_{i,k}^{(n)}\phi_{i,k}^{(n)}\right) - \frac{1}{K}\left(\sum_k \phi_{i,k}^{(n)}\right)^2 + \log\sum_k \exp\left(\phi_{i,k}^{(n)}\right)
\end{aligned}
$$

Charactering Herding Effects