北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University

NEURAL INFORMATION
PROCESSING SYSTEMS

Calibrating “Cheap Signals” in Peer Review without a Prior

Yuxuan Lu, Yuqing Kong

Peking University

NeurIPS 2023

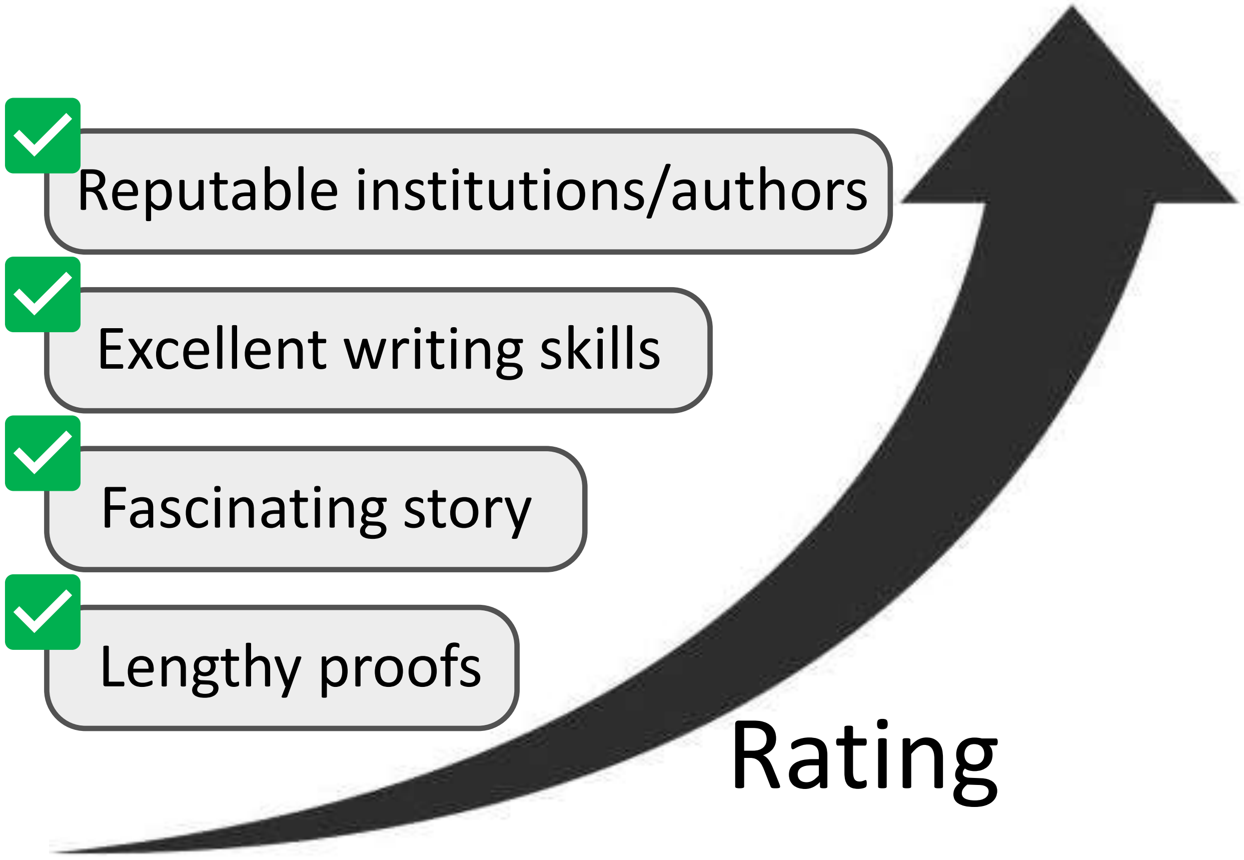
北京大学计算机学院
School of Computer Science

Bias in Peer Review

- Reviewer-specific bias:
  - Conflict of interest
  - Pre-existing Beliefs
  - Stringent or lenient standard
  - …

- Systematic bias
  - Identity-based bias
  - Resistance to emerging fields
  - Lack of qualified reviewers
  - …

Issue: Cheap Signals



- ✓ Reputable institutions/authors
- ✓ Excellent writing skills
- ✓ Fascinating story
- ✓ Lengthy proofs

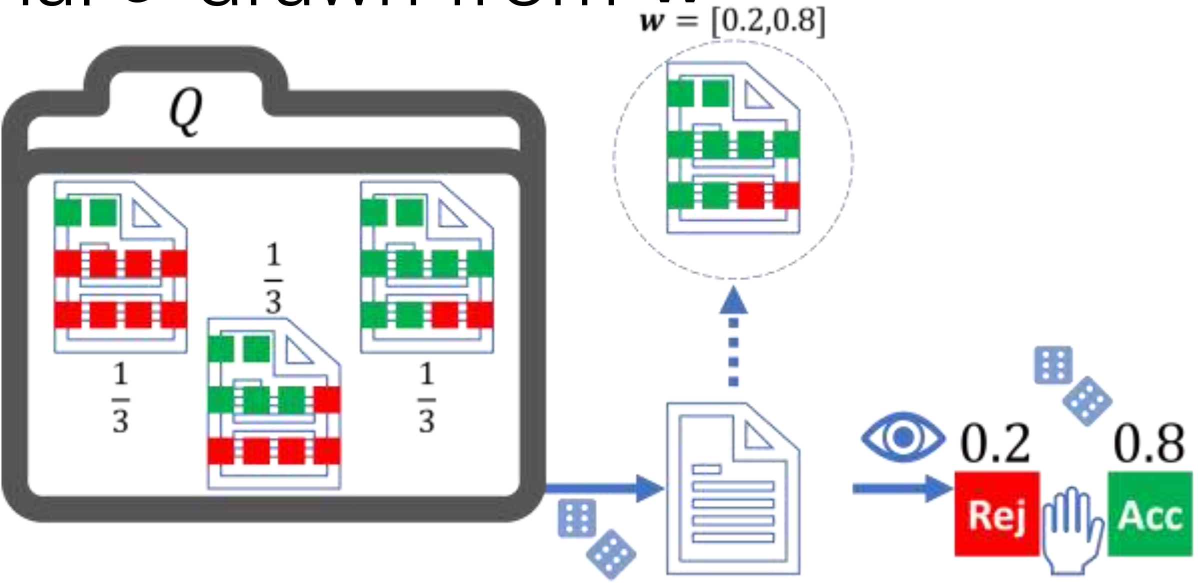
Rating

- Systematic bias
  - Identity-based bias
  - Resistance to emerging fields
  - Lack of qualified reviewers
  - …

Cheap signals

Modelling without Cheap Signal

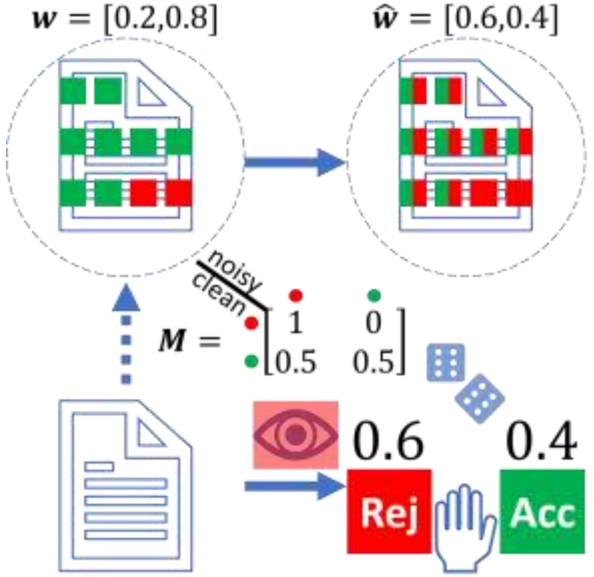
- The set of possible signals $\Sigma = \{0\ (\text{rej}), 1\ (\text{acc})\}$
- Paper state $\mathbf{w} \in \left\{ \begin{matrix} \text{bad}\ (\mathbf{w} = [.8, .2]) \\ \text{fair}\ (\mathbf{w} = [.5, .5]) \\ \text{good}\ (\mathbf{w} = [.2, .8]) \end{matrix} \right\}$
- Each reviewer receives i.i.d signal σ drawn from $\mathbf{w}$
- Prior $\boldsymbol{Q} = \frac{1}{3}\text{bad}, \frac{1}{3}\text{fair}, \frac{1}{3}\text{good}$



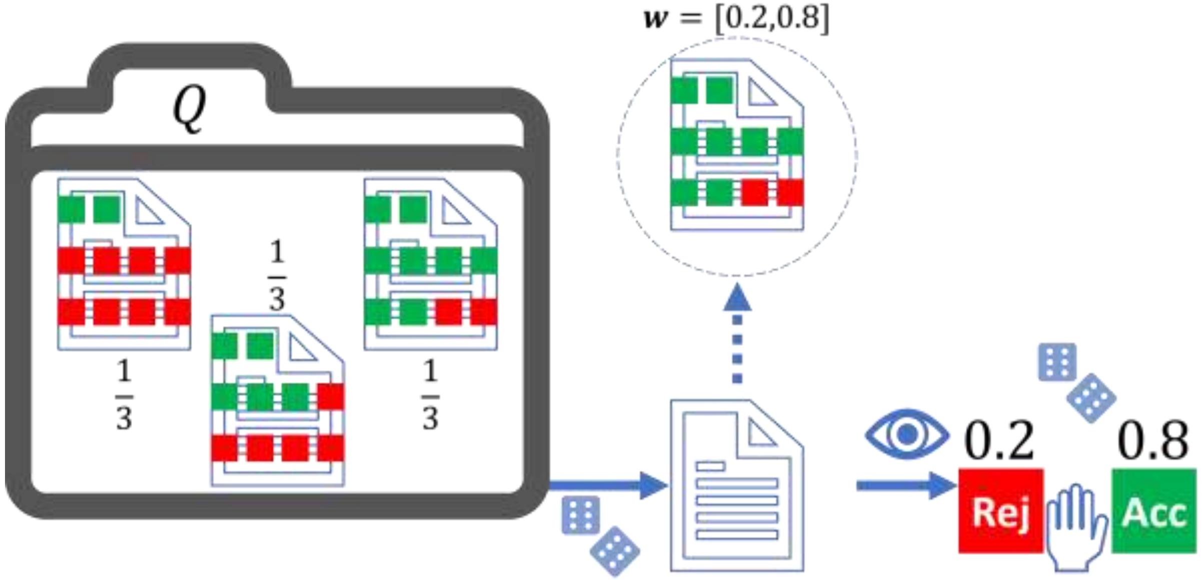
Modelling Cheap Signals

- Regard cheap signals as a bias operator M
  - Bias M alters reviewer's clean signal σ to a biased signal $\hat{\sigma} = M(\sigma)$
- Reviewer only obtains $\hat{\sigma}$ without realizing σ

Good paper + negative bias


$\boldsymbol{w} = [0.2,0.8]$
$\hat{\boldsymbol{w}} = [0.6,0.4]$
noisy
clean
$\boldsymbol{M} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}$
0.6
0.4
Rej
Acc

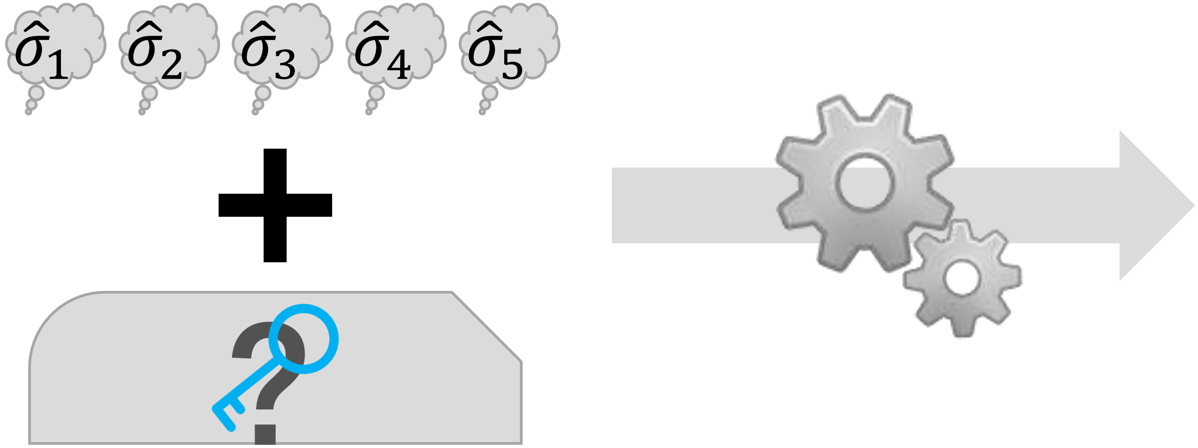
Ideal world with no bias



Target: Calibrating Cheap Signals

- We want a mechanism that, in a biased world, rank the quality of papers as if we have the clean signals.

- What additional information should we elicit?



$$f\left(\frac{\sigma_1+\sigma_2+\sigma_3+\sigma_4+\sigma_5}{5}\right)$$

Key Observations

Cheap signals affects reviewers' prior beliefs

$$\boldsymbol{Q} = \frac{1}{3}\text{bad}, \frac{1}{3}\text{fair}, \frac{1}{3}\text{good}$$

bad: $\mathbf{w} = [0.8, 0.2]$

fair: $\mathbf{w} = [0.5, 0.5]$

good: $\mathbf{w} = [0.2, 0.8]$

$$\widehat{\boldsymbol{Q}} = \frac{1}{3}\widehat{\text{bad}}, \frac{1}{3}\widehat{\text{fair}}, \frac{1}{3}\widehat{\text{good}}$$

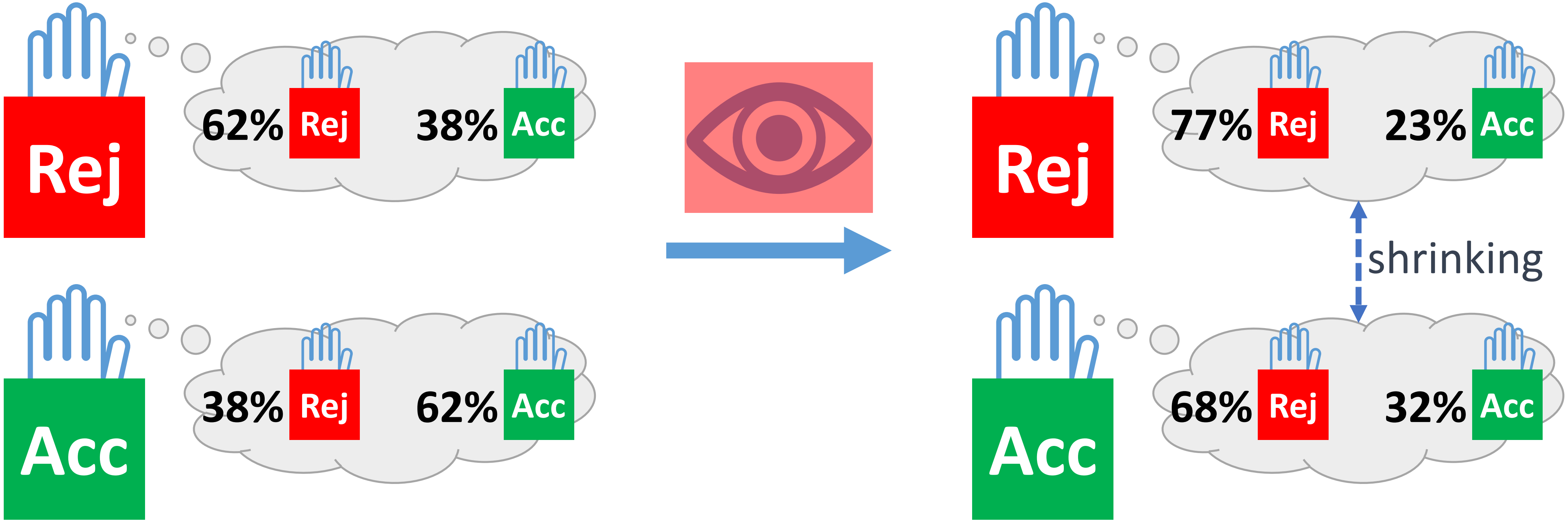
$\widehat{\text{bad}}$: $\widehat{\mathbf{w}} = [0.9, 0.1]$

$\widehat{\text{fair}}$: $\widehat{\mathbf{w}} = [0.75, 0.25]$

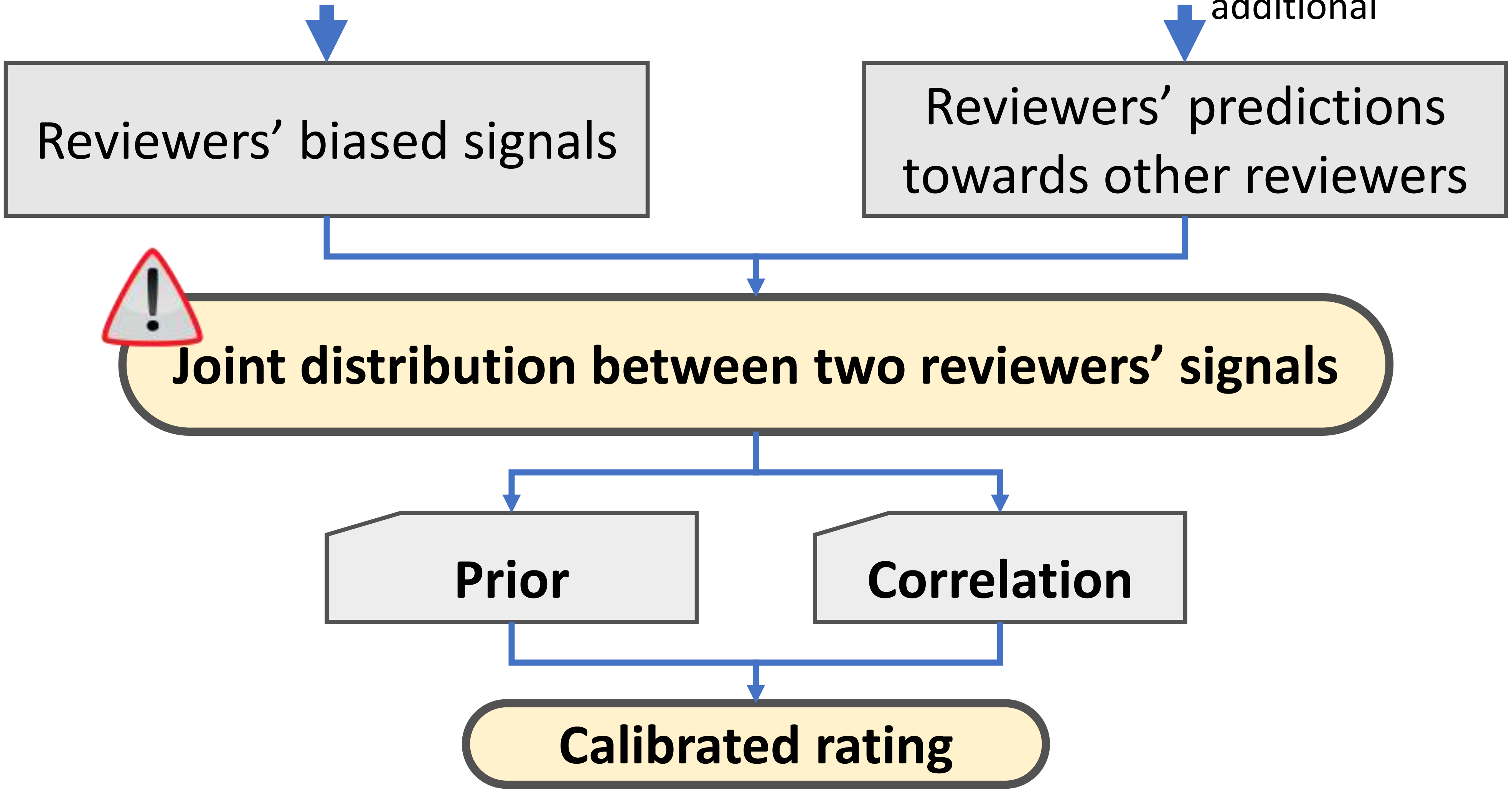
$\widehat{\text{good}}$: $\widehat{\mathbf{w}} = [0.6, 0.4]$

Key Observations

Cheap signals weaken reviewer feedback correlation



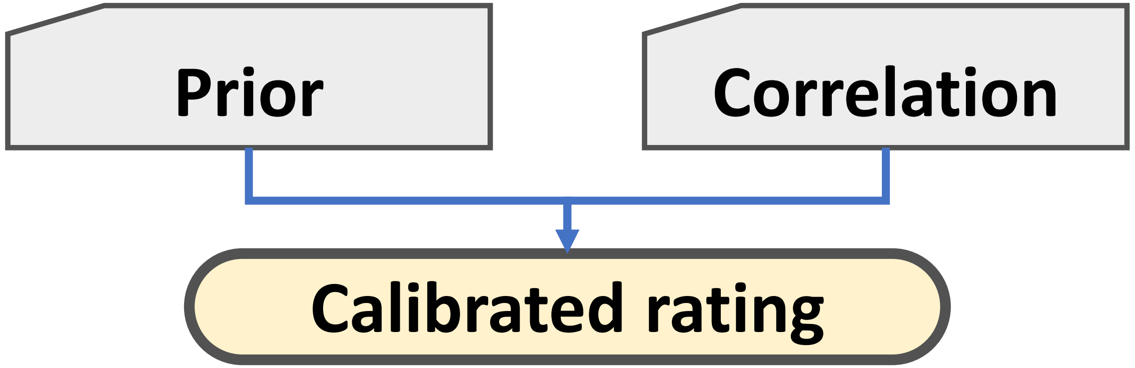
Main Idea: Calibration by Prediction



additional
Reviewers’ biased signals
Reviewers’ predictions towards other reviewers
Joint distribution between two reviewers’ signals
Prior
Correlation
Calibrated rating

Main Idea: Calibration by Prediction

Theorem (informal): the calibrated rating is an affine transformation of the true rating in expectation.



NEURAL INFORMATION PROCESSING SYSTEMS

Thank you for listening!

Contact: yx_lu@pku.edu.cn

Materials: https://yxlu.me/publication/peer_review_neurips23