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

- Bias
  - Reviewer-specific bias
    - Pre-existing Beliefs
    - Conflict of interest
    - Stringent or lenient standard
  - Systematic bias
    - Identity-based bias
    - Resistance to emerging fields
    - Lack of qualified reviewers

Issue: Cheap Signals

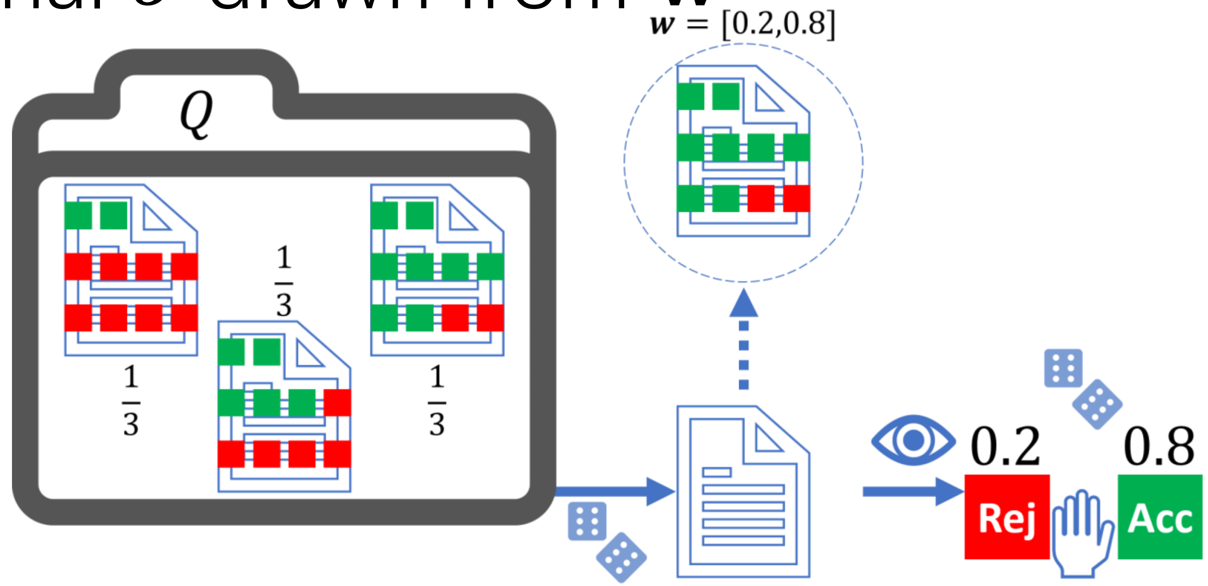
Cheap Signals

Systematic bias

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

Modelling without Cheap Signal

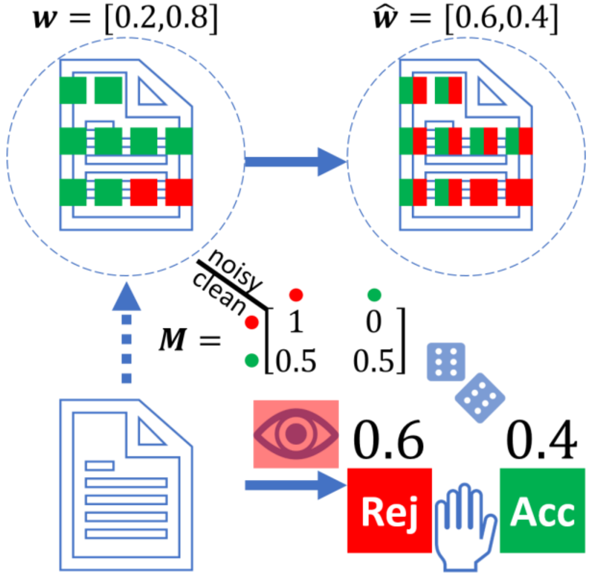
- The set of possible signals $\Sigma = \{0\ (\text{rej}), 1\ (\text{acc})\}$
- Paper state $\mathbf{w} \in \begin{cases} \text{bad } (\mathbf{w} = [.8, .2]) \\ \text{fair } (\mathbf{w} = [.5, .5]) \\ \text{good } (\mathbf{w} = [.2, .8]) \end{cases}$
- Each reviewer receives i.i.d signal σ drawn from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



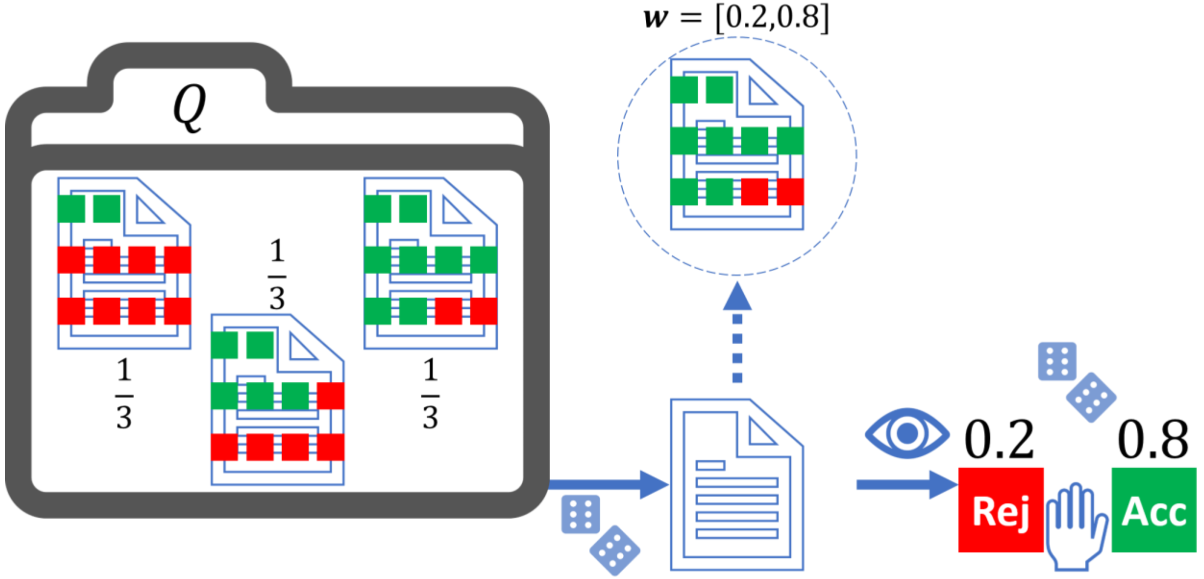
Modelling Cheap Signals

- 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

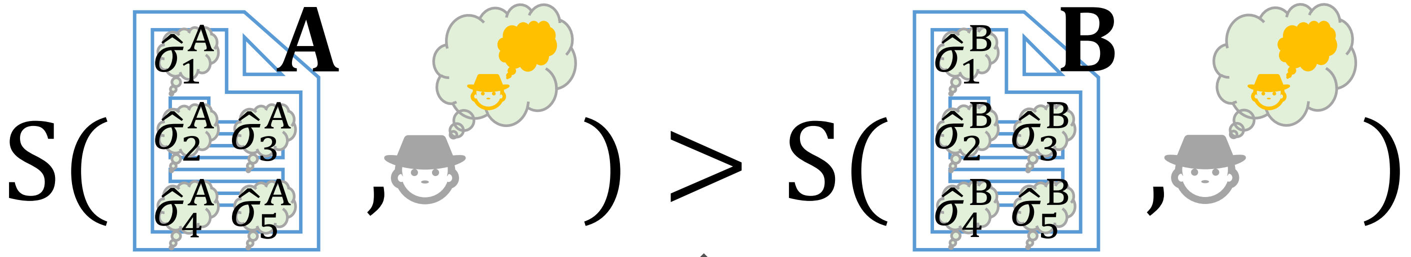


Ideal world with no bias



Target: Calibrating Cheap Signals

- We want a process that, in a biased world, rank the quality of papers as if there is no bias.
  - Our method: additionally collecting **second-order information** from reviewers.



$$\Updownarrow$$

$$\Sigma_{i=1}^{5}\sigma_i^{A} > \Sigma_{j=1}^{5}\sigma_j^{B}$$

Key Observations

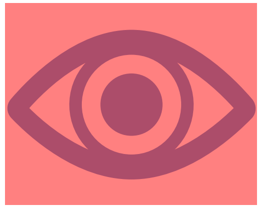
1. Cheap signals affects reviewers' ratings and prior beliefs in the same way

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

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

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

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



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

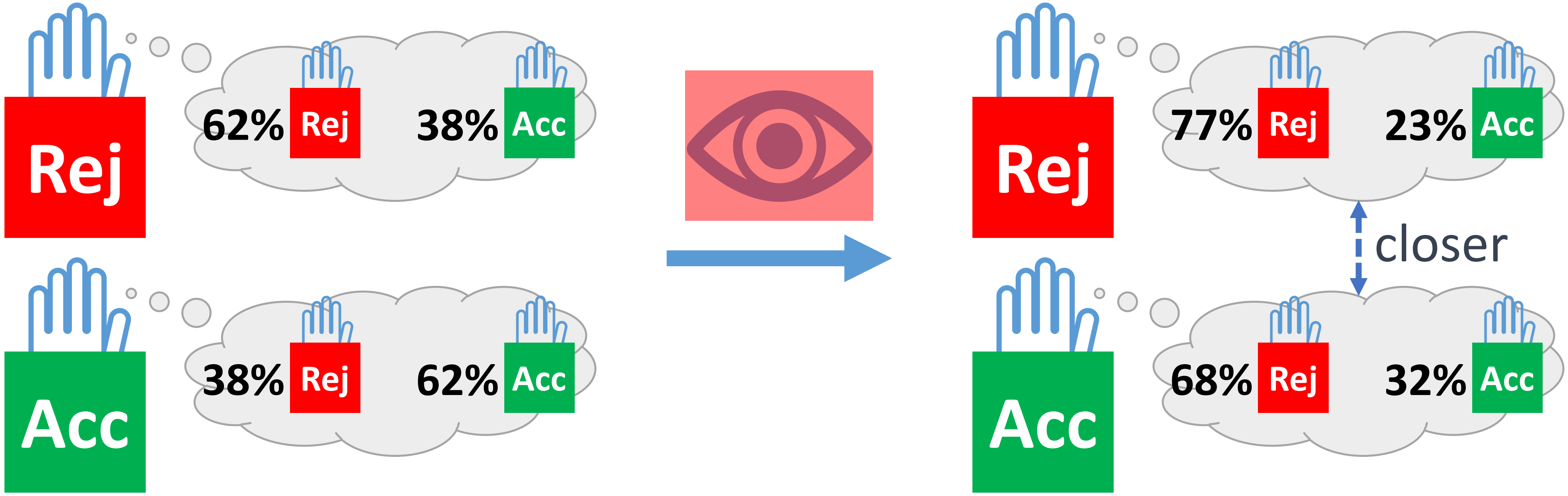
$\widehat{\text{bad}}$: $\widehat{\mathbf{w}} = [0.9, 0.1]$

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

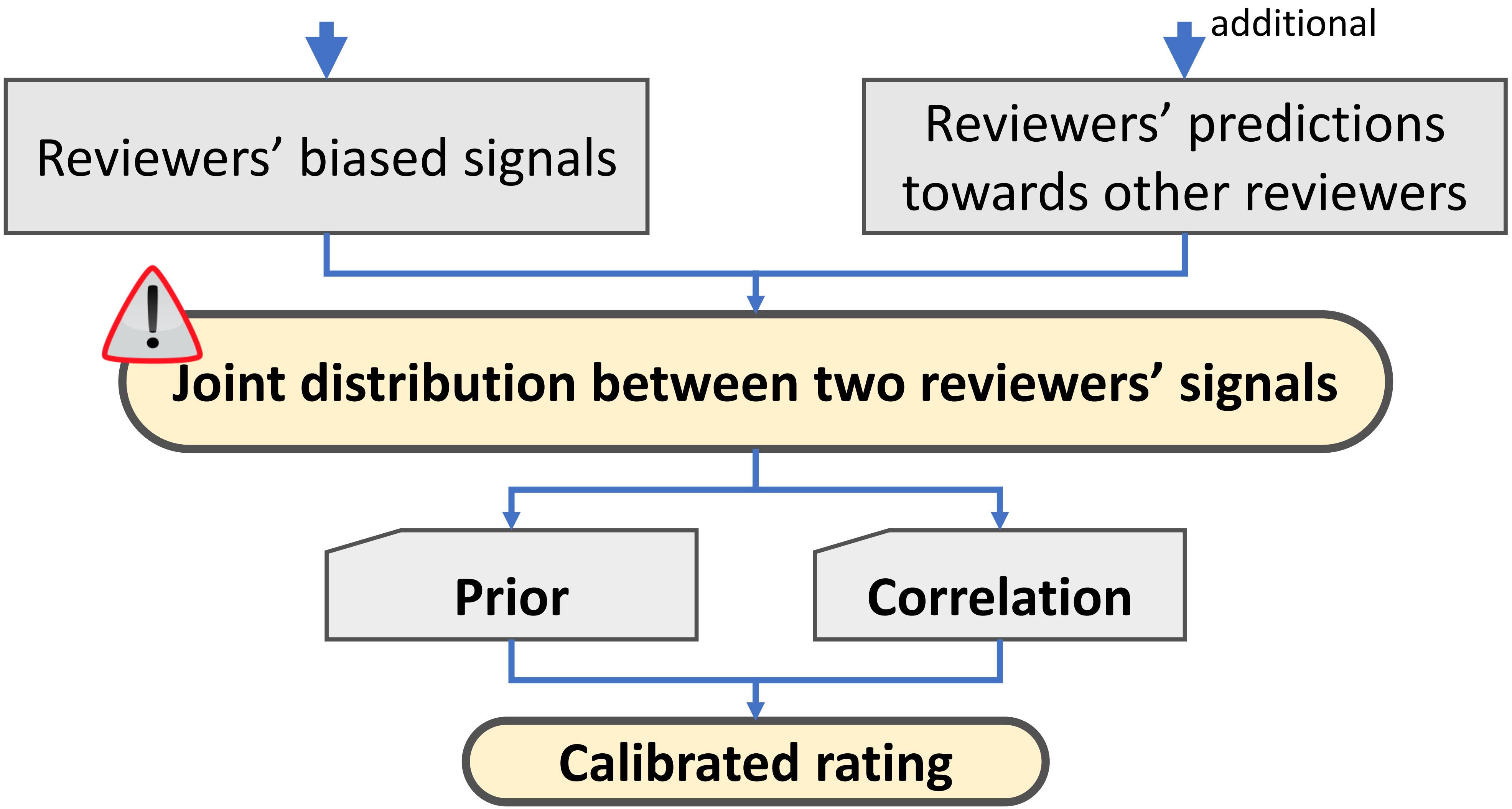
$\widehat{\text{good}}$: $\widehat{\mathbf{w}} = [0.6, 0.4]$

Key Observations

2. Systematic bias weaken the correlation of reviewers' feedback

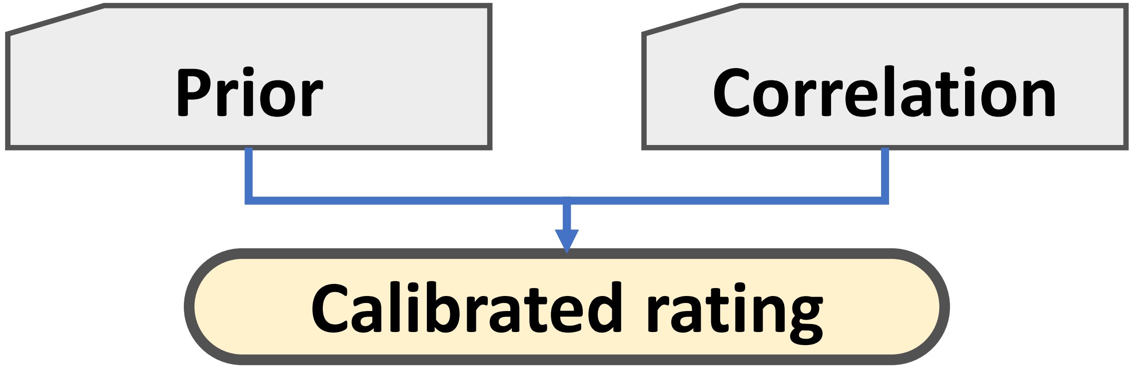


Main Idea: Calibration by Prediction



Main Idea: Calibration by Prediction

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



Thank you for listening!

Contact: yx_lu@pku.edu.cn

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