# Accounting for Confirmation Bias in Crowdsourced Label Aggregation

**Meric Altug Gemalmaz, Ming Yin**

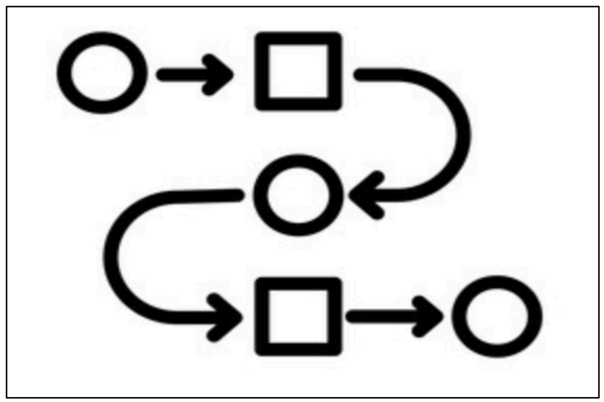
Purdue University

**IJCAI-21, Online, August 2021**

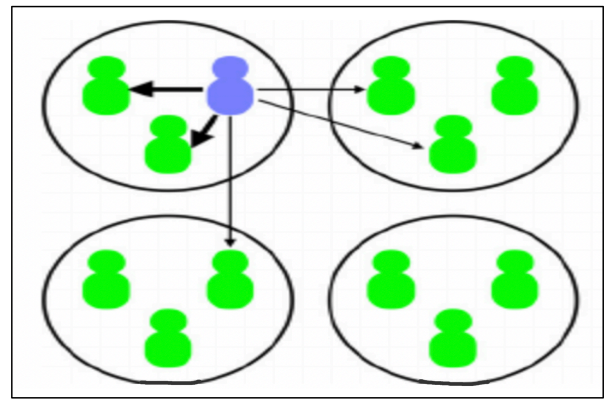
# Cognitive Biases in Crowdsourcing

**Workers are prone to a wide range of biases!**


In-batch annotation bias



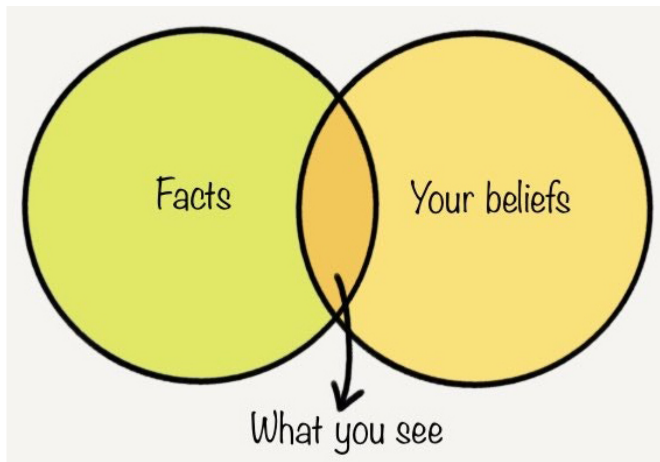
Sequential bias



In-group bias

# Confirmation Bias

**Another common type of cognitive bias is confirmation bias…**




Favoring information that confirms previously existing beliefs and values!

# Label Aggregation for Quality Control

**Redundancy-based strategy is often deployed.**

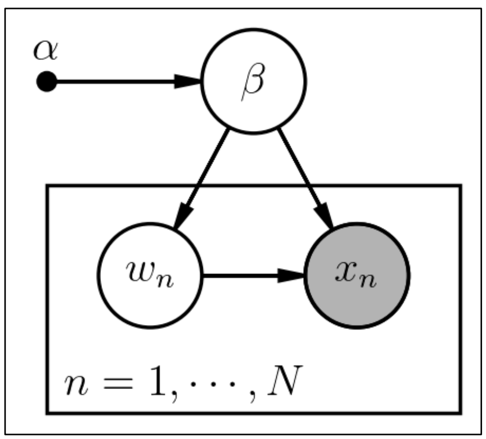
Majority Voting, Label Aggregation Algorithms;

- GLAD [Whitehill et al., 2009]
- Multi [Welinder et al., 2010]
- VI-BP [Liu et al., 2012]
- Minimax [Zhou et al., 2012]
- ZenCrowd [Demartini et al., 2012]
- CBCC [Venanzi et al., 2014]

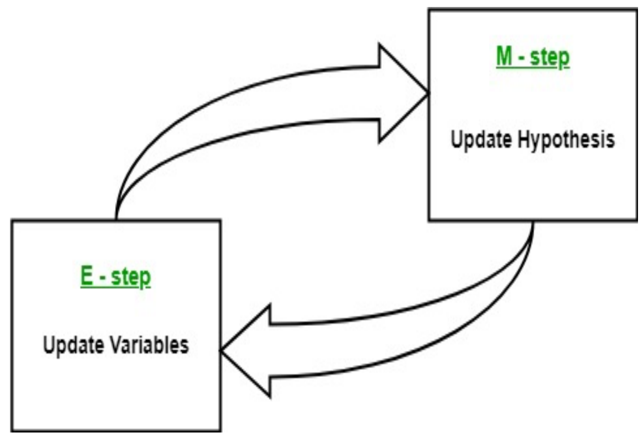
**Methods seldom take worker biases into account!**

# Our Approach

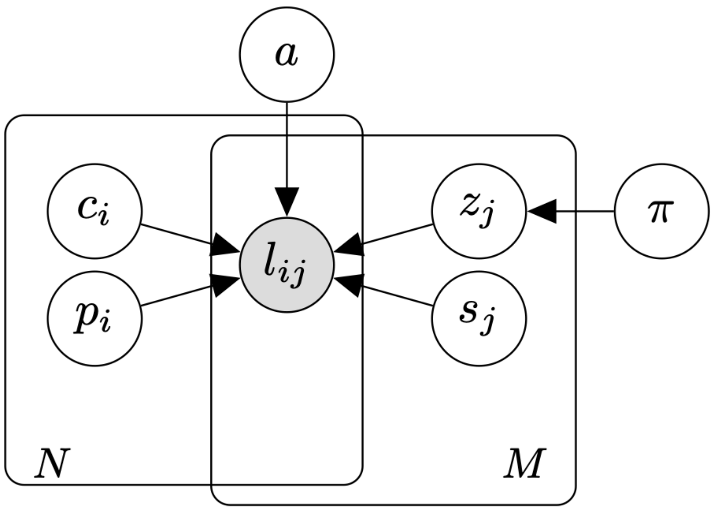
Model explicitly how worker's confirmation bias sneaks into annotations.




Design an algorithm based on the proposed model to reduce bias in the aggregated labels.



# Probabilistic Model of Label Generation




$$P(l_{ij} = 0 | c_i, p_i, s_j, z_j, a) = \frac{1}{e^{a[(1-p_i)(s_j-c_i)^2+p_i z_j]}}$$

$\mathbf{c_i} \in [0, 1]$ : the values of annotator **i**

$\mathbf{s_j} \in [0, 1]$ : the values of information contained in task **j**

$\mathbf{z_j} \in \{0, 1\}$ : ground truth label of the task **j**

$\boldsymbol{\pi}$ = P($\mathbf{z_j}$ = 0) : the prior probability for a task to have the preferable label

$\mathbf{p_i} \in [0, 1]$ : extent to which annotator **i** is subject to confirmation bias

$\mathbf{a} \in [0, +\infty)$ : annotators' base rate of providing the preferable label

# Model Inference

**We then use EM Algorithm to learn model parameters and make inference:**

**Expectation step**, compute the posterior probabilities for the ground truth of each task $\mathbf{z_j}$

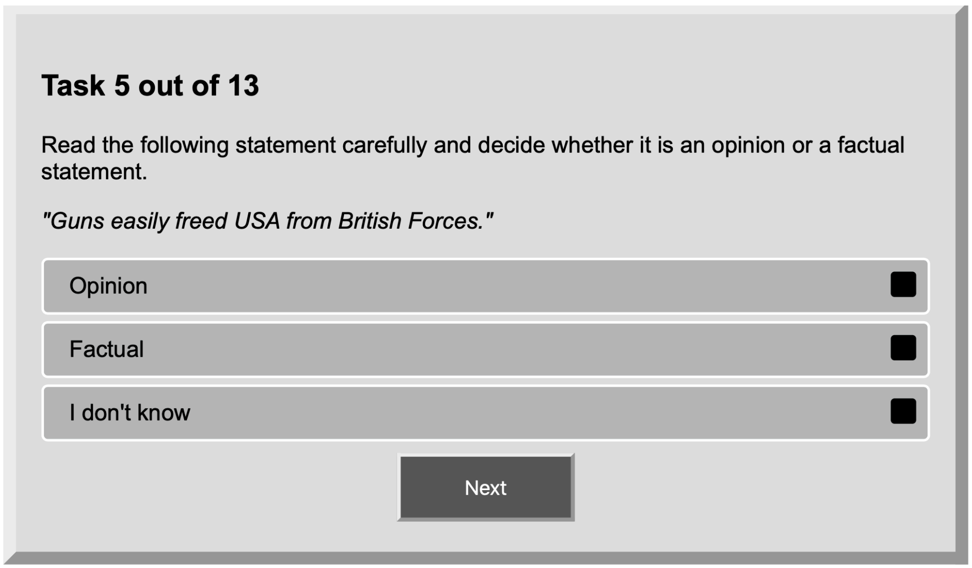
$$p(z_j|\mathbf{L}, \mathbf{c}, \mathbf{p}, \mathbf{s}, a, \pi) \propto p(z_j|\pi)p(\mathbf{L}|z_j, \mathbf{c}, \mathbf{p}, \mathbf{s}, a)$$
$$\propto p(z_j|\pi) \prod_{i \in W_j} p(l_{ij}|c_i, p_i, s_j, z_j, a)$$

**Maximization step,** search for optimal parameter values to maximize the auxiliary function *Q*

$$Q(\mathbf{c}, \mathbf{p}, \mathbf{s}, a, \pi) = E[\ln p(\mathbf{L}, \mathbf{z}|\mathbf{c}, \mathbf{p}, \mathbf{s}, a, \pi)]$$
$$= E[\ln \prod_j (p(z_j|\pi) \prod_{i \in W_j} p(l_{ij}|c_i, p_i, s_j, z_j, a))]$$
$$= \sum_j E[\ln p(z_j|\pi)] + \sum_{l_{ij} \in \mathbf{L}} E[\ln p(l_{ij}|c_i, p_i, s_j, z_j, a)]$$

# MTurk Experiment Task: Subjective Labeling

Label a statement as either “Opinion” or “Factual”

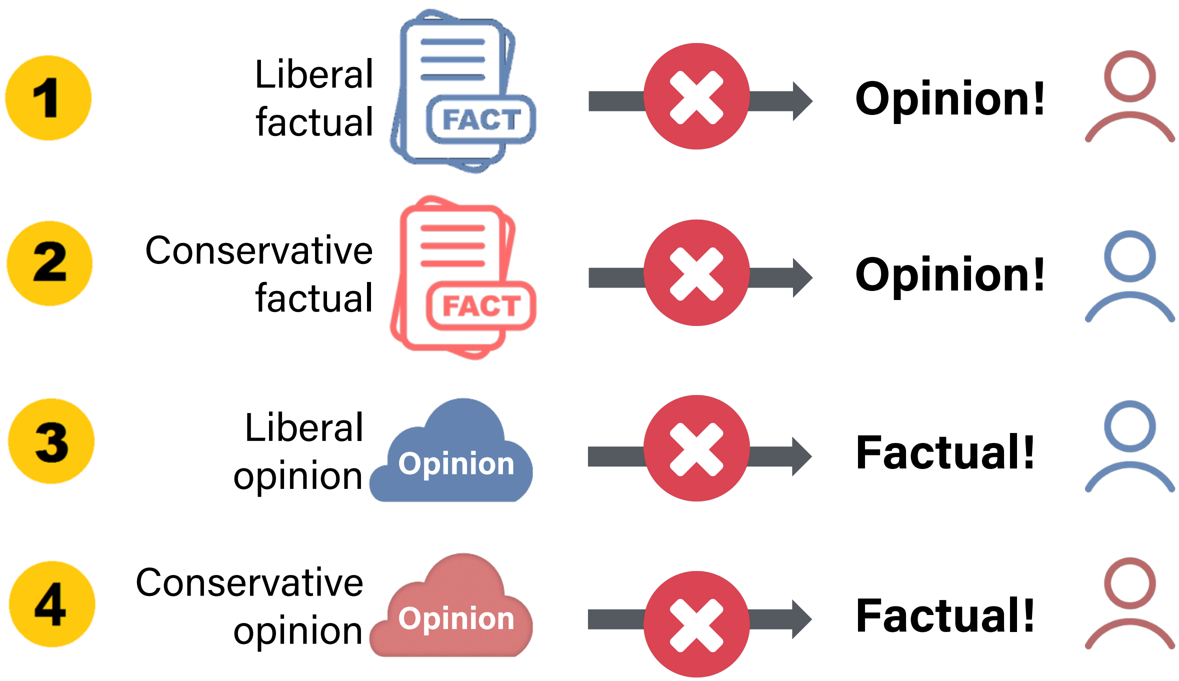



**12** statements (6 liberal + 6 conservative)

**110** workers complete our HIT (57 liberal, 42 conservative, and 11 neutral)

**110 × 12 = 1320** labels generated by workers (8% of the labels are IDK)

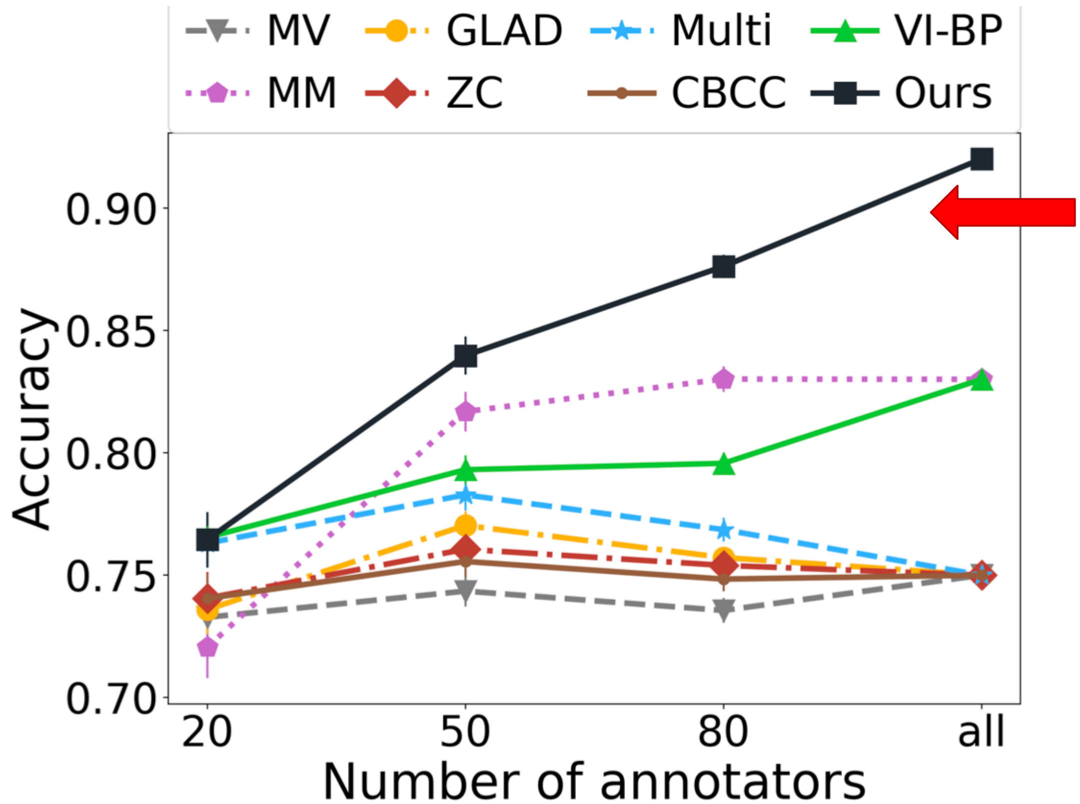
# Do Workers Exhibit Confirmation Bias in This Task?




**Liberal bias score:**
Normalized Type 2 or 3 error rate – Normalized Type 1 or 4 error rate

Workers holding liberal values have larger liberal bias scores!

# Can Our Algorithm Increase Aggregation Accuracy?



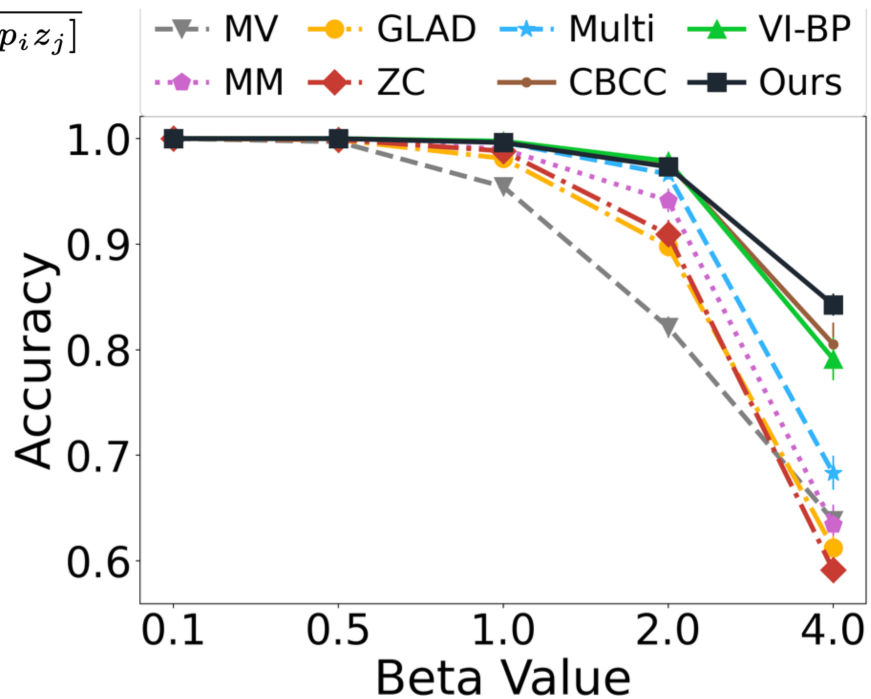

Our algorithm almost always achieves the highest accuracy!

# The Impact of Confirmation Bias Degree

$$P(l_{ij}=0|c_i,p_i,s_j,z_j,a)=\frac{1}{e^{a[(1-p_i)(s_j-c_i)^2+p_iz_j]}}$$

**25** workers, $\mathbf{p_i}$ ~ Beta(1, β)

- The larger β is, the more workers tend to subject to confirmation bias

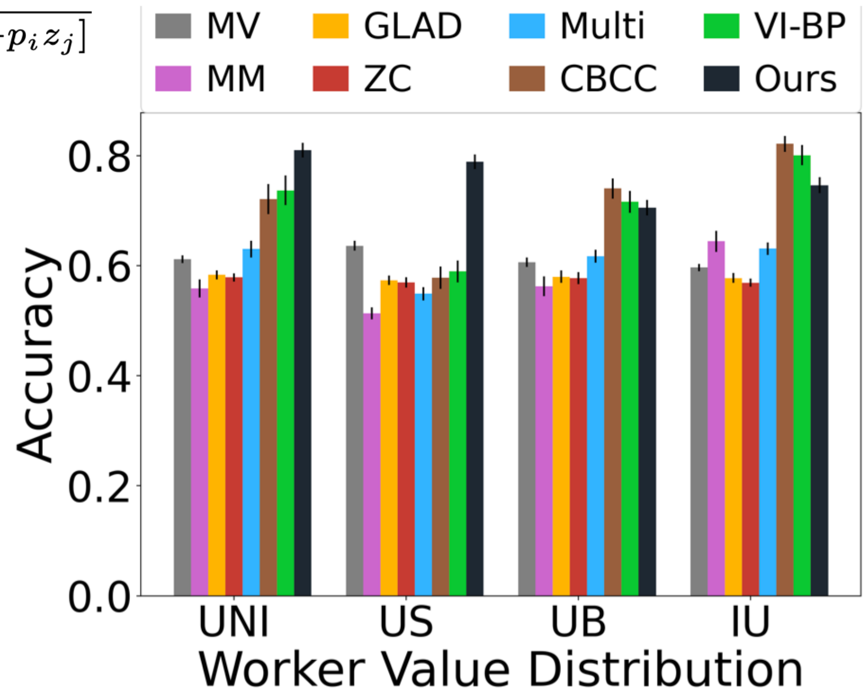


# The Impact of the Distribution of Worker's Values

$$P(l_{ij} = 0 | c_i, p_i, s_j, z_j, a) = \frac{1}{e^{a[(1-p_i)(s_j-c_i)^2 + p_i z_j]}}$$

**25** workers, $\mathbf{c_i} \sim \text{Beta}(a, b)$

- *Uniform* (a=b=1)
- *U-Shape* (a=b=0.5)
- *Unbalanced* (a=1, b=2)
- *Inverse-U shape* (a=b=2)

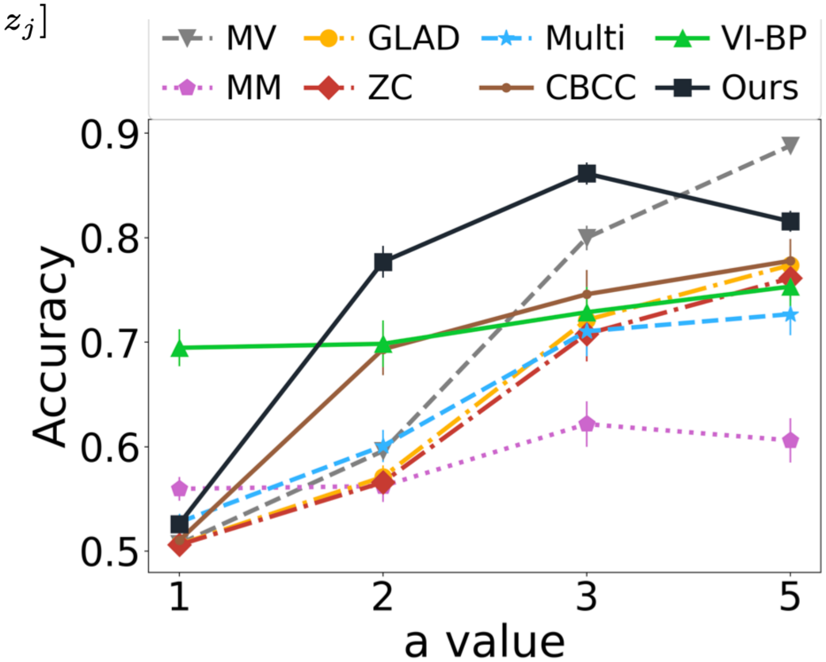


# The Impact of Base Rate of the Preferable Label

$$P(l_{ij} = 0 | c_i, p_i, s_j, z_j, a) = \frac{1}{e^{a[(1-p_i)(s_j-c_i)^2+p_i z_j]}}$$

**25** workers, a ∈ {1, 2, 3, 5}

- The larger a is, the base chance for workers to provide the preferable label is lower



# Summary

- We propose a new label aggregation method that accounts for a worker's cognitive bias (confirmation bias).

- We evaluate the effectiveness of our approach through both MTurk experiment and simulation, and showed that it has the largest advantage when the workers exhibit more confirmation bias and when the distribution of workers' values is more dispersed or even polarized.

- Shows that accounting for cognitive biases can greatly improve label aggregation accuracy!

Scan the code to check out our paper!