ARTIFICIAL INTELLIGELLIGENCE IN CRIMINAL JUSTICE

Fair Jury Selection & Deliberation

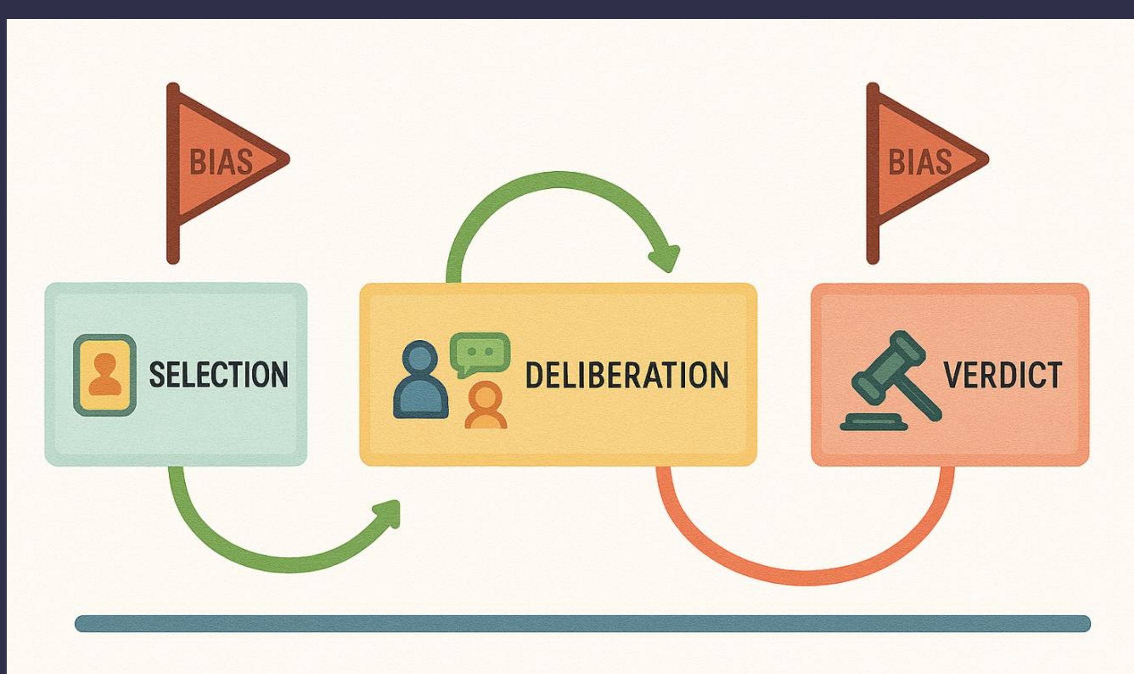
RIT Rochester Institute of Technology

Leona Joseph | Mandira Sawkar | Dr. Mihail Barbosu

MSAI, RIT

Abstract

- Introduces a novel framework combining optimization and game theory.
- Aims to reduce biases in jury selection and deliberation.
- Uses demographic weights and behavioral modeling for fairness.
- Supports real-world implementation through simulations and data tools.
- Useful as both a theoretical and practical contribution to AI and justice.



Introduction

- Jury trials often skewed by systemic and interpersonal biases.
- Issues in jury selection:
  - Demographic under-representation
  - Sampling inequalities and low response rates
- Issues in jury deliberation:
  - Dominant personalities monopolizing discussion
  - Hidden influence and groupthink
- Our framework addresses both components with scientific modeling.

Methodology

Jury Selection
Fairness-Constrained Optimization

- Formulate selection as an optimization problem with fairness constraints.
- Key variables:
  - $x_i=1$ if person i selected, 0 otherwise
  - N: Eligible individuals
  - G: Demographic groups
  - w_g: Sampling weight for group g
  - p_g: Estimated non-response rate
- Constraints:
  - Total jurors selected = S
  - Each group's minimum representation set using weights and response probabilities
- Algorithm Steps:
  - Compile eligible list using source data
  - Assign sampling weights
  - Randomly select jurors within group
  - Adjust for expected non-response
  - Iteratively refine selection pool

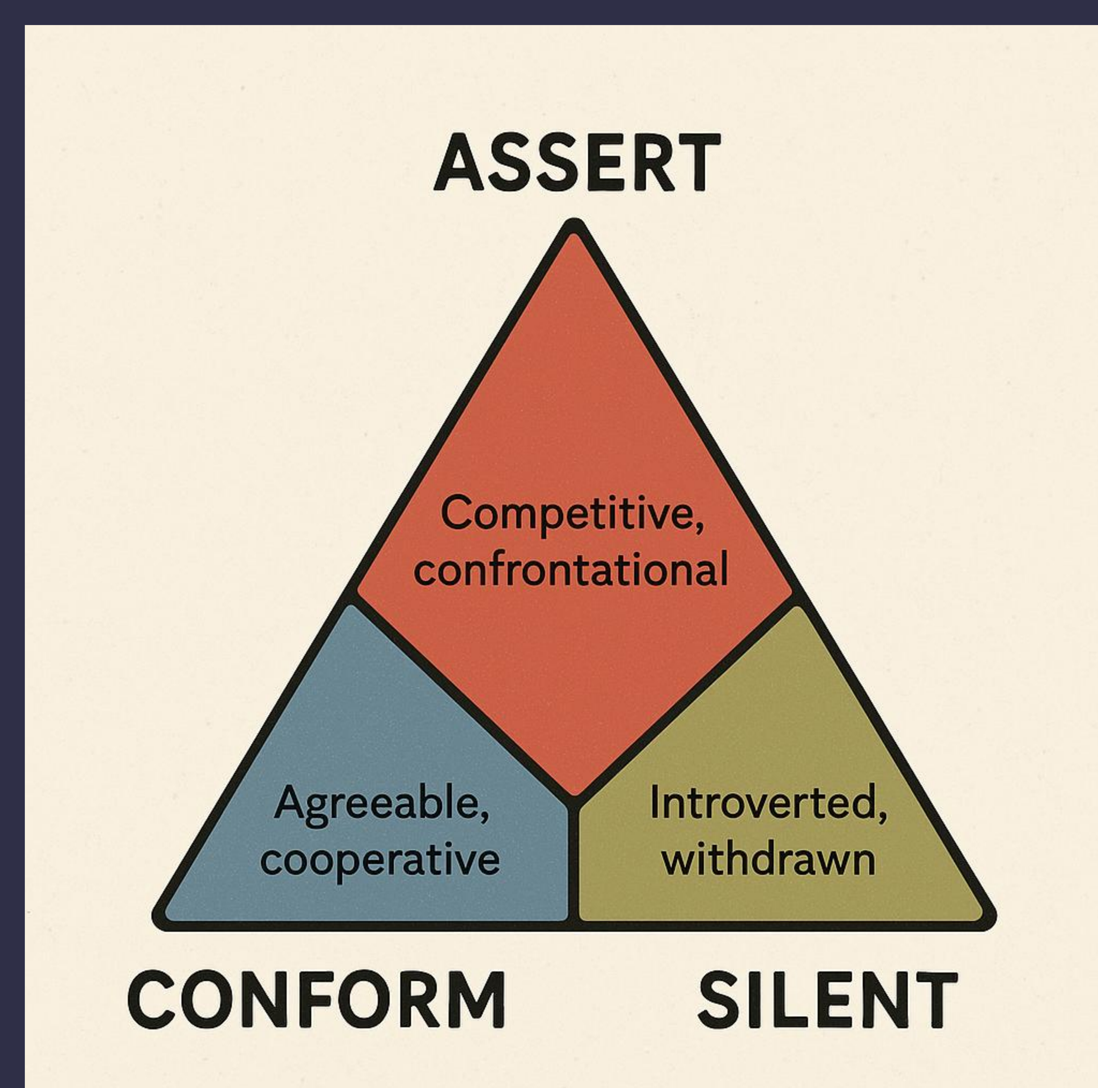
$$\min_x \sum_{g \in G} w_g \left| \frac{\sum_{i \in G_g} x_i}{S} - D_g \right|$$

Methodology

Juror Interaction: Game-Theoretic Modeling

- Deliberation modeled as a non-cooperative game.
- Jurors choose one of three strategies:
  - Assert (A) – speak up and influence
  - Conform (C) – agree with group
  - Silent (S) – stay quiet
- Payoff Functions:
  - Include belief strength (B_i), effort aversion (E_i), susceptibility (S_i)
  - Affected by parameters γ (effort cost), δ (social cost), λ (social reward)
- New Equilibrium: Interaction Equilibrium (*IE*)
  - Extends Nash Equilibrium to include social influence
  - Captures cascading effects of influence across juror networks

$$U_i(k) = \begin{cases} B_i - \gamma E_i + \delta(1 - S_i) & \text{if } k = A \\ \lambda S_i - \gamma E_i & \text{if } k = C \\ -\gamma E_i & \text{if } k = S \end{cases}$$



Potential

- Simulation Insights
  - Jury selection meets demographic quotas within error bounds.
  - Overcomes systemic under-representation by boosting under-sampled groups.
- Interaction Analysis
  - Strong influence from dominant jurors can shift group consensus.
  - Jurors with high susceptibility more likely to conform.
- Mock Trials
  - Used in simulation settings like "People v. Canning" case
  - Feedback collected via Likert scales, peer ratings, and observed behavior
- Influence Mapping
  - Influence coefficients (I_{jk}) derived from post-trial surveys
  - Social Network Analysis (SNA) validates these mappings

Conclusion

Framework advances justice by:

- Enhancing fairness in jury selection
- Modeling deliberation behavior with influence dynamics
- Applicable to real-world systems:
  - Jury simulations
  - AI-assisted court evaluations
  - Research on group decision-making
- Contributes new tools for understanding and correcting bias in civic institutions.

Acknowledgements

Thanks to:

- Harvard Law School Mock Trial Association (HLSMTA)
- Illinois, Maine, and New York State Bar Associations
- Devine et al. (2001), Asch (1956) for foundational behavioral research

Inspired by simulations like People v. Canning