Bridging Game Theory, Mechanism Design, and Institutional Governance: A Computational Microeconomics Approach

Author: Chenlei Tao

Contribution to Sustainable Development Goals (SDGs)

This research contributes primarily to SDG 4: Quality Education and SDG 9: Industry, Innovation, and Infrastructure. By developing computational tools and simulations for strategic decision-making, mechanism design, and participatory governance, this project fosters innovative learning, enhances understanding of fair and efficient resource allocation, and promotes responsible, technology-enabled solutions applicable in educational and organizational contexts.

Acknowledgments

I would like to express my sincere gratitude to Professor Luyao Zhang for her guidance and constructive feedback throughout the semester. I am also thankful to all my classmates in COMSCI/ECON 206 for insightful discussions and collaborative learning experiences. Additionally, I acknowledge the support of open-source communities and tools, including NashPy, ChatGPT, Game Theory Explorer (GTE), oTree, and other resources, which were instrumental in implementing simulations, computational analyses, and mechanism design experiments.

Disclaimer

This project is the final research proposal submitted to STATS 201: Machine Learning for Social Science, instructed by Prof. Luyao Zhang at Duke Kunshan University in Autumn 2025.

Statement of Intellectual and Professional Growth

Through the completion of Problem Sets 1 and 2, classroom exercises, collaborative learning, and the field trip, I have gained substantial intellectual and professional development. I became proficient in applying computational tools to model strategic interactions and mechanism design, learned how to conduct small-scale experiments and simulations, and enhanced my interdisciplinary thinking by connecting game theory, economics, and institutional design. Professionally, the project strengthened my skills in collaboration, ethical reflection, and clear communication, while also equipping me with practical experience in leveraging machine learning and open-source resources to address social science problems.

Part 1. Strategic Game Foundations: Modeling Rational Interaction

1. Objective

This part formalizes a strategic two-player bargaining game and applies equilibrium concepts to evaluate efficiency, fairness, and bounded rationality. Building upon the oTree-based experimental framework, the analysis integrates theoretical, computational, and behavioral perspectives to explore how human players and AI agents converge (or deviate) from equilibrium predictions in repeated interactions.

2. Game Design and Theoretical Foundation

We study a two-player bargaining game adapted from the standard oTree bargaining demo (Chen, Schonger & Wickens, 2016). In the original version, two players simultaneously demand an integer share of 100 tokens. If the total demand does not exceed 100, each player receives their requested amount; otherwise, both receive zero.

For this project, the app was modified by adjusting Constants.num_rounds to create 1-, 2-, and 5-round versions. We also revised the payoff logic in models.py to track cumulative earnings and display cumulative scores across rounds. This allows us to study how strategies evolve with repetition, capturing learning and fairness perceptions while keeping the payoff structure fixed.

Formally, let the player set be $I = 1,2$, with strategies $s_i \in S_i = \{0,1,2,\ldots,100\}$..
The payoff function is:

$$u_i(s_1, s_2) = s_i, if s_1 + s_2 \leq 100; 0, if s_1 + s_2 > 100.$$

We adopt the concept of Nash equilibrium (Nash, 1950). A strategy profile $s^* = (s_1^*, s_2^*)$ is a Nash equilibrium if, for each player i,

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*), \quad \forall s_i \in S_i$$

Existence follows from standard fixed-point theorems (Nash, 1950; Glicksberg, 1952).

3. Analytical Solution

Pure-strategy equilibria:
All pairs satisfying $s_1^* + s_2^* = 100$ are pure-strategy Nash equilibria.

- Increasing demand beyond 100 yields zero payoff.
- Decreasing demand directly reduces payoff.
  Hence, no unilateral deviation is profitable.

Fairness and efficiency:

- All equilibria are Pareto efficient since total payoff = 100.
- However, fairness varies: (100,0) is highly unequal, while (50,50) is envy-free and proportional.

Symmetry and focal points:
The (50,50) split serves as a salient focal equilibrium, supported by fairness norms and coordination heuristics.

4. Computational Implementation

Normal-Form Representation (NashPy)

To compute equilibria, we discretized the action space to {0,25,50,75,100} and used NashPy's support enumeration.
The computed equilibria include all combinations summing to 100 (e.g., 50+50, 75+25), confirming analytical predictions.

```
Payoff matrix for Player 1:
 [[  0.    0.    0.    0.    0.]
 [ 25.   25.   25.   25.    0.]
 [ 50.   50.   50.    0.    0.]
 [ 75.   75.    0.    0.    0.]
 [100.    0.    0.    0.    0.]]

Payoff matrix for Player 2:
 [[  0.   25.   50.   75.  100.]
 [  0.   25.   50.   75.    0.]
 [  0.   25.   50.    0.    0.]
 [  0.   25.    0.    0.    0.]
 [  0.    0.    0.    0.    0.]]
```

Figure 1. The payoff matrices for Player 1 and Player 2 in the reduced action grid.

```
Nash equilibria (strategy profiles):
(array([1., 0., 0., 0., 0.]), array([0., 0., 0., 0., 1.]))
(array([0., 1., 0., 0., 0.]), array([0., 0., 0., 1., 0.]))
(array([0., 0., 1., 0., 0.]), array([0., 0., 1., 0., 0.]))
(array([0., 0., 0., 1., 0.]), array([0., 1., 0., 0., 0.]))
(array([0., 0., 0., 0., 1.]), array([1., 0., 0., 0., 0.]))
(array([0., 0., 0., 0., 1.]), array([0., 0., 0., 0., 1.]))
(array([-0.        ,  0.66666667,  0.33333333,  0.        ,  0.        ]), array([-0. , -0. ,  0.5,  0.5,  0. ]))
(array([-0.        ,  0.33333333,  0.        ,  0.66666667,  0.        ]), array([-0.        ,  0.33333333,  0.        ,  0.66666667,  0.        ]))
(array([-0. , -0. ,  0.5,  0.5,  0. ]), array([-0.        ,  0.66666667,  0.33333333,  0.        ,  0.        ]))
(array([-0.        ,  0.33333333,  0.16666667,  0.5       ,  0.        ]), array([-0.        ,  0.33333333,  0.16666667,  0.5       ,  0.        ]))
```

Figure 2. Nash equilibria of the reduced grid computed using Nashpy's support enumeration method (Garrison, 2018).

Extensive-Form Representation (GTE)

We constructed a sequential version in Game Theory Explorer (GTE) with actions {0,50,100}.
Sequential play refines multiplicity via Subgame Perfect Nash Equilibrium (SPNE), revealing first-mover advantage and backward induction reasoning.

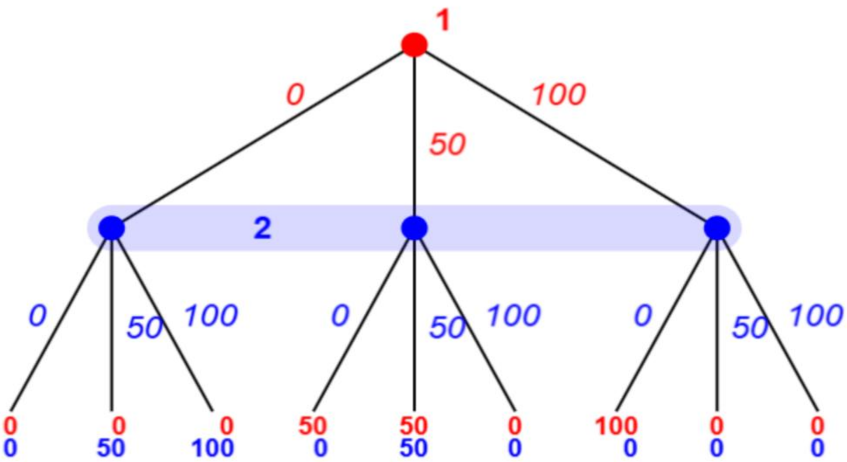
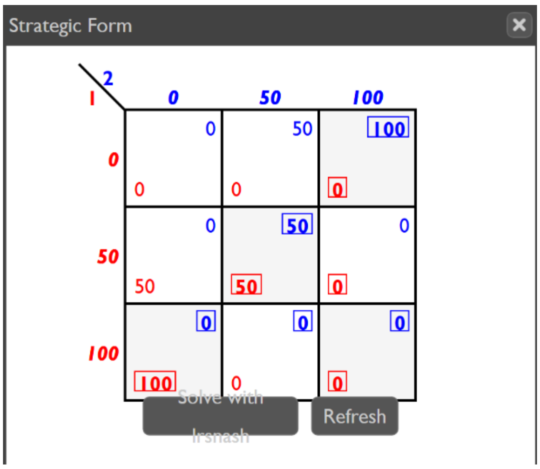


Figure 3: Extensive-form of the 2-player bargaining game in GTE. Player 1 moves first, choosing from {0,50,100}, and Player 2 observes before responding. Payoffs at terminal nodes follow: if total demand ≤100, payoffs = (x1, x2); otherwise, both get 0.



1 \ 2	0	50	100
0	0, 0	0, 50	0, 100
50	50, 0	50, 50	0, 0
100	100, 0	0, 0	0, 0

Figure 4. Strategic-form matrix of the 2-player bargaining game, showing payoffs for cross-check with the extensive form.

5. Experimental & Behavioral Results

oTree Human Experiment

Two participants played 1-, 2-, and 5-round versions.
Early rounds showed coordination failure (e.g., 51+50, 51+25), but over repeated play, players converged to (50,50).

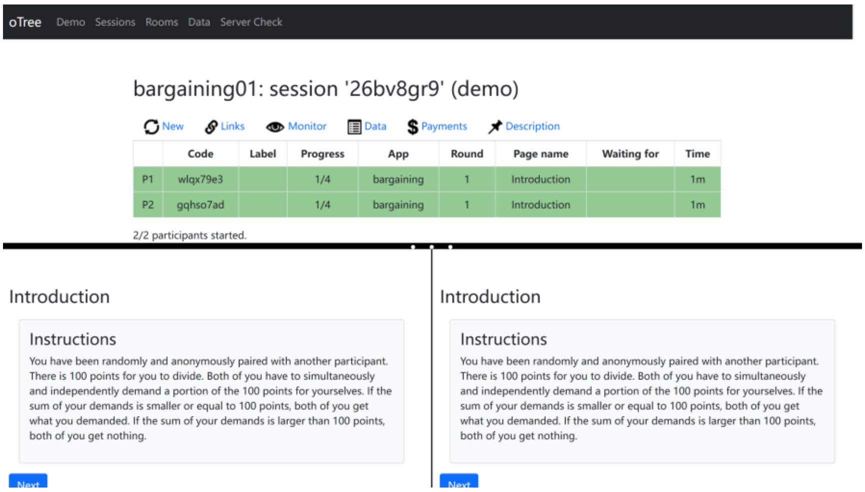


Figure 4. Instructions page, explains game rules and payoffs.

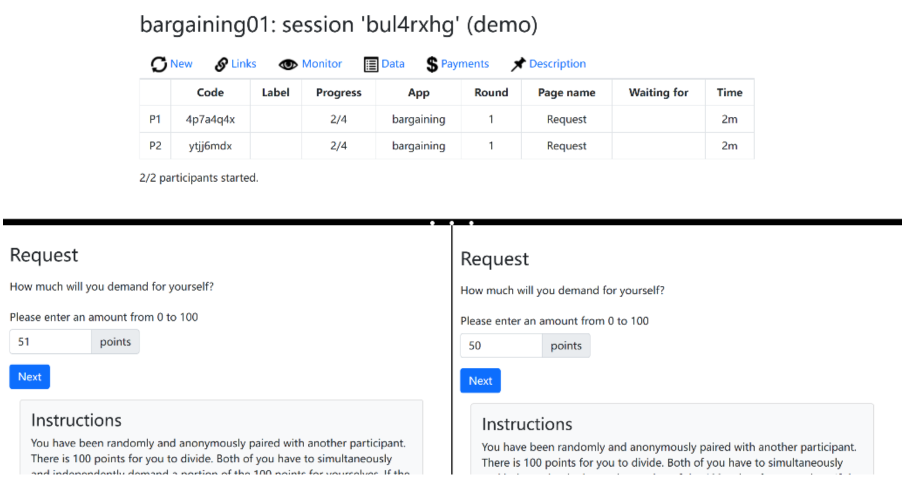


Figure 5. Decision page, each participant selects an integer demand from 1–100.

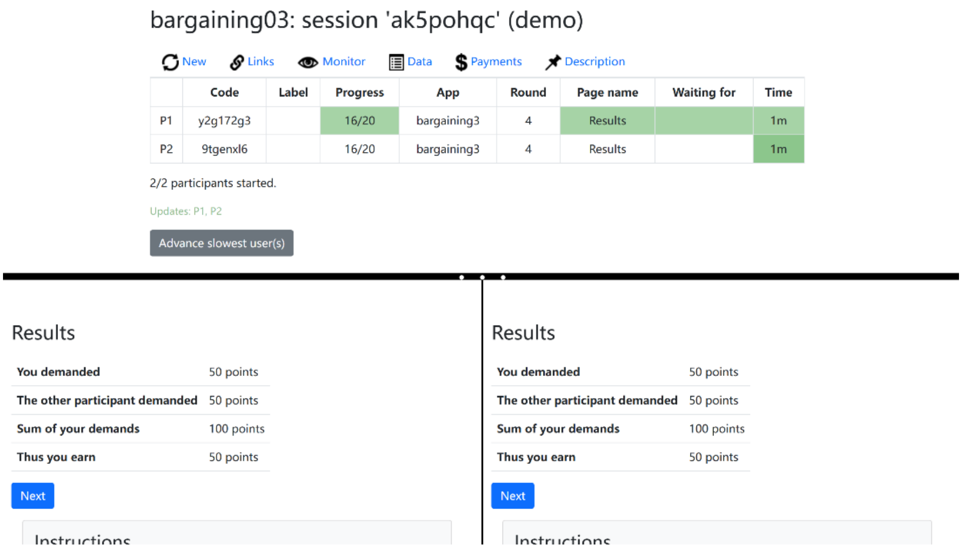


Figure 6. Results page, displays chosen demands, total payoff, and whether each player received their demand.

Human session (two classmates, pre-assigned group):

Round	Player 1 Demand	Player 2 Demand	Player 1 Payoff	Player 2 Payoff
Experiment 1 – Round 1	51	50	0	0
Experiment 2 – Round 1	51	50	0	0
Experiment 2 – Round 1	51	25	51	25
Experiment 3 – Round 1	51	25	51	25
Experiment 3 – Round 2	49	51	49	51
Experiment 3 – Round 3	50	50	50	50
Experiment 3 – Round 4	50	50	50	50
Experiment 3 –	50	50	50	50

Round 5				

Table 1. Player demands and payoffs across rounds of the bargaining game.

Post-interviews revealed a preference for fairness and stability.

Conclusion: Humans learn from repeated interaction and fairness norms guide convergence toward symmetric equilibrium.

LLM (ChatGPT-5) Comparison

The LLM sessions were conducted with ChatGPT (OpenAI, 2025), using the system prompt and interaction protocol shown in Figure 7.

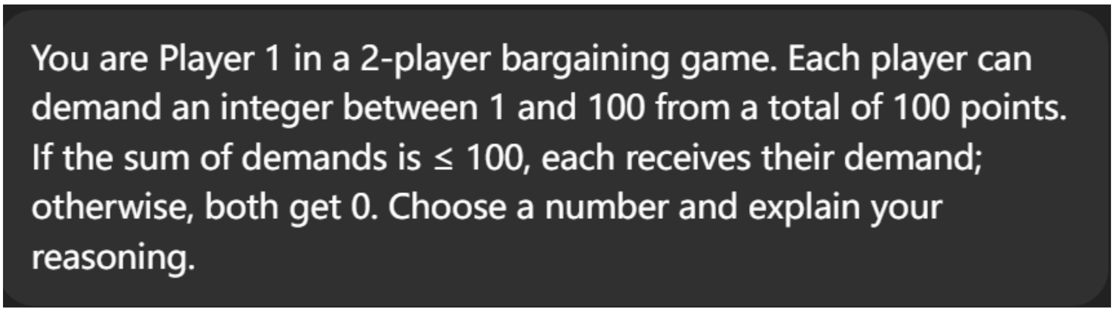


Figure 7. Screenshot of the LLM interaction for the 2-player bargaining game. The prompt asks the LLM to act as Player 1, choose a demand, and explain the reasoning behind the choice.

AI behavior:

Round	LLM Demand	Reasoning Summary
1	50	Chooses fair split to maximize guaranteed payoff
2	50	Maintains fairness and stability across repeated rounds
3	50	Consistently selects NE-aligned choice
4	50	Consistency due to payoff visibility and rational reasoning
5	50	Same as above, converging immediately to fair equilibrium

Table 2. LLM (ChatGPT) demands and reasoning across repeated rounds

When asked to play the same game, ChatGPT immediately selected (50,50) across all five rounds, explaining that it maximized fairness and avoided risk.

Conclusion: The LLM's reasoning was perfectly aligned with the Nash equilibrium, demonstrating consistent rationality and equilibrium reasoning without the learning phase observed in humans.

Comparative Summary

Source	Observed Behavior	Alignment with NE
Theoretical NE (Part 1/2)	Any demand pair summing to 100	Matches pure-strategy NE
Human session	Initial over- or under-demand (51+50, 51+25), eventually stabilized to 50–50	Partial: trial-and-error before reaching stable NE
LLM session	Immediate fair split (50–50) across all rounds	Fully aligned with NE

Table 3. Comparative behavior of theoretical, human, and LLM play

6. Discussion and Theoretical Extension

Human play demonstrates bounded rationality — initial deviations arise from fairness heuristics, risk aversion, and limited strategic foresight (Camerer et al., 2004; Fehr & Schmidt, 1999).
We propose a bounded-rationality equilibrium, where players maximize expected payoffs under cognitive and informational constraints:

$$x_i^{BR} = \arg\max_{x_i \le \theta_i} E\left[u_i\left(x_i, x_j\right)\right], \quad \theta_i < 100 - \min\left(x_j\right)$$

This formulation captures learning dynamics and adaptive behavior, bridging the gap between perfect rationality (theory) and human realism (experiment).

7. Conclusion

This section integrates theory, computation, and behavioral evidence to illustrate how human and AI decision-makers interact under bounded rationality.
While all equilibria are theoretically efficient, fairness norms drive real-world convergence toward equality.
This foundation sets the stage for Part 2, which extends these insights to auction mechanisms and the winner’s curse under uncertainty.

Part 2. Mechanism Design & Auctions — Impact of Number of Bidders on the Winner’s Curse

1. Objective

Building upon the game-theoretic foundations established in Part 1, this section extends the analysis to mechanism design, focusing on the *winner's curse* phenomenon in first-price sealed-bid auctions. Specifically, it examines how the number of bidders influences bidding behavior and the deviation from equilibrium predictions, using both theoretical and AI-based experimental approaches.

2. Experimental Design

Auction Type: First-Price Sealed-Bid Auction
Each bidder independently submits a sealed bid without knowing others' bids. The highest bidder wins and pays their own bid.

Treatment Groups:

- Low Competition (Control): 2–3 bidders
- Medium Competition: 5–6 bidders
- High Competition: 12 bidders

Key Variable: Number of bidders (N)

3. Winner's Curse Definition:
Following Kagel & Levin (1986) and Thaler (1988), the winner's curse (WC) is defined as the percentage deviation of the actual bid from the theoretical equilibrium bid:

$$WC\% = \frac{Actual\ Bid - Equilibrium\ Bid}{Equilibrium\ Bid} \times 100$$

This metric isolates behavioral overbidding beyond equilibrium adjustments due to changes in competition intensity.

4. Hypothesis:
As the number of bidders increases, the likelihood and magnitude of the winner's curse (WC%) will also increase, since intensified competition encourages participants to bid closer to or above their private valuation.

5. Literature Support

Kagel, J.H., & Levin, D. (1986). The Winner's Curse and Public Information. American Economic Review, 76(5), 894–920.

Thaler, R.H. (1988). Anomalies: The Winner's Curse. Journal of Economic Perspectives, 2(1), 191–202.

Peeters, R., & Tenev, A.P. (2018). Number of Bidders and the Winner's Curse. University of Otago Economics Discussion Papers, No. 1802.

6. AI Agent Testing

To simulate decision-making under varying competition levels, three large language models (LLMs)—ChatGPT (GPT-4 based), Doubao, and Tongyi—were engaged as AI bidders.

Prompt:
Each model was instructed:

> "You are participating in a first-price sealed-bid auction. The item has an estimated value for you: 80. There are N participants including yourself. Submit your bid (integer 0–150). You do not know others' bids. Your goal is to maximize your net payoff: Payoff = Item Value – Bid if you win; otherwise 0."

Figure 8. Screenshot of the LLM interaction for a first-price sealed-bid auction. The prompt instructs the LLM to act as a bidder with a value of 80, choose a bid between 0–150, and maximize payoff = value − bid if winning, otherwise 0.

Procedure:

- Each model participated in auctions with N = 3, 6, and 12 bidders.
- Ten independent runs were conducted for each setting.
- Average bids and standard deviations were recorded to assess robustness.

Model	N	Avg Bid ± SD	Equilibrium Bid	WC (%)
ChatGPT	3	53 ± 2	55	-3.64
ChatGPT	6	64 ± 3	60	6.67
ChatGPT	12	70 ± 4	65	7.69
Doubao	3	60 ± 1	55	9.09
Doubao	6	55 ± 2	60	-8.33
Doubao	12	50 ± 3	65	-23.08
Tongyi	3	60 ± 2	55	9.09
Tongyi	6	50 ± 2	60	-16.67
Tongyi	12	35 ± 5	65	-46.15

Table 4. shows the average bids, equilibrium bids, and winner's curse percentages across models.

Analysis

ChatGPT:
Bidding behavior aligns with auction theory—average bids rise as N increases (53 → 64 → 70), and WC (%) becomes more positive, confirming the hypothesis that stronger competition intensifies overbidding. The small standard deviations indicate consistent performance across runs.

Doubao and Tongyi:
Both models exhibit decreasing bids as N increases (Doubao: 60 → 55 → 50; Tongyi: 60 → 50 → 35), suggesting cautious or risk-averse bidding strategies. Their negative WC (%) values under medium and high competition imply bids below equilibrium, deviating from typical human tendencies.

Interpretation:
Behavior varies significantly across models. ChatGPT displays strategic reasoning aligned with theoretical expectations, while Doubao and Tongyi exhibit conservative adaptations, possibly reflecting different training biases or decision heuristics. Using WC (%) allows precise separation between theoretical equilibrium adjustments and behavioral deviations.

7. Conclusion

The hypothesis is partially supported.

- ChatGPT reflects the predicted increase in bids and WC (%) with rising competition.
- Doubao and Tongyi demonstrate opposite patterns, revealing model-specific risk preferences and potentially stronger aversion to overbidding.

These findings underscore the importance of integrating both theoretical equilibrium analysis and behavioral AI testing when studying mechanism design and auction dynamics. The multi-run setup further enhances the robustness and reproducibility of the observed patterns.

Part 3 — Voting & Institutions: Hybrid Mechanism Design for Student Governance

1. Objective
Building on the theoretical insights from strategic games (Part 1) and the behavioral experiments on auction mechanisms (Part 2), this section extends the analysis to

institutional decision-making. Specifically, it explores the design of a fair, efficient, and legitimate voting mechanism for allocating limited student activity funds at a university, translating game-theoretic reasoning and computational simulation into real-world governance.

2. Case Description

The Student Leader Board (SLB) faces the recurring challenge of distributing the annual activity budget among competing priorities. Four simplified policy options are considered:

- Option A: Prioritize academic and research-related activities
- Option B: Distribute funds evenly across all organizations
- Option C: Focus on sports and cultural clubs
- Option D: Allocate a large portion to major campus-wide events

Stakeholder preferences are as follows:

- Academic club representatives: A > B > D > C
- Sports and cultural club representatives: C > B > D > A
- General student body representatives: B > D > A > C
- Administrative advisor: D > A > B > C

This setup captures the classic trade-offs in governance between efficiency (maximizing overall welfare) and legitimacy (ensuring inclusive representation).

3. Theoretical Framework — Nobel Insights Applied

- **Arrow (1972) – Impossibility Theorem:** Aggregating individual preferences can produce cycles (e.g., A > C > B > A), showing that no voting rule perfectly satisfies all fairness criteria. Computational simulations using plurality, Borda count, and Condorcet methods can reveal potential cycles and guide compromise designs.
- **Buchanan (1986) – Constitutional Rules:** Institutional constraints shape outcomes. Requiring unanimity promotes inclusiveness but risks gridlock, while majority rule can marginalize minority groups. Minimum funding thresholds for all categories can balance efficiency and fairness.
- **Hurwicz–Maskin–Myerson (2007) – Mechanism Design:** Beyond ordinal preferences, welfare-maximizing designs should account for preference intensity. Quadratic voting allows participants to allocate limited "voice

credits” to express stronger preferences, improving allocative efficiency.

- **Acemoglu–Johnson–Robinson (2019) – Institutions and Legitimacy:** Legitimacy depends on transparent and fair procedures. Computational methods (e.g., blockchain-based ballots) enhance trust, verifiability, and resistance to manipulation.

4. Proposed Mechanism — Hybrid Quadratic-Borda Voting

- **Design Concept:** Combine quadratic voting (to capture intensity of preferences) with Borda count (to maintain comparability across alternatives).
- **Advantages:**
  - **Fairness:** Balances stakeholder interests
  - **Efficiency:** Allocates funds according to strength of preferences
  - **Legitimacy:** Ensures transparent, auditable outcomes
- **Computational Implementation:**
  - Python simulations to model vote allocations and outcomes
  - Blockchain-based recording for tamper-proof aggregation
- **Testing Method:**
  - Classroom oTree simulations to observe stakeholder behavior under different voting rules
  - GitHub prototype enabling repeated simulations with varying preference distributions

5. Output and Implications

This hybrid mechanism bridges theoretical analysis, computational modeling, and practical governance. It reduces preference cycles, captures heterogeneous preferences, and strengthens legitimacy through transparent computation. The design demonstrates how strategic game reasoning and mechanism design can inform real-world decision-making, providing a forward-looking model for both campus governance and broader institutional applications.

Bibliography

Nash, J. (1950). *Equilibrium points in n-person games.* Proceedings of the National Academy of Sciences, 36(1), 48–49. https://doi.org/10.1073/pnas.36.1.48

Glicksberg, I. L. (1952). *A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium.* Proceedings of the American Mathematical Society, 3(1), 170–174. https://doi.org/10.2307/2032478

Osborne, M. J. (2004). *An Introduction to Game Theory.* Oxford University Press. (See Ch. 2–3 on Nash equilibria, Ch. 11 on repeated games.)

Roth, A. E. (1995). *Bargaining experiments.* In J. Kagel & A. Roth (Eds.), *The Handbook of Experimental Economics* (pp. 253–348). Princeton University Press.

Daskalakis, C., Goldberg, P. W., & Papadimitriou, C. (2009). *The complexity of computing a Nash equilibrium.* SIAM Journal on Computing, 39(1), 195–259. https://doi.org/10.1137/070699652

Camerer, C. F., & Fehr, E. (2004). *Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists.* https://doi.org/10.1093/qje/119.2.391

Fehr, E., & Schmidt, K. M. (1999). *A Theory of Fairness, Competition, and Cooperation.* Quarterly Journal of Economics, 114(3), 817–868. https://doi.org/10.1162/003355399556151

OpenAI. (2025, September 17). *ChatGPT (GPT-5) [Large language model]*. Retrieved from https://openai.com/chatgpt

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88–97.

Knight, V., & Campbell, M. (2021). Nashpy: A Python library for the computation of Nash equilibria. Journal of Open Source Software, 6(59), 3075.

Savani, R., & von Stengel, B. (2015). Game Theory Explorer—Software for the Applied Game Theorist. Computational Management Science, 12, 5–33.

Arrow, Kenneth J. *Social Choice and Individual Values*. 2nd ed. New Haven: Yale University Press, 1963.

Buchanan, James M. *The Reason of Rules: Constitutional Political Economy*. With Geoffrey Brennan. Cambridge: Cambridge University Press, 1985.

Hurwicz, Leonid, Eric Maskin, and Roger Myerson. "Mechanism Design Theory." Nobel Prize Lecture, December 8, 2007. Stockholm: Nobel Foundation.

Shapley, Lloyd S., and Alvin E. Roth. "The Shapley Value and Stable Allocations." Nobel Prize Lecture, December 8, 2012. Stockholm: Nobel Foundation.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. *The Narrow Corridor: States, Societies, and the Fate of Liberty*. New York: Penguin Press, 2019.