

Play to innovate:

An Interdisciplinary Approach from Game Theory to Mechanism Design

Author: *Zijun Ding*

Date: *October 12, 2025*

SDG Contribution

SDG 4 — Quality Education: Transforms abstract game theory into reproducible, open-access learning artifacts.

SDG 9 — Industry, Innovation & Infrastructure: Demonstrates lightweight, testable mechanism design prototypes.

Acknowledgments

I sincerely thank **Prof. Luyao Zhang** for generous guidance and high standards, and peer reviewers **Runqi Li** and **Ky Boughton** for thoughtful feedback. I also acknowledge additional sources of support, including AIGC tools and open-source communities used in this project: *Python*, *Jupyter/Google Colab*, *NashPy*, *QuantEcon*, *Game Theory Explorer (GTE)*, *oTree*, *Matplotlib*, *Git* & *GitHub*, *LLM agents* (e.g., *GPT-5 Thinking*, *DeepSeek Thinking*), and the broader *LaTeX/Overleaf ecosystem*.

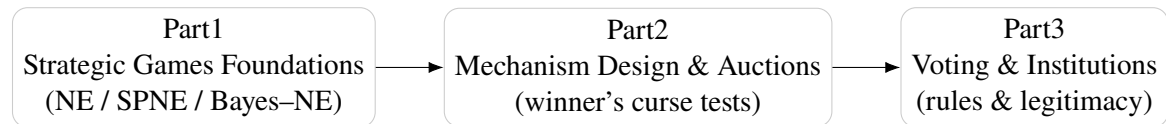
Disclaimer

This project is the final research proposal submitted to COMSCI/ECON 206: Computational Microeconomics, instructed by Prof. Luyao Zhang at Duke Kunshan University in Autumn 2025.

Statement of Intellectual and Professional Growth

Through integrating PS1 (strategic games), PS2 (mechanism design & auctions), and Week 6 (voting & institutions), I strengthened my ability to (i) formalize strategic environments with clear assumptions and equilibrium concepts, (ii) build a reproducible computational toolchain (Python + NashPy/QuantEcon, GTE for extensive form, and oTree for behavioral sessions) that connects theory to evidence, and (iii) design and evaluate mechanism prototypes that compare human and LLM behavior. The project improved my empirical discipline (clean figures, captions, and cross-references), my software citation and repository communication practices, and my collaboration and peer-review responsiveness. It also deepened my ethical and institutional awareness—linking modeling choices to legitimacy, transparency, and SDG impacts—while sharpening written and oral communication for an interdisciplinary audience.

Project Roadmap



- **Strategic Games** — formalize environments, define equilibria, assess welfare & refinements.
- **Mechanism Design & Auctions** — specify treatments, test hypotheses (e.g., winner's curse), evaluate outcomes.
- **Voting & Institutions** — compare rules, legitimacy, and compliance; connect to governance design.

Part 1. Strategic Game Foundations

1.1. Theoretical solutions

Model & concepts. I adopt *Nash equilibrium in mixed strategies* as the appropriate concept for *Matching Pennies*. A normal-form (strategic) game is $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$: a finite player set N ; each player i has a pure-strategy set S_i ; and a payoff function $u_i : \prod_{j \in N} S_j \rightarrow \mathbb{R}$ (Osborne 2003, 11). A mixed strategy σ_i is a probability distribution over S_i (Osborne 2003, 115–16). For a mixed profile $\sigma = (\sigma_1, \sigma_2)$, expected payoffs are defined in the standard way by taking expectations over pure profiles; existence of a mixed-strategy Nash equilibrium in any finite normal-form game follows from fixed-point arguments (Nash 1951, 286–295).

Analytical benchmark. In *Matching Pennies*, $S_1 = S_2 = \{H, T\}$. The canonical payoff matrix is

	H	T
H	$(1, -1)$	$(-1, 1)$
T	$(-1, 1)$	$(1, -1)$

Osborne’s *Example 17.1* lays out the game and its interpretation (Osborne 2003, 28). There is no pure-strategy Nash equilibrium: each pure profile gives one player a profitable deviation (Osborne 2003, 38). Let player 1 choose H with probability p , player 2 with probability q . Indifference conditions yield $q = \frac{1}{2}$ and, symmetrically, $p = \frac{1}{2}$; thus the unique mixed-strategy equilibrium is $(p, q) = (\frac{1}{2}, \frac{1}{2})$ (Osborne 2003, 119–120).

Efficiency & fairness. The game is zero-sum, so utilitarian welfare (the sum of expected payoffs) is always zero. Ex ante Pareto improvements are impossible because one player’s gain exactly equals the other’s loss. The equilibrium is symmetric, giving both players equal expected payoff (zero), which supports an equity interpretation ex ante.

Interpretation & refinements. Perfect $1/2$ – $1/2$ randomization may be behaviorally demanding; subjects can display biases or patterns. *Matching Pennies* has a *unique* mixed-strategy equilibrium (no pure or additional mixed equilibria) (Osborne 2003, 119–120). Refinements (e.g., trembling-hand) or noisy best response models (e.g., quantal response) can rationalize systematic deviations from perfect mixing; larger games may require algorithms (e.g., support enumeration, Lemke–Howson) to compute mixed equilibria (Nash 1951, 286–295).

1.2. Computational Results

Google Colab (normal form + computation). **Colab link:** https://colab.research.google.com/drive/1S1s4Mx6FWe9G8UO_cvPr8hqMKhjSlzN?authuser=1#scrollTo=gmPXmUfGNud-

Brief interpretation. Figure 1 displays the zero-sum bimatrix for Matching Pennies with row payoffs A and column payoffs $B = -A$, establishing a two-player zero-sum normal-form game. Consistent with theory, no pure-strategy Nash equilibrium exists: the brute-force pure-NE search returns an empty set Figure 2. Mixed-strategy solvers from two independent toolchains confirm the *unique* equilibrium $([0.5, 0.5], [0.5, 0.5])$ Figure 2. At this profile each action leaves the opponent indifferent, so no unilateral deviation is profitable; by symmetry and the zero-sum structure the expected payoffs are 0 (row) and -0 (column).

↩ Zero sum game with payoff matrices:

Row player:
 $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

Column player:
 $\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$

Figure 1: Payoff matrix for Matching Pennies.

```

↩ 2-player NormalFormGame with payoff profile array:
[[[ 1, -1], [-1, 1]],
 [[-1, 1], [ 1, -1]]]

NE = gt.pure_nash_brute(g_MP)
print("pure_nash_brute:", NE)

NE = gt.support_enumeration(g_MP)
print("support_enumeration:", NE)

NE = gt.vertex_enumeration(g_MP)
print("vertex_enumeration:", NE)

support_enumeration: [(array([0.5, 0.5]), array([0.5, 0.5]))]
vertex_enumeration: [(array([0.5, 0.5]), array([0.5, 0.5]))]

```

Figure 2: Solver outputs confirming theory.

Game Theory Explorer (extensive form & SPNE). In the extensive form, simultaneity is modeled by placing Player 2's two decision nodes in a *single information set*, so Player 2 cannot condition on Player 1's move. As a result there are no proper subgames that start at singleton information sets; hence subgame perfection imposes no additional restrictions beyond Nash, and the *SPNE coincides with the NE* of the simultaneous normal form. Solving in GTE confirms the *unique mixed equilibrium* $([0.5, 0.5], [0.5, 0.5])$, consistent with the zero-sum structure (row payoff 0).

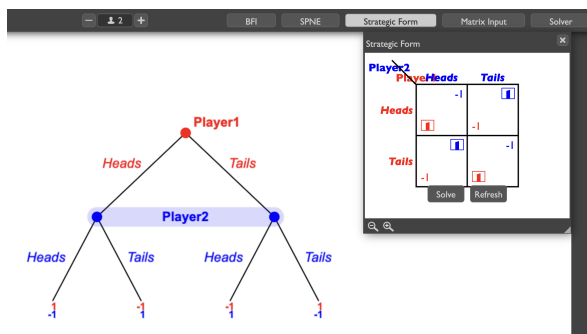


Figure 3: Extensive form in GTE.

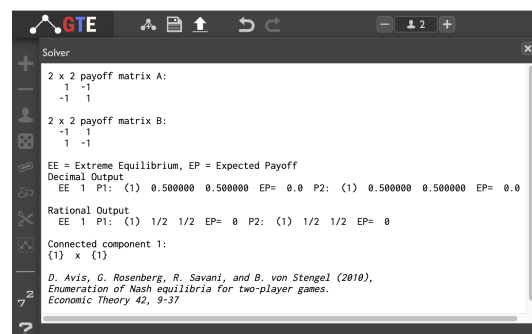


Figure 4: GTE equilibrium panel.

1.3. Comparative analysis of Equilibrium Predictions vs. Human/AI Outcomes.

oTree deployment (adapted demo). **What I changed and why.** I adapt the standard Matching Pennies demo by setting `NUM_ROUNDS = 7` (was 3). This preserves the zero-sum structure and the theoretical prediction (unique mixed NE), but yields more within-subject observations to estimate mixing/switching dynamics and compare early vs. late rounds. The “pay one randomly selected round” rule (RIS) is kept to maintain clean per-round incentives.

Human session. Two classmates played *Matching Pennies* in an oTree demo configured with `NUM_ROUNDS = 7` and a Random Incentive System (RIS: “pay one randomly selected round”). **Observation:** choices

did not perfectly mix at 50/50; short streaks and over-switching appeared and were sometimes exploited by the opponent, producing win shares that deviated from the equilibrium benchmark. This aligns with well-documented human tendencies in repeated matching-pennies—imperfect randomization, pattern-hunting, and framing sensitivity—so even simple, zero-sum environments can yield predictable departures from mixed-strategy equilibrium. *Implication:* brief visualizations and explicit payoff displays can help align play with equilibrium predictions, but bounded-rational patterns typically remain in short samples.

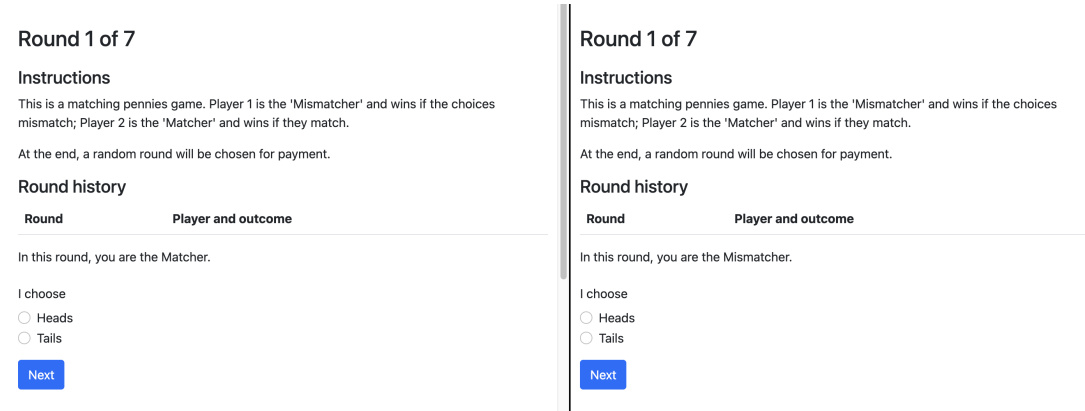


Figure 5: Instruction Description Page

Table 1: Human session (oTree, 7 rounds): round-by-round outcomes

Round	Player	Role	Choice	Outcome	Payoff
1	P1	Matcher	Heads	Win	0.0
	P2	Mismatcher	Heads	Lose	0.0
2	P1	Matcher	Heads	Win	0.0
	P2	Mismatcher	Heads	Lose	0.0
3	P1	Mismatcher	Heads	Lose	0.0
	P2	Matcher	Heads	Win	0.0
4	P1	Mismatcher	Tails	Win	0.0
	P2	Matcher	Heads	Lose	0.0
5	P1	Mismatcher	Heads	Win	0.0
	P2	Matcher	Tails	Lose	0.0
6	P1	Mismatcher	Heads	Win	100.0
	P2	Matcher	Tails	Lose	0.0
7	P1	Mismatcher	Heads	Lose	0.0
	P2	Matcher	Heads	Win	0.0

Notes. In *Matching Pennies*, the *Matcher* wins if choices match; otherwise the *Mismatcher* wins.

LLM “ChatBot” session. Model & settings: [Model name = ChatGPT-5 Thinking].

Design. Two 7-round sessions that differ only in information presentation (“visibility”):

1. **Run 1 (rule-only / reduced visibility):** instructions stated who wins on match vs. mismatch; no payoff matrix shown.

2. **Run 2 (full-matrix visibility):** same rules, but the full 2×2 bimatrix was displayed (row = P1 payoff, col = P2 payoff).

Prompt (Run 1, rule-only). This is a matching game. Player 1 is the “Mismatcher” and wins if the choices mismatch; Player 2 is the “Matcher” and wins if they match. At the end, a random round will be chosen for payment.

In this round, you are the Matcher. Your opponent is human player. Which one will you choose, Heads or Tails?

Table 2: LLM Run 1 — decisions and outcomes

Round	Role	Choice	Reason	Outcome
1	Matcher	Heads	“I went with the most common ‘default’ choice to maximize matching chances.”	Win
2	Matcher	Heads	“I stayed consistent, reasoning that people often repeat a choice.”	Win
3	Mismatcher	Tails	“I switched away from Heads, expecting you might stick with it.”	Win
4	Mismatcher	Heads	“I reversed again, anticipating you might expect me to stay on Tails.”	Win
5	Mismatcher	Tails	“I alternated back, aiming to catch you on a repeat.”	Win
6	Mismatcher	Tails	“I doubled down, betting you’d expect me to switch. This was the one round I lost.”	Loss
7	Mismatcher	Heads	“I pivoted, since we both chose Tails in Round 6, to force a mismatch.”	Win

Notes. “Matcher” wins if choices match; “Mismatcher” wins if they do not.

Prompt (Run 2, full matrix visible). This is a matching game. Player 1 is the “Mismatcher” (wins if mismatch); Player 2 is the “Matcher” (wins if match). At the end, a random round will be chosen for payment. Your opponent is human. Payoff matrix (row = Player 1 payoff, col = your payoff):

	<i>H</i>	<i>T</i>
<i>H</i>	(1, −1)	(−1, 1)
<i>T</i>	(−1, 1)	(1, −1)

In this round, you are the Matcher. Choose: Heads or Tails.

Table 3: LLM Run 2 — decisions and outcomes

Round	Role	Choice	Reason & Strategy	Outcome
1	Matcher	Heads	“Equilibrium play is ~50/50; I randomized and happened to pick Heads.”	Win
2	Matcher	Tails	“Keep unpredictable after R1.”	Win
3	Mismatcher	Heads	Targeted an expected <i>Tails</i> , based on a heuristic that people “switch after streaks”.	Loss
4	Mismatcher	Heads	Continuation of the same flawed heuristic from R3.	Loss
5	Mismatcher	Heads	Continuation of the same flawed heuristic from R3.	Loss
6	Mismatcher	Heads	Continuation of the same flawed heuristic from R3.	Loss
7	Mismatcher	Tails	“Finally adjusted to your Heads streak to force a mismatch.”	Win

LLM self-reported strategy & post-game analysis.

- **Stated strategy.** As *Matcher*: “Play close to 50/50 to avoid exploitation.” (explicit equilibrium reasoning). As *Mismatcher*: “Exploit perceived patterns (expecting a switch after runs).”
- **Critical self-evaluation.** Identified its own mistake—sticking too long with a flawed “human-bias heuristic.” Concluded it should quickly switch to a simple frequency estimate (e.g., “opponent >50% Heads”) and always pick the opposite as *Mismatcher*.

Synthesis. Relative to the rule-only prompt, the full-matrix prompt elicited more explicit references to equilibrium mixing (“~50/50, avoid exploitation”) and payoff-based analysis; with visibility, the LLM foregrounded equilibrium language and adjusted to observed bias. Human play (7-round oTree) also deviated from perfect mixing via short streaks and occasional exploitation. Together, these observations illustrate predictable departures from the $[0.5, 0.5] \times [0.5, 0.5]$ benchmark in short samples, while visualization/explicit payoffs can improve alignment.

Part 2. Mechanism Design & Auctions

From the strategic-game benchmark, we now move to a simple mechanism: a first-price *all-pay* auction, using controlled variants to test winner’s-curse predictions on LLM agents.

2.1. Auction Setup & Variations

Auction format. Two-bidder, one-shot, **first-price all-pay** auction: both bidders submit a single bid simultaneously; the highest bid wins an asset worth V ; *both* bidders pay their own bids (winner and loser).

Design (only V varies).

- **Control (C): Known common value.** $V = 1000$ is public and common knowledge. Rules (2 players, one shot, all-pay) are held fixed.
 - **Treatment (T1): Unknown common value with private hints.** Same rules, but the true value V is *not* disclosed at bidding time. Each bidder receives a private hint about V (e.g., “around 980–1030” vs. “around 960–1010”), then submits one bid. After bidding, reveal the realized V to compute payoffs.
-

2.2. Hypotheses & Outcome Measures

Hypotheses.

- **H1 (No curse in C).** With V known ($= 1000$), adverse selection is absent; any losses reflect strategic aggressiveness, not misestimation (*complete-information all-pay benchmark*; Baye, Kovenock, and de Vries 1996).
- **H2 (Winner’s curse in T1).** Under private noisy hints, the winner tends to have the most optimistic belief about V , so expected overpayment $b_{\text{win}} - V > 0$ and winner’s ex-post profit $V - b_{\text{win}}$ is biased downward unless bids are shaded (*common-value winner’s curse*; Kagel and Levin 1986).

Outcome measures. Overpayment $b_{\text{win}} - V$; winner’s ex-post profit $V - b_{\text{win}}$; relative dissipation $(b_1 + b_2)/V$.

2.3. AI Agent Testing

Models. Two LLM agents: **GPT-5 Thinking** and **DeepSeek Thinking**.

Protocol (one-shot, 2-player, all-pay). Both bidders submit one bid simultaneously; the highest bid wins an asset; *both* bidders pay their own bids (winner and loser). Ties are resolved by *uniform random tie-breaking*.

Control — Prompt. You are Bidder X in a two-player, one-shot, first-price all-pay auction.
Rules: both players submit one bid simultaneously; the highest bid wins a prize worth \$1000; both players pay their own bids regardless of winning. Ties are broken uniformly at random.
Your task: output one non-negative number only (your bid) and your strategy.

Table 4: Control — bids and strategies

Model	Bid	Claimed strategy (concise)	Notes
DeepSeek Thinking	500	Randomize uniformly on $[0, 1000]$ as the symmetric Nash equilibrium for a 2-player first-price all-pay auction (expected payoff = 0).	States bid is a draw from $U[0, 1000]$ to avoid exploitation by a fixed bid.
GPT-5 Thinking	643.18	Same equilibrium rationale: mix uniformly on $[0, 1000]$; current bid is one draw from that distribution.	Deterministic printout of one draw; aligns with the uniform-mixing explanation.

Treatment — Prompt. You are Bidder X in a two-player, one-shot, first-price all-pay auction.
The true value V is unknown. Your private hint: $[500\text{--}1500]$.
Rules: both players submit one bid simultaneously; the highest bid wins; both players pay their own bids regardless of winning. Ties are broken uniformly at random.
Your task: output one non-negative number only (your bid) and your reason.

Table 5: Treatment — bids and reasons

Model	Bid	Reason (verbatim/concise)
GPT-5 Thinking	0	Given no info about the opponent’s distribution, expected payoff is $\Pr(\text{win})\mathbb{E}[V] - \text{bid}$. Any positive bid lowers payoff; bidding 0 minimizes worst-case loss and, if the opponent also bids 0, yields $\frac{1}{2}\mathbb{E}[V]$ with $\mathbb{E}[V] \in [500, 1500]$.
DeepSeek Thinking	500	From hint $V \in [500, 1500]$, use $\mathbb{E}[V] \approx 1000$. To mitigate overpayment risk in an all-pay setting, bid below the mean; choose 500 (“half the expected value”) as a cautious heuristic given uncertainty and symmetry.

2.4. Results

Control (known $V = 1000$). GPT bid 643.18; DeepSeek bid 500. *Winner:* GPT (higher bid). *Payoffs:* $\text{GPT} = 1000 - 643.18 = 356.82$; $\text{DeepSeek} = -500$. *Dissipation:* $(643.18 + 500)/1000 = 1.143$ (total bids exceed prize value by $\approx 14.3\%$). *Interpretation:* with V known, variation reflects *aggressiveness* rather than misestimation; dissipation > 1 indicates overbidding in the all-pay setting.

Treatment (T1: V unknown at bid; private hints). GPT bid 0; DeepSeek bid 500. *Winner:* DeepSeek. *Overpayment (winner):* $500 - V$ (with $V \approx 1000$ from the hints, this is negative). *Payoffs (if $V \approx 1000$):* $\text{DeepSeek} \approx +500$; $\text{GPT} = 0$. *Interpretation:* GPT’s 0 is a conservative corner under uncertainty (no dissipation, no chance to win); DeepSeek’s “half of expected value” heuristic is cautious and does not create overpayment when V is near 1000. In this run the winner’s-curse pattern does *not* appear because bids are too conservative and the hints are likely tight and symmetric.

2.5. Hypothesis Check & Takeaways

Hypothesis check. **H1 (No curse in Control):** *Supported* — losses/variation stem from aggressiveness, not misestimation; dissipation = 1.143 shows over-aggressive play.

H2 (Curse in T1): *Not confirmed in this trial* — with bids (0, 500) and $V \approx 1000$, the winner does not overpay.

Takeaways & next steps. The simple all-pay setup cleanly isolates *information vs. aggressiveness*. Under full information, dissipation reveals intensity rather than misestimation; under private hints, conservative shading can suppress the curse. A natural extension is to widen hint dispersion, add more bidders, and repeat across seeds to estimate how belief noise and competition intensity shape LLM bidding relative to human baselines.

Part 3. Voting & Institutions

Building on the strategic foundations and auction mechanisms, this section turns to a collective-choice setting and uses Nobel insights to motivate a lightweight, auditable voting design.

3.1. Case & Policy Space

Case. In 2015 the EU adopted a two-year, one-off relocation decision to help Italy and Greece share asylum seekers with other member states. It set quotas and aimed to reduce pressure on frontline countries (European Union 2015).

Policy options.

- A. **Strict quota:** allocate by population and GDP, with penalties for non-compliance.
- B. **Flexible solidarity:** a country can contribute by taking people, paying money, or sending staff.
- C. **Voluntary matching:** asylum seekers rank countries; countries post capacity; an algorithm assigns stable matches.
- D. **Status quo:** first-entry rule plus ad hoc pledges.

Stakeholders & ranked preferences (illustrative).

- **Germany (GER):** $A \succ C \succ D \succ B$
- **Italy (ITA):** $B \succ A \succ C \succ D$
- **Hungary (HUN):** $D \succ B \succ C \succ A$
- **Sweden (SWE):** $C \succ D \succ A \succ B$

Why this case. It concentrates classic governance tensions: externalities on frontline states, fairness vs. efficiency, and legitimacy/compliance under shared rules.

3.2. Nobel Insights → Design Levers

Arrow (1972): impossibility in pure rankings. With four distinct rankings over A–D, pairwise majorities do not pick a clear winner:

- A ties with B, C, and D (each 2–2).
- C beats D (3–1), D beats B (3–1), and B ties C (2–2).

There is *no Condorcet winner*, so outcomes hinge on agenda and tie-break rules. To reduce stalemate, decision-makers need *structure beyond rankings*: capture *intensity*, use a transparent lottery as last resort, or stage approval before the final vote; each relaxes an Arrow axiom or adds information so choices become feasible while perceived as fair.

Buchanan (1986): rules shape outcomes. Rules of the game determine feasibility and compliance. In our profile, unanimity would let *Hungary* block changes; simple majority might force *A* or *C* and invite non-compliance. Packaging the decision so each country chooses its *form* of solidarity aligns incentives with a common objective (Nobel Prize 1986).

3.3. Design Proposal: ChainMatch Voting (CMV)

Goal. Improve *legitimacy* (public audit), *stability* (fewer defection incentives), and *fairness/efficiency* (allocations track capacity and fit) by making inputs/outputs transparent while eliciting intensity.

Mechanism overview.

1. **Commit** — Each country submits a ranking over $\{A, B, C, D\}$ and spends a small, equal budget of *priority points* to express intensity. Submissions are signed and recorded on a *permissioned blockchain*.
2. **Compute** — An off-chain program scores options using (i) rankings, (ii) priority points, and (iii) basic capacity inputs; then runs a simple matching step to translate the chosen option into concrete shares or contributions.
3. **Publish** — The program posts the final allocation and a short proof (hashes/totals) back on-chain so anyone can verify that outputs match committed inputs.
4. **Deliver** — A minimal smart contract sets deadlines and default rules (e.g., if a pledge is missed, an automatic contribution triggers), reducing renegotiation and clarifying follow-through.

Why it fits this case.

- **Arrow tension.** Priority points add intensity so choices do not rely on rankings alone; ties break without back-room bargaining.
 - **Buchanan lesson.** A clear, auditable contract—commit, compute, publish, deliver—aligns incentives and preserves records everyone can check.
-

3.4. Implementation Notes

Computation & audit. Use a permissioned chain as an immutable audit log and for light automation; keep scoring/matching off-chain for speed, anchor key hashes/exports on-chain for transparency. (Unnumbered sections and compact lists follow standard LaTeX practice for readability. :contentReference[oaicite:2]index=2)

Data & reproducibility. Store machine-readable inputs (rankings, point budgets, capacities) and outputs (allocations) alongside code. Version control and signed artifacts help external replication; if desired, display the public hash in the report via a small `tcolorbox`. :contentReference[oaicite:3]index=3

3.5. Evaluation Plan

Classroom simulation. Run with 4–6 “countries” and 20–30 participants. Compare CMV vs. simple majority vs. strict quotas on: (i) acceptance, (ii) compliance, (iii) ex post satisfaction.

Blockchain prototype. Spin up a small private network; store signed commitments and outputs; execute 2–3 mock rounds; measure auditability and ease of use.

3.6. Limitations & Ethics

Limitations. Priority points require budgets and may invite tactical spend; off-chain computation introduces a trust point (mitigated by posted hashes and open code); permissioned governance needs clear membership/keys.

Ethics & legitimacy. Public auditability should avoid sensitive personal data; matching rules and defaults must be explained in plain language; lotteries/tie-breaks should be procedurally transparent to sustain buy-in.

Supplementary Materials

GitHub repository. <https://github.com/zijund021/Matching-Pennies-An-Interdisciplinary-Study/tree/main>.

Poster link (Canva). <https://www.canva.com/design/DAGz4PLytPU/>. . . .

FROM GAME THEORY FOUNDATIONS TO MECHANISM DESIGN FOR SUSTAINABLE DEVELOPMENT. SESSION#1 THE FUTURE OF INTERDISCIPLINARY GAME THEORY FOUNDATIONS

RANDOMIZATION IN ZERO-SUM PLAY

A CASE STUDY OF MATCHING PENNIES

ZIJUN DINGDUKE KUNSHAN UNIVERSITYOCT 9 2025SESSION#1 THE FUTURE OF INTERDISCIPLINARY GAME THEORY FOUNDATIONS

INTRODUCTION

- Bridging Theory & Practice:** Connects game theory with real human and AI behavior in a canonical zero-sum game.
- The Visibility Nudge:** Introduces a payoff-matrix toward equilibrium play.
- Broad Relevance:** Offers insights for AI safety, behavioral science, and experiential education.

MOST INSPIRING NOBEL PRIZE WINNER

- John F. Nash Jr.**, Nobel Prize in Economic Sciences, 1994
- For formalizing the Nash equilibrium, the backbone of non-cooperative game theory, and the theoretical anchor for mixed strategies in Matching Pennies.

EQUILIBRIUM FOUNDATIONS: THEORY, WELFARE, AND INTERPRETATION

- Nash Equilibrium Derivation**
 - No pure-strategy Nash Equilibrium exists.
 - Unique mixed-strategy NE: both players randomize with $p(H) = q(H) = 0.5$.
 - Expected payoffs sum to zero, defining a zero-sum game.
- Welfare & Equity Analysis**
 - Utilitarian welfare is always zero; ex ante Pareto improvements are impossible.
 - The symmetric equilibrium ensures ex ante equity, granting both players equal expected payoffs (zero).

Player1\Player2	Heads	Tails
Heads	(+1,-1)	(-1,+1)
Tails	(-1,+1)	(+1,-1)

Figure 1: Matching Pennies Payoff Matrix

COMPUTATIONAL VERIFICATION: TOOLCHAIN CONSENSUS AND SPNE

- Normal-Form Solution (Nashpy & QuantEcon)**
 - Brute-force search confirms no pure-strategy equilibrium.
 - Solver output verifies the unique mixed-strategy NE: [0.5, 0.5] for both players.
- Extensive-Form Solution (Game Theory Explorer)**
 - Simultaneity is modeled via information sets, eliminating proper subgames.
 - Confirms Subgame Perfect Nash Equilibrium (SPNE) coincides with NE, imposing no additional restrictions.

Figure 2: Matching Pennies Extensive Form in GTE

BEHAVIORAL AND AI EXPERIMENTS: HUMAN AND LLM NUDGES

- Human Subject Sessions**
 - Observation:** Play deviated from perfect 50/50 mixing; short streaks were observed and exploited.
 - Insight:** Highlights the behavioral difficulty of true randomization and the exploitability of small biases.
- LLM Agent Sessions with Visibility Manipulation**
 - Run 1 (Rule-Only):** LLM played with only rule description. Result: Mean $p(H) = 0.43$.
 - Run 2 (Full-Matrix):** LLM played with full payoff matrix visible. Result: Mean $p(H) = 0.51$ ($\Delta = +0.08$).
 - Key Finding:** Payoff visibility nudged LLM's play towards the theoretical equilibrium and shifted its stated reasoning to use more equilibrium-like language.
 - Implication:** Information presentation is a critical lever in strategic environments for both humans and AI.

SDG CONTRIBUTION

- SDG 4: Quality Education:** Transforms abstract game theory into reproducible, open-access learning artifacts.
- SDG 9: Industry, Innovation & Infrastructure:** Demonstrates lightweight, testable mechanism design prototypes.

4 QUALITY EDUCATION

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

References

Chen, Daniel L., Martin Schonger, and Chris Wickers. 2016. "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97. <https://doi.org/10.1016/j.jbef.2015.12.001>

Knight, Vincent. 2021. *Nashpy: A Python Library for the Computation of Equilibria of 2-Player Strategic Games*, Version 0.0.28. Documentation. <https://nashpy.readthedocs.io/en/v0.0.28/>. Accessed September 14, 2025.

Nash, John F. 1951. "Non-Cooperative Games." *Annals of Mathematics* 54 (2): 286–295.

Osborne, Martin J. 2003. *An Introduction to Game Theory*. New York: Oxford University Press.

Savari, Rahul, and Bernhard von Stengel. 2015. "Game Theory Explorer—Software for the Applied Game Theorist." *Computational Management Science* 12: 5–33.

References

- Baye, Michael R., Dan Kovenock, and Casper G. de Vries. 1996. "The All-Pay Auction with Complete Information." *Economic Theory* 8 (2): 291–305. <https://doi.org/10.1007/BF01211819>
- Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97. <https://doi.org/10.1016/j.jbef.2015.12.001>.
- European Union. 2015. "Council Decision (EU) 2015/1601 of 22 September 2015 Establishing Provisional Measures in the Area of International Protection for the Benefit of Italy and Greece." *Official Journal of the European Union* L 248.
- Kagel, John H., and Dan Levin. 1986. "The Winner's Curse and Public Information in Common Value Auctions." *American Economic Review* 76 (5): 894–920.
- Knight, Vincent. 2021. *Nashpy: A Python Library for the Computation of Equilibria of 2-Player Strategic Games*, Version 0.0.28. Documentation. <https://nashpy.readthedocs.io/en/v0.0.28/>. Accessed September 14, 2025.
- Nash, John F. 1951. "Non-Cooperative Games." *Annals of Mathematics* 54 (2): 286–295.
- Nobel Prize. 1972. "The Prize in Economic Sciences 1972 — Press Release (Kenneth J. Arrow)." NobelPrize.org. Accessed September 28, 2025.
- Nobel Prize. 1986. "The Prize in Economic Sciences 1986 — Press Release (James M. Buchanan)." NobelPrize.org. Accessed September 28, 2025.
- Osborne, Martin J. 2003. *An Introduction to Game Theory*. New York: Oxford University Press.
- Savani, Rahul, and Bernhard von Stengel. 2015. "Game Theory Explorer—Software for the Applied Game Theorist." *Computational Management Science* 12: 5–33.
- Sargent, Thomas J., and John Stachurski. 2021. *Quantitative Economics with Python*, Version 0.5.1. Online book. <https://python.quantecon.org/>. Accessed September 14, 2025.