# Digital Twin–Driven Trust-Aware Strategy Modeling in Asymmetric Multi-Agent Social Games

Angela Li
*Applied Mathematics & Physics*
*Stony Brook University*
Stony Brook, NY
angela.li.4@stonybrook.edu

David Li
*Katz School of Science and Health*
*Yeshiva University*
New York, NY
david.li@yu.edu

***Abstract*—In dynamic social environments, strategic interactions are shaped by evolving trust, information asymmetry, and the potential for deception. This paper introduces a novel framework for modeling and optimizing agent behavior in multi-agent games that emulate social settings, where individuals must decide how much private information to share, whether to probe others, and how to strategically mislead—all under varying trust conditions. We propose a trust-sensitive, information-theoretic game model with asymmetric agents, each endowed with a personalized digital twin: a cognitive engine that estimates opponent behavior, predicts deception, updates trust beliefs, and optimizes strategies using projected gradient learning. Our model captures trust dynamics as smooth, continuous updates influenced by observed honesty and integrates them into players' utility computation. Analytical and simulation results demonstrate the emergence of strategic behaviors such as cooperation, opportunistic probing, and adaptive deception. By visualizing strategy and trust evolution across agent populations, we reveal interpretable patterns of social behavior. The proposed architecture offers a scalable and explainable tool for studying trust and influence in socially embedded systems, with applications in cybersecurity, decentralized coordination, and adversarial communication.**

***Index Terms*—Digital twin, multi-agent systems, game theory, trust dynamics, information gain, deception modeling, Nash learning, strategy optimization, adaptive systems, asymmetric games**

## I. Introduction

Strategic interaction in real-world environments often requires agents to make critical decisions under asymmetric information, evolving trust, and competitive incentives. Whether in cybersecurity, decentralized collaboration, or adversarial marketplaces, players must continually balance how much private information to reveal, how much to probe others for secrets, and whether to engage in deception. These actions are further complicated by the interplay of trust dynamics, which modulate both the effectiveness and the risk of communication and probing.

Classical game theory provides powerful tools for modeling rational behavior under uncertainty, yet traditional formulations often assume symmetric agents, static strategies, or perfect observability. In contrast, human decision-makers adapt dynamically: they infer trust from behavior, simulate potential outcomes, and selectively adjust their strategies over time. To capture these cognitive layers, we introduce a new model that extends the strategic information-sharing game to an adaptive, asymmetric, trust-sensitive, and digitally augmented setting.

At the core of our approach is the use of a digital twin for each player: a computational model that maintains beliefs about other players, updates internal trust estimators, simulates strategy outcomes, and performs utility-based optimization. Each digital twin acts as a cognitive engine, enabling the player to reason through "what-if" scenarios, adaptively refine its probing and deception levels, and optimize utility in competitive, uncertain environments.

We first formalize a trust-weighted multi-agent game where each player selects a triple-strategy profile: $\alpha$ (information sharing), $\beta$ (probing effort), and $\delta$ (deception). We propose the framework to fully asymmetric agents with individual cost sensitivities, personalized trust update dynamics, and heterogeneous reward structures. Finally, we construct a Nash learning algorithm that allows each player's digital twin to compute personalized gradients and adaptively converge toward equilibrium.

Our contributions are threefold:

- A mathematical formulation of the asymmetric trust-aware information game with deception and probing.
- A digital twin architecture that supports individualized simulation, belief updating, and best-response optimization.
- A symbolic and numerical solution pipeline, including a gradient-based Nash solver and population-level simulation.

This framework advances the modeling of real-world strategic systems where cognitive, deceptive, and cooperative dynamics are deeply intertwined.

## II. Related Work

This work draws upon and integrates insights from several domains: multi-agent game theory, trust-aware systems, deception modeling, and the emerging use of digital twins for decision-making.

### A. Multi-Agent Game Theory and Strategy Learning

Traditional game-theoretic models have long explored optimal strategies in competitive settings with incomplete or asymmetric information. Notable examples include Bayesian

games, signaling games, and evolutionary dynamics. In particular, repeated games with private information have been used to model scenarios where agents learn and adapt over time [1]. However, these models typically assume simplified or symmetric structures that limit their applicability to heterogeneous populations with personalized cost and trust models. Our approach extends these foundations by introducing per-agent heterogeneity and continuous gradient-based adaptation.

Recent work in multi-agent reinforcement learning (MARL) further explores learning best responses in dynamic environments, using techniques like fictitious play, policy gradients, or Q-learning [2]. While powerful, most MARL frameworks do not explicitly model trust evolution or deception. Our model closes this gap by embedding trust-sensitive utilities and strategic information filtering into the learning process.

### B. Trust and Deception Modeling in Agent Systems

Trust and deception have been studied in agent-based modeling as key mechanisms influencing cooperation, reputation, and information flow. Computational trust models (e.g., REGRET [3], FIRE [4]) define trust metrics based on past behavior, while game-theoretic models examine incentives for truthful vs. deceptive behavior in signaling games [5]. Recent work has focused on learning trust-aware policies in social dilemmas and negotiation settings [6], [7].

Our framework incorporates deception as a continuous strategy variable and allows trust to evolve endogenously based on inferred honesty. Unlike binary or threshold-based models, our trust update rules are smooth and personalized, allowing fine-grained adaptation to strategic misalignment or emerging cooperation.

### C. Digital Twins for Cognitive and Adaptive Systems

The concept of digital twins originated in engineering and IoT systems, where a physical object is mirrored by a real-time, data-driven simulation [8]. More recently, digital twins have been proposed for intelligent control, smart cities, and even digital humans [9]. In multi-agent contexts, digital twins are now being explored as internal decision models that enable real-time adaptation and forecasting [10].

Our work is one of the first to formalize digital twins as embedded simulation agents in competitive strategy games. Each digital twin estimates trust, predicts opponent behavior, performs utility optimization, and enables model-predictive reasoning over time. This supports adaptive and personalized strategy learning, even in adversarial and asymmetric conditions.

### D. Nash Learning and Gradient-Based Best Responses

Recent research on learning in games has explored the use of differentiable utility functions and policy gradients to approximate Nash equilibria in continuous action spaces [11]. Our model adopts a similar approach by using closed-form utility gradients to guide each agent's best-response updates. However, we augment this process with digital twin–mediated opponent modeling and trust dynamics, yielding a richer and more interpretable behavior space.

## III. Methodology

This section provides a complete mathematical formulation of the proposed multi-agent game under asymmetric information, integrating trust dynamics, deception modeling, and digital twin–guided strategy optimization.

### A. Game Structure and Player Strategies

Let $\mathcal{N} = \{1, 2, \ldots, N\}$ be the set of players. The interaction is modeled as a repeated game with discrete rounds indexed by $t = 0, 1, 2, \ldots$. In each round $t$, every player $i \in \mathcal{N}$ selects a strategy vector:

$$\pi_i^t = (\alpha_i^t, \beta_i^t, \delta_i^t) \in [0,1]^3, \quad \text{subject to} \quad \alpha_i^t + \beta_i^t + \delta_i^t \leq 1, \tag{1}$$

where:

- $\alpha_i^t$: degree of information sharing,
- $\beta_i^t$: probing effort to extract information from others,
- $\delta_i^t$: deception level to obfuscate shared information.

The constraint $\alpha_i^t + \beta_i^t + \delta_i^t \leq 1$ enforces a bounded strategy budget, ensuring that players cannot exceed their total behavioral allocation in each round.

For a given player $i \in \mathcal{N}$, we denote by $\pi_{-i}^t$ the joint strategy profile of all other players:

$$\pi_{-i}^t := \left\{ \pi_j^t \,\middle|\, j \in \mathcal{N},\, j \neq i \right\}. \tag{2}$$

Each player $i$ is further characterized by the following *individualized parameters*:

- $c_{1i} > 0$: Cost coefficient for information sharing.
- $c_{2i} > 0$: Cost coefficient for probing.
- $c_{3i} > 0$: Cost coefficient for deception.
- $\lambda_i > 0$: Sensitivity to marginal gains in information, be used in (10).
- $R_i > 0$: Maximum attainable reward, be used in (11).

The strategy execution incurs a quadratic cost for each behavior. The total cost function for player $i$ in round $t$ is defined as:

$$C_i^t(\pi_i^t) = c_{1i}(\alpha_i^t)^2 + c_{2i}(\beta_i^t)^2 + c_{3i}(\delta_i^t)^2, \tag{3}$$

which penalizes extreme or imbalanced behavior. Quadratic form ensures convexity and smooth gradients for learning.

### B. Deception Modeling

Deception is explicitly modeled as a strategic decision variable $\delta_i^t \in [0,1]$ for each player $i$, representing the degree to which the player misrepresents or distorts the information they share in round $t$. Rather than treating deception as a binary act, we allow it to vary continuously, enabling more nuanced behaviors such as partial truth-telling or subtle misdirection.

The deception-modulated shared information is modeled as:

$$\tilde{\alpha}_i^t = \alpha_i^t(1 - \delta_i^t), \tag{4}$$

where $\tilde{\alpha}_i^t$ denotes the *effective honesty* of the information shared by player $i$.

This transformation appears in two critical places:

1) In the player's own information gain, which is penalized if their shared data is deceptive.

2) In trust updates by others, where players observing high $\alpha_i^t$ and low $\delta_i^t$ are more likely to increase their trust in player $i$.

Deception incurs a quadratic cost $c_{3i}(\delta_i^t)^2$, capturing the idea that misleading others may reduce future information access, reputation, or utility, even if short-term gains are realized. Thus, players must trade off between short-term strategic advantage and long-term relational loss.

### C. Digital Twin Architecture

Each player $i \in \mathcal{N}$ is paired with a dedicated digital twin player $\mathcal{T}_i$, which functions as a personalized, autonomous optimization engine. The digital twin continuously monitors the environment, simulates possible actions, predicts outcomes, and updates strategies in real time. Its architecture consists of five key computational modules:

*1) Belief and Opponent Modeling:* The digital twin maintains a predictive belief distribution over the strategies of all other players $\hat{\pi}^t_{-i}$. For each opponent $j \neq i$, the digital twin constructs an empirical distribution $\hat{P}^t_{ij}(\pi_j)$, from which it derives estimates of their likely actions:

$$\hat{\alpha}^t_{ij} = \mathbb{E}_{\hat{P}^t_{ij}}[\alpha_j], \quad \hat{\delta}^t_{ij} = \mathbb{E}_{\hat{P}^t_{ij}}[\delta_j]. \tag{5}$$

- $\hat{\alpha}^t_{ij}$: Estimated level of information sharing by player $j$, as perceived by player $i$'s digital twin $\mathcal{T}_i$.
- $\hat{\delta}^t_{ij}$: Estimated deception level of player $j$, as inferred by player $i$'s digital twin $\mathcal{T}_i$.

These expectations are used both in trust updates and to anticipate strategic responses. Belief updates may be Bayesian or based on moving averages of historical strategy traces.

The effective honesty (4) of the information shared by player $j$ inferred by digital twin player $\mathcal{T}_i$ becomes

$$\tilde{\hat{\alpha}}^t_{ij} = \hat{\alpha}^t_{ij}(1 - \hat{\delta}^t_{ij}). \tag{6}$$

*2) Trust Dynamics:* In a competitive game where information can be shared, hidden, or falsified, the concept of *trust* plays a central role in determining how much weight a player assigns to information received from others. We define a dynamic trust variable $\tau^t_{ij} \in [0,1]$ as the level of trust that player $i$ has in player $j$ at round $t$. This trust governs how much player $i$ is willing to rely on information shared by player $j$.

The soft trust update rule is defined as:

$$\begin{aligned} \tau^{t+1}_{ij} &= (1-\eta_i)\cdot \tau^t_{ij} + \eta_i \cdot \tilde{\hat{\alpha}}^t_{ij} \\ &= (1-\eta_i)\cdot \tau^t_{ij} + \eta_i \cdot \hat{\alpha}^t_{ij}(1-\hat{\delta}^t_{ij}), \end{aligned} \tag{7}$$

where $\eta_i \in (0,1)$ is player $i$'s trust responsiveness parameter, which controls how quickly trust adjusts based on new evidence. This update rule is recursive and smooth, capturing both short-term fluctuations and long-term reputational effects. A high value of $\eta_i$ implies that player $i$ quickly adjusts trust based on new observations, whereas a low $\eta_i$ means the player is conservative in updating trust and requires more consistent behavior to change beliefs.

High trust increases when another player shares more ($\hat{\alpha}^t_{ij}$ high) and deceives less ($\hat{\delta}^t_{ij}$ low). Conversely, low or negative updates occur when a player appears evasive or dishonest.

Trust scores directly affect player $i$'s perceived information gain via:

$$\sum_{j \neq i} \tau^t_{ij} \cdot \tilde{\hat{\alpha}}^t_{ij}, \tag{8}$$

which is a key term in the overall information gain expression $\mathcal{I}^t_i$ in (9). Thus, trust is not just a passive measure of opinion, but an active modifier of reward-relevant information flow in the game.

This trust mechanism enables personalized, memory-aware adaptation, which is critical in environments with persistent deception, variable incentives, and evolving alliances.

*3) Information Value Modeling:* Player $i$'s effective information gain at round $t$ is defined as:

$$\begin{aligned} \mathcal{I}^t_i &= a_i \tilde{\alpha}^t_i + a_i \sum_{j \neq i} \tau^t_{ij} \cdot \tilde{\hat{\alpha}}^t_{ij} + b_i \beta^t_i \\ &= a_i \alpha^t_i (1-\delta^t_i) + a_i \sum_{j \neq i} \tau^t_{ij} \cdot \hat{\alpha}^t_{ij}(1-\hat{\delta}^t_{ij}) + b_i \beta^t_i, \end{aligned} \tag{9}$$

where $a_i$ and $b_i$ control the strength of self-contributed and probed information. The first term is self-contributed honest information. Players can earn reward by sharing honest information with others, and they reward themselves for their own honesty in the payoff function (10). The second term measures the trust-weighted received info from others. The last term is probing-based information gain.

Each player receives a reward proportional to their information advantage, computed via a softmax function:

$$P^t_i = \frac{e^{\lambda_i \mathcal{I}^t_i}}{\sum_{k=1}^N e^{\lambda_k \mathcal{I}^t_k}}, \tag{10}$$

where:

- $\lambda_i$: Information sensitivity coefficient for player $i$, reflecting how steeply their reward responds to marginal increases in information.
- $\mathcal{I}^t_i$: Total trusted and acquired information in round $t$, including own honest sharing, trusted input from others, and probing returns.

This softmax formulation satisfies the following properties:

- Players are rewarded not for absolute information, but for relative information advantage, i.e., the softmax formulation induces strategic competition: increasing $\mathcal{I}^t_i$ improves $P^t_i$, but not in isolation—it depends on how $\mathcal{I}^t_i$ compares to others.
- As $\lambda_i \to 0$, payoffs become evenly distributed; as $\lambda_i \to \infty$, the highest-information player monopolizes the reward.

*4) Utility Modeling:* The player's objective is to choose a strategy $\pi^t_i$ that maximizes net utility, which depends on:

- The amount and quality of acquired information in belief estimator in (5),

- The competitive softmax-adjusted payoff share defined in (10),
- The incurred cost $C_i^t(\pi_i^t)$ defined in (3).

Since $P_i^t$ depends on both the player $i$'s own strategy $\pi_i^t$ and the digital twin's belief $\hat{\pi}_{-i}^t$ over opponents' current strategies $\pi_{-i}^t$ defined in (2), the actual utility received by player $i$ is then:

$$\begin{aligned}\mathcal{U}_i^t :&= \mathcal{U}_i^t(\pi_i^t, \pi_{-i}^t) \\ &\approx \mathcal{U}_i^t(\pi_i^t, \hat{\pi}_{-i}^t) \\ &= P_i^t \cdot R_i - C_i^t(\pi_i^t) \\ &= P_i^t \cdot R_i - c_{1i}(\alpha_i^t)^2 - c_{2i}(\beta_i^t)^2 - c_{3i}(\delta_i^t)^2,\end{aligned} \tag{11}$$

where the second term reflects the personalized cost structure for each dimension of strategic behavior.

*5) Nash Learning via Best Response Dynamics:* In a multi-agent game with asymmetric players and evolving trust dynamics, each player's optimal strategy must account for both their own expected utility and the behavior of others. We define a best-response learning framework in which each player's digital twin adapts its strategy over time using personalized gradient-based updates.

The digital twin seeks to maximize cumulative discounted utility:

$$\max_{\{\pi_i^t\}_{t=1}^T} \sum_{t=1}^{T} \gamma^{t-1} \mathbb{E}_{\pi_{-i}^t} \left[ \mathcal{U}_i^t(\pi_i^t, \pi_{-i}^t) \right], \tag{12}$$

where $\gamma \in (0, 1]$ is a discount factor that reduces the weight of future utilities.

However, direct optimization of this long-horizon objective is often computationally infeasible due to recursive dependence on other players' future actions. To address this, we adopt a *best-response learning approximation* that updates each player's strategy using a sampled or estimated utility gradient that includes future rewards:

$$\pi_i^{t+1} = \text{Proj}_\Delta \left( \pi_i^t + \hat{\eta}_i \cdot \widehat{\nabla}_{\pi_i^t} \sum_{k=t}^{T} \gamma^{k-t} \mathbb{E}_{\pi_{-i}^k} \left[ \mathcal{U}_i^k \right] \right), \tag{13}$$

where:

- $\hat{\eta}_i$ is the learning rate for player $i$,
- $\widehat{\nabla}$ represents an estimated gradient of cumulative utility, computed via finite-horizon rollouts, local value-function approximation, or model-predictive simulation within the digital twin,
- $\text{Proj}_\Delta$ denotes projection onto the simplex $\{\pi \in [0,1]^3 : \alpha + \beta + \delta \leq 1\}$.

This formulation reflects an adaptive best-response rule in which each player updates their strategy with respect to an internally simulated belief about opponent behavior and the long-term impact of current actions.

In practice, we implement this myopic learning rule using a single-step temporal approximation as a first-order approximation to the policy gradient of (12):

$$\pi_i^{t+1} = \text{Proj}_\Delta \left( \pi_i^t + \hat{\eta}_i \cdot \nabla_{\pi_i^t} \mathcal{U}_i^t(\pi_i^t, \hat{\pi}_{-i}^t) \right), \tag{14}$$

where

$$\begin{aligned}\nabla_{\pi_i^t} \mathcal{U}_i^t &= \left( \frac{\partial \mathcal{U}_i^t}{\partial \alpha_i^t}, \frac{\partial \mathcal{U}_i^t}{\partial \beta_i^t}, \frac{\partial \mathcal{U}_i^t}{\partial \delta_i^t} \right) \\ &= \begin{pmatrix} \lambda_i R_i P_i^t (1 - P_i^t) a_i (1 - \delta_i^t) - 2c_{1i}\alpha_i^t \\ \lambda_i R_i P_i^t (1 - P_i^t) b_i - 2c_{2i}\beta_i^t \\ -\lambda_i R_i P_i^t (1 - P_i^t) a_i \alpha_i^t - 2c_{3i}\delta_i^t \end{pmatrix}\end{aligned} \tag{15}$$

See the derivation in the appendix A. Over time, as players respond to one another's adjustments, this learning process converges toward a local Nash equilibrium, provided that updates are smooth, utility gradients are bounded, and learning rates decay appropriately.

The use of digital twins enables each player to simulate and plan proactively, allowing them to adapt not only to observed payoffs but also to their beliefs about the evolving population of opponents.

## IV. Simulation and Visualization of Strategic Dynamics

We simulate a population of $N = 5$ heterogeneous players interacting over $T = 50$ rounds under the digital twin–driven adaptive strategy framework. Each player begins with a randomly initialized strategy vector $\pi_i^0 = (\alpha_i^0, \beta_i^0, \delta_i^0)$ projected onto the unit simplex. Cost coefficients $(c_{1i}, c_{2i}, c_{3i})$, information sensitivity $\lambda_i$, and learning rate $\eta_i$ are independently drawn from uniform distributions, ensuring asymmetry in player preferences and learning behaviors.

At each round, each digital twin performs a best-response update based on the local utility gradient $\nabla_{\pi_i^t} U_i^t(\pi_i^t, \hat{\pi}_{-i}^t)$ and projects the result back onto the feasible simplex. Opponent strategies are estimated using prior-round behavior, and trust scores $\tau_{ij}^t$ are updated based on perceived honesty: $\hat{\alpha}_{ij}^t(1-\hat{\delta}_{ij}^t)$.

### A. Strategy Evolution

Figure 1 shows the evolution of sharing ($\alpha$), probing ($\beta$), and deception ($\delta$) strategies over time for all five players. We observe considerable heterogeneity in final strategic profiles:

- Some players converge to high $\alpha$ and low $\delta$ values, reflecting cooperative behavior.
- Others emphasize probing ($\beta$) and minimize sharing, signaling strategic opportunism.
- Deception evolves adaptively; players with lower trust tend to increase $\delta$ to distort shared information.

### B. Utility and Trust Dynamics

Figure 2 and Figure 3 illustrate how players' utility and trust values evolve over time:

In utility evolution, players with stable, trust-building behaviors achieve consistently higher net utility over time. Conversely, erratic or deceptive players experience volatile or suppressed payoffs.

- Initial low utility: All players begin cautiously due to lack of trust and minimal shared information.
- Differentiation emerges: As some players begin to share honestly and others adapt, utilities diverge.

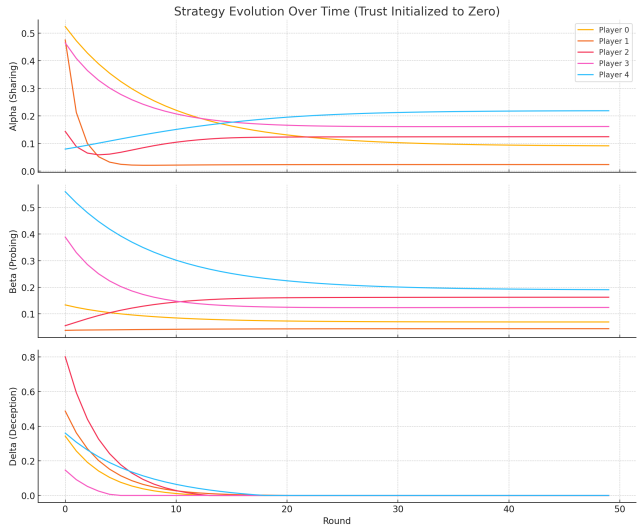


Fig. 1: Strategy evolution over time for all players. Each subplot corresponds to one strategic axis: $\alpha$ (sharing), $\beta$ (probing), and $\delta$ (deception).

- Stability forms: Over time, strategies stabilize and consistent utility trajectories emerge—indicating convergence toward locally optimal behaviors.

In trust dynamics, trust values are updated dynamically based on each player's perceived honesty. Some players maintain high trust with most peers, while others become marginalized due to high $\delta$ or low $\alpha$ behavior.

- Each subplot at position $(i, j)$ shows how player $i$'s trust in player $j$ evolves from round 0 to 49.
- Since trust started at zero, growth occurs only through observed honest sharing and low deception.
- We can see asymmetric and heterogeneous trust formation, with some player pairs stabilizing at high trust and others remaining low or fluctuating.

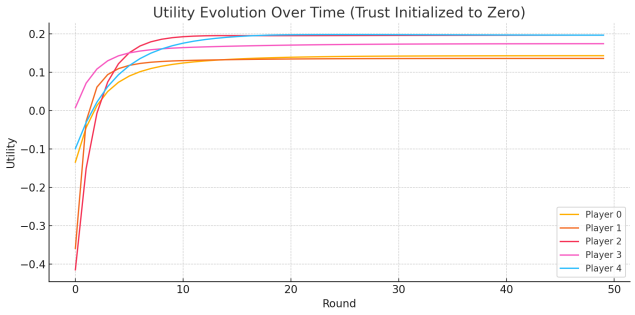


Fig. 2: Utility evolution over 50 rounds.

### *C. Symmetric Equilibrium Surfaces with Varying Sensitivity and Deception Cost*

Assume a symmetric population of $N$ players with identical parameters, we derive the closed-form symmetric equilibrium strategies $\alpha^*$, $\beta^*$, and $\delta^*$ as functions of the two key parameters:

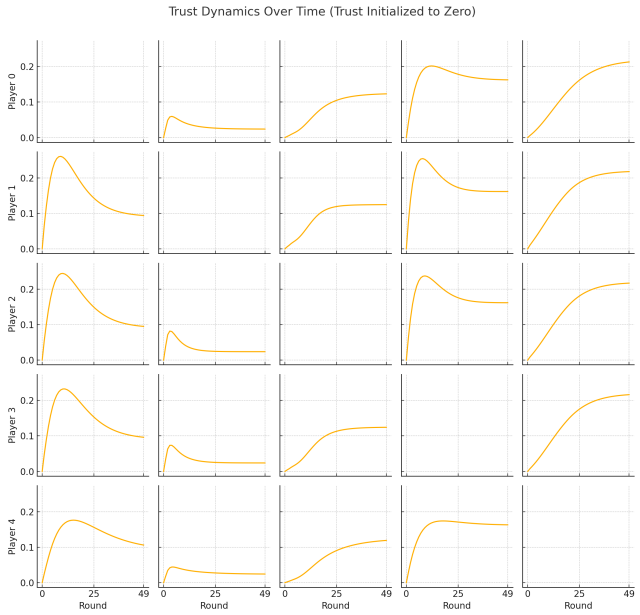


Fig. 3: Pairwise trust dynamics between all players over time.

- $\lambda$: the information sensitivity, controlling the weight of information gain in determining reward share.
- $c_3$: the cost of deception, regulating the incentive for strategic misrepresentation.

Then we get the final symmetric equilibrium system. See the derivation in the appendix B.

$$\begin{cases} \beta^* = \dfrac{\lambda R(N-1)b}{2N^2 c_2}, \\ \delta^* = \dfrac{\theta a \alpha^* + 2\theta a(N-1)(\alpha^*)^2}{2c_3 + 2\theta a(N-1)(\alpha^*)^2}, \\ \alpha^* \text{ is computed as the root of} \\ \qquad \theta \cdot \left[a(1-\delta^*) + 2a(N-1)\alpha^*(1-\delta^*)^2\right] = 2c_1\alpha^*, \\ \theta = \dfrac{\lambda R(N-1)}{N^2}. \end{cases} \tag{16}$$

We numerically compute the equilibrium strategies across a grid of values $(\lambda, c_3) \in [0.1, 3] \times [0.1, 5]$ and visualize the results using 3D surface plots with contour overlays in Figure 4.

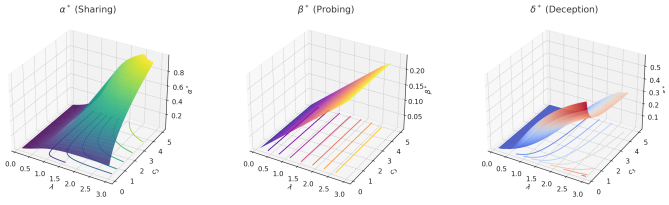


Fig. 4: Symmetric equilibrium strategies as functions of information sensitivity $\lambda$ and deception cost $c_3$. Left: Sharing $\alpha^*$, Center: Probing $\beta^*$, Right: Deception $\delta^*$. Contour lines illustrate gradient transitions.

- $\alpha^*$ increases with $c_3$, as higher deception cost discourages obfuscation and encourages honest sharing. It tends to decrease with $\lambda$, since higher sensitivity to information reward motivates secrecy.
- $\beta^*$ increases linearly with $\lambda$ and remains invariant to $c_3$, since probing does not directly interact with deception incentives.
- $\delta^*$ decreases as $c_3$ rises and increases with $\lambda$, reflecting that players deceive more when deception is cheap and information is valuable.

These numerical results confirm the expected monotonic and non-linear relationships in the symmetric equilibrium framework, and illustrate the strategic trade-offs embedded in the cost-reward structure.

## V. Conclusion

This paper presents a comprehensive framework for modeling and optimizing strategic behavior in trust-sensitive, asymmetric multi-agent games through the integration of digital twins. Each agent faces the challenge of deciding how much private information to share, whether to probe others for secrets, and how much to deceive—while adapting to dynamic interpersonal trust.

The proposed model formulates each agent's behavior as a continuous strategy vector with individualized cost parameters, trust responsiveness, and information sensitivity. Trust is treated as a dynamic, learned quantity that modulates each agent's perceived information gain. A digital twin is paired with every agent to simulate strategic interactions, estimate opponent honesty, and perform gradient-based policy updates.

We derive closed-form expressions for utility gradients, and propose a projected gradient ascent learning rule to approximate Nash equilibria. Simulations reveal emergent behavioral diversity, including adaptive deception, trust-driven cooperation, and information-maximizing opportunism. Equilibrium surfaces show how deception cost and information sensitivity jointly shape optimal strategies.

Overall, the results demonstrate that digital twin–guided agents can learn socially interpretable and strategically stable behaviors, with trust functioning as an internal regulatory mechanism. This framework offers promising applications in decentralized negotiation, autonomous multi-agent planning, and adversarial communication settings. Future work will address partial observability, uncertainty-aware trust dynamics, and real-world deployment in distributed autonomous systems.

## References


[1] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.

[2] K. Zhang, Z. Yang, and T. Basar, "Multi-agent reinforcement learning: A selective overview," *arXiv preprint arXiv:1911.10635*, 2019.

[3] J. Sabater and C. Sierra, "REGRET: Reputation in gregarious societies," in *Proc. of the 5th International Conference on Autonomous Agents*, 2001.

[4] T. D. Huynh, N. R. Jennings, and N. R. Shadbolt, "An integrated trust and reputation model for open multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 13, no. 2, pp. 119–154, 2006.

[5] V. Crawford and J. Sobel, "Strategic information transmission," *Econometrica*, vol. 50, no. 6, pp. 1431–1451, 1982.

[6] A. Lerer and A. Peysakhovich, "Maintaining cooperation in complex social dilemmas using deep reinforcement learning," *arXiv preprint arXiv:1707.01068*, 2017.

[7] Bowen Baker, "Emergent Reciprocity and Team Formation from Randomized Uncertain Social Preferences," in *NeurIPS*, 2020.

[8] M. Grieves, "Digital twin: manufacturing excellence through virtual factory replication," *White Paper, Florida Institute of Technology*, 2015.

[9] A. Fuller, Z. Fan, C. Day, and D. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE Access*, vol. 8, pp. 108952–108971, 2020.

[10] Y. Lu et al., "Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues," *Robotics and Computer-Integrated Manufacturing*, 61. 10.1016/j.rcim.2019.101837.

[11] D. Balduzzi, S. Racanière, J. Martens, J. Foerster, T. Tuyls, and T. Graepel, "The mechanics of n-player differentiable games," in *ICML*, 2018.

[12] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387–434, 2005.

[13] I. Rahwan et al., "Machine behaviour," *Nature*, vol. 568, no. 7753, pp. 477–486, 2019.

[14] J. N. Foerster et al., "Learning to communicate with deep multi-agent reinforcement learning," in *NeurIPS*, 2016.

[15] M. Carroll et al., "On the utility of learning about humans for human-AI coordination," in *NeurIPS*, 2019.


## Appendix

### A. Derivation of Equation (15)

Equation (15) provides the partial derivatives of player $i$'s utility $U_i^t$ with respect to their strategy components $\alpha_i^t$, $\beta_i^t$, and $\delta_i^t$, under the assumption that the player is optimizing against an estimated belief $\hat{\pi}_{-i}^t$ of opponents' strategies. We reproduce the utility function from Equation (11):

$$U_i^t = P_i^t \cdot R_i - c_{1i}(\alpha_i^t)^2 - c_{2i}(\beta_i^t)^2 - c_{3i}(\delta_i^t)^2$$

where $P_i^t$ is the softmax-based reward share given by:

$$P_i^t = \frac{e^{\lambda_i \mathcal{I}_i^t}}{\sum_{k=1}^N e^{\lambda_k \mathcal{I}_k^t}}$$

and $\mathcal{I}_i^t$ is the player's total perceived information gain:

$$\mathcal{I}_i^t = a_i \alpha_i^t (1 - \delta_i^t) + a_i \sum_{j \neq i} \tau_{ij}^t \hat{\alpha}_j^t (1 - \hat{\delta}_j^t) + b_i \beta_i^t$$

We now derive the gradient $\nabla_{\pi_i^t} U_i^t$ with respect to each component $\pi_i^t = (\alpha_i^t, \beta_i^t, \delta_i^t)$, applying the chain rule and treating $P_i^t$ as a softmax function:

*a) 1. Derivative with respect to $\alpha_i^t$::* First, compute the derivative of $\mathcal{I}_i^t$ with respect to $\alpha_i^t$:

$$\frac{\partial \mathcal{I}_i^t}{\partial \alpha_i^t} = a_i (1 - \delta_i^t)$$

Then use the softmax derivative:

$$\frac{\partial P_i^t}{\partial \mathcal{I}_i^t} = \lambda_i P_i^t (1 - P_i^t)$$

So,

$$\frac{\partial U_i^t}{\partial \alpha_i^t} = \lambda_i R_i P_i^t (1 - P_i^t) \cdot a_i (1 - \delta_i^t) - 2 c_{1i} \alpha_i^t$$

*b) 2. Derivative with respect to $\beta_i^t$::*

$$\frac{\partial \mathcal{I}_i^t}{\partial \beta_i^t} = b_i \quad \Rightarrow \quad \frac{\partial U_i^t}{\partial \beta_i^t} = \lambda_i R_i P_i^t (1 - P_i^t) \cdot b_i - 2c_{2i}\beta_i^t$$

*c) 3. Derivative with respect to $\delta_i^t$::*

$$\frac{\partial \mathcal{I}_i^t}{\partial \delta_i^t} = -a_i \alpha_i^t \quad \Rightarrow \quad \frac{\partial U_i^t}{\partial \delta_i^t} = -\lambda_i R_i P_i^t (1-P_i^t) \cdot a_i \alpha_i^t - 2c_{3i}\delta_i^t$$

*d) Final Expression::* Combining all three components, we obtain the full utility gradient:

$$\nabla_{\pi_i^t} U_i^t = \begin{pmatrix} \lambda_i R_i P_i^t (1 - P_i^t) a_i (1 - \delta_i^t) - 2c_{1i}\alpha_i^t \\ \lambda_i R_i P_i^t (1 - P_i^t) b_i - 2c_{2i}\beta_i^t \\ -\lambda_i R_i P_i^t (1 - P_i^t) a_i \alpha_i^t - 2c_{3i}\delta_i^t \end{pmatrix}$$

which is exactly the expression given in Equation (15).

### *B. Symmetric Equilibrium Derivation*

We derive the symmetric equilibrium under the assumption that all players are homogeneous and adopt identical strategies. Let $\pi_i = (\alpha, \beta, \delta)$ for all $i \in \mathcal{N}$. We use the utility formulation and trust dynamics defined in Section III.

*1) Assumptions:* All players share identical parameters:

- Strategy: $\alpha, \beta, \delta \in [0, 1]$
- Cost coefficients: $c_1, c_2, c_3$
- Reward: $R$, Information coefficients: $a, b$
- Information sensitivity: $\lambda$
- Number of players: $N$

Trust is symmetric and steady-state:

$$\tau_{ij} = \alpha(1 - \delta) \quad \text{for all } i \neq j$$

*2) Information Gain:* Each player's total information gain is:

$$\mathcal{I}_i = a\alpha(1 - \delta) + a(N - 1)\alpha^2(1 - \delta)^2 + b\beta$$

*3) Softmax Utility (Symmetric):* Since all players are identical, each receives the same softmax probability:

$$P_i = \frac{e^{\lambda \mathcal{I}_i}}{\sum_{k=1}^{N} e^{\lambda \mathcal{I}_k}} = \frac{1}{N}$$

and the utility simplifies to:

$$U_i = \frac{R}{N} - c_1\alpha^2 - c_2\beta^2 - c_3\delta^2$$

To derive equilibrium strategy profiles, we allow infinitesimal deviation and compute first-order conditions (FOCs) assuming the player maximizes:

$$U_i = \frac{e^{\lambda \mathcal{I}_i}}{\sum_k e^{\lambda \mathcal{I}_k}} R - c_1\alpha^2 - c_2\beta^2 - c_3\delta^2$$

*4) Gradient Components:* Define $\theta = \frac{\lambda R(N-1)}{N^2}$. Then:

*a) FOC w.r.t. $\beta$::*

$$\frac{\partial U_i}{\partial \beta} = \theta b - 2c_2\beta = 0 \Rightarrow \boxed{\beta^* = \frac{\theta b}{2c_2}}$$

*b) FOC w.r.t. $\delta$::*

$$\delta^* = \frac{\theta a\alpha + 2\theta a(N-1)\alpha^2}{2c_3 + 2\theta a(N-1)\alpha^2}$$

*c) FOC w.r.t. $\alpha$::* Solve numerically from:

$$\theta \left[a(1 - \delta) + 2a(N - 1)\alpha(1 - \delta)^2\right] = 2c_1\alpha$$

*5) Final System:*

$$\beta^* = \frac{\lambda R(N-1)b}{2N^2 c_2}$$

$$\delta^* = \frac{\theta a\alpha^* + 2\theta a(N-1)(\alpha^*)^2}{2c_3 + 2\theta a(N-1)(\alpha^*)^2}$$

$\alpha^*$ solves: $\theta \cdot \left[a(1 - \delta^*) + 2a(N - 1)\alpha^*(1 - \delta^*)^2\right] = 2c_1\alpha^*$

This derivation yields a system of coupled nonlinear expressions, solvable numerically, and used in Section IV for equilibrium surface visualization.