AI4Games: A General Strategy Search Framework for Evolutionary Games

Authors: Wang Hongyu, Wang Long, Wang Long

Date: 2025-08-29T00:00:00+00:00

Abstract

Systematically mining strategies with long-term evolutionary advantages in multi-agent systems is a key challenge in evolutionary game theory, complex systems science, and artificial intelligence research. This paper proposes and implements a unified strategy search framework—AI4Games, which systematically transforms the strategy construction problem into a reinforcement learning-driven strategy mining task. The framework abstracts five general steps: strategy representation, behavioral interaction, reward construction, strategy optimization, and strategy selection, thereby constructing a generalizable and reusable technical pathway applicable to different game environments and behavioral modeling tasks. As validation, we apply AI4Games to the evolutionary iterated prisoner's dilemma and successfully discover a two-step memory strategy with bilateral reciprocity structure (MTBR), whose behavioral rules are concise and interpretable, demonstrating significant payoff capability and evolutionary stability across multiple game environments. This strategy is not manually designed but automatically generated by the framework, showcasing AI4Games' capability to emerge high-quality behavioral patterns in high-dimensional strategy spaces. The proposal of AI4Games not only enhances the level of strategy modeling in evolutionary games but also demonstrates the methodological value of artificial intelligence in strategy evolution, mechanism design, and complex behavioral system modeling. As an extension of the AI for Science paradigm in the field of game theory, AI4Games provides theoretical support and a tool foundation for promoting interdisciplinary intelligent modeling under the "AI+" background.

Full Text

AI4Games: A General Strategy Discovery Framework for Evolutionary Games

Hongyu Wang[1], Long Wang[1],*
[1] Center for Systems and Control, Peking University, Beijing 100871, China

Abstract

Discovering strategies with long-term evolutionary advantages in multi-agent systems is a fundamental problem at the intersection of evolutionary game theory, complex systems science, and artificial intelligence. This paper presents a general strategy discovery framework, AI4Games, which systematically transforms strategy design into a reinforcement learning–driven optimization task. The framework abstracts five generalizable components—strategy representation, interaction design, reward construction, optimization, and evaluation—forming a reusable pipeline applicable to diverse game-theoretic and behavioral modeling settings.

To validate the framework, we apply AI4Games to the evolutionary Iterated Prisoner's Dilemma and successfully uncover a memory-two bilateral reciprocity strategy (MTBR) that emerges naturally from training. MTBR exhibits interpretable behavioral rules, robust performance across heterogeneous opponents, and strong evolutionary stability. Its emergence as a non-predefined outcome highlights the framework's capability in navigating high-dimensional strategy spaces and discovering effective behavioral patterns.

AI4Games advances strategy modeling beyond hand-crafted heuristics and exemplifies the methodological contribution of AI for Science (AI4S) in the game-theoretic domain. It provides both a theoretical foundation and a practical tool for cross-disciplinary modeling under the "AI+" national initiative.

Keywords: evolutionary game; reinforcement learning; strategy discovery; multi-agent system; AI for science

Corresponding author: Long Wang, E-mail: longwang@pku.edu.cn

Introduction

In recent years, artificial intelligence (AI) has been profoundly reshaping social structures, research paradigms, and economic logic, emerging as a core driving force behind a new wave of technological revolution and industrial transformation. Its influence has expanded from auxiliary computation and information processing to the ontological innovation of scientific research methodologies, propelling the research system from an "experience-driven" to an "intelligence-driven" paradigm. In August 2025, the State Council officially released the

"Opinions on Deepening the Implementation of the 'AI Plus' Action," calling for accelerated exploration of an "AI-driven new paradigm for scientific research" and explicitly supporting AI empowerment throughout the entire process of technology development, engineering implementation, and scientific discovery. The document emphasizes promoting AI's interdisciplinary traction and systematic collaborative innovation, aiming to achieve original scientific breakthroughs from "0 to 1," accelerate the cultivation of new quality productive forces, and build intelligent research infrastructure. This national strategy aligns closely with the increasingly prominent "AI for Science (AI4S)" paradigm in international academia.

In 2021, multiple domestic and international AI researchers proposed that AI should serve as a crucial driver for scientific research, assisting or even reconstructing traditional research paradigms [**?**]. In 2023, a group of renowned scholars and technology organization representatives published an article in *Nature* that systematically mapped AI's pathways in scientific discovery, highlighting its significant advantages in data-driven hypothesis generation, experimental design optimization, and complex systems modeling, and positioning it as a novel bridge connecting theory and practice [**?**]. It has become clear that AI is no longer merely a technical tool for enhancing research efficiency but is becoming an ontological force participating in scientific discovery itself. Domestic scholars have also proposed a top-level research layout for "AI4S," calling for the construction of an intelligent research system covering multiple fundamental disciplines [**?**]. This series of works marks a paradigm shift in scientific research from "human-dominated" to "human-machine collaboration," providing robust theoretical support and a technological engine for Chinese-style modernization and independent scientific innovation.

Currently, the deep integration of AI across multiple scientific research directions continues to demonstrate disruptive potential. For instance, in biomedicine, AI has been widely applied to drug screening, protein structure prediction, and clinical pathway optimization. In 2025, domestic scholars noted in *Nature Medicine* that AI is poised to reshape the entire new drug development process, significantly reducing both development cycles and failure rates [**?**]. In physics, from quantum computing and particle simulation to materials discovery, AI tools are assisting physicists in solving complex modeling problems that are intractable for traditional methods [**?**]. The practical landscape of AI for Science has expanded across numerous fundamental and interdisciplinary fields, gradually shifting from an auxiliary tool to a key methodology and becoming a core engine for driving original innovation.

However, despite the breakthrough progress of the "AI for Science" research model in fields such as biology, materials, and physics, it remains relatively weak in social behavior modeling and decision-making mechanism research, particularly lacking a systematic AI research framework for game theory. In fact, game models are ubiquitous in many critical real-world issues, such as resource allocation in epidemic prevention and control, cooperation incentive mechanisms

on social platforms, attack-defense games in cybersecurity, and even multi-agent coordination and adversarial games among AI systems themselves. These problems can all be abstracted as repeated interactive choices among individuals under bounded rationality and feedback effects—precisely the core scenario that evolutionary game theory focuses on.

In this context, discovering game strategies that can persist long-term in dynamic environments and possess evolutionary stability holds not only theoretical value but also direct relevance to understanding and guiding mechanisms of cooperation, trust, punishment, and reward in social systems. Traditional strategies mostly rely on manual construction and heuristic design, which, while interpretable, struggle to systematically cover complex strategy spaces or adapt to dynamically changing game structures. Therefore, there is an urgent need to establish an "AI4Games" framework with interpretability, transferability, and scalability, enabling AI to drive scientific discovery not only in natural sciences like biology and physics but also to mine decision-making mechanisms with practical guiding significance in social behavior and agent systems. This direction not only fills the theoretical gap of AI in game modeling but also aligns with the "AI Plus" national strategy's emphasis on cross-disciplinary research paradigm innovation.

2 Strategy Discovery in Repeated Games

The evolutionary mechanisms of cooperation and defection are core concerns across multiple fields, including complex systems science, economics, and behavioral ecology [**?**]. Among various game models, repeated games provide a fundamental framework for revealing how individuals form stable behavioral patterns through long-term interactions. In reality, individuals rarely interact only once; instead, they gradually build trust and develop strategies through prolonged engagement, making behavioral evolution dependent not only on immediate payoffs but also driven by long-term adaptability. Evolutionary game theory takes strategies as the basic unit of population evolution, investigating which behavioral patterns can survive and spread within a population under mechanisms such as natural selection, learning, and imitation. From this perspective, long-term advantageous strategies are not merely those that encourage cooperation but also include mechanisms capable of punishment and defense when facing defection. Consequently, such strategies must flexibly respond to different types of opponents in complex and dynamic environments to maximize average payoffs. Thus, systematically discovering strategies with long-term evolutionary advantages in complex multi-agent environments has become a key challenge for advancing evolutionary game theory toward automation and generalization. This process involves not only strategy generation itself but also adaptive evaluation and mechanistic analysis of behavioral evolution, representing a crucial step toward achieving the goal of "AI-driven behavioral modeling."

Previous research has proposed various classic strategies (such as TFT, GTFT, WSLS, etc.) that perform well in pairwise interactions within specific environ-

ments [**?**]. These strategies rely on human expert knowledge for construction and typically offer good intuitive interpretability, yet human expertise struggles to systematically cover vast strategy spaces or adapt to changing strategic demands in complex environments [**?**]. Therefore, developing a general strategy discovery framework to automatically identify strategies with long-term evolutionary advantages is of great significance for advancing research in evolutionary game theory. Beyond strategy design itself, numerous studies have examined how game environments and population structures influence cooperation evolution, such as the role of feedback mechanisms in cooperation formation [**?**] and the regulation of cooperation maintenance and propagation by complex network structures [**?**]. Additionally, research has shown that reputation-based partner selection mechanisms can effectively promote cooperative behavior in social networks [**?**]; other work has explored transition mechanisms of game rules themselves during evolution, demonstrating how strategies achieve spread and stability in dynamic game contexts [**?**]; and recent analyses of strategy evolution on higher-order network structures have further revealed the important role of complex topologies in cooperative behavior evolution [**?**]. These studies underscore the complexity and structural dependency of strategy discovery problems, further highlighting the necessity of developing general-purpose strategy search methods.

In recent years, reinforcement learning research and applications have gradually expanded. Particularly in scenarios with high-dimensional strategy spaces and complex behavioral feedback, reinforcement learning demonstrates superior adaptability and discovery capabilities compared to traditional analytical methods [**?**]. Against this backdrop, we propose and implement the AI4Games framework, aiming to provide a universal and systematic strategy search methodology for evolutionary games. Centered on reinforcement learning and combined with customized payoff functions, the framework enables efficient search and optimization within vast strategy spaces. Unlike strategy design that relies on human experience, AI4Games proposes a set of general strategy search principles for transforming specific game tasks into forms solvable by reinforcement learning.

The general principles of the AI4Games framework are as follows:

1. **Strategy Representation and Encoding Design**: First, strategic behaviors in evolutionary games are abstracted into state-action mapping structures processable by reinforcement learning. By setting memory length and action sets, candidate strategies are formally encoded to construct the strategy space.

2. **Behavioral Interaction and Experience Collection**: Purposefully design training environments containing representative opponent strategies, enabling agents to accumulate experience through representative interactive feedback, conduct adaptive evaluation of agents, and obtain feedback that effectively drives strategy improvement.

3. **Objective Function and Feedback Mechanism**: Design and construct specific reward structures according to task requirements to guide strategies toward desired behavioral patterns, such as improving payoffs, resisting exploitation, or maintaining cooperation.
4. **Strategy Exploration and Optimization Mechanism**: Utilize reinforcement learning methods for continuous optimization in the strategy space, adjusting the balance between exploration and exploitation during learning to enhance strategy performance.
5. **Strategy Evaluation and Screening Mechanism**: Construct systematic evaluation criteria to identify outstanding strategies from candidates and verify their long-term dominance in evolutionary dynamics.

These five steps constitute the universal backbone of the AI4Games framework, applicable to various evolutionary game tasks. For specific problems, one only needs to transform them into the construction items required by these five steps to conduct strategy discovery with AI4Games. Figure 1 [Figure 1: see original paper] illustrates the logical relationships and functional divisions among the five modules of the AI4Games framework, with each component detailed in subsequent sections.

To validate the effectiveness of AI4Games, we showcase a typical output from evolutionary repeated games—the Memory-Two Bilateral Reciprocity (MTBR) strategy. This strategy emerges naturally during training with a simple and interpretable response logic. MTBR demonstrates high payoffs across various adversarial environments and can significantly improve overall average payoffs in both non-evolutionary and evolutionary simulations, while also achieving dominance in evolving populations. This result proves that AI4Games can successfully mine evolutionarily advantageous strategies from complex strategy spaces.

As an exploratory attempt of the AI4S paradigm in the domain of game theory, AI4Games is not merely an algorithmic implementation but rather a systematic and generalizable research framework for strategy discovery. Its proposal expands the theoretical boundaries of artificial intelligence in complex behavior modeling. This paper proceeds from the general design of AI4Games to elaborate its concrete implementation in evolutionary games, further demonstrating how the framework facilitates the spontaneous emergence and screening of strategies in multi-agent games, and finally showcasing its search capabilities through the behavioral characteristics and evolutionary performance of MTBR. The research demonstrates that AI4Games, as a strategy discovery platform, possesses strong extensibility and generality, providing new tools and research pathways for future explorations of high-dimensional memory strategies and more complex game structures.

3 The AI4Games Framework: Strategy Discovery via Reinforcement Learning

This section systematically introduces our proposed intelligent strategy search framework, AI4Games, a unified methodological platform for evolutionary game tasks. The framework aims to systematize and formalize the strategy discovery process, with reinforcement learning as its technical core, maintaining high interpretability while possessing generality and scalability. Its design philosophy serves not only specific game problems but can also be regarded as a universal tool for AI intervention in behavioral modeling research. Unlike traditional methods that construct strategies based on human experience, AI4Games systematically transforms the strategy search problem into a reinforcement learning task, automatically mining dominant strategies with evolutionary advantages from complex strategy spaces through structured training and evaluation mechanisms. The AI4Games framework follows the five general principles proposed in the previous section; below we sequentially introduce its concrete implementation for discovering dominant strategies in the evolutionary iterated prisoner's dilemma environment. Mining dominant strategies in repeated games faces two core challenges: first, combinatorial explosion of the strategy space (especially when considering memory strategies), and second, the multi-round game interactions and feedback required for strategy evaluation. Multi-agent Q-learning methods provide a "self-evolving without manual construction" approach for strategy discovery in complex game systems. On one hand, its state-action value-based encoding can systematically express strategy rules under limited memory; on the other hand, its experience replay and reward update mechanisms allow agents to gradually optimize their response strategies through multi-round games, thereby adapting to diverse opponents and achieving stable convergence. Compared to manual design relying on expert knowledge, multi-agent Q-learning methods are more amenable to systematic and algorithmic implementation and are better suited for embedding into cross-disciplinary "AI-driven scientific modeling" platforms.

3.1 Strategy Representation and Encoding Design

Each agent's strategy is represented in the form of a Q-table. The Q-table records the mapping between all possible historical states (i.e., action combinations from past rounds) and Q-values for all current actions (in the Prisoner's Dilemma, cooperation or defection, two options). For the case of two-step memory, the state space includes the action combinations of both self and opponent from the two previous interactions, and the strategy selects the action with the highest Q-value for each state. For a two-player repeated game with M possible actions and memory length ℓ, there are $N_{\text{state}} = M^{2\ell} + M^2$ possible states. Specifically, in our iterated Prisoner's Dilemma, we set agents to have two-step memory, involving $2^4 + 2^2 = 20$ states.

In this paper, a state refers to the interaction history used by an agent to make decisions. For an agent with two-step memory, decisions are based only on

the actions taken by that agent and its opponent in the previous two rounds. Specifically, the relevant information includes the opponent's action two rounds ago, the agent's action two rounds ago, the opponent's action in the last round, and the agent's action in the last round. We represent the state as a four-element tuple: (opponent's action two rounds ago, agent's action two rounds ago, opponent's last action, agent's last action). Note that because interaction history is limited at the beginning of the game and agents cannot yet accumulate a full ℓ-step memory, the N_{state} formula includes the additional states needed. Each agent is assigned an independent $N_{\text{state}} \times M$ Q-table, where the entry in row i and column j reflects the agent's expected long-term cumulative reward for choosing action j in state s_i (see Figure 2 [Figure 2: see original paper]).

3.2 Behavioral Interaction and Experience Collection

Our training environment contains a total of 98 individuals, including 49 agents with empty Q-tables and 49 sparring individuals with preset strategies. During training, the reinforcement learning agents engage in repeated games with these sparring individuals. The sparring population includes equal numbers of TFT, GTFT, WSLS, Hold-a-Grudge, Fool-Me-Once, GradualTFT, and OmegaTFT. In the iterated Prisoner's Dilemma game used for training, the payoff for each round depends on the action combination of both parties, with specific settings as follows: If both cooperate, each receives payoff $R = 2$; If one cooperates and one defects, the cooperator receives payoff $S = 0$, and the defector receives payoff $T = 3$; If both defect, each receives payoff $P = 0.1$. This parameter combination intentionally reduces the payoff for mutual defection, thereby weakening the attractiveness of defection and providing a more favorable environment for the emergence of cooperative strategies.

3.3 Objective Function and Feedback Mechanism

Figure 2: Schematic diagram of discovering dominant strategies in evolutionary repeated games using the AI4Games framework. We consider a population consisting of $N_a = 49$ agents and $N_m = 49$ sparring partners. Each agent is equipped with an independent $N_{\text{state}} \times M$ Q-table, where N_{state} represents the number of all possible states and M represents the number of available actions. Each sparring partner carries a preset artificial strategy, with seven categories established. In each iteration, two individuals $p1$ and $p2$ are randomly selected from the agent and sparring pools to engage in an L-round iterated Prisoner's Dilemma game. In round t, agent $p1$ consults its own Q-table based on the current state $s_{p1,t}$ (i.e., the action combinations of both parties in the most recent ℓ interactions) and selects an action. Action selection employs an ϵ-greedy strategy to balance exploration and exploitation. Sparring partners always act according to their preset fixed strategies. After completing the L-round game, the weighted payoff $W_{p1}(s_{p1,t}, a_{p1,t})$ corresponding to each action is calculated based on the agent's historical payoffs and interaction outcomes. The agent then updates its Q-table using the Bellman equation (see Equation 3).

In Q-learning, agents optimize strategies by maximizing expected payoffs. To more effectively encourage long-term cooperation in repeated games and guide strategies to resist exploitation by others, we design an objective function for updating the Q-table. We define each agent pX's state at time step t as $s_{pX,t}$, action as $a_{pX,t}$, and payoff for the action as $U_{pX,t}$. Simultaneously, we define the agent's average payoff in this repeated game as $\bar{U}_{pX}$, and its opponent's average payoff as $\bar{U}_{p-X}$. The objective function is defined as:

$$W_{pX}(s_{pX,t}, a_{pX,t}) = \begin{cases} \theta U_{pX,t} + (1-\theta)\bar{U}_{pX} & \text{if } \bar{U}_{pX} \geq \bar{U}_{p-X} \\ \theta U_{pX,t} & \text{if } \bar{U}_{pX} < \bar{U}_{p-X} \end{cases}$$

where $\theta \in [0, 1]$ is a parameter that adjusts the weight between immediate current payoff and long-term cooperative payoff. The definition of average payoff is: $\bar{U}_{pX} = (1/L)\sum_{t=1}^{L} U_{pX,t}$, where L is the number of rounds in the repeated game. Subsequently, the Q-table is updated as follows:

$$\text{New } Q_{pX}(s_{pX,t}, a_{pX,t}) = Q_{pX}(s_{pX,t}, a_{pX,t}) + \alpha[W_{pX}(s_{pX,t}, a_{pX,t}) + \gamma \max_{a'} Q_{pX}(s', a') - Q_{pX}(s_{pX,t}, a_{pX,t})]$$

where α is the learning rate, γ is the discount factor, and s' is the next state. We believe this design can effectively balance short-term instantaneous payoffs with stable long-term benefits from cooperation, thereby helping agents better survive in evolving populations and enhancing the overall cooperation level of the evolutionary population.

3.4 Strategy Exploration and Optimization Mechanism

In each training round, an agent engages in a 20-round repeated game with an opponent and updates its Q-table based on the payoffs. The Q-learning related parameters are as follows: learning rate $\alpha = 0.2$, discount factor $\gamma = 0.5$, and preference parameter in weighted payoff $\theta = 0.8$. This parameter combination demonstrates good stability and balance in experiments, accommodating both the convergence speed of reinforcement learning, the emphasis on future payoffs, and the tendency for long-term cooperation in games. Training employs an ϵ-greedy strategy for exploration, with ϵ gradually decaying to a small value as rounds increase to achieve a transition from exploration to convergence.

3.5 Strategy Evaluation and Screening Mechanism

After training is completed, we evaluate the agent strategies obtained through the AI4Games framework to assess their stability and cooperation level across different game environments. If a strategy exhibits high average payoffs against multiple types of opponents and can persist long-term in evolutionary simulations, it is considered to have the potential to become a dominant strategy. In

practice, we conduct multiple independent training runs and evaluate the results to screen for strategies that are behaviorally stable, structurally simple, and outstanding in performance. Among numerous strategies, we ultimately extract one with a clear structure and interpretable behavioral rules—the Memory-Two Bilateral Reciprocity (MTBR) strategy described below.

4 Analysis of AI4Games Framework Output: Decision Logic and Evolutionary Advantages of the Memory-Two Bilateral Reciprocity Strategy

To validate the strategy mining capability of the AI4Games framework in complex game tasks, this section showcases its representative output in the classic iterated Prisoner's Dilemma environment. This task is not only widely used in evolutionary studies of cooperative behavior but has also become a standard benchmark for validating strategy evolution algorithms due to its simple structure and flexible parameter control. We select this task as a typical application scenario to demonstrate how AI4Games can automatically emerge strategies with evolutionary advantages from high-dimensional strategy spaces. Although the AI4Games framework itself is applicable to repeated game modeling with arbitrary memory lengths and action set sizes, to more clearly characterize the structural features of its output strategies, we focus on a more challenging experimental setting: iterated Prisoner's Dilemma games with a limited number of rounds (20 rounds). In this setting, the temptation of short-term payoffs is stronger, posing greater resistance to the formation of long-term reciprocal behavior. We adopt the classic parameter configuration: $R = 3, S = 0, T = 5, P = 1$, a parameter combination with typical "defection advantage" that thus places higher demands on a strategy's fault tolerance and recovery mechanisms.

4.1 The Memory-Two Bilateral Reciprocity Strategy

Through analysis and evaluation of the trained agent strategies, we obtained a high-performance agent strategy. This strategy makes decisions based on two-step memory and exhibits good cooperation ability and payoff advantage when interacting with various sparring partners. We term it the "Memory-Two Bilateral Reciprocity" (MTBR) strategy. The core decision logic of MTBR is as follows: - When the opponent defects in the first round while the agent cooperates, MTBR continues to cooperate in the second round. This choice demonstrates fault tolerance and can encourage the opponent to reciprocate cooperation. - When both parties have defected in the past two rounds, MTBR proactively switches to cooperation, attempting to break the persistent defection pattern. - In other cases, MTBR mimics the opponent's action from the previous round, i.e., executing the "TFT" strategy.

These rules are not manually designed through mathematical analysis or inspired by biology but are automatically formed by the agent through multi-round reinforcement learning. In terms of strategy structure, MTBR possesses

both the ability to identify cooperation trends and incorporates mechanisms of forgiveness and retaliation, enabling flexible adjustment of responses when facing different types of strategies. To further demonstrate its behavioral characteristics, we conducted a visual analysis of the game process between two MTBR strategies. As shown in Figure 3 [Figure 3: see original paper], when initial states are inconsistent (e.g., one cooperates, one defects), MTBR can quickly recover and establish stable reciprocal cooperation; whereas in the same scenario, the TFT strategy falls into repeated "cooperate-defect" cycles, and GradualTFT requires many more rounds to restore cooperation. Figure 3b shows the scenario where both strategies defect in the first round. Both MTBR and GradualTFT can restore cooperation within a short time, while TFT remains trapped in persistent mutual defection. These behavioral differences indicate that MTBR possesses stronger fault tolerance and recovery capabilities, able to rapidly guide cooperation formation without requiring complex structures. This capability is particularly important in noise-free repeated games and lays the foundation for its subsequent evolutionary stability.

4.2 The Addition of MTBR Strategy Can Enhance Population Payoffs

This section examines the impact of adding the MTBR strategy on individual payoffs and population average payoff within a population with fixed strategy composition. We construct two different strategy combination environments and evaluate its ability to promote cooperation by comparing game payoffs before and after introducing MTBR. First, consider Strategy Set 1, containing seven classic strategies: GradualTFT, OmegaTFT, TFT, GTFT (tolerance 0.3), Fool-Me-Once, WSLS, and Hold-a-Grudge. In Figure 4a [Figure 4: see original paper], the purple bars represent the average payoffs when strategies within this set play against each other. After introducing the MTBR strategy (see blue bars), the average payoff of each strategy in the set increases significantly, and the overall population average payoff also improves substantially. Notably, the MTBR strategy achieves the second-highest payoff in this environment, just below GradualTFT, with minimal difference between them. GradualTFT is the most challenging opponent for MTBR in this environment, yet both can stably maintain cooperation during games. This result indicates that MTBR not only performs well itself but also enhances the overall payoff level of the entire population.

We then introduce Strategy Set 2, adding eight typical zero-determinant (ZD) strategies to Set 1. ZD strategies are characterized by their ability to control opponents' payoffs and are strongly exploitative in various scenarios. Figure 4b shows that without MTBR, the addition of ZD strategies reduces the population average payoff from 2.15 to 1.93, indicating that ZD strategies (exploitative strategies) clearly undermine population cooperation. However, after introducing MTBR into this environment, the population average payoff rises again, and the MTBR strategy itself obtains payoffs far above the average level in interactions with other strategies. These results demonstrate that MTBR is

not only highly competitive in repeated games but can also enhance the overall cooperation level of the population under non-evolutionary conditions.

4.3 Evolutionary Performance of MTBR in Well-Mixed Populations

The previous results demonstrate MTBR's payoff advantage under fixed strategy sets; this section further examines its long-term evolutionary capabilities in evolving mixed populations. We construct an evolutionary system where individuals initially randomly select either MTBR or any strategy from Strategy Set 2, with strategy propagation subsequently achieved through repeated games and strategy imitation processes. In each generation, all individuals in the population are randomly paired to play 20-round iterated Prisoner's Dilemma games. The game matrix is consistent with previous sections: $R = 3, S = 0, T = 5, P = 1$. All individuals obtain average payoffs after completing the games, after which an individual in the population may imitate another individual's strategy, with the imitation probability determined by the payoff difference between the two through the following formula:

$$p_{i \to j} = \frac{1}{1 + \exp[-\delta(\bar{U}_j - \bar{U}_i)]}$$

where δ represents the selection intensity, controlling the degree to which payoff differences affect imitation probability. When δ approaches 0, selection becomes random; when δ is large, strategy propagation depends more heavily on payoff differences. Figure 5 [Figure 5: see original paper] shows comparative results for two evolutionary scenarios. If the initial population does not include MTBR, the evolutionarily stable state of the population consists of GTFT0.3, GradualTFT, and ZDGTFT2, with a population average payoff of approximately 2.900 (blue lines in Figures 5a, 5c). After introducing MTBR (Figure 5b), its frequency rises rapidly and gradually replaces other strategies, eventually taking over the entire population, with average payoff increasing to 2.938 (red line in Figure 5c). Further analysis reveals that GradualTFT and MTBR obtain identical payoffs when interacting with each other. However, compared to GradualTFT, MTBR can achieve higher payoffs when interacting with itself, thus gradually replacing the former during evolution. Meanwhile, in mixed populations containing multiple exploitative strategies, MTBR demonstrates stable ability to resist exploitation and can continuously guide the population toward higher cooperation levels. These results not only verify MTBR's evolutionary advantages but also demonstrate that AI4Games possesses the capability to automatically generate interpretable strategies, reflecting its effectiveness and generality as a strategy search framework.

5 Conclusion and Outlook

This paper proposes and implements a unified framework for evolutionary game strategy discovery—AI4Games—which systematically transforms the

traditional process of strategy construction relying on human experience into reinforcement learning-driven strategy search tasks. The framework abstracts five generalizable steps: strategy representation, behavioral interaction, reward construction, strategy optimization, and strategy evaluation, forming a reusable and extensible strategy discovery pipeline applicable to diverse game modeling and evolutionary mechanism research tasks. To validate the framework's effectiveness, we apply it to the evolutionary iterated Prisoner's Dilemma environment and successfully discover a bilateral reciprocity strategy with a two-step memory structure (MTBR). This strategy is not prescriptively generated but spontaneously emerges from complex strategy spaces, possessing clear and interpretable behavioral rules and good evolutionary stability, demonstrating AI4Games' capability to automatically produce high-quality strategies without human intervention. This result not only verifies the functional effectiveness of AI4Games but also highlights its potential as an intelligent research tool.

From a broader perspective, the proposal of AI4Games represents a leap for artificial intelligence in game modeling and behavioral strategy research from "auxiliary tool" to "problem framework." It is not only a strategy discovery method but also a novel research paradigm connecting game theory, evolutionary dynamics, and intelligent decision-making, providing a generalized tool platform for complex systems modeling in the AI4S era. Its methodological thinking also aligns closely with the national "AI Plus" action plan's emphasis on cross-disciplinary collaborative innovation, high-dimensional problem modeling, and intelligent mechanism design. In summary, AI4Games not only provides solutions for specific game tasks but also drives a paradigm shift in strategy evolution research from rule design to intelligent modeling. It will serve as an important foundation for future cross-disciplinary research on "AI Plus Game Theory," providing a replicable and generalizable new pathway for promoting intelligent scientific modeling, understanding evolutionary social mechanisms, and designing complex behavioral systems.

References

[1] XU Y, LIU X, CAO X, et al. Artificial intelligence: A powerful paradigm for scientific research[J]. The Innovation, 2021, 2(4).

[2] WANG H, FU T, DU Y, et al. Scientific discovery in the age of artificial intelligence[J]. Nature, 2023, 620(7972): 47-60.

[3] LI G J. Review of milestone achievements in AI4S[J]. Computing, 2025, 1(4): 6-15.

[4] ZHANG K, YANG X, WANG Y, et al. Artificial intelligence in drug development[J]. Nature Medicine, 2025, 31(1): 45-59.

[5] JIAO L, SONG X, YOU C, et al. AI meets physics: a comprehensive survey[J]. Artificial Intelligence Review, 2024, 57(9): 256.

[6] NOWAK M A. Five Rules for the Evolution of Cooperation[J]. Science, 2006, 314(5805).

[7] WANG L, FU F, CHEN X J, et al. Evolutionary games and self-organizing cooperation[J]. Systems Science and Mathematics, 2007(03): 330-343.

[8] WANG L, FU F, CHEN X J, et al. Evolutionary games on complex networks[J]. Journal of Intelligent Systems, 2007(02): 1-10.

[9] HILBE C, CHATTERJEE K, NOWAK M A. Partners and rivals in direct reciprocity[J]. Nature Human Behaviour, 2018, 2(7): 469-477.

[10] AXELROD R, HAMILTON W D. The Evolution of Cooperation[J]. Science, 1981, 211(4489): 1390-1396.

[11] NOWAK M, SIGMUND K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game[J]. Nature, 1993, 364(6432): 56-58.

[12] PRESS W H, DYSON F J. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent[J]. Proceedings of the National Academy of Sciences, 2012, 109(26): 10409-10412.

[13] GRÜNE-YANOFF T. Evolutionary game theory, interpersonal comparisons and natural selection: a dilemma[J]. Biology & Philosophy, 2011, 26(5): 637-654.

[14] ADAMI C, SCHOSSAU J, HINTZE A. Evolutionary game theory using agent-based methods[J]. Physics of Life Reviews, 2016, 19: 1-26.

[15] WANG L, WU T, ZHANG Y L. Feedback mechanisms in co-evolutionary games[J]. Control Theory and Applications, 2014(07): 823-836.

[16] WANG L, CONG R, LI K. Feedback mechanisms in the evolution of cooperation[J]. Science China: Information Sciences, 2014(12): 1495-1514.

[17] WANG L, FU F, CHEN X J, et al. Group decision-making on complex networks[J]. Journal of Intelligent Systems, 2008(02): 95-108.

[18] WANG L, WU B, DU J M, et al. Analysis of propagation behavior on complex networks[J]. Science China: Information Sciences, 2020, 50(11).

[19] FU F, HAUERT C, NOWAK M A, et al. Reputation-based partner choice promotes cooperation in social networks[J]. Physical Review E, 2008, 78: 026117.

[20] SU Q, MCAVOY A, WANG L, et al. Evolutionary dynamics with game transitions[J]. Proceedings of the National Academy of Sciences, 2019, 116(51): 25398-25404.

[21] SHENG A, SU Q, WANG L, et al. Strategy evolution on higher-order networks[J]. Nature Computational Science, 2024, 4(4): 274-284.

[22] LITTMAN M L. Markov games as a framework for multi-agent reinforcement learning[C]// Machine Learning Proceedings 1994. San Francisco, CA: Morgan Kaufmann, 1994: 157-163.

[23] HU J, WELLMAN M P. Multiagent reinforcement learning: theoretical framework and an algorithm[C]//Proceedings of the International Conference on Machine Learning: vol. 98. 1998: 242-250.

[24] CHALKIADAKIS G, BOUTILIER C. Coordination in multiagent reinforcement learning: a Bayesian approach[C]//Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '03). New York, NY, USA: Association for Computing Machinery, 2003: 709-716.

[25] MATIGNON L, LAURENT G J, FORT-PIAT N L. Independent reinforcement learners in cooperative Markov games: A survey regarding coordination problems[J]. The Knowledge Engineering Review, 2012, 27(1): 1-31.

[26] NOWÉ A, VRANCX P, HAUWERE Y M D. Game Theory and Multiagent Reinforcement Learning[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 441-470.

[27] HARPER M, KNIGHT V, JONES M, et al. Reinforcement Learning Produces Dominant Strategies for the Iterated Prisoner's Dilemma[J]. PLOS ONE, 2017, 12(12): e0188046.

Author Contributions: Hongyu Wang and Long Wang: conceived the research idea, designed the research plan, conducted experiments, drafted the manuscript, and revised the manuscript.

Note: Figure translations are in progress. See original paper for figures.