参赛学生姓名：苏骏岳

中学：北京十一学校

省份：北京

国家/地区：中国

指导老师姓名：王艺臻

指导老师单位：北京十一学校

论文题目：Cooperation Mechanisms in the Iterated Prisoner's Dilemma under a Finite-Memory Satisfaction-Switching Rule

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Cooperation Mechanisms in the Iterated Prisoner's Dilemma under a Finite-Memory Satisfaction-Switching Rule

Junyue Su

Beijing National Day School, Beijing, China

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

Abstract

This paper studies how cooperation emerges and persists in the iterated Prisoner's Dilemma when players are boundedly rational, have finite memory, and make decision errors. Existing work has largely treated memory weighting, aspiration levels, and decision errors as separate objects of study; how these three elements interact within a single, tractable framework remains underexplored.

We model the interaction as a satisfaction-switching rule in which each player weights the most recent two payoffs and switches action when dissatisfied. With two-period memory, the model is represented by an exact 16-state Markov chain, allowing both long-run stationary analysis and finite-round transient analysis. Starting from this rule, we characterize the parameter regions in which this rule reduces to win-stay, lose-shift (WSLS), produces low-cooperation trapping, or generates high-frequency oscillation. Under the standard Prisoner's Dilemma payoff matrix used in this paper and with $A = 2$, the WSLS-equivalent region is $\alpha > 3/4$, which achieves a long-run cooperation rate of 0.8645 at $\varepsilon = 0.05$ in this model. This region remains stable across tested decision-error levels $\varepsilon \in [0, 0.3]$, although the cooperation rate declines as ε increases. We further use the cooperation rate, the $CCCC$ probability, and the actual switching rate to distinguish stable reciprocity from high-frequency oscillation.

Monte Carlo simulations are used to validate the analytical results, and extensions to longer memory and adaptive aspiration test the robustness of the mechanism. As an exploratory application, we implement a related mechanism in a large language model agent with external memory and lightweight hypergraph retrieval, improving memory use, state consistency, conflict handling, and multi-round task performance in the tested scenarios. Overall, the results suggest that cooperation should be evaluated not only by long-run averages but also by stability, switching behavior, and finite-time dynamics.

Keywords

Iterated Prisoner's Dilemma, Finite Memory, Satisfaction-Switching Rule, Cooperation Mechanism, Markov Chain, Monte Carlo Simulation, Large Language Model Agent

CONTENTS

I Introduction 4

II Literature Review 7

III Model Setup 10

IV Long-Run Stationary Analysis 15

V Finite-Round Dynamics and Error Analysis 20

VI Simulation and Extensions 22

VII LLM-Based Agent Experiment 26

VIII Discussion and Conclusion 29

Appendix A: Supplementary Materials 33

References 55

Acknowledgements 57

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

I. Introduction

The Prisoner's Dilemma provides the standard framework for studying the tension between individual incentives and collective welfare. In the one-shot game, defection is the strictly dominant strategy for both players, leading to the unique Nash equilibrium (D, D), even though mutual cooperation would yield a strictly higher payoff for both. When the game is repeated over a finite number of rounds, the classical backward-induction argument predicts that cooperation unravels from the final round: since defection is optimal in the last round, the same logic applies to the second-to-last round, and so on, until the unique subgame-perfect equilibrium is defection in every round. This result holds regardless of how long the game is, as long as the horizon is finite and commonly known.

This prediction, however, is not consistent with the widespread cooperation observed in laboratory experiments and real-world settings. A growing literature has therefore relaxed the assumption of full rationality, allowing players to follow simple behavioral rules based on limited memory and satisfaction. In this paper, we study one such rule: a finite-memory satisfaction-switching rule, in which players weight their recent payoffs and switch actions when dissatisfied. This rule requires neither full history nor backward induction, and it provides a tractable framework for understanding how cooperation can emerge under bounded rationality.

The study of cooperation in repeated games has a long history. Repeated-game theory explains how future interaction can support cooperation despite short-run incentives to defect [1], [2]. Relational-contract models connect this logic to continuing economic relationships [17], [18], while experimental evidence shows that equilibrium feasibility alone does not ensure cooperation [16]. A complementary behavioral approach builds on bounded rationality and satisficing [5]: players retain an action after satisfactory outcomes and revise it after disappointment. Win-stay, lose-shift and aspiration-based learning provide important precedents for this mechanism [3], [6]–[8].

Research on direct reciprocity further shows that the consequences of memory depend on both its length and the strategy class [12]–[14]. Building on these strands, we examine a particular two-period satisfaction-switching rule in which the recency weight α, aspiration threshold A, and execution-error probability ε jointly determine behavior. Our focus is on how this interaction shapes transient adjustment and long-run outcomes, and on distinguishing sustained mutual cooperation from alternation that produces a similar average cooperation rate.

2026 S.T. Yau High School Science Award
仅用于2026区域科学论文公示

An application extension connects the behavioral rule to language-model agents with external memory. Memory-based agents and Graph-R1 motivate the use of stored experience and structured, multi-turn retrieval [25], [30]. We combine these ideas with an external satisfaction state machine and supervised tool-use training. This experiment examines the implementation of the mechanism in language-based tasks; it does not establish an economic equilibrium or human behavioral validity.

Thus, this paper addresses the following research questions:

- **RQ1:** How do the two-period memory weight α and the aspiration level A jointly affect long-run cooperation, stable reciprocity, and behavioral switching?
- **RQ2:** For a fixed aspiration level $A = 2$, which memory-weight regions generate high cooperation under different error levels ε, and how are these regions related to WSLS-equivalent behavior?
- **RQ3:** Can the analytical results from the exact Markov chain be validated by Monte Carlo simulations, and can they be extended to memory lengths $k > 2$?
- **RQ4:** When the aspiration level adapts to realized payoffs, does cooperation improve relative to the fixed-threshold benchmark?

We model the interaction as a finite-memory satisfaction-switching rule in the iterated Prisoner's Dilemma. With two-period memory, the system is represented by an exact 16-state Markov chain, from which we derive both the long-run stationary distribution and the finite-round transient dynamics. We then validate the analytical results through Monte Carlo simulations and extend the analysis to longer memory lengths ($k = 2, \ldots, 8$) and a bidirectional adaptive aspiration rule. Finally, we implement a related mechanism in a large language model agent as an exploratory engineering application inspired by the theoretical rule.

Our main findings are as follows. Under the numerical payoff matrix used in this paper and for a fixed aspiration level $A = 2$, the high-cooperation region $\alpha > 3/4$ is the $A = 2$ slice of the WSLS-equivalent wedge, and it achieves a long-run cooperation rate of 0.8645 at $\varepsilon = 0.05$. This region remains stable across tested decision-error levels. Across the parameter space, the same finite-memory rule can also produce low-cooperation trapping or high-frequency oscillation. We compare these regions using both long-run stationary behavior and finite-time dynamics, showing that a high average cooperation rate does not necessarily imply stable reciprocity. The Monte Carlo validation confirms the analytical results (87 of 87 absolute errors below 0.015), the longer-memory extension preserves the high-recent-weight mechanism, and the best observed

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

adaptive-aspiration mean on the scanned grid does not exceed the fixed-threshold benchmark.

The main contributions of this paper are summarized below.

- **Mechanism identification.** Starting from a finite-memory satisfaction-switching rule, we characterize when the rule reduces to WSLS, when it produces low-cooperation trapping, and when it generates high-frequency oscillation. We compare these regions in both long-run stationary behavior and finite-time dynamics, and show that high average cooperation can arise from stable reciprocity or frequent behavioral switching. We introduce the joint use of the long-run cooperation rate, the $CCCC$ probability, and the actual switching rate to distinguish these two mechanisms. Within the WSLS-equivalent wedge, the older payoff does not affect decisions, so the high-cooperation plateau is not evidence of an advantage over one-period WSLS.
- **Methodological framework.** We provide a unified analytical and computational framework that separates the finite-round transient dynamics from the long-run stationary analysis, distinguishes the zero-error Cesàro average from the low-positive-error stationary distribution, and validates the exact Markov chain results against independent Monte Carlo simulations.
- **Exploratory application.** We implement a related mechanism in a large language model agent and evaluate it in multi-round real-file tasks. External memory and hypergraph retrieval improve measured task performance in the tested scenarios. Behavior-data training raises macro-average category accuracy and measured satisfaction, while overall round-level accuracy is unchanged. These results do not establish an independent stability benefit of the state machine or validate the game's cooperation mechanism.

The rest of the paper is organized as follows. Section II reviews the related literature on cooperation in repeated games, aspiration-based learning, finite-memory strategies and applications on large language models. Section III presents the model setup, including the payoff structure, the state space, the satisfaction-switching rule, and the performance measures. Section IV analyzes the long-run stationary behavior of the 16-state Markov chain. Section V examines the finite-round dynamics and the effect of decision errors. Section VI reports the Monte Carlo validation, the longer-memory extension, and the adaptive aspiration analysis. Section VII presents the LLM-based agent experiment. Section VIII concludes and discusses the limitations and future work.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

II. Literature Review

This study connects repeated-game cooperation with aspiration-based learning and finite-memory decision rules. Its economic interpretation concerns how agents keep repeated relationships by using experience to decide if outcomes are satisfactory. The language-model extension draws additionally on memory, structured retrieval and tool-use training. These connections motivate the model and show which parts of the application come from the game-theoretic mechanism.

A. Repeated Cooperation and Economic Relationships

Axelrod [1] shows that reciprocal strategies are a way to get cooperation in the iterated Prisoner's Dilemma. Fudenberg and Maskin [2] describe under which conditions patient players can keep individually rational payoffs in repeated games. These findings explain when cooperation can happen with the incentives while leaving room for a behavioral account of how players get to and keep a cooperative pattern.

The distinction between feasibility and behavior matters empirically. Dal Bó and Fréchette [16] find that cooperation in repeated games needs conditions that are stricter than equilibrium feasibility alone. This motivates studying adjustment rules alongside run outcomes. The simulations in this paper check a mathematical mechanism; they do not give the human-subject evidence that experimental economics gives.

Relational contracts provide an economic setting for repeated cooperation. Baker, Gibbons, and Murphy [17] examine how firm boundaries influence the stability of agreements, and Levin [18] studies how incentives work in sustaining relationships. These studies help us see the model as a stylized account of cooperation between recurring trading partners. The model does not show contract design, bargaining or market institutions directly. Its economic point is to examine an experience-based adjustment mechanism that could operate within these relationships.

B. Satisficing, Aspiration Levels, and Learning

Simon [5] provides the bounded-rationality foundation for decisions that are built around outcomes. In this model, A is the payoff standard used to judge experience and α shows how much weight is put on the latest payoff. The parameter ε stands for an execution error compared to the intended action. These are interpretations of model parameters rather than estimates of individual preferences or biases.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Nowak and Sigmund [3] demonstrate the effectiveness of win-stay, lose-shift in their evolutionary comparison. This rule is a direct precedent for retaining satisfactory actions and changing unsatisfactory ones. Our specification extends the evaluation window to two weighted payoffs, so satisfaction depends on recent payoff history rather than only the immediately preceding outcome.

Karandikar et al. [6] study evolving aspirations and cooperation among satisficing players. Bendor et al. [7] review aspiration-based learning with different sources of aspiration formation, and Posch et al. [8] analyze the efficiency of adapting aspiration levels. Together, these studies motivate treating the satisfaction threshold as a substantive behavioral component. They also motivate comparing a fixed threshold with an adaptive rule without assuming that adaptation must improve performance.

Learning models offer further points of comparison. Macy and Flache [9] examine learning dynamics in social dilemmas, while Camerer and Ho [10] develop experience-weighted attraction learning. Our rule checks a weighted payoff history against a threshold; it does not keep the action-attraction system of that model. Its simplicity lets us do a state-space analysis of how recall weights, aspirations and errors interact.

C. Memory Structure and Decision Errors

Memory constraints can be represented in different ways. Monte [11] studies a player with finitely many memory states in zero-sum repeated games with one-sided incomplete information. That setting illustrates constraints on information processing, but does not establish cooperation results for the Prisoner's Dilemma considered here. Our memory instead stores a fixed window of realized joint actions.

Within direct reciprocity, Baek et al. [12] compare reactive and memory-one strategies, showing why the information retained by a strategy matters. Press and Dyson [4] identify strategies that enforce linear payoff relations, illustrating the analytical power of finite-memory representations. Their payoff-control mechanism is distinct from the satisfaction-based switching studied here.

Longer memory does not have a uniform effect across all strategy spaces. Glynatsi et al. [13] study conditional cooperation with longer memory, including reactive and counting-based strategies. Lesigang et al. [14] establish reduced-memory best responses under specified conditions in repeated additive games. Their result does not imply that additional memory is generally irrelevant. These works motivate checking both how long memory is and the rule that summarizes

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

remembered experience. Our longer-memory extension therefore looks at a family of weighted satisfaction rules not at all possible strategies of that memory length.

Errors also affect the interpretation of long-run behavior. Young [15] studies convention selection under small perturbations, providing a related perspective on the distinction between unperturbed dynamics and small-noise limits. In our model, the treatment of $\varepsilon = 0$ follows from the transition matrix itself: multiple recurrent classes or periodic behavior can make initialization and averaging relevant. We consequently distinguish the zero-error Cesàro average from the stationary distribution at positive error, without assuming that the two limits coincide.

D. External Memory and Language-Model Agents

The application uses cognitive concepts as design motivations. Tulving's distinction between episodic and semantic memory [20] motivates separating interaction records from condensed summaries. Ebbinghaus [21] supplies a historical basis for studying retention and forgetting, while Collins and Loftus [19] motivate retrieval through spreading activation. These connections do not empirically determine the application's capacity, decay rate, or retrieval depth; those remain engineering choices evaluated in the implementation.

Retrieval-augmented generation connects language models to external information [22]. GraphRAG uses graph structure and community summaries for query-focused summarization [23]. For interactive agents, ReAct interleaves reasoning with actions [24], Generative Agents organize stored experiences and reflections [25], and Reflexion uses linguistic feedback retained in memory to improve subsequent attempts [26]. These approaches motivate external state and feedback-sensitive behavior, while differing in their tasks and adaptation mechanisms.

Graph-R1, presented at ICML 2026 [30], is particularly relevant to the retrieval component. It combines lightweight knowledge hypergraphs with multi-turn retrieval and end-to-end reinforcement learning. Our implementation draws on the structured-retrieval idea, but maintains its hypergraph and retrieval procedure as deterministic external services. Tool-use decisions are trained on rule-based oracle examples through supervised fine-tuning, rather than through Graph-R1's reinforcement-learning procedure.

The training implementation uses low-rank adaptation [27], quantized fine-tuning [28], and LlamaFactory [29]. These are supporting methods, not new algorithms proposed by this paper. The application evaluates whether training improves the use of an otherwise fixed external mechanism. The satisfaction state machine is inspired by the game model, whereas the hypergraph is a separate

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

retrieval component. Success on language-based tasks does not by itself validate cooperation predictions for human economic agents.

E. Positioning of the Present Study

The preceding literature motivates a focused question: within a specified satisfaction-switching rule, how do recency weighting, aspiration thresholds, and execution errors jointly shape cooperation? We address this question through an exact 16-state representation of two-period memory, combining transient analysis with long-run calculations. The contribution concerns this mechanism and its parameter-dependent behavior, rather than a claim that finite memory, aspiration learning, or noisy cooperation is itself new.

The analysis reports cooperation frequency alongside the probability of two consecutive mutual-cooperation outcomes and the actual switching rate. These measures distinguish persistence from alternation and help interpret whether an apparently cooperative relationship is stable. They do not constitute a complete welfare analysis: payoff comparisons also require the distribution of joint outcomes and the stage-game payoff structure. Monte Carlo checks, longer-memory simulations, and adaptive-aspiration experiments examine the internal consistency and scope of the proposed mechanism. The language-model experiment then explores its implementation as an external behavioral controller with memory and retrieval support.

III. Model Setup

A. Basic Assumptions

The model is built on the following assumptions; the payoff matrix, memory index, switching rule, and error process are defined in the following subsections.

1) **Repeated Prisoner's Dilemma Structure.** Two players engage in a finitely repeated Prisoner's Dilemma with simultaneous actions in each round.
2) **Finite memory.** Each player uses only the most recent $k = 2$ rounds of actions and payoffs.
3) **No backward induction.** Players do *not* condition their decisions on the remaining number of rounds. The decision rule is fixed, time-invariant, and independent of the remaining horizon. This rules out finite-horizon backward induction by assumption, not by cognitive limitation.
4) **Weighted memory mechanism.** Each player aggregates the two remembered payoffs into a scalar memory index.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

5) **Satisfaction-switching decision rule.** Each player compares the memory index with an aspiration level and then either repeats or switches the previous action.
6) **Decision error.** Intended actions may be reversed by independent execution errors.
7) **Independent decision-making.** Players cannot communicate, form binding agreements, or condition decisions on the opponent's private intention.
8) **Analytical approach.** The $k = 2$ case is analyzed exactly using a 16-state Markov chain. Longer memory lengths and extensions are examined via Monte Carlo simulations.

B. The Prisoner's Dilemma

Consider a two-player simultaneous-move game in which two players choose between the two actions: cooperate (C) or defect (D). The payoff structure satisfies the standard Prisoner's dilemma conditions:

$$T > R > P > S, \quad 2R > T + S \tag{1}$$

where T is the temptation payoff (defecting while the other cooperates), R is the reward for mutual cooperation, P is the punishment for mutual defection, and S is the sucker's payoff (cooperating while the other defects).

TABLE I: General Payoff Matrix of the Prisoner's Dilemma

	Player 2	
	Cooperate (C)	Defect (D)
Player 1		
Cooperate (C)	(R, R)	(S, T)
Defect (D)	(T, S)	(P, P)

Following the standard parametrization in the literature, we set $T = 5$, $R = 3$, $P = 1$, $S = 0$, which satisfy the conditions in Eq. (1). All numerical thresholds reported below, including the $\alpha > 3/4$ cutoff at $A = 2$, are conditional on this payoff matrix and on the stated error specification. The numerical payoff matrix is:

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

TABLE II: Numerical Payoff Matrix

	Player 2	
	Cooperate (C)	Defect (D)
Player 1		
Cooperate (C)	$(3, 3)$	$(0, 5)$
Defect (D)	$(5, 0)$	$(1, 1)$

In the one-shot game, defection is a strictly dominant strategy for both players, yielding the unique Nash equilibrium (D, D) with payoffs $(1, 1)$. However, in repeated interactions, cooperation can be sustained through conditional strategies and reciprocity [1], [2]. This paper investigates how finite memory and satisfaction-based adaptation affect the emergence and stability of cooperation.

C. Players and Strategies

Two bounded-rational players engage in a finite repeated Prisoner's Dilemma over t rounds, where t denotes the fixed horizon. Let $r = 1, \ldots, t$ denote the current game round. The outcomes of rounds 1 and 2 provide the initial two-period history, while the satisfaction-switching rule governs decisions in rounds $r = 3, \ldots, t$. In each round, players choose their action $a_i(r) \in \{C, D\}$. Their payoff $u_i(r)$ is determined by the action pair according to Table II.

D. State Space

In each round, the joint action is:

$$z_r = (a_1(r), a_2(r)) \in \ CC, CD, DC, DD \tag{2}$$

With two-period memory, the state before round r is:

$$X_r = (z_{r-1}, z_{r-2}) \tag{3}$$

which yields $4 \times 4 = 16$ states. The next state is:

$$X_{r+1} = (z_r, z_{r-1}) \tag{4}$$

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

E. Finite Memory and Weighted Recall

The two players have limited memory and can only recall the most recent $k = 2$ rounds of history. Memory is encoded through a weighted average of past payoffs:

$$\begin{aligned} m_i(X_r) \coloneqq m_i(r) &= \alpha u_i(r-1) + (1-\alpha)u_i(r-2), \\ r &= 3, \ldots, t, \qquad 0 \le \alpha \le 1, \end{aligned} \tag{5}$$

where α is the weight assigned to the most recent payoff, $1-\alpha$ is the weight assigned to the earlier payoff.

F. Satisfaction-Switching Decision Rule

Player i compares the weighted memory $m_i(X_r)$ with an aspiration level $A \in [S, T]$, while we set the benchmark $A = 2$ in the intermediate range $P < A < R$. The decision rule is:

$$a_i^*(r) = \begin{cases} a_i(r-1), & \text{if } m_i(X_r) \ge A \quad \text{(satisfied)}, \\ \text{opposite of } a_i(r-1), & \text{if } m_i(X_r) < A \quad \text{(unsatisfied)}. \end{cases} \tag{6}$$

Let $b_i(X_r) \in \{0, 1\}$ indicate whether player i plans to cooperate in round r, where $b_i(X_r) = 1$ if $a_i^*(r) = C$ and 0 otherwise. Equivalently:

$$b_i(X_r) = 1 \quad \text{if and only if} \quad \begin{cases} a_i(r-1) = C \text{ and } m_i(X_r) \ge A, \\ a_i(r-1) = D \text{ and } m_i(X_r) < A. \end{cases} \tag{7}$$

G. Decision Error

Following the common usage in the literature, we set the probability of decision error as $\varepsilon \in (0, 1/2)$, where each player fails to execute the intended action with this probability. The actual probability of cooperating is:

$$p_i(X_r) = \varepsilon + (1 - 2\varepsilon) b_i(X_r) \tag{8}$$

Thus, if the player intends to cooperate ($b_i = 1$), the actual probability of cooperation is $1 - \varepsilon$; if the player intends to defect ($b_i = 0$), the actual probability of cooperation is ε.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

H. Transition Probabilities

Depending on the current state X_r, the probabilities of the four possible next-period action pairs are the following:

$$P(CC \mid X_r) = p_1 p_2, \tag{9}$$

$$P(CD \mid X_r) = p_1(1 - p_2), \tag{10}$$

$$P(DC \mid X_r) = (1 - p_1)p_2, \tag{11}$$

$$P(DD \mid X_r) = (1 - p_1)(1 - p_2). \tag{12}$$

These probabilities define a row-stochastic 16×16 transition matrix $P(\alpha, A, \varepsilon)$.

I. Performance Measures

Let $\mu(n) = \mu_0 P^n$ denote the distribution over states after n transitions, where n denotes the number of transition steps after the complete two-period memory state has been initialized. Equivalently, for the original game-round index, $n = r - 2$. Here μ_0 is the initial distribution over the 16 two-period histories (assumed uniform: $\mu_0(i) = 1/16$).

Define $c(X)$ as the fraction of cooperative actions in the most recent period of state X:

$$c(X) = \begin{cases} 1, & z_{r-1} = CC, \\ 1/2, & z_{r-1} = CD \text{ or } DC, \\ 0, & z_{r-1} = DD. \end{cases} \tag{13}$$

The point-wise cooperation rate after the n transitions is:

$$\rho(n) = \mu(n) \cdot c \tag{14}$$

The cumulative average cooperation rate after the n transitions is:

$$\bar{\rho}(n) = \frac{1}{n} \sum_{j=1}^{n} \rho(j) \tag{15}$$

For the positive-error system, the long-run cooperation rate is obtained from the stationary distribution:

$$\rho_\infty = \pi \cdot c, \quad \pi P = \pi. \tag{16}$$

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

The zero-error case and the low-positive-error limit are discussed separately in Section V.

To distinguish stable reciprocity from mere high average cooperation, we additionally track:

- $\pi(CCCC)$: the stationary probability of the state where both players cooperated in both of the most recent two periods (a proxy for sustained reciprocity);
- σ_{actual}: the actual switching rate, i.e., the fraction of players who change actions relative to the previous period, accounting for decision errors.

A complete notation table is provided in Appendix A-A.

IV. Long-Run Stationary Analysis

Using the 16-state Markov chain defined in Section III-D, we now analyze the long-run behavior of the positive-error system through the stationary distribution π satisfying $\pi P = \pi$. The special case $\varepsilon = 0$ and its relation to the low-positive-error limit are treated in Section V. We then examine how the long-run cooperation rate varies with the memory weight α and the aspiration level A.

A. Stationary Distribution

For each $\varepsilon > 0$, the 16-state chain is irreducible and aperiodic, so it has a unique stationary distribution π satisfying $\pi P = \pi$. Full transition-matrix and state-probability diagnostics for the baseline parameter setting are reported in the appendix.

B. Parameter Scan: α at Fixed $A = 2$

We first fix the aspiration level at $A = 2$ and scan the memory weight α under $\varepsilon = 0.05$. This one-dimensional scan provides the clearest view of the stepped platform structure before we examine the full (α, A) space.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

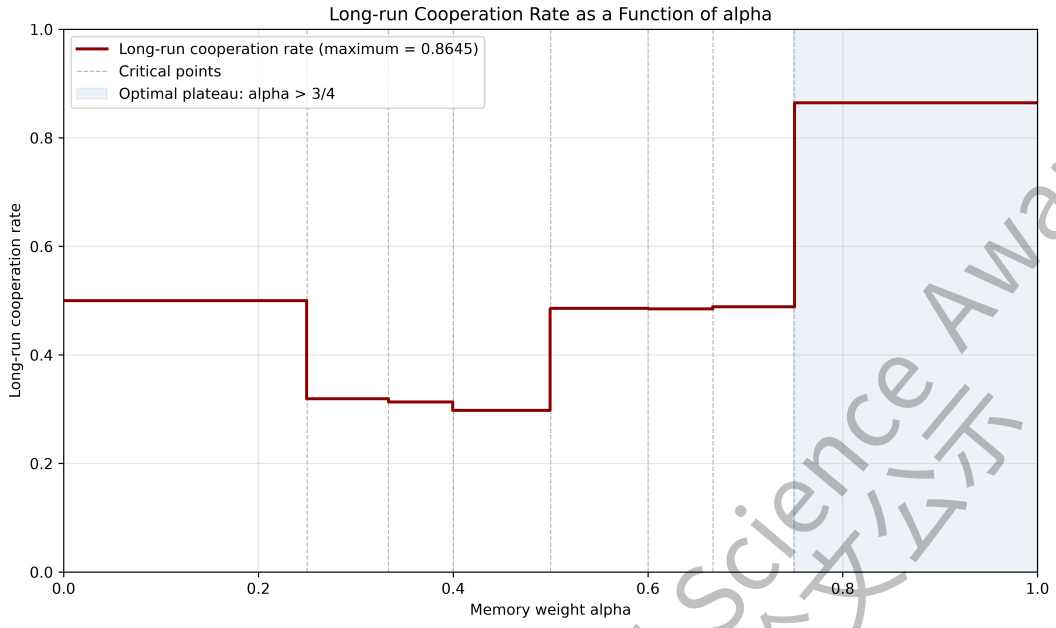


Fig. 1: Cooperation rate under $A = 2$ and $\varepsilon = 0.05$.

Figure 1 has a stepped shape, with critical values $\frac{1}{4}$, $\frac{1}{3}$, $\frac{2}{5}$, $\frac{1}{2}$, $\frac{3}{5}$, $\frac{2}{3}$, and $\frac{3}{4}$. Because the aspiration level A is fixed, crossing a critical value changes whether the weighted memory $m_i(X_r)$ is above or below A, and therefore changes the induced switching pattern. Under the current payoff matrix, $A = 2$, and $\varepsilon = 0.05$, the high-cooperation platform begins at $\alpha > 3/4$ and reaches a long-run cooperation rate of 0.8645.

C. Parameter Scan: (α, A) Phase Diagram

The fixed-A scan shows the high-cooperation platform at $A = 2$, but the aspiration level A itself is a key behavioral variable that determines how difficult it is for players to remain satisfied with the current action. Thus, a joint scan over α and A is needed to show how the $A = 2$ platform fits into the broader parameter space, and to test whether high average cooperation always corresponds to stable reciprocity rather than frequent behavioral switching.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

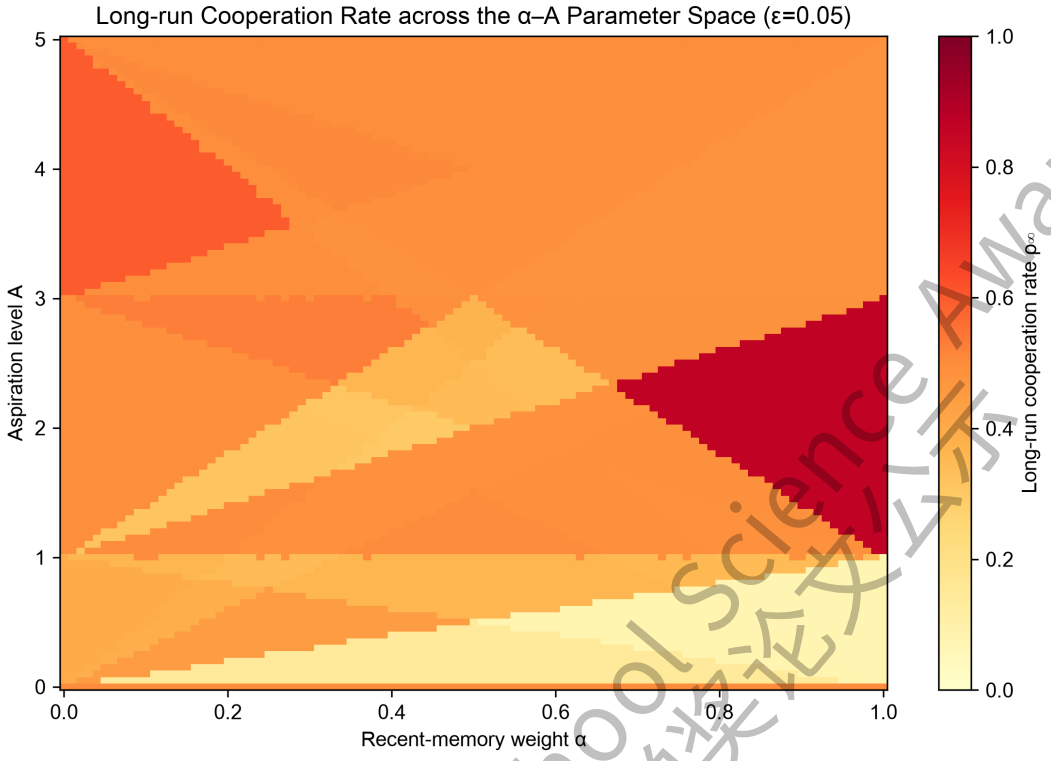


Fig. 2: Cooperation rate scan over α and A with $\varepsilon = 0.05$

Figure 2 reports the long-run cooperation rate ρ_∞ over α and A. The high-cooperation plateau includes the wedge $5 - 4\alpha < A \leq 3\alpha$. To obtain this boundary, note that the largest weighted payoff with a most recent losing payoff ($S = 0$ or $P = 1$) is $\alpha P + (1-\alpha)T = 5 - 4\alpha$, whereas the smallest with a most recent winning payoff ($R = 3$ or $T = 5$) is $\alpha R + (1-\alpha)S = 3\alpha$. Within this wedge, every losing outcome induces switching and every winning outcome induces staying, independently of the older payoff. The decision rule is therefore exactly the classic win-stay, lose-shift (WSLS) rule [3], with the same decision errors. The wedge is nonempty for $\alpha > 5/7$; under the current payoff matrix, its $A = 2$ slice gives $\alpha > 3/4$. Thus, the optimal plateau does not demonstrate an advantage of two-period memory over one-period WSLS. High-A settings such as $(\alpha, A) = (0.5, 4.5)$ instead produce cooperation near 0.5 through frequent switching. To distinguish these patterns, we additionally examine the stationary probability of $CCCC$ (Figure 3) and the actual switching rate (Figure 4).

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

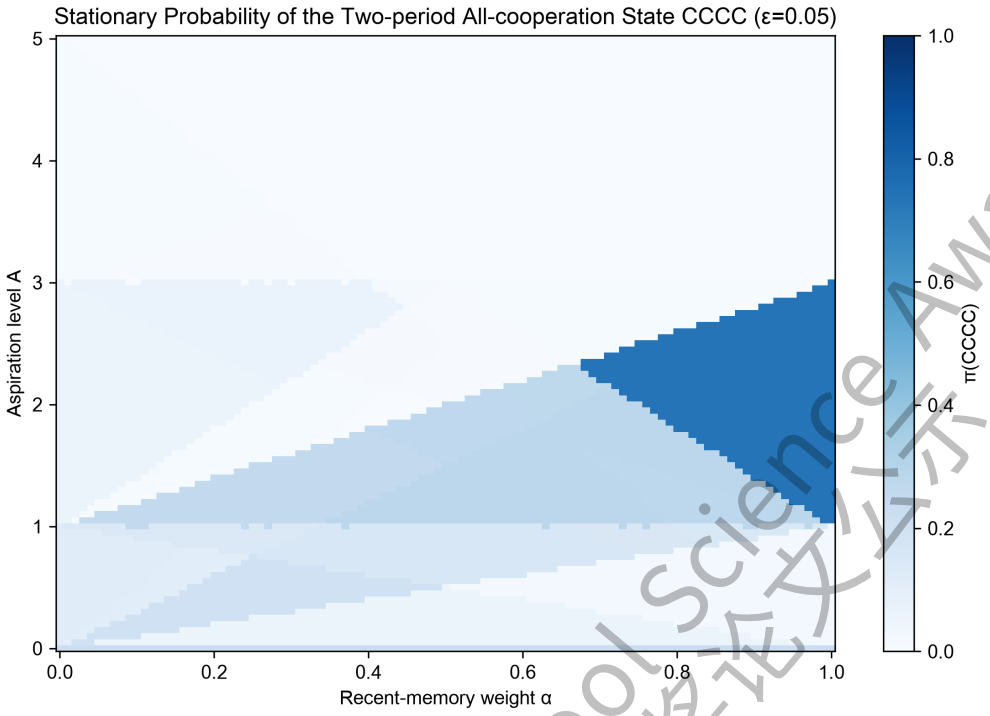


Fig. 3: Stationary probability of $CCCC$ scan over α and A with $\varepsilon = 0.05$

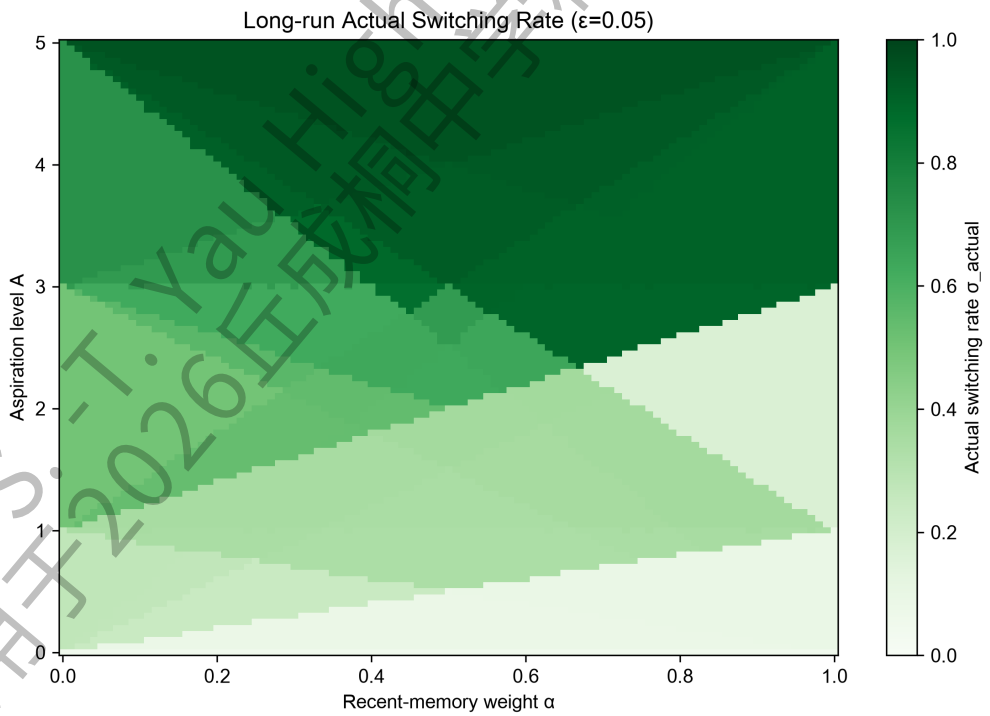


Fig. 4: Actual switching rate scan over α and A with $\varepsilon = 0.05$

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Figures 3 and 4 illustrate this contrast. At the representative high-A point $(\alpha, A) = (0.5, 4.5)$, the $CCCC$ probability is low and the switching rate is 0.9494. At $(0.8, 2)$, inside the WSLS-equivalent wedge, the $CCCC$ probability is high and the switching rate is 0.1720. These numerical values describe the selected points rather than entire rectangular regions.

These contrast patterns reveal a key distinction: high average cooperation can arise from two qualitatively different mechanisms. The first mechanism, observed in the WSLS-equivalent wedge, corresponds to stable reciprocity, where players remain in mutual cooperation ($CCCC$) for most of the time with little behavioral switching. The second mechanism, illustrated by the high-A representative point, generates high cooperation only as a time average, driven by frequent alternation between cooperation and defection rather than by sustained mutual cooperation.

Therefore, interpreting cooperation solely through the long-run average rate ρ_∞ can be misleading. A complete evaluation must jointly consider the cooperation rate, the stationary probability of $CCCC$, and the actual switching rate.

D. Four Representative Regions

To summarize the information in Figure 2, we choose four representative points.

Table III summarizes their long-run cooperation rates, $CCCC$ probabilities, actual switching rates, and the underlying behavioral mechanisms.

TABLE III: Four representative points ($\varepsilon = 0.05$).

Region	(α, A)	ρ_∞	$\pi(CCCC)$	σ_{actual}	**Mechanism**
Low cooperation	$(0.3, 2)$	0.3192	0.0110	0.5331	Low cooperation, moderate switching
Stable cooperation	$(0.8, 2)$	0.8645	0.7373	0.1720	$CCCC$-dominant stable reciprocity
High-A fluctuation	$(0.5, 4.5)$	0.4997	0.0006	0.9494	Frequent C/D alternation
Low-A lenient	$(0.8, 0.5)$	0.0786	0.0255	0.0921	Stuck in low-cooperation states

The four points reveal distinct behavioral mechanisms. The stable cooperation region ($\alpha = 0.8, A = 2$) achieves the highest cooperation rate with the highest $CCCC$ probability among the four points and a low switching rate. The high-A fluctuation region ($\alpha = 0.5, A = 4.5$), despite a moderate cooperation rate of approximately 0.5, exhibits nearly zero $CCCC$ probability and a switching rate close to one, confirming that its average cooperation is driven by frequent alternation. The low cooperation region ($\alpha = 0.3, A = 2$) and the low-A lenient region ($\alpha = 0.8, A = 0.5$)

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

both yield low cooperation, but through different channels: the former through moderate switching, and the latter through inertia that traps players in defective states.

V. Finite-Round Dynamics and Error Analysis

While the stationary analysis in Section IV characterizes the asymptotic behavior of the system and serves as a base model, real-world interactions are often finite in duration.

This section examines the transient dynamics of the same 16-state Markov chain over finite horizons, with the state space, payoff structure, and rules unchanged from Section IV. We focus on the representative early-stage trajectory and the distinction between zero-error and low-positive-error behavior. Detailed point-wise, cumulative, speed, and convergence diagnostics are reported in the appendix.

A. *Early Transition (from $n = 1$ to $n = 5$)*

The first few transitions exhibit the most dramatic changes in cooperation rates, as the system moves away from the uniform initial distribution toward the long-run behavior.

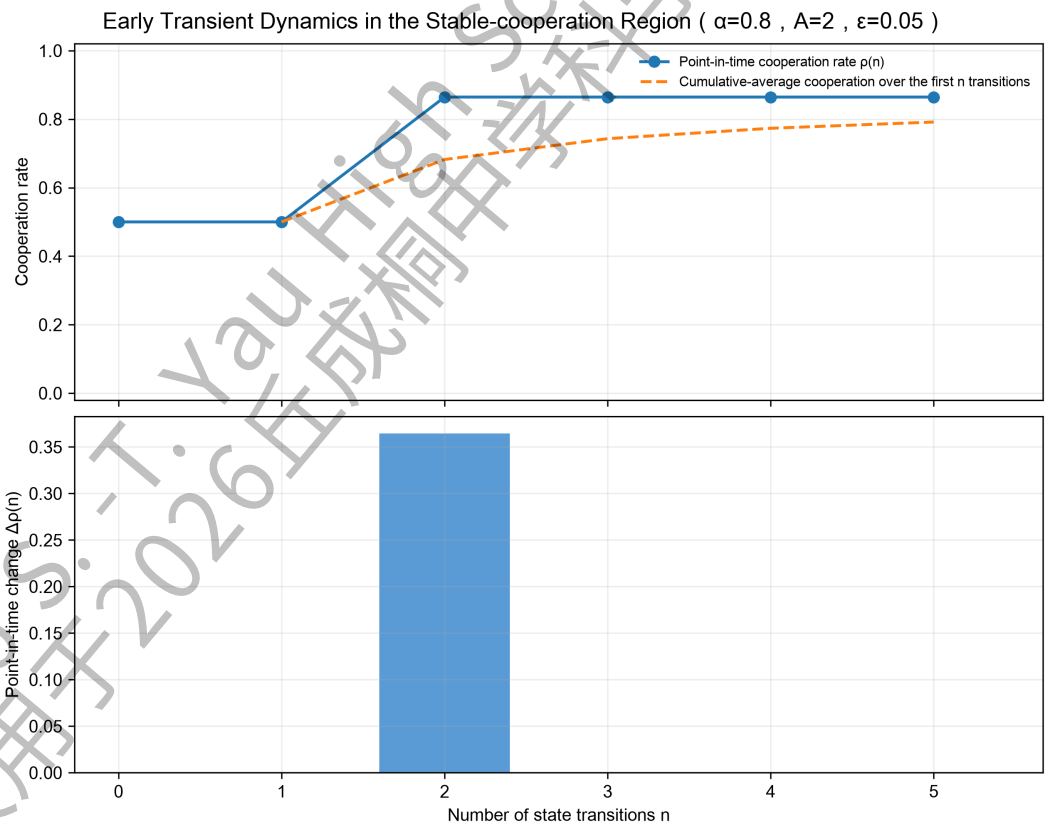


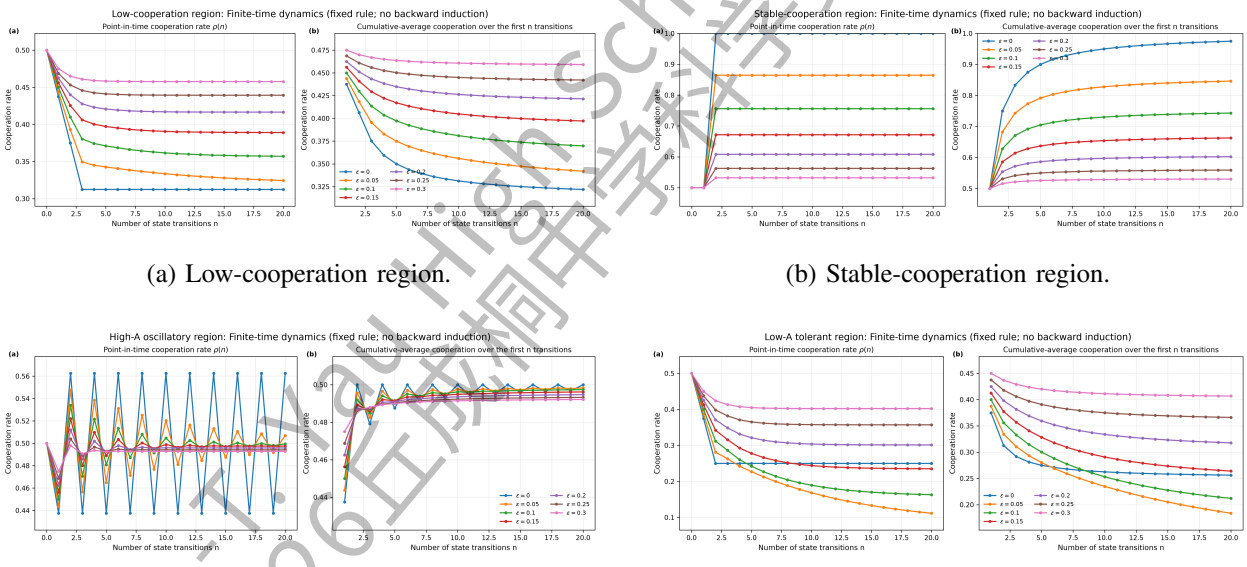
Fig. 5: Early-stage cooperation change under $\alpha = 0.8, A = 2$

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Figure 5 focuses on the early-stage cooperation change over transitions $n = 1$ to $n = 5$ for the stable-cooperation region ($\alpha = 0.8, A = 2$). We observe that the cooperation rate increases rapidly from $n = 1$ to $n = 2$ and stabilizes thereafter. That is, the system quickly converges to a high cooperation rate after the first finite memory is formed, and the subsequent transitions maintain this level with minor fluctuations.

B. Representative Finite-Time Dynamics

The stable-cooperation point shows rapid formation of cooperation, but the four representative regions have different finite-time paths. Figure 6 compares the point-wise and cumulative cooperation rates for these regions. The contrast confirms that a similar average cooperation level can hide different mechanisms: stable cooperation is sustained through a high point-wise rate, whereas the high-A oscillatory region is driven by repeated alternation. The cumulative average also adjusts more slowly than the point-wise rate because it retains the influence of the early transitions.



(a) Low-cooperation region.

(b) Stable-cooperation region.

(c) High-A oscillatory region.

(d) Low-A tolerant region.

Fig. 6: Point-wise and cumulative cooperation rates for the four representative regions. Note that the cumulative-average curves start at $n = 1$ since no transition has occurred at $n = 0$.

C. Error Analysis Across α and ε

Having established the finite-time behavior at a fixed error level $\varepsilon = 0.05$, we now examine how decision errors affect both the long-run reference and the transient dynamics. We fix

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

the aspiration level at $A = 2$ and scan the memory weight $\alpha \in [0, 1]$ over error levels $\varepsilon \in \{0, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3\}$.

The main text focuses on the conceptual distinction between the zero-error system and the low-positive-error limit. The full α-ε phase graph, convergence-time comparison, early-stage change, and four-region error scan are reported in the appendix.

1) Zero-Error vs. Low-Positive-Error Limit: A fundamental distinction arises between the deterministic system at $\varepsilon = 0$ and the randomized system at any $\varepsilon > 0$. When $\varepsilon = 0$, the Markov chain may be reducible or periodic, and its long-run behavior depends on the initial distribution μ_0 through the Cesàro time average:

$$\bar{\rho}^0 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mu_0 P^j \cdot c. \tag{17}$$

When $\varepsilon > 0$, no matter how small, the chain becomes irreducible and aperiodic, with a unique stationary distribution $\pi(\varepsilon)$ that is independent of μ_0.

The two limits need not coincide:

$$\lim_{\varepsilon\to 0^+} \pi(\varepsilon) \cdot c \neq \bar{\rho}^0. \tag{18}$$

This discontinuity arises because the zero-error system can remain trapped in closed communicating classes that are only connected through cases when $\varepsilon > 0$, which is shown in Figure 7.

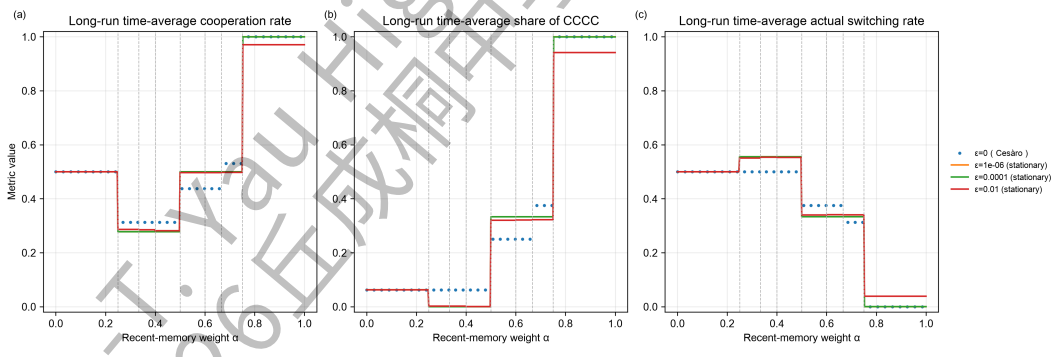


Fig. 7: Comparison of long-run outcomes at $\varepsilon = 0$ (Cesàro average) and at three small positive error levels (stationary values) with $A = 2$. Panels show (a) cooperation rate, (b) the share of CCCC, and (c) actual switching rate.

VI. Simulation and Extensions

The previous section established the analytical results of the two-period ($k = 2$) memory model with fixed aspiration level using Markov chain analysis. In this section, we first validate

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

the analytical results through Monte Carlo simulations, ensuring that the state encoding, payoff mapping, error execution, and index conventions are correctly implemented. Then, we extend the analysis beyond the baseline $k = 2$ model in two directions: longer memory lengths ($k = 2, \ldots, 8$) and a bidirectional adaptive aspiration rule. These extensions test whether the core mechanism identified in the baseline model remains robust.

A. Monte Carlo Validation (29 points, 87 comparisons)

We use Monte Carlo simulations to verify the analytical results computed from the Markov chain are not artifacts of implementation errors. This verification is done under parameter settings $\varepsilon = 0.05$, $A = 2$ and $k = 2$, with α values sampled from the interval $[0, 1]$.

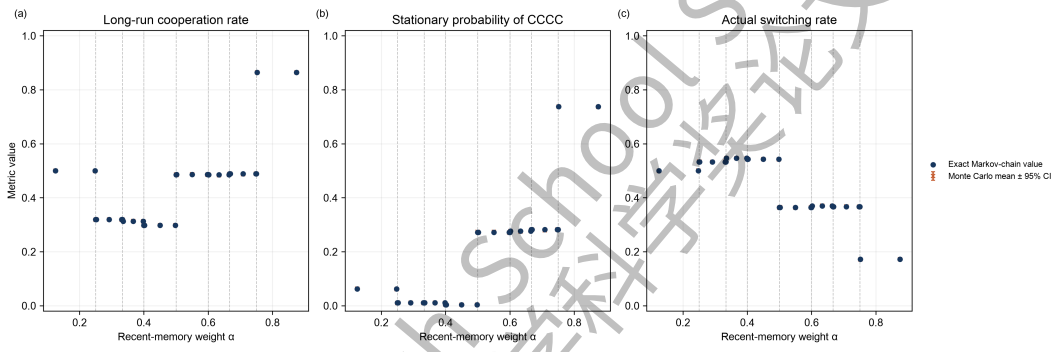


Fig. 8: Monte Carlo validation of analytical results. For each sampled α value, the blue point gives the exact value obtained from the Markov-chain calculation, while the orange error bar gives the Monte Carlo mean with a 95% confidence interval. The three panels report (a) the long-run cooperation rate, (b) the stationary probability of $CCCC$, and (c) the actual switching rate.

In Figure 8, the blue points show the exact Markov chain values as functions of α, and the orange error bars show the Monte Carlo means with nominal 95% confidence intervals. Note that most confidence intervals are smaller than the marker size and are therefore visually hidden by the exact-value markers.

Across 29 α points and 3 metrics (87 comparisons in total), 82 exact values fall inside the corresponding nominal 95% confidence intervals, and all 87 absolute errors are below the pre-specified tolerance of 0.015. The confidence intervals are not adjusted for multiple comparisons and are used solely as a consistency check.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

B. Extension to Longer Memory (k=2,...,8)

The baseline model assumes $k = 2$, which keeps the state space at $4^2 = 16$. For longer memory, the exact state space rapidly grows as 4^k, making full enumeration impractical. We then use Monte Carlo simulations to explore the effects of longer memory lengths $k = 2, \ldots, 8$ on cooperation dynamics.

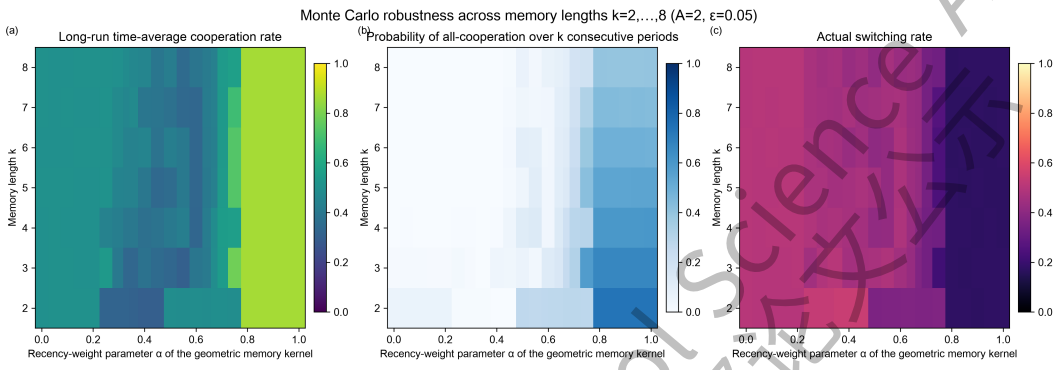


Fig. 9: Robustness of the high-recent-weight mechanism across different memory lengths. The three panels correspond to (a) long-run cooperation rate, (b) probability of mutual cooperation over k consecutive rounds, and (c) actual switching rate.

The results in Figure 9 show that the maximum cooperation rate remains approximately 0.865 for every k from 2 to 8, confirming that the high-recent-weight mechanism is robust across the examined memory lengths. The probability of k-period all-cooperation declines as k increases, because the event itself becomes more strict with longer memory; this decline should not be interpreted as a deterioration of cooperation.

C. Adaptive Aspiration

The baseline model assumes a fixed aspiration level $A = 2$. We now allow A to adjust in response to payoffs using a bidirectional exponential tracking rule:

$$A_i(r+1) = (1-\lambda)A_i(r) + \lambda u_i(r),$$
$$r = 3, \ldots, t-1 \quad \lambda \in [0,1], \tag{19}$$

where λ is the adaptation rate and $u_i(r)$ is the realized payoff of player i in round r. A payoff below the current aspiration lowers A; a payoff above it raises A. The case $\lambda = 0$ recovers the fixed-threshold benchmark $A = 2$. The update order is: (i) determine the action using the current

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公开

history and current A, (ii) apply the decision error and realize the payoff, and (iii) update A for the next round.

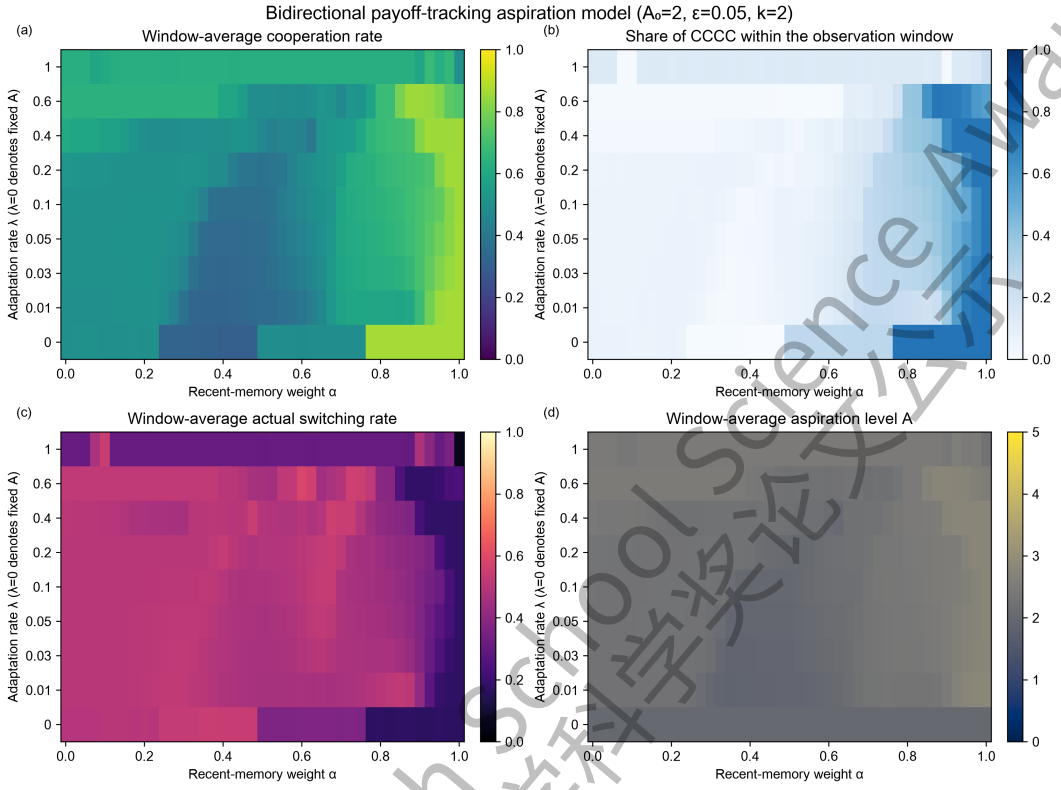


Fig. 10: Bidirectional payoff-tracking aspiration model ($A_0 = 2$, $\varepsilon = 0.05$, $k = 2$). Panel (a): window-average cooperation rate. Panel (b): share of $CCCC$ within the observation window. Panel (c): window-average actual switching rate. Panel (d): window-average aspiration level A. The row $\lambda = 0$ denotes the fixed-A benchmark.

Figure 10 reports the window-average cooperation rate, the share of $CCCC$, the actual switching rate, and the average aspiration level over the full (α, λ) domain. We get to know the following results:

First, the best fixed-threshold baseline achieves a cooperation rate of approximately 0.8651, and no positive adaptation rate on the scanned grid has a higher observed mean. This is a numerical comparison, not a formal statistical test or a claim about all possible adaptive rules. This may be attributed to the case when the aspiration level turns out to be too high, which leads to frequent dissatisfaction and switching; when the aspiration level turns out to be too low, the players will be more tolerant to defection, which can also lead to a low cooperation rate.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Second, high cooperation is concentrated near $\alpha \to 1$, where recent payoffs dominate. The fixed-A benchmark in this regime is equivalent to WSLS; with adaptive aspiration, changing thresholds can alter its stay/switch decisions. The comparison therefore concerns payoff-based satisfaction learning rather than direct imitation of the opponent's action.

Third, at high α and $\lambda = 1$, cooperation drops substantially, the $CCCC$ share declines, and the actual switching rate rises.

The adaptive rule therefore does not uniformly improve cooperation. In the low-α region, where the fixed-threshold baseline already performs poorly, moderate adaptation can raise cooperation by lowering the aspiration level toward realized payoffs. In the high-α region, where the baseline already achieves high cooperation, slow adaptation preserves the benchmark, but fast adaptation destabilizes the $CCCC$-dominant structure and reduces cooperation substantially.

VII. LLM-Based Agent Experiment

A. Motivation and Research Question

The theoretical model, its simulation verification, and its extensions establish the internal robustness of the satisfaction-switching mechanism under finite memory. A further question is whether this mechanism can be instantiated in realistic, language-based decision-makers. We therefore conduct an experiment in which external finite memory and a lightweight hypergraph are connected to Qwen3.6-27B, and the model is fine-tuned via 4-bit QLoRA to learn when to record, retrieve, query the graph, clarify, and refuse unsafe operations.

The experiment tests a specific claim: under an identical external cognitive mechanism, does behavior-data training make the model call these mechanisms more stably and more evenly across behavior categories? It is not intended to test whether the base model's general capability improves.

B. Three-Layer Architecture

The system combines a language-model decision layer with deterministic external mechanisms. Finite memory and the satisfaction state machine are inspired by the iterated Prisoner's Dilemma model analyzed in the previous sections, whereas lightweight hypergraph retrieval is used as an additional engineering component. A 4-bit QLoRA adapter is trained on 980 oracle-synthetic tool-decision samples to decide when to invoke these mechanisms; the mechanisms themselves are not trained into the model weights. Figure 11 summarizes the relation between the language-model

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

decision layer and the external deterministic mechanism layer, and the full three-layer description is given in Appendix A-D.

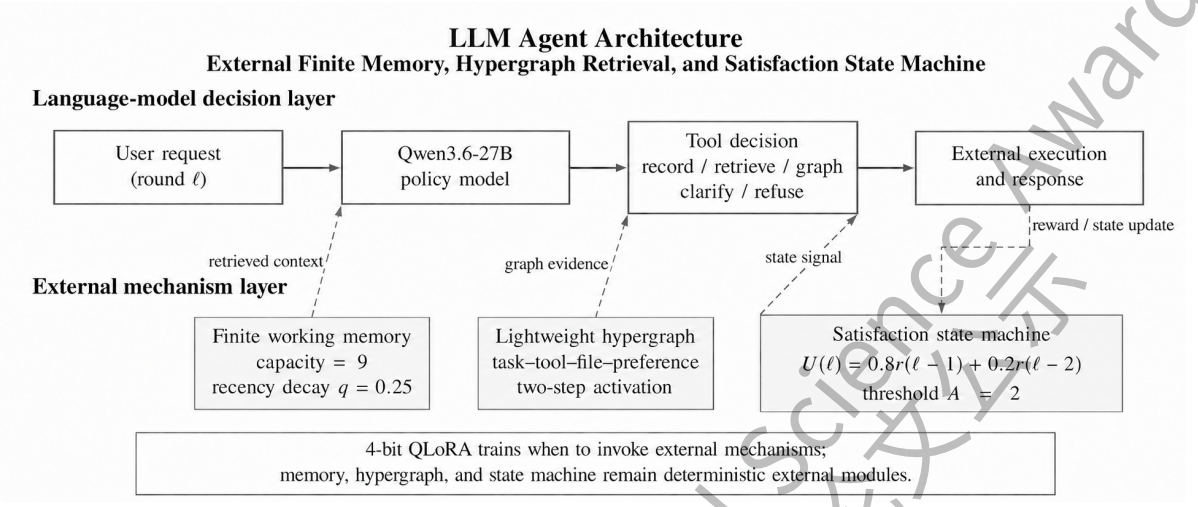


Fig. 11: Architecture of the LLM agent experiment. The language model chooses when to call tools and external mechanisms; finite memory, hypergraph retrieval, and the satisfaction state machine are implemented outside the model.

C. Experimental Design

We evaluate the base and trained models on three multi-round file-based scenarios under decoding parameters, context window, and tool-step limit. A rule-based oracle generates the training samples, and deterministic assertions score each task round. The full experimental pipeline, scenario and data details are provided in Appendix A-D.

D. Results and Analysis

1) Main Condition Comparison: Seven conditions are designed in total, the main text reports four conditions: the base model only (C0), the base model with a mechanism prompt (C1), the base model with the full external mechanism (C2), and the LoRA-980 adapter with the same external mechanism (C4). C2 and C4 are the primary training comparison. The complete seven-condition design, results, and component ablations are reported in Appendix A-D.

Table IV reports the four-condition evaluation results. The improvement from C0 to C1 shows that an explicit mechanism prompt has a substantial practical effect (round accuracy $0.196 \rightarrow 0.578$; macro-average category accuracy $0.152 \rightarrow 0.520$). The further improvement

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

from C1 to C2 shows that an executable external mechanism is more effective than an empty tool interface (round accuracy $0.578 \rightarrow 0.633$; macro accuracy $0.520 \rightarrow 0.560$).

TABLE IV: Four-condition evaluation results.

Condition	Valid runs	Round accuracy	Macro-avg category accuracy	Satisfaction rate
C0 (base, no mechanism)	9	0.1963	0.1523	0.8333
C1 (prompt only)	9	0.5778	0.5203	0.8630
C2 (base + external)	9	0.6333	0.5603	0.8704
C4 (LoRA-980 + external)	9	0.6333	0.6818	0.9556

2) Core Causal Comparison: C2 vs. C4: The core causal comparison is C2 versus C4, since both use the same base model version, scenarios, decoding parameters, context window, and external mechanism; the only difference is the 980-sample LoRA adapter loaded in C4.

The overall round accuracy is identical at 0.6333. The paired bootstrap difference is 0, with a 95% confidence interval of $[-0.0667, 0.0667]$, which includes zero. We therefore cannot claim that QLoRA significantly improves overall round accuracy.

However, other round-level metrics show consistent improvement. The macro-average category accuracy rises from 0.5603 to 0.6818 ($+0.1215$), and the satisfaction rate rises from 0.8704 to 0.9556 ($+0.0852$). These results show higher measured macro-average category accuracy and satisfaction in the tested scenarios. They do not, by themselves, establish greater behavioral stability or generalize beyond these tasks.

An interactive demo about the results above and other supplement materials of this exploratory application are described in Appendix A-D.

E. Limitations and Future Work

The LLM experiment has the following limitations. First, the evaluation uses three public scenarios with three seeds each, and C3 has only 8 valid runs, so statistical power is limited. Second, the scenarios are separated from the training templates but are not hidden tests; blind evaluation remains necessary. Third, the satisfaction threshold $A = 2$ and the memory parameters are only partially calibrated, so the state-machine branch may understate the effect of dissatisfaction

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

and switching. Finally, C5 and C6 are component ablations on the untrained base model, so they should be read as diagnostic comparisons rather than independent model variants. [1]

As for future work, the experiment can be extended in three directions. First, hidden tests and blind evaluation should be used to test whether the mechanism generalizes beyond the public scenarios. Second, the satisfaction threshold, reward scale, memory length, and memory capacity should be calibrated systematically rather than fixed manually. Third, the same architecture can be extended from a single language-model agent to multiple agents with separate finite memories and satisfaction states, allowing us to test whether a shared hypergraph can serve as a coordination device among bounded-rational agents.

The LLM experiment provides exploratory engineering evidence: external memory and hypergraph retrieval improve measured task performance in the tested scenarios, while QLoRA raises macro-average category accuracy and satisfaction without changing overall round accuracy. These single-agent tasks are not a two-player Prisoner's Dilemma, the two-round switching tolerance modifies the theoretical rule, and hypergraph retrieval is an additional component. The experiment therefore illustrates an application inspired by the model, rather than validating its economic cooperation mechanism.

VIII. Discussion and Conclusion

A. Main Findings

This paper has studied cooperation in the iterated Prisoner's Dilemma under a finite-memory satisfaction-switching rule. The main findings are summarized as follows.

RQ1. The memory weight α and the aspiration level A jointly determine both the level and the nature of cooperation. Stable reciprocity requires not only a high cooperation rate but also a high probability of the $CCCC$ state and a low switching rate. For $k = 2$ and $A = 2$, the interval $\alpha > 3/4$ achieves all three simultaneously under the current payoff matrix and baseline error specification.

RQ2. Under the numerical payoff matrix used in this paper and for a fixed aspiration level $A = 2$, the high-cooperation region is the interval $\alpha > 3/4$, which is the $A = 2$ slice of the WSLS-equivalent wedge. It achieves a long-run cooperation rate of 0.8645 at $\varepsilon = 0.05$ in this

[1] C3, C5, and C6 are reported in Appendix A-D.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

model. Within that specification, the same region remains visible across the tested error levels $\varepsilon \in [0, 0.3]$, although increasing error reduces the cooperation rate on the plateau.

RQ3. The analytical results from the exact 16-state Markov chain are validated by Monte Carlo simulations: across 87 comparisons, all absolute errors are below 0.015, and 82 of 87 exact values fall inside the nominal 95% confidence intervals. The high-recent-weight mechanism remains effective when memory is extended to $k = 2, \ldots, 8$.

RQ4. On the scanned grid, the best positive adaptation rate ($\lambda = 0.01$) has an observed cooperation rate of 0.8642, compared with 0.8651 for the fixed-threshold benchmark. No formal test of this difference is performed, so statistical equivalence is not established. At high α, fast adaptation ($\lambda = 1$) reduces cooperation and increases switching; its effect differs in the low-α region.

B. Theoretical Implications

Our results carry three theoretical implications.

First, finite memory does not by itself prevent backward induction; what prevents it in our model is the behavioral assumption that players do not condition their decisions on the remaining horizon. This distinction matters: the model isolates the effect of memory weighting and satisfaction adaptation from the effect of forward-looking calculation, and shows that cooperation can emerge without either full history or backward induction.

Second, under this payoff matrix, the high-cooperation plateau is exactly WSLS-equivalent in the wedge $5 - 4\alpha < A \leq 3\alpha$, including the $A = 2$ slice $\alpha > 3/4$ [3]. The older payoff does not affect decisions there, so this plateau is not evidence that two-period memory outperforms one-period WSLS. The contribution is the characterization of how weighted satisfaction rules change behavior across parameter regions, including departures from WSLS.

Third, a high cooperation rate does not necessarily imply stable reciprocity. The high-A region achieves an average cooperation rate near 0.5, but this average is generated by frequent alternation between cooperation and defection, not by sustained mutual cooperation. This distinction would be invisible if we reported only the long-run cooperation rate.

Fourth, decision error does not have a unidirectional effect on cooperation. It disrupts the existing structure—whether that structure is cooperative, defective, or oscillatory. The direction of the effect depends entirely on the nature of the zero-error system, and the zero-error limit need not coincide with the low-positive-error stationary distribution.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

C. Methodological Implications

Methodologically, this paper makes three contributions.

First, we separate the finite-round transient dynamics from the long-run stationary analysis. The point-wise cooperation rate, the cumulative average cooperation rate, and the long-run cooperation rate answer different questions, and reporting only one of them can be misleading.

Second, we distinguish the zero-error Cesàro time average from the low-positive-error stationary distribution. These two limits are not interchangeable, and treating them as equivalent can hide the closed communicating classes that trap the deterministic system.

Third, we validate the exact Markov chain results against independent Monte Carlo simulations, and we extend the analysis to longer memory and adaptive aspiration. This combination of exact computation, simulation, and extension provides a replicable framework for studying finite-memory cooperation.

Fourth, we carry out the exploratory LLM application in which we extend the mechanism from repeated-game analysis to multi-round language-model interaction. In the tested scenarios, external memory and lightweight hypergraph retrieval improve round accuracy and macro-average category accuracy, while behavior-data training further improves category-balanced behavior and measured satisfaction without improving overall round accuracy. Supplementary task-level metrics also indicate higher task completion, higher requirement coverage, higher tool-use accuracy, and a lower rework rate. This suggests a possible mechanism-design perspective for improving human–LLM interaction and coordination within multi-agent systems.

D. Limitations and Future Work

Theoretical model. The exact model includes only two players; network information sharing is not yet incorporated. Results depend on the specific payoff values and the uniform initial distribution. For $k > 2$, only a one-parameter geometric memory kernel is examined.

Simulation and extensions. The Monte Carlo validation uses 29 α points; larger-scale validation could further reduce sampling uncertainty. The adaptive aspiration analysis does not include a formal paired test between the best adaptive grid point and the fixed-threshold benchmark.

LLM experiment. The LLM experiment is limited by a small number of scenarios and seeds, an uncalibrated satisfaction threshold, and the absence of hidden tests and system-level stress tests. Its conclusions provide exploratory engineering evidence, not validation of the Prisoner's Dilemma cooperation predictions.

2026 C.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

Several directions remain open.

First, the two-player baseline could be extended to heterogeneous populations with different α, A, and λ across players.

Second, network information sharing could be introduced as an independent extension.

Third, the LLM experiment could be extended with hidden tests, blind evaluation, and system-level stress testing.

Fourth, the satisfaction parameters in the LLM state machine could be calibrated through ablation studies.

Fifth, the mechanism could be extended to multi-agent systems, with multiple agents sharing a common hypergraph. This would test whether the satisfaction-switching rule sustains coordination among multiple language-based agents in a shared task environment.

E. Conclusion

This paper has analyzed a finite-memory satisfaction-switching rule in the iterated Prisoner's Dilemma without full history and without backward induction. It identifies when the rule reduces to WSLS, when it produces low-cooperation trapping, and when it generates high-frequency oscillation. For $k = 2$ and $A = 2$, the high-cooperation interval $\alpha > 3/4$ persists across the tested error levels under the current payoff matrix. Longer-memory simulations retain high cooperation at sufficiently large recent weights, without establishing an identical high-cooperation interval or an advantage over one-period WSLS. Monte Carlo simulation supports the exact computations; the LLM experiment is a separate exploratory application of the model. More generally, the results suggest that cooperation research should not rely on a single long-run average: only by jointly examining the point-wise and cumulative cooperation rates, the stationary probability of sustained reciprocity, and the actual switching rate, can we distinguish stable cooperation from frequent alternation.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

APPENDIX A
SUPPLEMENTARY MATERIALS

A. Notation Table

TABLE V: Notation and definitions.

Symbol	Definition
T, R, P, S	Prisoner's Dilemma payoffs
t	Total number of repeated-game rounds (fixed horizon)
r	Current repeated-game round index, $r = 1, \ldots, t$
n	Transition step after two-period memory is initialized
$a_i(r)$	Action of player i in round r (C or D)
$u_i(r)$	Payoff of player i in round r
k	Memory length
α	Weight on the most recent payoff in two-period memory
A	Aspiration level
$m_i(r)$	Weighted payoff from recent memory in round r
ε	Probability of decision error
z_r	Action pair in round r
X_r	Markov state formed by the most recent two action pairs before round r
$p_i(X_r)$	Actual probability that player i cooperates in state X_r
$P(\alpha, A, \varepsilon)$	16×16 row-stochastic transition matrix
μ_0	Initial distribution over 16 states
$\mu(n)$	Distribution over states after n transitions
π	Stationary distribution (for $\varepsilon > 0$)
ρ_∞	Long-run cooperation rate
$\rho(n)$	Point-in-time cooperation rate after n transitions
$\bar{\rho}(n)$	Cumulative average cooperation rate over the first n transitions
$\pi(CCCC)$	Stationary probability of the $CCCC$ state
σ_{actual}	Actual switching rate, accounting for decision error

B. Theoretical and Finite-Time Supplementary Diagnostics

Baseline Transition and Stationary Diagnostics: We first examine the transition structure of the 16-state chain. Figure 12 visualizes the 16×16 transition matrix $P(\alpha, A, \varepsilon)$ for the baseline parameters $\alpha = 0.5$, $A = 2$, and $\varepsilon = 0.05$. Each row corresponds to a current state and sums to one. Each state has exactly four possible successors, corresponding to the four possible action pairs in the next round.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

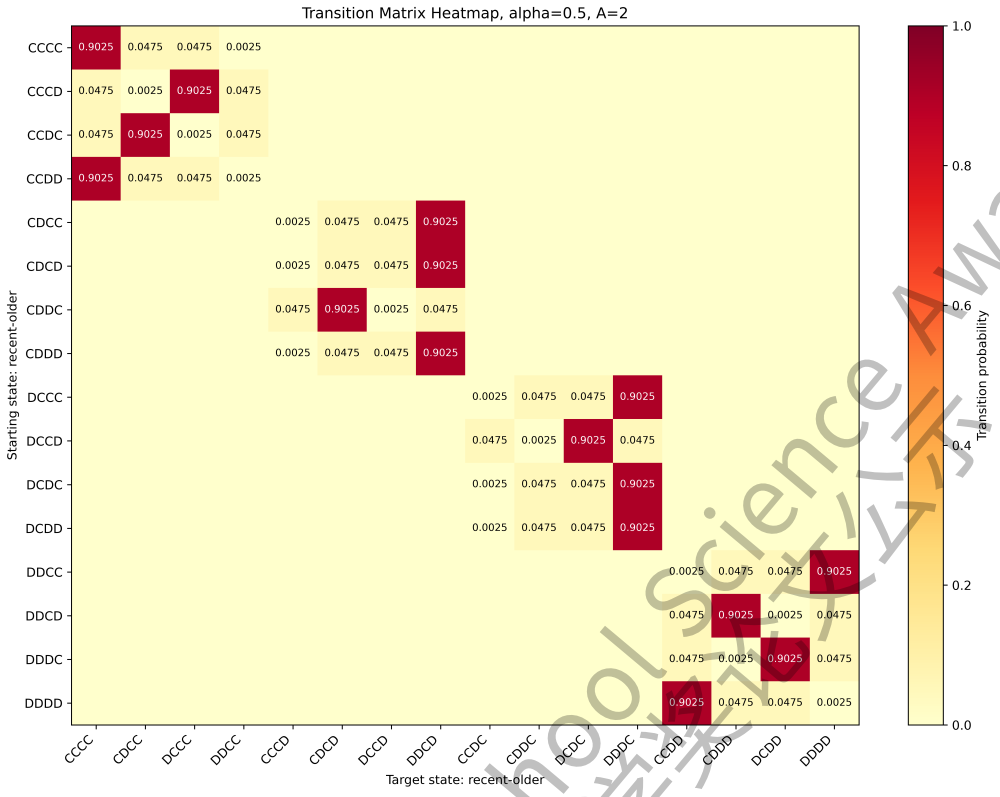


Fig. 12: 16×16 transition matrix for $\alpha = 0.5$, $A = 2$, $\varepsilon = 0.05$. States are ordered with the most recent period first. Darker cells indicate higher transition probabilities.

With this transition structure, we compute the unique stationary distribution π satisfying $\pi P = \pi$. Table VI reports the stationary probabilities for the 16 states

TABLE VI: Stationary distribution of the 16 states ($\alpha = 0.5$, $A = 2$, $\varepsilon = 0.05$).

State	Probability	State	Probability
CCCC	0.271429	DCCC	0.015033
CCCD	0.000831	DCCD	0.008306
CCDC	0.000831	DCDC	0.015781
CCDD	0.029236	DCDD	0.143605
CDCC	0.015033	DDCC	0.000831
CDCD	0.015781	DDCD	0.157807
CDDC	0.008306	DDDC	0.157807
CDDD	0.143605	DDDD	0.015781

According to Table VI, the states $CCCC$, $DDCD$, and $DDDC$ have high probabilities, while

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

the other states have relatively low probabilities under this parameter setting.

Supplementary Point-wise and Cumulative Phase Maps: Starting from the uniform initial distribution μ_0 defined in Section III, we compute the distribution after n transitions as $\mu(n) = \mu_0 P^n$. The point-wise cooperation rate $\rho(n) = \mu(n) \cdot c$ captures the level of cooperation in a specific transition count n, while the cumulative average $\bar{\rho}(n) = \frac{1}{n} \sum_{j=1}^{n} \rho(j)$ reflects the overall cooperation accumulated over the first n transitions.

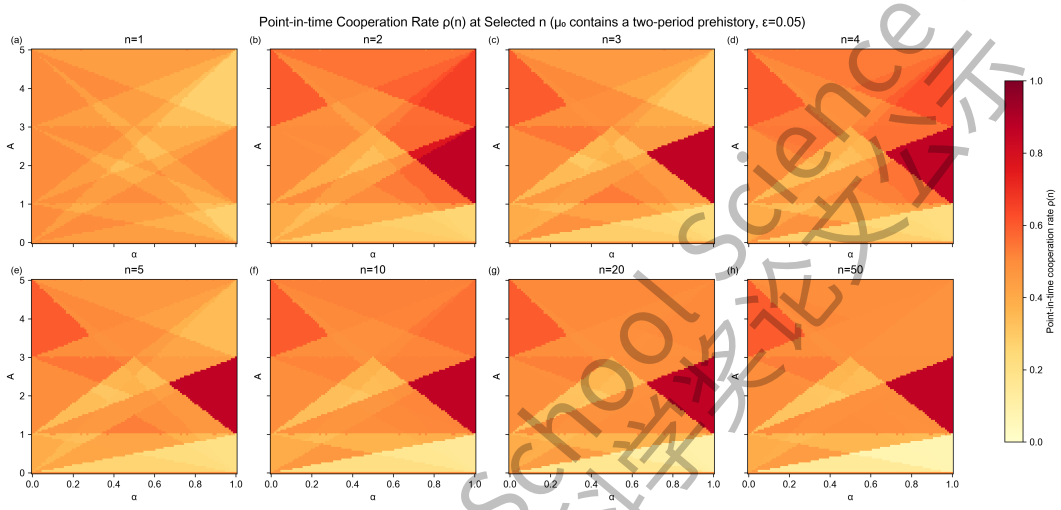


Fig. 13: Point-wise cooperation rate over α and A

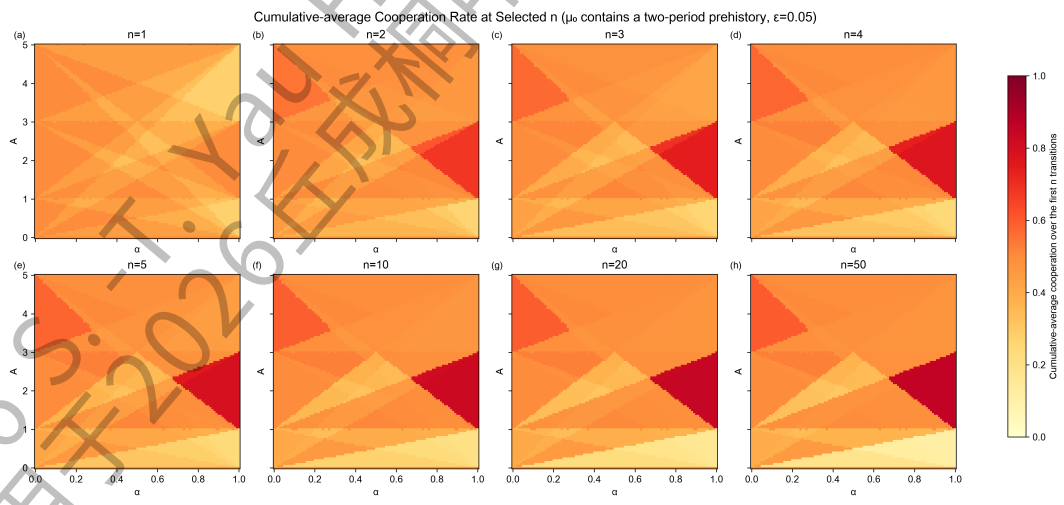


Fig. 14: Cumulative cooperation rate over α and A

Figure 13 and Figure 14 show $\rho(n)$ and $\bar{\rho}(n)$ for $n = 1, 2, \ldots, 50$ over the full (α, A) space.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

We observe that $\rho(n)$ reaches its long-run platform much earlier than $\bar{\rho}(n)$, because the cumulative average retains the influence of the initial uniform distribution over the 16 states. This distinction is important since reporting only the final point-wise cooperation rate may overstate the average performance over a finite horizon.

Supplementary Cooperation-Formation Speed: To quantify how quickly the cooperation emerges or decays, we define the cooperation formation speed as the discrete-time change rate:

$$v(n_1, n_2) = \frac{\rho(n_2) - \rho(n_1)}{n_2 - n_1}, \quad n_2 > n_1. \tag{20}$$

Positive $v(n_1, n_2)$ indicates that cooperation is forming; negative $v(n_1, n_2)$ indicates that cooperation is decaying; $v(n_1, n_2) \approx 0$ indicates little net change between the two endpoints, which alone does not establish convergence.

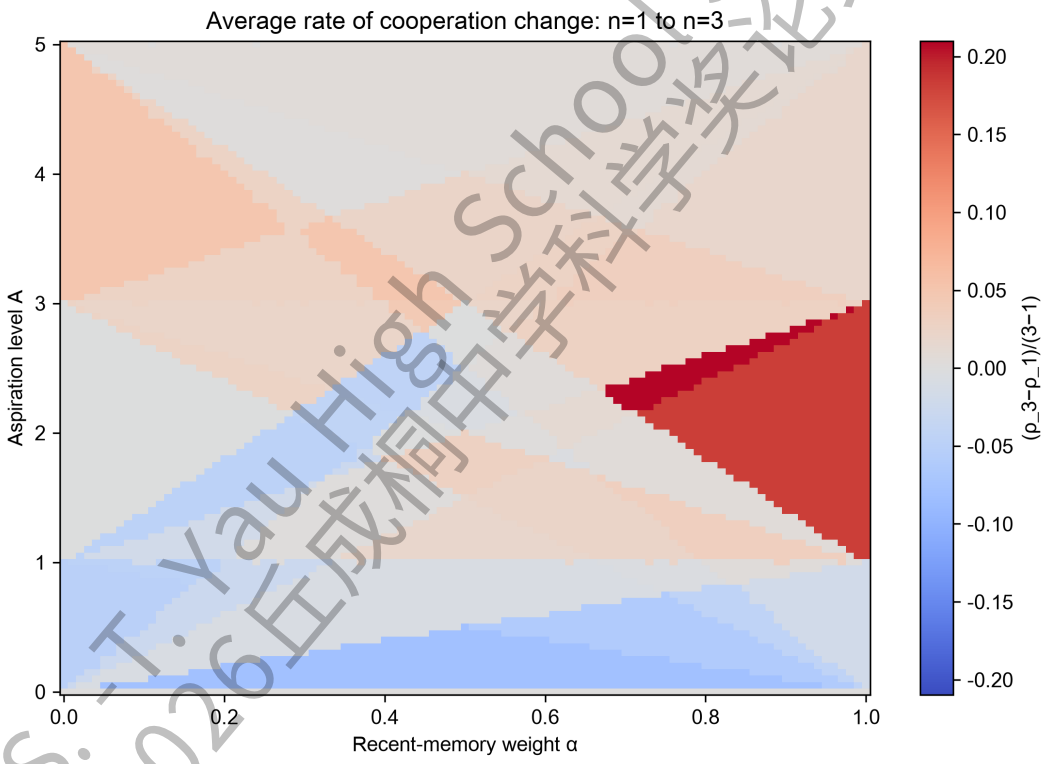


Fig. 15: Cooperation formation speed from $n_1 = 1$ to $n_2 = 3$ over α and A

Figure 15 reports $v(n_1, n_2)$ over the early stage ($n = 1$ to $n = 3$) across the (α, A) space. The WSLS-equivalent wedge ($5 - 4\alpha < A \leq 3\alpha$) exhibits strong positive speed in this window, confirming that cooperation emerges rapidly once the two-period memory is established. In

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

contrast, the low-α, high-A region shows near-zero or negative speed, reflecting its inability to form stable cooperation even in the early transitions.

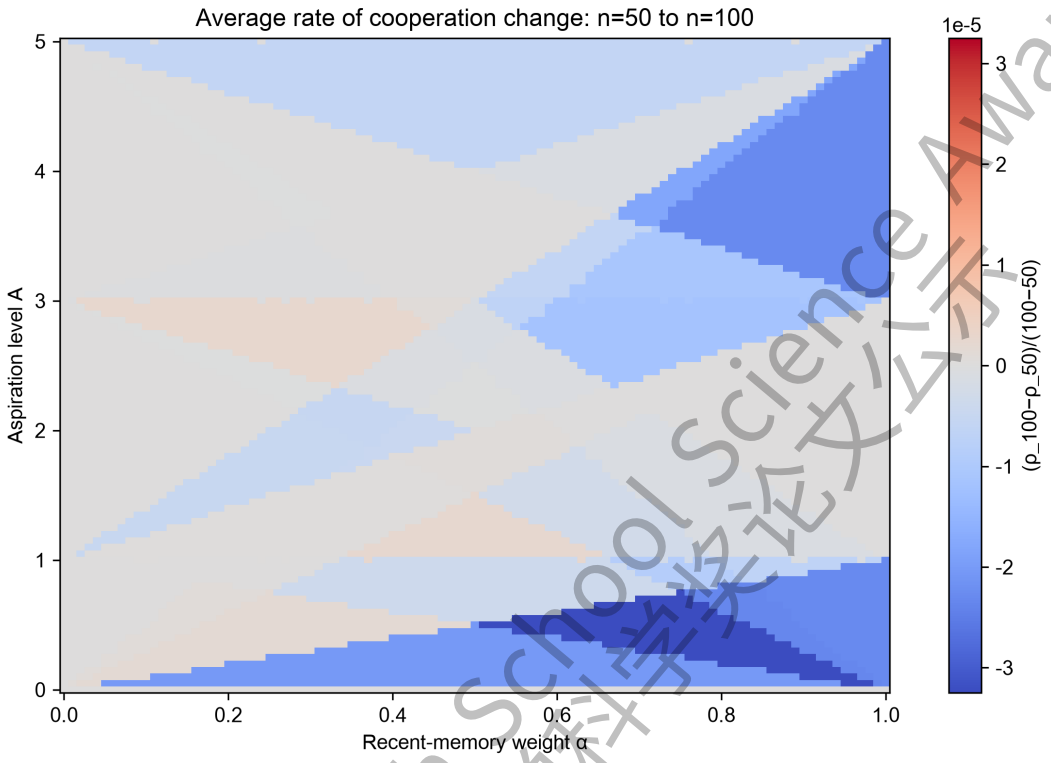


Fig. 16: Cooperation formation speed from $n_1 = 50$ to $n_2 = 100$ over α and A

To examine whether the formation process still exists beyond the early transitions, Figure 16 reports $v(n_1, n_2)$ over the later stage ($n = 50$ to $n = 100$) across the (α, A) space. The speed is close to zero across most of the parameter space, indicating little net change between these endpoints; convergence requires the separate checks below. The remaining non-zero regions coincide with the high-A fluctuation zone, where cooperation continues to oscillate even at large n, consistent with the high switching rate observed in Figure 4.

Supplementary Convergence-Time Diagnostics: We define the convergence time as the first transition step n such that the point-wise cooperation rate remains within a tolerance δ of its long-run value for all subsequent transitions:

$$\tau_\rho(\delta) = \min\{n \geq 0 : |\rho(m) - \rho_\infty| < \delta,\ \forall m \geq n\}. \tag{21}$$

We use $\delta = 0.01$ as the primary criterion and verify robustness with $\delta = 0.001$.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Because the cooperation rate is a scalar summary of the full state distribution, it may stabilize before the distribution itself does. Thus, to capture this distinction, we also report the total variation (TV) mixing time of the full 16-state distribution:

$$\tau_{TV}(\delta) = \min\left\{n \geq 0 : \|\mu(m) - \pi\|_{TV} < \delta,\ \forall m \geq n\right\}, \tag{22}$$

where $\|\cdot\|_{TV}$ denotes the total variation distance and π is the stationary distribution.

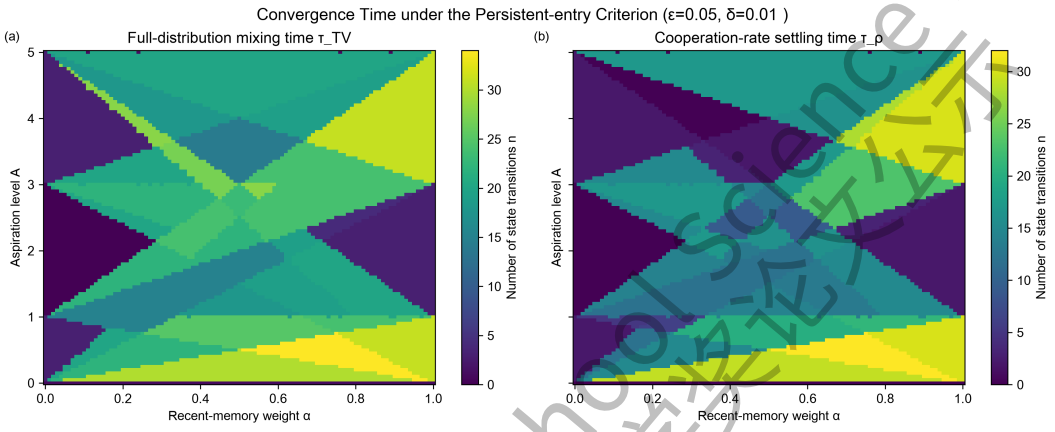


Fig. 17: Convergence time under criterion $\delta = 0.01$ over α and A

Figure 17 reports both τ_{TV} and τ_ρ over the (α, A) space for $\varepsilon = 0.05$ and $\delta = 0.01$. We observe the following features:

First, the representative stable-cooperation setting $(\alpha, A) = (0.8, 2)$ converges quickly: $\tau_{TV} = 3$ and $\tau_\rho = 2$. The system reaches its long-run cooperation level after just two transitions, and the full state distribution stabilizes one step later.

Second, convergence is slowest in parts of the high-α, low-A region, and is also relatively slow in parts of the high-A region. At $\delta = 0.01$, the scanned maximum is $\tau_{TV} = 34$ and $\tau_\rho = 32$. These regions reflect different mechanisms: low-A trapping and slowly decaying high-A oscillations, respectively.

Supplementary Convergence-Time Check: To further verify that these patterns are not artifacts of the tolerance choice, we also compute convergence times under a stricter criterion $\delta = 0.001$ (Figure 18).

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

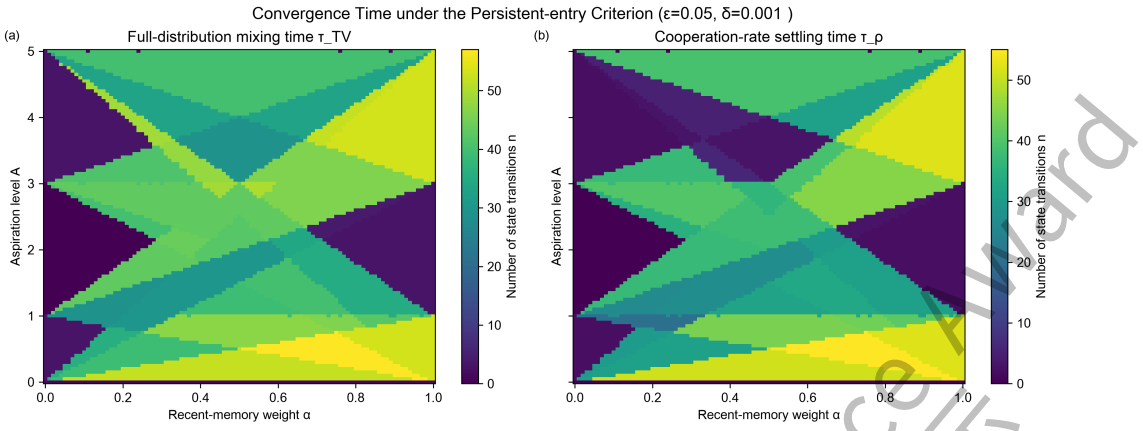


Fig. 18: Convergence time under criterion $\delta = 0.001$ over α and A

As it is shown, the distribution of fast and slow regions remain the same, although the absolute values of τ increase as expected (the criterion becomes stricter). This confirms that the convergence time ranking across parameter regions is robust.

Supplementary Error-Scan Diagnostics:

α-ε phase diagram of long-run cooperation as reference: Before analyzing the transient effects of decision errors ε, we first report the long-run cooperation rate ρ_∞ as a function of both α and ε for $A = 2$. In Figure 19, the high-cooperation interval $\alpha > 3/4$ remains unchanged across the tested error levels $\varepsilon \in [0, 0.3]$ under the current payoff matrix and $A = 2$. As ε increases, unintended actions become more frequent, reducing the cooperation rate on this plateau from 0.997 at $\varepsilon = 0.001$ to 0.532 at $\varepsilon = 0.3$. Thus, the location of the high-cooperation interval is stable within this specific scan, while the absolute level of cooperation is sensitive to decision errors.

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

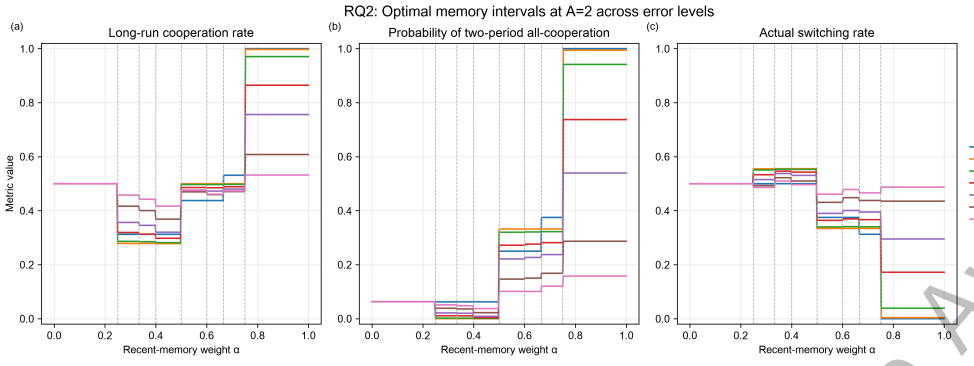


Fig. 19: Long-run cooperation rate ρ_∞ (panel a), probability of $CCCC$ (panel b), and actual switching rate σ_{actual} (panel c) as functions of α and ε at $A = 2$. The vertical dashed lines mark the theoretical critical values of α.

Convergence time as a function of ε: While the location of the high-cooperation interval $\alpha > 3/4$ is stable in the preceding $A = 2$ error scan under the current payoff matrix, the convergence time varies dramatically with ε. Figure 20 reports the convergence time τ_{TV} and τ_ρ as functions of ε for the four representative regions defined in Section IV-D.

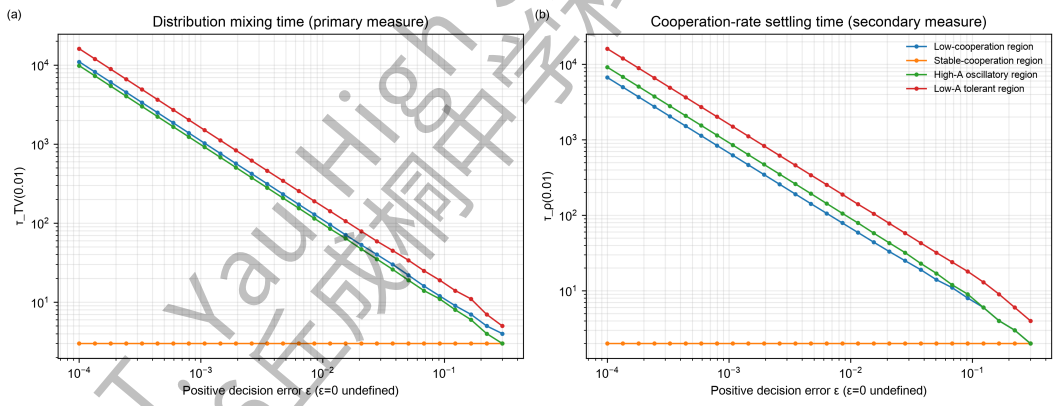


Fig. 20: Convergence time τ_{TV} (panel a) and τ_ρ (panel b) as functions of ε for the four representative regions.

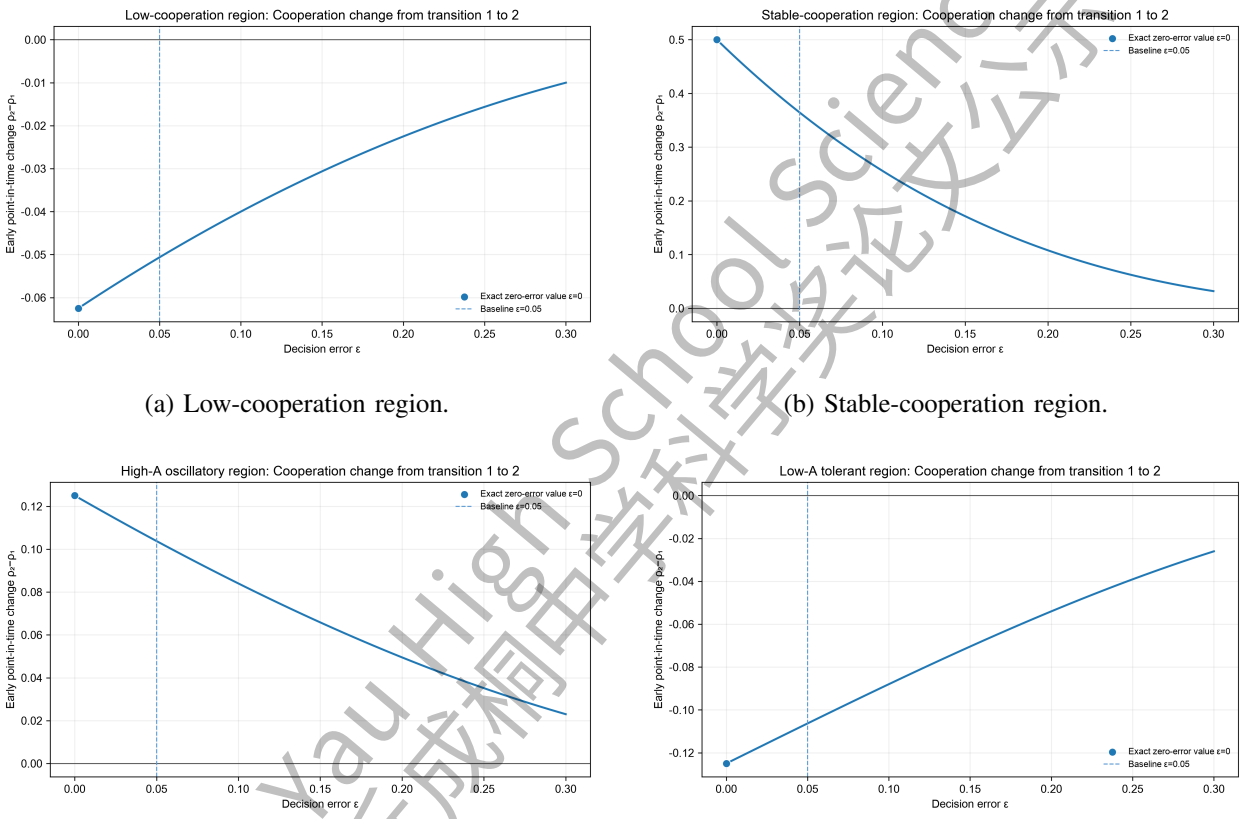
We observe the following features:

First, in the stable cooperation region ($\alpha = 0.8$, $A = 2$), convergence is fast and stable ($\tau_{TV} = 3$) across all $0 < \varepsilon \leq 0.3$, indicating that the $CCCC$-dominant structure is robust to moderate noise.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

Second, for extremely small ε (e.g., $\varepsilon = 10^{-4}$), the convergence time can be orders of magnitude larger than at $\varepsilon = 0.05$. This occurs because the chain behaves like a nearly closed-certain system (similar to the case $\varepsilon = 0$) before eventually mixing to the unique stationary distribution. The small error acts as a rare noise that only gradually connects previously isolated state classes.

Early-stage cooperation change under different ε: Decision errors also affect the early-stage dynamics. Figure 21 reports the early cooperation change $\rho(2) - \rho(1)$ as a function of ε for the four representative regions.



(a) Low-cooperation region.

(b) Stable-cooperation region.

(c) High-A oscillatory region.

(d) Low-A tolerant region.

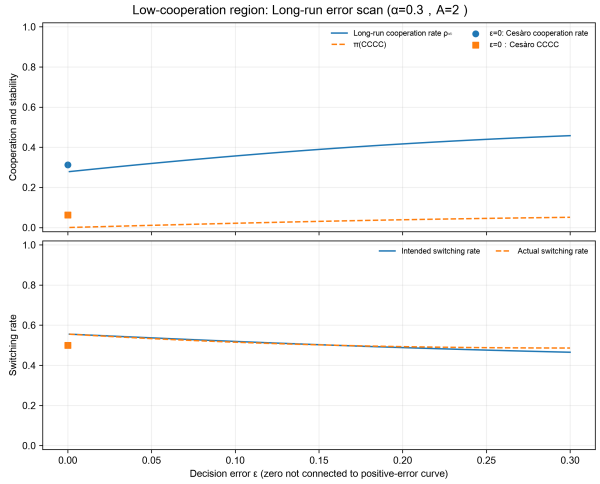
Fig. 21: Early-stage cooperation change $\rho(2) - \rho(1)$ as a function of ε for the four representative regions. The dashed vertical line marks the baseline $\varepsilon = 0.05$.

In the stable cooperation region, larger errors weaken the positive early jump, because the error interrupts the rapid formation of mutual cooperation. In the low-cooperation region, larger errors reduce the negative early decline, because noise occasionally flips defection into cooperation.

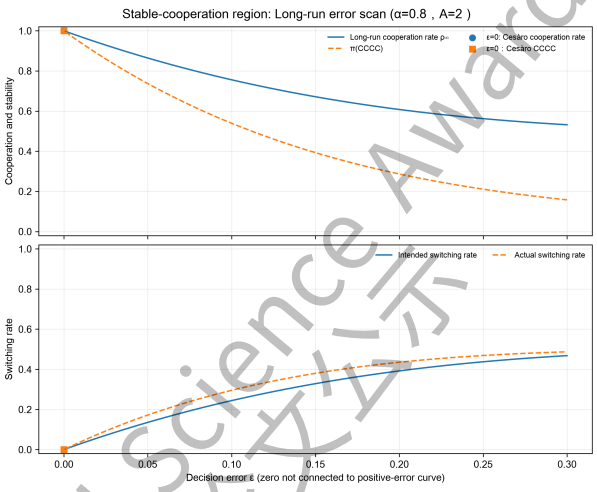
Four representative regions with ε-scan: Finally, we examine how the four representative regions respond to decision errors across the full range of ε. Figure 22 reports the long-run

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

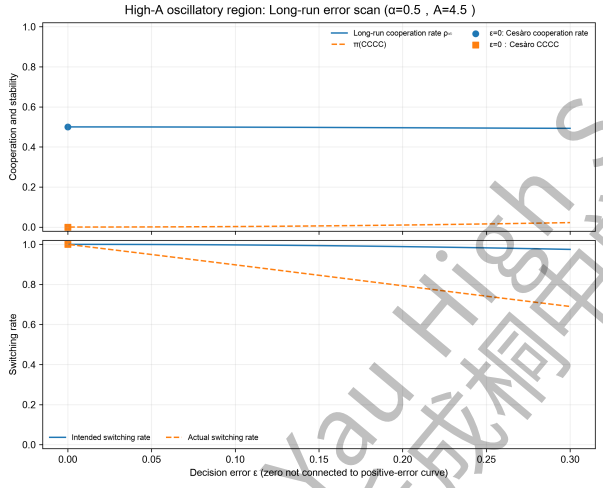
cooperation rate, the $CCCC$ probability, and the actual switching rate as functions of ε for each region.



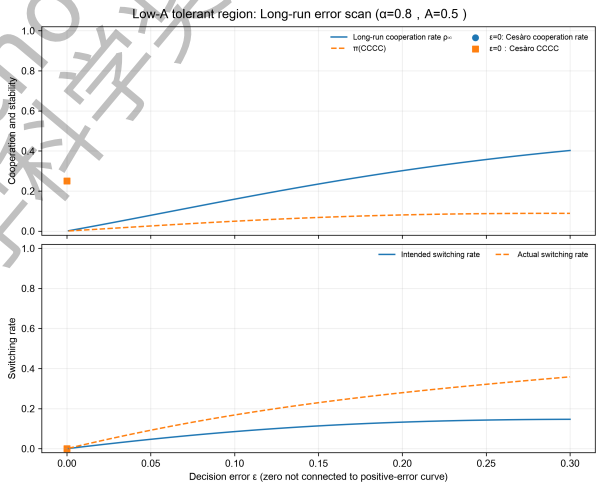
(a) Low-cooperation region.



(b) Stable-cooperation region.



(c) High-A oscillatory region.



(d) Low-A tolerant region.

Fig. 22: Long-run cooperation rate, stationary probability of $CCCC$, and actual switching rate as a function of ε for the four representative regions.

Table VII summarizes the direction of the error effect in each region.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

TABLE VII: Effect of increasing ε across four representative regions.

Region	(α, A)	ρ_∞	$\pi(CCCC)$	**Mechanism**
Low cooperation	$(0.3, 2)$	Increases	Increases	Noise breaks low-cooperation trapping
Stable cooperation	$(0.8, 2)$	Decreases	Decreases	Noise disrupts $CCCC$ cycles
High-A fluctuation	$(0.5, 4.5)$	Stable	Stable	Noise damps deterministic oscillations
Low-A lenient	$(0.8, 0.5)$	Increases	Increases	Noise helps exit low-cooperation states

From Figure 22 and Table VII, we can conclude that decision error ε does not have a unidirectional effect on cooperation. It disrupts the existing structure—whether that structure is cooperative, defective, or oscillatory—and the direction of the effect depends entirely on the nature of the $\varepsilon = 0$ system.

C. Simulation and Extension Details

Simulation Implementation Details: All three experiments share the same path simulator. Each path starts from a uniformly random k-period exogenous history, with the most recent period placed first. At each round, the simulator computes the weighted memory payoff for each player according to the Prisoner's Dilemma payoff matrix, determines the intended action according to the satisfaction-switching rule, applies the decision error with probability ε, and then updates the history by appending the new outcome and removing the earliest period. A warm-up phase precedes the observation window, and all reported metrics are computed within the observation window.

Monte Carlo Sampling Details: The sampling points cover the theoretical critical values, their immediate neighborhoods (± 0.002), and the midpoints of each critical interval, yielding 29 distinct α values after de-duplication.

For each α, we simulate 6,000 independent paths with a warm-up of 300 transitions and an observation window of 400 transitions. We compare three quantities between the exact chain and the simulation: the long-run cooperation rate, the stationary probability of $CCCC$, and the actual switching rate. The exact values of the three metrics are computed from the Markov chain in the same way as in Section IV, and the Monte Carlo estimates are obtained as the mean across paths, with the standard error defined as the standard deviation of the path-level metrics divided by the square root of the number of paths. Nominal 95% confidence intervals are constructed as Monte Carlo mean $\pm$ 1.96 $\times$ standard error.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Longer-Memory Simulation Details: As for memory weights, we adopt a one-parameter geometric memory kernel:

$$w_j = \frac{q^{j-1}}{\sum_{i=1}^{k} q^{i-1}}, \quad j = 1, \ldots, k, \tag{23}$$

where $j = 1$ indexes the most recent round. For $0 < \alpha < 1$, let $q = (1-\alpha)/\alpha$ and at $\alpha = 0$, set $w_k = 1$; at $\alpha = 1$, set $w_1 = 1$. All other weights are zero at these endpoints. When $k = 2$, this kernel reduces exactly to $[\alpha, 1-\alpha]$. For $k > 2$, the most recent weight is $w_1 = 1/\sum_{i=1}^{k} q^{i-1}$, which generally differs from α; the $k = 2$ threshold $\alpha > 3/4$ cannot be transferred unchanged. All examined memory lengths share the $\alpha = 1$ one-period benchmark, so a similar maximum across k does not by itself show a benefit from additional memory. The parameter $\alpha > 0.5$ implies more weight on recent payoffs, $\alpha = 0.5$ corresponds to equal weighting, and $\alpha < 0.5$ places more weight on earlier payoffs. The extreme case $\alpha = 1$ corresponds to a pure one-period memory, where only the most recent payoff matters and all earlier payoffs are ignored; $\alpha = 0$ corresponds to a pure k-period memory, where only the earliest payoff matters and all later payoffs are ignored.

We run simulations for $k = 2, \ldots, 8$, scanning α from 0 to 1 in steps of 0.05, with the seven theoretical critical values added explicitly. Each (k, α) grid point uses 3 independent seeds with 600 paths per seed, with a warm-up and observation window of 500 transitions each. Cooperation-rate uncertainty is taken as the larger of the within-seed path standard error and the between-seed standard error.

Adaptive-Aspiration Simulation Details: Since A becomes a continuous state variable, the adaptive model is no longer a finite 16-state chain. We therefore rely on simulation and do not refer to the observed window average as an exact stationary distribution. The scan covers $\alpha \in [0, 1]$ with step 0.025 (41 points) and $\lambda \in \{0, 0.01, 0.03, 0.05, 0.1, 0.2, 0.4, 0.6, 1\}$. Each grid point uses 3 independent seeds with 600 paths per seed, with a warm-up and observation window of 600 transitions each.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

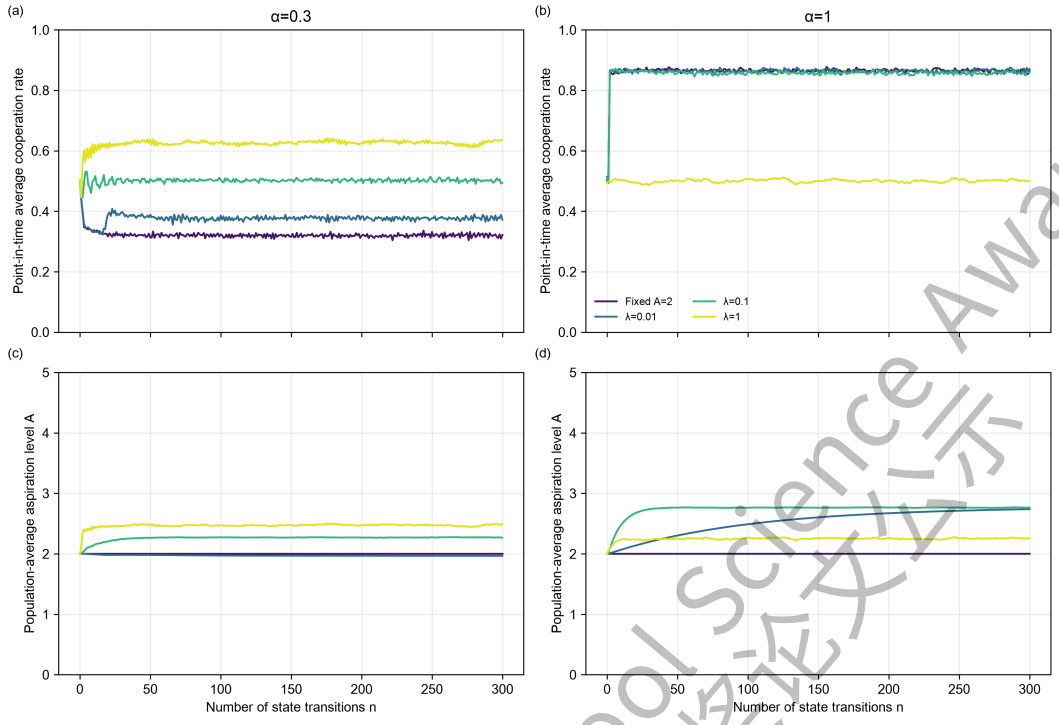


Fig. 23: Transient trajectories of the point-in-time average cooperation rate (panels a and b) and the population-average aspiration level A (panels c and d) for $\alpha = 0.3$ (left) and $\alpha = 1.0$ (right), with $\lambda \in \{0, 0.01, 0.1, 1\}$. Fast adaptation destabilizes the high-cooperation structure at $\alpha = 1$.

Adaptive-Aspiration Trajectory Diagnostics: To understand why fast adaptation destabilizes cooperation, Figure 23 reports the transient trajectories of the cooperation rate and the population-average aspiration level for two representative memory weights, $\alpha = 0.3$ and $\alpha = 1.0$, with $\lambda \in \{0, 0.01, 0.1, 1\}$.

At $\alpha = 0.3$, the fixed-threshold benchmark settles at a low cooperation level around 0.32. Moderate adaptation ($\lambda = 0.01$ or 0.1) raises the cooperation rate to approximately 0.40 and 0.50, respectively, because the aspiration level drifts downward toward the realized payoffs. At $\lambda = 1$, the aspiration level fluctuates around 2.5 and cooperation rises to about 0.62.

At $\alpha = 1.0$, the fixed-threshold benchmark already achieves a high cooperation rate of 0.86. Slow adaptation ($\lambda = 0.01$ or 0.1) preserves this level, but fast adaptation ($\lambda = 1$) drives the aspiration level toward 2.2 and reduces cooperation to approximately 0.50. This confirms that fast adaptation destabilizes the $CCCC$-dominant structure when the baseline mechanism is already performing well.

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

D. LLM Experiment Details

The LLM experiment uses Qwen3.6-27B with 4-bit QLoRA (rank 16, alpha 32, dropout 0.05), trained on 980 oracle-synthetic samples covering seven behavior categories. The training configuration is reported in Table VIII.

TABLE VIII: QLoRA training configuration.

Parameter	**Setting**
Base model	Qwen3.6-27B (approximately 52–54 GB in 4-bit form)
Server	RTX 6000D 84 GB
Training framework	LLaMA-Factory 0.9.5, transformers 5.6.0, PEFT 0.18.0
Quantization	4-bit NF4 with double quantization
LoRA rank / alpha	16 / 32
LoRA dropout	0.05
Cutoff length	2048
Epochs	1
Batch size / gradient accumulation	1 / 8
Learning rate	5×10^{-5}
Scheduler	cosine
Warmup ratio	0.05
Weight decay	0.01
Optimizer	paged_adamw_8bit
Primary training seed	3407
Supplementary training seeds	5701, 8501
Inference seeds	20260825, 20260826, 20260827

Three-Layer Architecture: The system consists of three layers with distinct origins and responsibilities.

- **Layer 1: Finite memory + satisfaction-switching state machine.** This layer is derived from the satisfaction-switching rule in the iterated Prisoner's Dilemma. Working memory holds at most 9 items with recency decay $q = 0.25$ and forgetting threshold 0.05. At the beginning of task round ℓ, the satisfaction utility is computed from the two most recent completed rewards: $U(\ell) = 0.8r(\ell - 1) + 0.2r(\ell - 2), \quad \ell \geq 3$ with aspiration level $A = 2$. The round reward is $r(\ell) = 5p(\ell)$, where $p(\ell)$ is the fraction of passed checks. When $U(\ell) \geq A$, the state remains *efficient*; after two consecutive unsatisfied rounds, the state switches to *clarify*. This layer is implemented as a deterministic external service, not learned by the model.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

- **Layer 2: Lightweight hypergraph + spreading activation.** This layer is related to graph-based retrieval-augmented generation, especially GraphRAG-style retrieval and recent agentic extensions such as Graph-R1 [23], [30]. It is *not* derived from the theoretical model of iterated Prisoner's Dilemma. Each interaction forms one hyperedge connecting multiple entities (task, tool, file, preference). Retrieval uses two-step spreading activation [19], propagating from a query seed along hyperedges without embedding. This layer is also a deterministic external service.
- **Layer 3: Model training (4-bit QLoRA).** This is the only trained component. The model is fine-tuned on 980 oracle-synthetic tool-decision samples, learning *when* to record, retrieve, query, clarify, and refuse unsafe operations. The mechanism itself (Layers 1 and 2) remains external and is never trained into the weights.

External Memory and Lightweight Hypergraph: The external memory consists of a finite working memory and a gist summary. Working memory holds at most 9 items; each item stores importance, write time, last access time, rehearsal count, entity, and topic tags. Memory weight decays with age at rate $q = 0.25$; successful retrieval produces limited rehearsal reinforcement. When weight falls below 0.05, the item enters the forgetting process. High-importance content can be compressed into a gist summary of at most 6 items.

When new information conflicts with old memory, the system retains the conflict relation and requires the model to clarify when necessary, rather than overwriting unconditionally. The lightweight hypergraph holds at most 60 nodes and 40 hyperedges, placing entities, files, tasks, decisions, constraints, and memory items into a single relational structure. Queries propagate at most two steps, recovering information relevant to the current task but outside the recent dialogue window.

The memory design draws on two cognitive-science concepts: Tulving's distinction between episodic and semantic memory, where working memory serves as episodic storage and the gist summary as semantic storage; and Ebbinghaus's forgetting curve, implemented as the recency decay $q = 0.25$ with a retrieval-triggered reinforcement mechanism.

Satisfaction State Machine: After each round, a verifier scores file existence, JSON validity, field content, tool usage, and security boundaries. The round reward is $r(\ell) = 5p(\ell)$, where $p(\ell)$ is the fraction of passed checks. The recent two rewards form the utility $U(\ell) = 0.8r(\ell - 1) + 0.2r(\ell - 2), \quad \ell \geq 3$ with $A = 2$. When $U(\ell) \geq 2$, the state is marked as satisfied. After two consecutive unsatisfied rounds, the state switches from *efficient* to *clarify*; one satisfied round in

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

clarify returns the state to *efficient*.

This two-round tolerance is an empirical design choice: the theoretical model switches whenever $U(\ell) < A$, but the LLM implementation requires two consecutive unsatisfied rounds to reduce excessive switching. This tolerance is intended to limit excessive switching; the reported comparisons do not isolate a reduction in rework caused by this design choice.

Experimental Design: The system uses a "language model decision + external deterministic execution" architecture. Qwen3.6-27B handles task understanding, tool selection, parameter generation, and answer generation, while a Python program maintains four external components: a finite working memory with gist summarization, a conflict manager, a lightweight hypergraph for cross-entity retrieval, and a satisfaction state machine.

The experiment proceeds in six steps. First, we construct three multi-round file-based evaluation scenarios: Orion engineering delivery, PipelineNova data pipeline, and AtlasReview research writing. Second, a deterministic rule-based oracle generates behavior-decision examples for the external memory, hypergraph, clarification, and security actions. Third, the 980-sample dataset is used to train a 4-bit QLoRA adapter, while the external memory, hypergraph, and satisfaction state machine remain deterministic and unchanged. Fourth, we evaluate seven conditions (C0–C6) under the same scenarios, decoding parameters, context window, and tool-step limit. Fifth, deterministic assertions check whether each round satisfies the required files, content, tool-use, clarification, and security constraints. Finally, the results are aggregated into round-level metrics, task-level diagnostics, and component ablations.

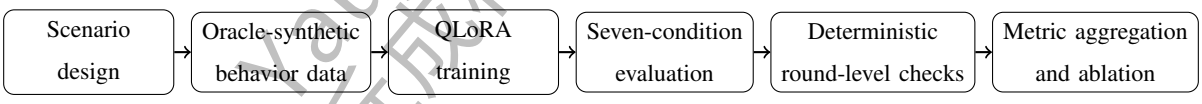


Fig. 24: Experimental pipeline for the exploratory LLM application.

The training data is generated by a deterministic rule-based oracle that simulates the external mechanism and produces tool-decision samples. The oracle constructs system states, user requests, and standard assistant behaviors from fixed scenario templates, covering seven behavior categories (record, recall, no_call, conflict, graph, clarify, security), 140 samples each in the 980-sample version. Splits are separated by scenario family, source project, and entity namespace to reduce template and entity leakage. A 280-sample pilot precedes the 980-sample formal training, and an independent validation script checks the generated data without importing the generator.

2026 Yau High School Science Award 仅用于2026年丘成桐中学科学奖公示

LLM Evaluation Scenarios: The LLM evaluation uses three multi-round file-based scenarios.

- **Orion engineering delivery (18 rounds).** This scenario simulates an engineering project handoff. The model must record project requirements, update structured files, recall earlier constraints, handle conflicts between new and old instructions, and decide when clarification is needed.
- **PipelineNova data pipeline (30 rounds).** This scenario simulates a data-pipeline maintenance task. The model must track dataset fields, transformation steps, validation requirements, configuration changes, and cross-file dependencies across multiple rounds.
- **AtlasReview research writing (30 rounds).** This scenario simulates a research-writing workflow. The model must organize notes, maintain citation-like records, update summaries, resolve inconsistencies, and preserve earlier writing constraints over a long interaction.

Each round requires the model to create or modify local files, and deterministic assertions evaluate file existence, format validity, required content, forbidden tool use, clarification behavior, and sandbox-boundary compliance.

Evaluation Metrics: We use deterministic, round-level checks to evaluate the LLM agent. **Round accuracy** is the fraction of task rounds in which all required assertions are passed. **Macro-average category accuracy** first computes accuracy within each behavior category and then averages across categories, so that frequent categories do not dominate the score. **Satisfaction rate** is the fraction of rounds in which the external state machine marks the recent performance as satisfactory, that is, $U(\ell) \geq A$. **Rework rate** is the fraction of rounds requiring correction, repeated execution, or recovery from an earlier error. **Tool-use accuracy** measures whether the selected tool and its arguments satisfy the expected tool-use constraints.

For additional task-level measures, **task completion** summarizes the overall completion of the multi-round scenario, **requirement coverage** measures the fraction of specified requirements satisfied, and **artifact completion** measures whether the expected output files or structured artifacts are produced and remain valid. These measures are descriptive diagnostics computed by the evaluation scripts.

Seven-Condition Evaluation: Seven conditions are compared in a paired design. Here, the *external mechanism* refers to executable modules outside the language model, including finite working memory, the lightweight hypergraph, and the satisfaction state machine.

- **C0:** base model only; no mechanism prompt and no external mechanism.
- **C1:** base model with mechanism prompt only; no executable external mechanism.

2026 S.-T. Yau High School Science Award 应用于2026丘成桐中学科学奖论文公示

- **C2:** base model with the full external mechanism.
- **C3:** LoRA-980 adapter with mechanism prompt only; no executable external mechanism.
- **C4:** LoRA-980 adapter with the full external mechanism.
- **C5:** base model with the external mechanism but without the lightweight hypergraph.
- **C6:** base model with the external mechanism but without the satisfaction state machine.

Across all conditions, each evaluation round allows at most 4 tool steps, and the model receives only the most recent 4 messages.

Table IX reports the seven-condition evaluation results. The improvement from C0 to C1 shows that an explicit mechanism prompt has a substantial practical effect (round accuracy $0.196 \rightarrow 0.578$; macro-average category accuracy $0.152 \rightarrow 0.520$). The further improvement from C1 to C2 shows that an executable external mechanism is more effective than an empty tool interface (round accuracy $0.578 \rightarrow 0.633$; macro accuracy $0.520 \rightarrow 0.560$).

TABLE IX: Seven-condition evaluation results.

Condition	Valid runs	Round accuracy	Macro-avg category accuracy	Satisfaction rate
C0 (base, no mechanism)	9	0.1963	0.1523	0.8333
C1 (prompt only)	9	0.5778	0.5203	0.8630
C2 (base + external)	9	0.6333	0.5603	0.8704
C3 (LoRA-980, prompt only)	8	0.5958	0.6214	0.9167
C4 (LoRA-980 + external)	9	0.6333	0.6818	0.9556
C5 (external without hypergraph)	9	0.5070	0.4780	0.9440
C6 (external without state machine)	9	0.6110	0.5480	1.0000*

*For C6, the satisfaction rate is degenerate because the satisfaction state machine is removed.

LLM Component Ablation: Table X reports the component ablation on the untrained base model. Removing the hypergraph (C5) reduces the macro-average category accuracy from 0.560 to 0.478, while removing the satisfaction state machine (C6) reduces it only from 0.560 to 0.548. The larger observed macro-accuracy decrease occurs when the hypergraph is removed. An independent stability benefit of the state machine is not established: rework is 0.545 with the full mechanism, compared with 0.533 without the state machine and 0.459 without the hypergraph.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

TABLE X: Component ablation of the external mechanism (untrained base model).

Condition	Macro accuracy	Satisfaction rate	Rework rate
C2 (full external)	0.560	0.870	0.545
C5 (without hypergraph)	0.478	0.944	0.459
C6 (without state machine)	0.548	1.000*	0.533

*The satisfaction rate of C6 is a degenerate measurement: removing the state machine makes satisfaction permanently true, so this value does not indicate better performance.

LLM Parameter Calibration: We calibrate the reward memory length L on the development scenario (Orion, 18 rounds). The default $L = 2$ with weights $(0.8, 0.2)$ achieves an accuracy of 0.667. An $L = 3$ configuration with equal weights $(1/3, 1/3, 1/3)$ achieves 0.778, the highest among all tested configurations. Non-uniform $L = 3$ weights (e.g., $0.6, 0.25, 0.15$) perform worse (0.611), indicating that the gain from a longer memory depends on equal weighting across periods. This calibration is exploratory and should be validated on independent test scenarios.

Online Demo Interface: We provide an interactive web demo as supplementary material. It visualizes the three-layer architecture, the satisfaction-switching trajectory, the hypergraph spreading-activation process, the seven-condition comparison, the parameter-calibration sweep, and an additional website-building task. These demo views are used for transparency and illustration; the main empirical claims in the paper rely on the seven-condition comparison, the component ablation, and the parameter-calibration result reported above. The demo is available at:

- **Demo HTML:** `reproducibility_code/llm/demo/html/finmem_demo.html`
- **Demo data:** `reproducibility_code/llm/demo/html/demo_data.js`

All numbers in the demo are drawn from the same deterministic result files used in this paper, including final_task_completion.json, final_cooperation_metrics.json, and five_condition_report.json, so the demo and the paper are consistent.

E. Code Reproducibility

The accompanying code package contains the theoretical model, Monte Carlo simulations, extensions, LLM training and evaluation assets, and figure-generation scripts. It is organized in a top-level folder named `reproducibility_code/`. Its `README.md` maps the research components to their entry points; `SOURCE_MANIFEST.json` records file provenance and `verify_package.py` runs CPU-only consistency checks. The theoretical programs are implemented in Python in `theory/`. The main files are:

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

- `paper_model.py`: the model kernel, including state ordering, transition matrix construction, and metric computation.
- `run_all_research.py`: runs the RQ2 scan, zero-error analysis, Monte Carlo validation, longer-memory extension, and adaptive aspiration analysis.
- `run_complete_research.py`: regenerates the complete figure and data family used in the manuscript, including the long-run, finite-time, convergence, error-scan, memory-extension, and adaptive-aspiration outputs.
- `tests/test_paper_model.py`: unit checks for state enumeration, transition matrices, stationary distributions, and simulation consistency.

The preserved theoretical outputs are stored in `outputs/theory_data/` as CSV, NPZ, and JSON metadata files, and in `outputs/theory_figures/` as PNG figures. The recommended reproducibility sequence is: first run `python3 reproducibility_code/verify_package.py` from the paper folder; then, if `pytest` is available, run `python3 -m pytest tests/test_paper_model.py` inside `reproducibility_code/theory/`; finally run `run_complete_research.py` to regenerate the full theoretical output family. Historical plotting and exploration scripts are retained in `theory/historical_scripts/` as provenance, but the canonical theory entry points are the files directly under `theory/`.

The LLM experiment is stored under `llm/`. The corresponding reproducibility files are:

- `llm/finmem/`: local finite-memory, hypergraph, task-completion, cooperation-metric, synthetic-data, and multi-round evaluation scripts, including `generate_pilot_data.py`, `oracle_replay_check.py`, `validate_pilot_data.py`, and `run_multiturn_eval.py`.
- `llm/configs/`: QLoRA and inference configuration files.
- `llm/autodl_scripts/`: server-side scripts used for historical training, evaluation, ablation reporting, and diagnostic checks.
- `llm/reports/`: preserved aggregate reports, including `final_task_completion.json`, `final_cooperation_metrics.json`, and `five_condition_report.json`.
- `llm/demo/`: paper-facing demo assets. The formal public scenarios are stored in `llm/demo/scenarios/`, and the browser-demo HTML and data bundle are stored in `llm/demo/html/` and `llm/demo/web_data/demo_data.js`.
- `llm/figures_data/` and `llm/reproduce_llm_figures.py`: preserved figure data and the script for regenerating the LLM summary figures.

Full LLM reruns require the base model, a compatible GPU software environment, and trained adapters; tensor weights are not included in the code package. The package includes aggregate reports and figure data, so the reported tables and figures can be audited. Independent full reruns of every historical LLM condition require the model weights and evaluation environment, and may differ slightly because of model-serving and hardware differences.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

F. Supplementary Figures

The following figures are placed in the appendix because they are either complete versions of figures shown in the main text or auxiliary diagnostics.

$\alpha - \varepsilon$ Cooperation Rate Phase Graphs: Figure 25 reports the $\alpha - \varepsilon$ cooperation rate phase graph for $A = 2$.

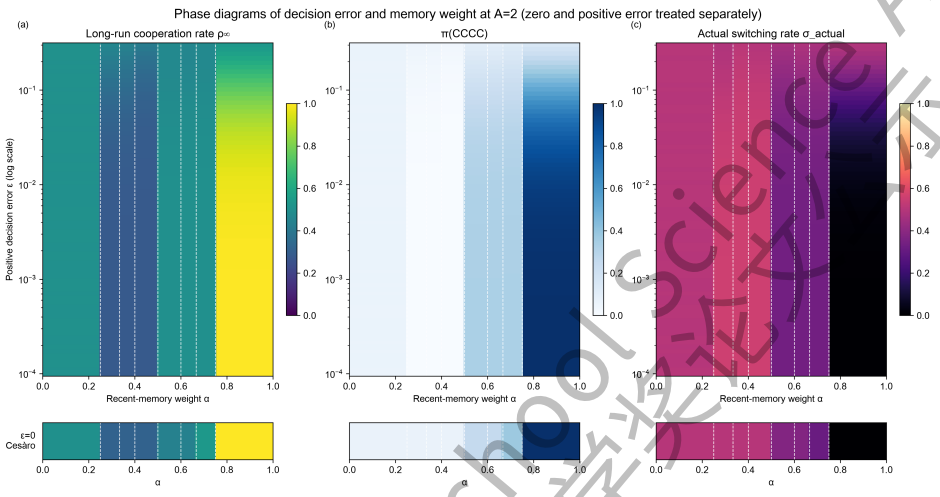


Fig. 25: $\alpha - \varepsilon$ phase graph of ρ_∞, $\pi(CCCC)$, and σ_{actual}.

Training Loss Curve: Figure 26 reports the training loss for the 280-sample pilot training.

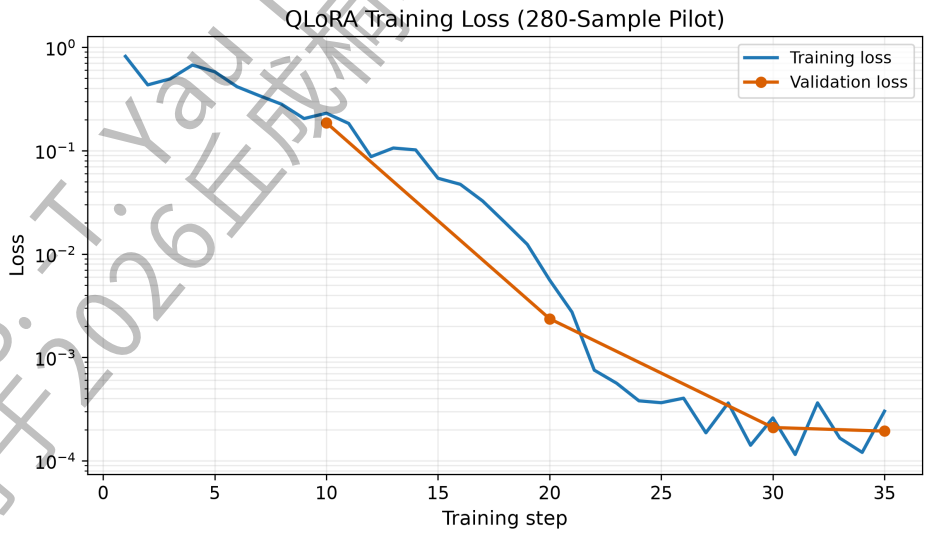


Fig. 26: Training loss for the 280-sample pilot training.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文评

Data Scaling Comparison: Figure 27 reports the comparison between the base model, LoRA-980, and LoRA-280.

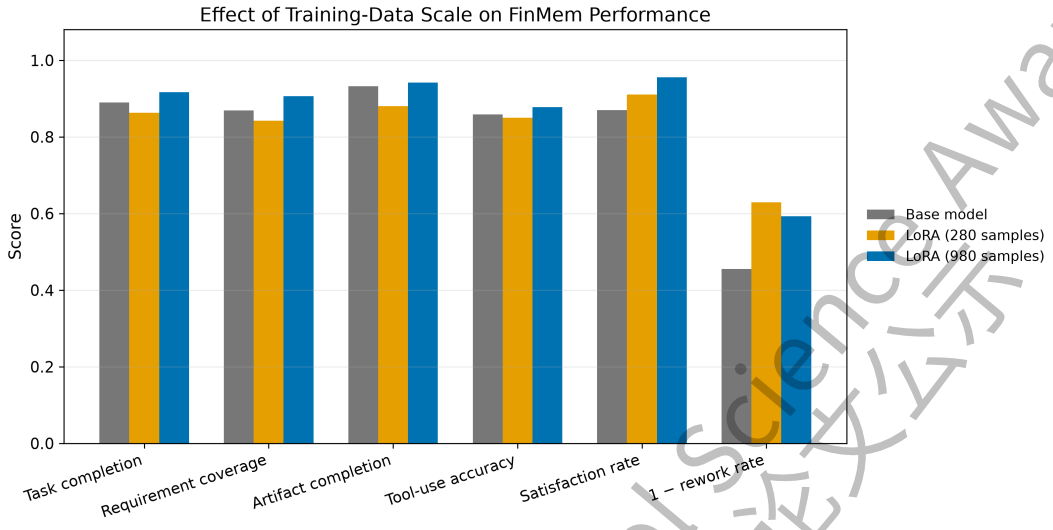


Fig. 27: Comparison of the base model, LoRA-980, and LoRA-280 on task completion, requirement coverage, artifact completion, tool-use accuracy, satisfaction rate, and one minus rework rate (higher is better).

Parameter Calibration: Memory Length Sweep: Figure 28 reports the memory length sweep on the development scenario (Orion, 18 rounds).

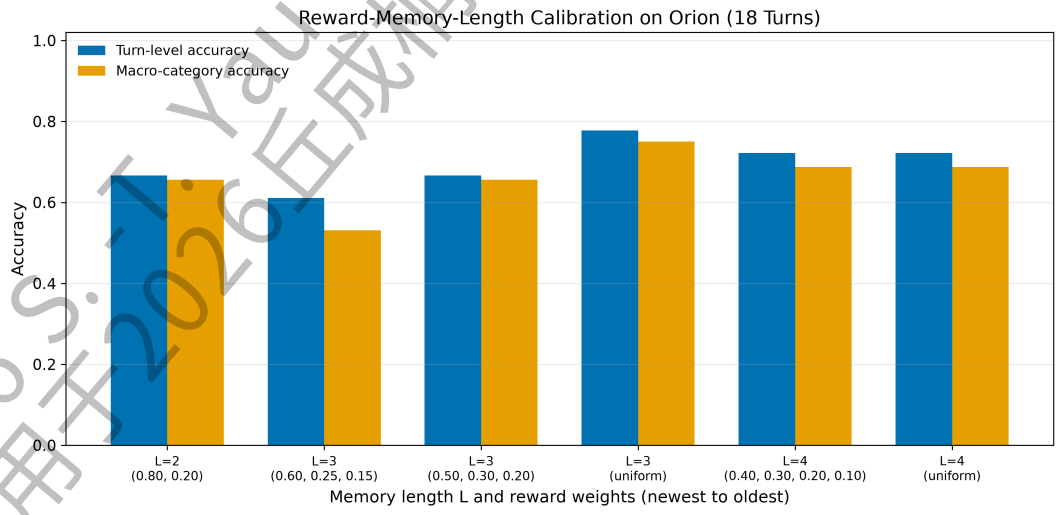


Fig. 28: Memory length L sweep on the development scenario. $L = 3$ with equal weights achieves the highest accuracy (0.778), compared with 0.667 for the default $L = 2$.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Website-Building Task Comparison: Figure 29 reports performance on the 20-round website-building task for the base model and the trained models.

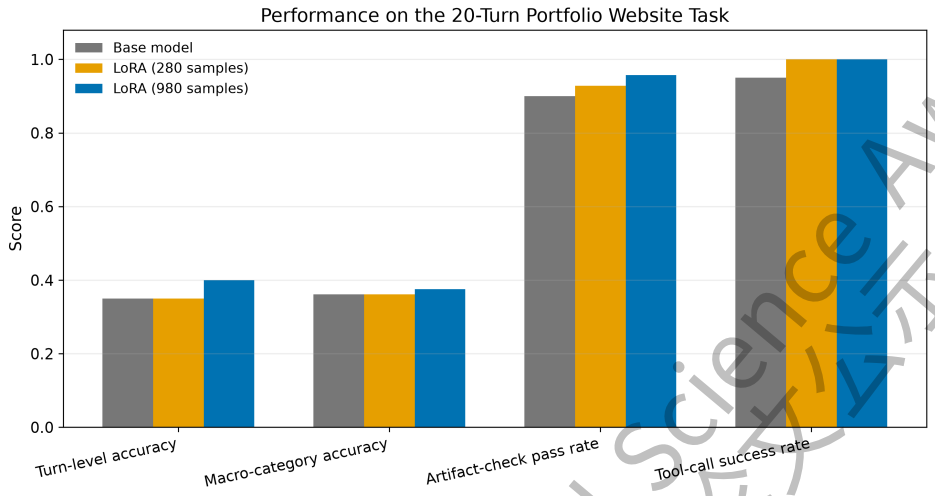


Fig. 29: Performance on the 20-round website-building task. This task is illustrative and does not support statistical conclusions.

REFERENCES

[1] R. Axelrod, *The Evolution of Cooperation*. New York, NY, USA: Basic Books, 1984.

[2] D. Fudenberg and E. Maskin, "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, vol. 54, no. 3, pp. 533–554, 1986, doi: 10.2307/1911307.

[3] M. A. Nowak and K. Sigmund, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game," *Nature*, vol. 364, pp. 56–58, 1993, doi: 10.1038/364056a0.

[4] W. H. Press and F. J. Dyson, "Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent," *Proceedings of the National Academy of Sciences*, vol. 109, no. 26, pp. 10409–10413, 2012, doi: 10.1073/pnas.1206569109.

[5] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99–118, 1955, doi: 10.2307/1884852.

[6] R. Karandikar, D. Mookherjee, D. Ray, and F. Vega-Redondo, "Evolving aspirations and cooperation," *Journal of Economic Theory*, vol. 80, no. 2, pp. 292–331, 1998, doi: 10.1006/jeth.1997.2379.

[7] J. Bendor, D. Mookherjee, and D. Ray, "Aspiration-based reinforcement learning in repeated interaction games: An overview," *International Game Theory Review*, vol. 3, nos. 2–3, pp. 159–174, 2001, doi: 10.1142/S0219198901000348.

[8] M. Posch, A. Pichler, and K. Sigmund, "The efficiency of adapting aspiration levels," *Proceedings of the Royal Society B: Biological Sciences*, vol. 266, no. 1427, pp. 1427–1435, 1999, doi: 10.1098/rspb.1999.0797.

[9] M. W. Macy and A. Flache, "Learning dynamics in social dilemmas," *Proceedings of the National Academy of Sciences*, vol. 99, suppl. 3, pp. 7229–7236, 2002, doi: 10.1073/pnas.092080099.

[10] C. Camerer and T.-H. Ho, "Experience-weighted attraction learning in normal form games," *Econometrica*, vol. 67, no. 4, pp. 827–874, 1999, doi: 10.1111/1468-0262.00054.

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

[11] D. Monte, "Learning with bounded memory in games," *Games and Economic Behavior*, vol. 87, pp. 204–223, 2014, doi: 10.1016/j.geb.2014.03.005.

[12] S. K. Baek, H.-C. Jeong, C. Hilbe, and M. A. Nowak, "Comparing reactive and memory-one strategies of direct reciprocity," *Scientific Reports*, vol. 6, Art. no. 25676, 2016, doi: 10.1038/srep25676.

[13] N. E. Glynatsi *et al.*, "Conditional cooperation with longer memory," *Proceedings of the National Academy of Sciences*, vol. 121, no. 50, Art. no. e2420125121, 2024, doi: 10.1073/pnas.2420125121.

[14] F. Lesigang, C. Hilbe, and N. E. Glynatsi, "Can I afford to remember less than you? Best responses in repeated additive games," *Economics Letters*, vol. 250, Art. no. 112300, 2025, doi: 10.1016/j.econlet.2025.112300.

[15] H. P. Young, "The evolution of conventions," *Econometrica*, vol. 61, no. 1, pp. 57–84, 1993, doi: 10.2307/2951778.

[16] P. Dal Bó and G. R. Fréchette, "The evolution of cooperation in infinitely repeated games: Experimental evidence," *American Economic Review*, vol. 101, no. 1, pp. 411–429, 2011, doi: 10.1257/aer.101.1.411.

[17] G. Baker, R. Gibbons, and K. J. Murphy, "Relational contracts and the theory of the firm," *The Quarterly Journal of Economics*, vol. 117, no. 1, pp. 39–84, 2002, doi: 10.1162/003355302753399445.

[18] J. Levin, "Relational incentive contracts," *American Economic Review*, vol. 93, no. 3, pp. 835–857, 2003, doi: 10.1257/000282803322157115.

[19] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing," *Psychological Review*, vol. 82, no. 6, pp. 407–428, 1975, doi: 10.1037/0033-295X.82.6.407.

[20] E. Tulving, "Episodic and semantic memory," in *Organization of Memory*, E. Tulving and W. Donaldson, Eds. New York, NY, USA: Academic Press, 1972, pp. 381–403.

[21] H. Ebbinghaus, *Memory: A Contribution to Experimental Psychology*, H. A. Ruger and C. E. Bussenius, Trans. New York, NY, USA: Teachers College, Columbia University, 1913. Original German edition published in 1885.

[22] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474.

[23] D. Edge *et al.*, "From local to global: A graph RAG approach to query-focused summarization," arXiv:2404.16130, 2024, doi: 10.48550/arXiv.2404.16130.

[24] S. Yao *et al.*, "ReAct: Synergizing reasoning and acting in language models," in *Proc. International Conference on Learning Representations (ICLR)*, 2023.

[25] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proc. 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, 2023, Art. no. 2, pp. 1–22, doi: 10.1145/3586183.3606763.

[26] N. Shinn *et al.*, "Reflexion: Language agents with verbal reinforcement learning," in *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[27] E. J. Hu *et al.*, "LoRA: Low-rank adaptation of large language models," in *Proc. International Conference on Learning Representations (ICLR)*, 2022.

[28] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient finetuning of quantized LLMs," in *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[29] Y. Zheng *et al.*, "LlamaFactory: Unified efficient fine-tuning of 100+ language models," in *Proc. 62nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2024, pp. 359–367, doi: 10.18653/v1/2024.acl-demos.38.

[30] H. Luo *et al.*, "Graph-R1: Towards agentic GraphRAG framework via end-to-end reinforcement learning," in *Proc. 43rd International Conference on Machine Learning (ICML)*, 2026.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

致谢页

由衷感谢指导老师王艺臻老师在本研究过程中给予的耐心指导，也感谢各位评委老师阅读本文。

本研究的选题源自王艺臻老师在校内经济课程中关于囚徒困境和重复博弈的讲解。通过课程学习，我了解到，在单轮囚徒困境中，完全理性的个体选择往往导致双方都无法获得理想结果；而在有限重复博弈中，逆向归纳法又使得合作难以稳定维持。然而，在现实经济活动中，长期商业合作、平台交易、供应链关系以及重复契约非常普遍；在人工智能应用场景中，用户与大语言模型之间、多个智能体之间的多轮互动和协作也十分常见。这促使我思考：如果放松完全理性和无限记忆的严格假设，在有限重复囚徒困境中，是否可能通过引入有限记忆机制和满意状态转换，实现稳定的合作?

王艺臻老师是我校内经济课的授课教师，此次指导为无偿学术指导。研究初期，我查阅了大量关于重复囚徒困境和有限记忆机制的文献，也尝试过机制设计和有限状态自动机等研究方法，但发现这些方法均存在一定局限。后来我与王老师进行了多次讨论，王老师建议我从最基础的两期记忆加权收益模型入手，引入满意阈值和决策误差，并利用马尔可夫链分析长期稳态和有限轮次的动态行为。在后续研究中，王老师主要在研究方向、模型合理性、经济学含义、论文结构和论文撰写等方面给予宝贵建议；具体的课题确定、研究问题细化、模型构建、公式推导、代码编写、实验进行、图表制作、文献整理和 LaTeX 格式调整和完整论文撰写由我完成。其中，代码调试、图表格式微调、文献整理、LaTeX 格式调整和英文语言润色在人工智能的辅助下完成。本研究中人工智能工具的使用事先获得了指导老师王艺臻老师的许可。

在理论建模部分，我负责建立有限记忆下的满意状态转换规则，推导记忆长度为 2 时的 16 状态马尔可夫链模型，分析长期稳态特性，包括计算长期合作率、CCCC 状态概率和实际转换率，同时进一步研究有限轮次下的瞬态动态。在数值实验部分，我编写程序进行参数扫描、蒙特卡洛模拟、记忆长度扩展和自适应满意阈值实验，以验证理论结果的稳健性。在应用拓展部分，我结合在校内学习的人工智能相关知识，将有限记忆与满意状态转换思想应用于大语言模型智能体训练，设计了外部记忆模块、轻量超图结构和满意状态机机制，训练了千问大模型如何调用这些机制，并整理了七组条件对比、组件消融和参数标定结果。

研究过程中遇到的主要困难有三点：第一，如何区分总博弈轮次与两期记忆形成后的状态转移次数。为解决这一问题，我在论文中明确区分，统一用 t 表示原始博弈轮次，用 n 表示记忆形成后的转移步数。第二，如何处理决策误差 $\varepsilon = 0$ 与 $\varepsilon > 0$ 的差异。由于 $\varepsilon > 0$ 时马尔可夫链具有唯一平稳分布，而 $\varepsilon = 0$ 时可能产生可约或周期结构，我在论文中分别讨论这两种情形，并增加了零误差与低正误差极限的对比分析。第三，如何判断“高平均合作率”是否真正代表稳定合作。为此，我除了计算长期合作率外，还引入 CCCC 状态概率和实际转换率，用以区分稳定互惠合作与频繁震荡行为。

本研究使用了 DeepSeek-V4-Flash 人工智能工具，使用时间为 2026 年 8 月，使用频率约为每周 1 到 2 次。AI 主要用于文献检索线索、LaTeX 环境配置和排版、英文表达润

2026 STS High School Student Science Award 仅用于2026年科普科学奖公示

色、代码调试、图表格式调整以及 Demo 界面与交互代码调试。部分文献整理和代码修改过程中，AI 曾提供修改建议。我对相关文献、代码和实验结果进行了逐项核验，并根据研究逻辑重新组织和修改。研究问题的确定、模型变量的选择、结果解释及最终结论由本人负责。

除王艺臻老师的无偿学术指导和上述 AI 工具辅助外，本研究未接受其他个人或机构在核心模型设计、数据分析、程序实现或论文撰写方面的参与。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示