参赛学生姓名：孙浩宸

中学：北京师范大学附属实验中学

省份：北京

国家/地区：中国/北方赛区

指导老师1姓名：马静

指导老师1单位：北京师范大学附属实验中学

指导老师2姓名：国萌

指导老师2单位：北京大学

论文题目：The Influence of the Shadow of the Future on the Evolution of Cooperative Strategies in Multi-Agent Systems Based on LLM Architecture in Repeated Games

Research Paper for the Computer Award
of the S.-T. Yau High School Science Award

The Influence of the Shadow of the Future on the Evolution of Cooperative Strategies in Multi-Agent Systems Based on LLM Architecture in Repeated Games

Author: Haochen Sun

Academic Supervisor: Jing Ma
External Supervisor: Meng Guo

November 6, 2025

Abstract

The emergence of Large Language Models (LLMs) as autonomous agents in strategic interactions necessitates a comprehensive understanding of their decision-making behavior through established game-theoretic frameworks. This study systematically investigates how the "shadow of the future"—a fundamental concept in repeated game theory—influences cooperative strategy evolution in LLM-based multi-agent systems. Through a rigorous experimental framework encompassing four distinct experiments across canonical game scenarios (Prisoner's Dilemma and Battle of the Sexes), we examine the effects of continuation probability disclosure, temporal structure variations, environmental dynamics, and strategic reasoning enhancement on cooperative outcomes. Our analysis of three state-of-the-art LLMs (ChatGPT, DeepSeek, and Kimi) reveals distinct behavioral "personalities": ChatGPT exhibits strategic adaptability with selective cooperation, Kimi demonstrates unconditional cooperativeness, and DeepSeek shows sophisticated conditional reciprocity. Counterintuitively, we find that explicit Strategic Chain-of-Thought reasoning fails to improve performance relative to baseline intuitive responses (59.7% vs 92.3% success rates), suggesting that reasoning complexity does not necessarily translate to strategic superiority. Furthermore, while cooperation remains robust under predictable environmental changes (99.2% cooperation rate), it becomes fragile under stochastic uncertainty. Information transparency effects prove context-dependent, enhancing coordination in asymmetric games while potentially enabling exploitation in social dilemmas. These findings establish foundational insights for deploying LLM agents in collaborative systems and highlight the critical importance of agent personality compatibility, environmental predictability, and reasoning mode selection in multi-agent strategic interactions.

Keywords: Large Language Models, Game Theory, Multi-Agent Systems, Cooperation, Shadow of the Future, Strategic Reasoning

Contents

1 Introduction

1.1 The Rise of LLMs as Strategic Agents in Multi-Agent Systems

The rapid evolution of large language models (LLMs) has fundamentally transformed artificial intelligence, demonstrating remarkable capabilities in complex reasoning [1], tool manipulation [2], and human-like interaction [3, 4]. This technological advancement has positioned LLMs as increasingly prevalent autonomous agents in strategic environments, from automated negotiation systems and algorithmic trading platforms to collaborative decision-support tools and multi-agent coordination systems [5, 6]. As these systems assume roles traditionally occupied by human decision-makers, the need for rigorous theoretical frameworks to understand their strategic behavior becomes paramount.

Game theory provides the mathematical foundation for analyzing strategic interactions among rational agents [7]. For decades, it has served as the cornerstone for understanding decision-making across economics, political science, and computer science [8, 9, 10], offering robust predictive models for agent behavior in scenarios ranging from coordination problems to complex negotiations [11, 12]. However, the emergence of LLM agents introduces novel computational characteristics that challenge traditional game-theoretic assumptions and necessitate new analytical approaches.

1.2 Current Limitations: The Need for Systematic Architecture

Existing research on LLM strategic behavior reveals significant methodological and theoretical gaps. While preliminary studies have documented instances of cooperation, reciprocity, and strategic deception in LLM interactions [13, 14], these findings remain fragmented and lack systematic grounding in established theoretical frameworks [15]. Current approaches largely consist of isolated experiments or anecdotal observations, preventing the development of comprehensive understanding.

More critically, empirical evidence demonstrates that LLMs exhibit inconsistent strategic performance, frequently deviating from equilibrium predictions without clear explanatory patterns [6, 16]. They display limited capacity for recursive reasoning about opponent beliefs—a fundamental requirement for sophisticated strategic thinking [17]—and demonstrate excessive sensitivity to contextual variations that is absent in human strategic behavior [18, 19].

The absence of a unified architectural framework for systematic evaluation has hindered progress in understanding when and why LLMs succeed or fail in strategic contexts. Existing multi-agent orchestration systems such as AutoGPT, CAMEL, and Voyager focus primarily on task execution and autonomous tool usage [14], rather than systematic strategic reasoning and game-theoretic evaluation. This gap necessitates a specialized architectural approach designed specifically for strategic interaction analysis.

1.3 Contributions: A Unified Multi-Agent Architecture for Game-Theoretic Evaluation

This paper introduces a novel multi-agent architecture specifically designed to bridge the gap between classical game theory and LLM strategic evaluation. Our approach represents the first unified framework dedicated to systematic assessment of LLM behavior in strategic environments, making several key architectural and methodological contributions:

Unified Multi-Agent Architecture: We propose a comprehensive system architecture that integrates hierarchical reasoning modules (Baseline, Chain-of-Thought, and Strategic Chain-of-Thought), memory-enhanced interaction mechanisms, and stochastic termination protocols. This architecture extends Empirical Game-Theoretic Analysis (EGTA) [20] to the LLM domain, providing a systematic platform for strategic behavior assessment rather than ad-hoc experimental approaches.

Hierarchical Cognitive Modules: Rather than treating reasoning approaches as mere prompt variations, our architecture conceptualizes Baseline, Chain-of-Thought (CoT), and Strategic Chain-of-Thought (SCoT) as distinct cognitive layers within a modular reasoning framework. This architectural design enables systematic evaluation of different reasoning depths and their impact on strategic performance, moving beyond simple prompt engineering to structured cognitive modeling.

EGTA-LLM Integration Framework: We introduce architectural innovations specifically designed to adapt Empirical Game-Theoretic Analysis for LLM environments, including: (1) probabilistic continuation mechanisms that implement the "Shadow of the Future" [21] in LLM contexts, (2) structured memory update protocols that maintain strategic context across interactions, and (3) cross-seed aggregation methods that account for LLM stochasticity in game-theoretic evaluation.

Paradigmatic Bridge: Our architecture serves as the first systematic connection between game-theoretic multi-agent systems and large language model evaluation, establishing a reusable platform that extends beyond specific games (Prisoner's Dilemma, Battle of the Sexes) to general strategic scenarios. This represents a paradigmatic shift from isolated LLM experiments to systematic strategic reasoning evaluation.

The proposed framework enables rigorous comparison of LLM strategic capabilities against established human behavioral patterns while providing mechanistic insights into the computational foundations of strategic decision-making in artificial agents. Through comprehensive evaluation across multiple model families and strategic environments, our architecture establishes new benchmarks for understanding and deploying LLM agents in strategic contexts, contributing to the development of more reliable and predictable artificial intelligence systems [22, 23].

This work represents a foundational contribution to the emerging field of strategic AI, providing both the theoretical framework and practical architecture necessary for systematic evaluation of LLM agents in game-theoretic contexts. By establishing this unified platform, we enable future research to build upon a solid architectural foundation rather than starting from isolated experimental approaches.

2 Related Work

2.1 Theoretical Foundations: Game Theory and Strategic Cooperation

The theoretical foundation of cooperation in multi-agent systems has been extensively studied through repeated game theory [11, 21]. Axelrod's seminal work [21] demonstrated that cooperation can emerge and persist through the "Shadow of the Future" mechanism, where future interaction prospects incentivize present cooperative behavior [24]. This concept is formally captured by the discount factor (δ) or continuation probability (p-value), representing the likelihood of future interactions [25].

The folk theorem establishes that cooperation becomes a Nash equilibrium when the discount factor exceeds a critical threshold, making long-term cooperation benefits outweigh short-term defection gains [26]. Higher continuation probabilities generally promote more stable cooperative outcomes [27, 28]. However, these theoretical predictions assume perfectly rational agents with complete information and unlimited computational capacity—assumptions that may not hold for modern AI systems [29, 30].

While these theoretical frameworks provide robust foundations for understanding strategic behavior, their application to LLM agents requires systematic architectural adaptations that can bridge classical game theory with the unique computational characteristics of large language models.

2.2 Multi-Agent Orchestration Systems: Task-Focused vs. Strategic-Focused Architectures

The landscape of multi-agent LLM systems has been dominated by frameworks designed for task execution and autonomous coordination. AutoGPT [31] pioneered autonomous task decomposition and execution, enabling LLM agents to interact with external tools and pursue complex objectives through iterative planning. CAMEL [32] introduced role-playing conversational frameworks that enable collaborative task completion through structured dialogue between specialized agents. Voyager [33] demonstrated sophisticated autonomous exploration and skill acquisition in complex virtual environments.

These systems excel at autonomous tool usage, task orchestration, and collaborative problem-solving, focusing primarily on achieving external objectives through agent coordination. Their architectures prioritize functionality, adaptability, and task completion efficiency. However, they are not specifically designed for systematic strategic reasoning evaluation or game-theoretic analysis [14].

Architectural Distinction: Unlike existing multi-agent orchestration systems that focus on task execution and autonomous coordination, our framework is purpose-built for strategic interaction analysis and game-theoretic evaluation. While AutoGPT, CAMEL, and Voyager optimize for task completion and environmental adaptation, our architecture is specifically designed to isolate, measure, and analyze strategic decision-making processes

under controlled theoretical conditions. This fundamental difference in design philosophy necessitates entirely different architectural approaches, evaluation metrics, and theoretical foundations.

2.3 LLM Strategic Behavior: From Anecdotal to Systematic Analysis

Recent research has begun exploring LLM cooperative behavior in multi-agent settings [13, 14]. Wu et al. (2024) [34] demonstrated that LLM agents can develop cooperative strategies in competitive environments without external incentives, revealing inherent capacity for strategic reasoning [1, 2]. However, this research primarily focuses on demonstrating whether spontaneous cooperation occurs, rather than systematically investigating underlying governing factors.

The work by Mozikov et al. (2024) [16] on emotional decision-making reveals that LLMs exhibit decision-making patterns deviating from purely rational calculations when subjected to emotional framing, suggesting behavioral characteristics analogous to human cognitive biases [35]. Most existing work focuses on externally induced biases through prompt engineering [36] or environmental manipulation.

Methodological Gap: Current approaches lack systematic architectural frameworks for rigorous strategic evaluation. Existing studies primarily employ ad-hoc experimental designs without unified theoretical grounding, preventing comprehensive understanding of when and why LLMs succeed or fail in strategic contexts. This methodological limitation has hindered progress toward reliable LLM deployment in strategic environments.

2.4 Empirical Game-Theoretic Analysis: From Traditional Agents to LLMs

Empirical Game-Theoretic Analysis (EGTA) represents the mature integration of game-theoretic principles with computational multi-agent systems. Lanctot et al. (2017) [23] established comprehensive frameworks for analyzing strategic interactions in complex environments, utilizing computational simulations to discover emergent strategies and meta-strategies [37]. Their methodology provides systematic approaches for understanding multi-agent behavior patterns that would be intractable through theoretical analysis alone [12].

The EGTA framework has proven valuable for analyzing systems where traditional analytical solutions are intractable due to complexity or non-standard agent capability assumptions [38]. However, this methodology has been primarily applied to reinforcement learning agents with well-defined utility functions and learning algorithms [39], rather than the stochastic, prompt-sensitive, and contextually dependent nature of LLM agents.

Architectural Innovation: This work represents the first systematic adaptation of EGTA methodology to LLM agents [40, 22], requiring novel architectural considerations for factors such as: (1) stochastic output generation and temperature sensitivity, (2) prompt engineering and reasoning paradigm integration [36, 41], (3) memory management and context maintenance across interactions, and (4) inherent model characteristics and emergent "personality" traits [1].

2.5 Reasoning Paradigms in Strategic Contexts: Beyond Prompt Engineering

The development of reasoning paradigms in LLMs has progressed from simple prompting to sophisticated multi-step approaches. Chain-of-Thought (CoT) reasoning [36] enables step-by-step problem decomposition, while Strategic Chain-of-Thought approaches [41] incorporate opponent modeling and recursive reasoning. However, existing work treats these approaches primarily as prompt engineering techniques rather than architectural components of strategic reasoning systems.

Current research focuses on demonstrating the effectiveness of various reasoning approaches in isolated contexts, without systematic integration into unified frameworks for strategic evaluation. The lack of architectural perspective prevents understanding of how different reasoning paradigms interact with strategic contexts and how they can be systematically leveraged for comprehensive behavioral analysis.

Architectural Contribution: Our framework reconceptualizes reasoning paradigms as hierarchical cognitive modules within a unified strategic analysis architecture. Rather than treating Baseline, CoT, and SCoT as mere prompt variations, we integrate them as distinct cognitive layers within a modular reasoning framework that enables systematic evaluation of reasoning depth effects on strategic performance. This architectural approach provides the foundation for mechanistic understanding of how different cognitive processes influence strategic decision-making in LLM agents.

2.6 Positioning: First Unified Architecture for LLM Game-Theoretic Evaluation

While existing research has made valuable contributions to understanding LLM behavior in strategic contexts, the field lacks a unified architectural framework that systematically bridges classical game theory with modern language model capabilities. Current approaches remain fragmented across task-focused orchestration systems, anecdotal strategic behavior studies, and isolated reasoning paradigm investigations.

Our work addresses this fundamental gap by introducing the first comprehensive architecture specifically designed for systematic game-theoretic evaluation of LLM agents. This architecture integrates established theoretical foundations with novel adaptations necessary for LLM characteristics, providing a reusable platform for rigorous strategic behavior analysis that extends beyond specific games to general strategic scenarios.

This architectural contribution represents a paradigmatic shift from isolated experimental approaches to systematic strategic reasoning evaluation, establishing the foundation for reliable understanding and deployment of LLM agents in strategic contexts.

3 Methodology

3.1 Unified Multi-Agent Architecture for Strategic Game-Theoretic Evaluation

We propose a novel architectural framework specifically designed for evaluating Large Language Model (LLM) strategic reasoning within game-theoretic contexts. This unified architecture extends Empirical Game-Theoretic Analysis (EGTA) to the LLM domain through the integration of hierarchical reasoning modules, memory-enabled interaction mechanisms, and probabilistic termination protocols. Unlike existing multi-agent orchestration systems focused on task execution and autonomy (e.g., AutoGPT, CAMEL, Voyager), our architecture is purpose-built for systematic evaluation of strategic reasoning capabilities across different cognitive processing depths.

The architectural framework comprises four core components: (1) Modular Cognitive Processing Layers that enable systematic evaluation of reasoning capabilities, (2) Strategic Game Environment Management that handles payoff computation and state transitions, (3) Adaptive Memory Systems that maintain interaction histories and enable strategic learning, and (4) Probabilistic Termination Mechanisms that implement both finite and infinite-horizon conditions essential for game-theoretic analysis.

Table 3.1 contrasts our framework with existing multi-agent systems, highlighting the architectural innovations required for game-theoretic evaluation.

Table 3.1: Architectural Comparison with Existing Multi-Agent Systems

System	Primary Focus	Reasoning Modules	Game-Theoretic Support
AutoGPT	Task execution	Single-layer	Limited
CAMEL	Role-playing dialogue	Context-dependent	No
Voyager	Environment exploration	Goal-oriented	No
Our Framework	Strategic evaluation	Hierarchical (3-tier)	Full support

3.2 Game Environment Architecture

The strategic evaluation platform supports two canonical game-theoretic environments that serve as representative testbeds for understanding how architectural components influence strategic behavior. The environment management system implements standardized payoff computation, state transition protocols, and outcome recording mechanisms that ensure consistency across all experimental conditions.

The Prisoner's Dilemma (PD) environment implements the classic social dilemma structure where agents face the fundamental tension between individual rationality and collective optimality. The symmetric payoff matrix, presented in Table 3.2, creates the canonical cooperation-defection dynamics essential for evaluating strategic reasoning under conflicting incentives.

Table 3.2: Prisoner's Dilemma Payoff Matrix

Agent 1 \ Agent 2	**Option J**	**Option F**
Option J	$(8, 8)$	$(0, 10)$
Option F	$(10, 0)$	$(5, 5)$

The Battle of the Sexes (BoS) environment addresses coordination challenges with preference conflicts over equilibrium selection. The asymmetric coordination structure, shown in Table 3.3, requires agents to simultaneously solve coordination problems and equilibrium selection challenges, providing insight into architectural performance under strategic interdependence.

Table 3.3: Battle of the Sexes Payoff Matrix

Agent 1 \ Agent 2	**Option J**	**Option F**
Option J	$(10, 7)$	$(0, 0)$
Option F	$(0, 0)$	$(7, 10)$

Both environments integrate seamlessly with the temporal structure components of our architecture, enabling systematic evaluation of how continuation probabilities and finite horizons influence strategic behavior across different reasoning modules.

3.3 Hierarchical Cognitive Processing Architecture

The core architectural innovation lies in the modular cognitive processing system that systematically varies reasoning depth while maintaining consistent interface protocols. These cognitive modules represent distinct architectural layers rather than simple prompt variations, enabling systematic evaluation of how reasoning sophistication affects strategic performance.

Table 3.4 presents the hierarchical specification of cognitive processing layers integrated within the architectural framework.

Table 3.4: Hierarchical Cognitive Processing Modules

Module	**Processing Layer**	**Architectural Function**
Baseline	Direct decision mapping	Natural reasoning baseline
CoT	Structured trace generation	Explicit reasoning scaffolding
SCoT	Game-theoretic analysis	Strategic reasoning integration

The **Baseline Module** serves as the foundational cognitive processing layer, capturing agents' natural strategic inclinations without additional analytical scaffolding. This module establishes the architectural baseline by processing game state information through the model's inherent reasoning capabilities, providing insight into default strategic behavior patterns embedded within pre-trained LLM architectures.

The **Chain-of-Thought (CoT) Module** implements structured reasoning protocols that require explicit trace generation before decision execution. This architectural layer systematically enhances cognitive processing through step-by-step analytical frameworks, following established protocols for improving LLM performance through structured reasoning pathways. The module integrates seamlessly with the broader architecture while maintaining consistency across different temporal conditions and agent configurations.

The **Strategic Chain-of-Thought (SCoT) Module** represents the most sophisticated cognitive processing layer, incorporating explicit game-theoretic analytical frameworks into the reasoning architecture. This module integrates Nash equilibrium identification, best response analysis, opponent modeling protocols, and strategic forecasting mechanisms directly into the cognitive processing pipeline. The SCoT architecture enables systematic evaluation of how game-theoretic reasoning sophistication affects strategic performance across different temporal structures.

The modular design enables flexible architectural configurations where reasoning modules can be systematically replaced or combined to evaluate different cognitive processing approaches. This architectural flexibility is essential for systematic EGTA evaluation, allowing researchers to isolate the effects of reasoning sophistication while controlling for other experimental variables.

3.4 LLM Agent Integration and Configuration

Our architectural framework supports integration of diverse LLM agents representing different training paradigms and architectural approaches. The agent integration system maintains consistent interface protocols while preserving the unique characteristics of different model architectures, enabling comparative evaluation across varied LLM capabilities.

The framework integrates three representative LLM architectures: ChatGPT (GPT-4) representing OpenAI's RLHF methodology, DeepSeek-V3 representing reasoning-focused training approaches, and Kimi-Large representing long-context and multilingual optimization paradigms. This diverse agent portfolio enables systematic evaluation of how different training approaches interact with our cognitive processing architecture.

Table 3.5 presents the standardized configuration parameters that ensure methodological rigor across all agent types and experimental conditions.

Table 3.5: Unified Agent Configuration Parameters

Parameter	Value
Temperature	0.7
Top-p	0.9
Max tokens	500
Random seed management	Systematic variation
API timeout	30 seconds
Retry attempts	3
Model versions	GPT-4-turbo-2024-04-09, DeepSeek-V3-API, Kimi-Large-API

Agent assignment protocols implement systematic role counterbalancing across experimental conditions to control for potential systematic biases. Each agent receives consistent role identities (Agent 1 or Agent 2) throughout experimental sessions, with assignments rotated systematically across conditions to ensure that observed strategic differences reflect genuine architectural effects rather than role-specific artifacts.

3.5 Temporal Structure Implementation and EGTA Adaptation

A critical architectural innovation involves the systematic implementation of temporal structures essential for game-theoretic analysis. Our framework incorporates both

finite and infinite-horizon mechanisms that enable rigorous evaluation of how temporal parameters interact with cognitive processing modules.

The continuation probability mechanism (p-value) implements the "shadow of the future" concept through probabilistic termination protocols. This architectural component communicates temporal parameters through structured interfaces that explicitly convey both numerical probabilities and their strategic implications for long-term optimization. The implementation employs cryptographically secure random number generation with systematic seed management to ensure genuine uncertainty while maintaining experimental reproducibility.

The finite horizon implementation (T-value) creates common knowledge of predetermined interaction lengths, enabling evaluation of backward induction reasoning within the architectural framework. This component provides perfect information about interaction timelines while maintaining consistency with game-theoretic assumptions about rational strategic behavior.

Table 3.6 details the temporal structure implementation within the architectural framework.

Table 3.6: Temporal Structure Implementation Specifications

Component	Implementation	Architectural Integration
p-value	Probabilistic continuation (0.1, 0.5, 0.9)	Structured prompt interfaces with strategic implications
T-value	Deterministic termination (5, 10, 20 rounds)	Common knowledge protocols with perfect information
Termination Protocol	Secure random generation (Seed range: 0-10)	Systematic seed management for reproducibility

3.6 Architectural System Integration

Figure 3.1 illustrates the comprehensive architectural integration that orchestrates all system components within the unified framework. The architecture demonstrates the systematic flow from input processing through cognitive modules to decision execution and memory integration, highlighting the modular design that enables systematic EGTA evaluation.

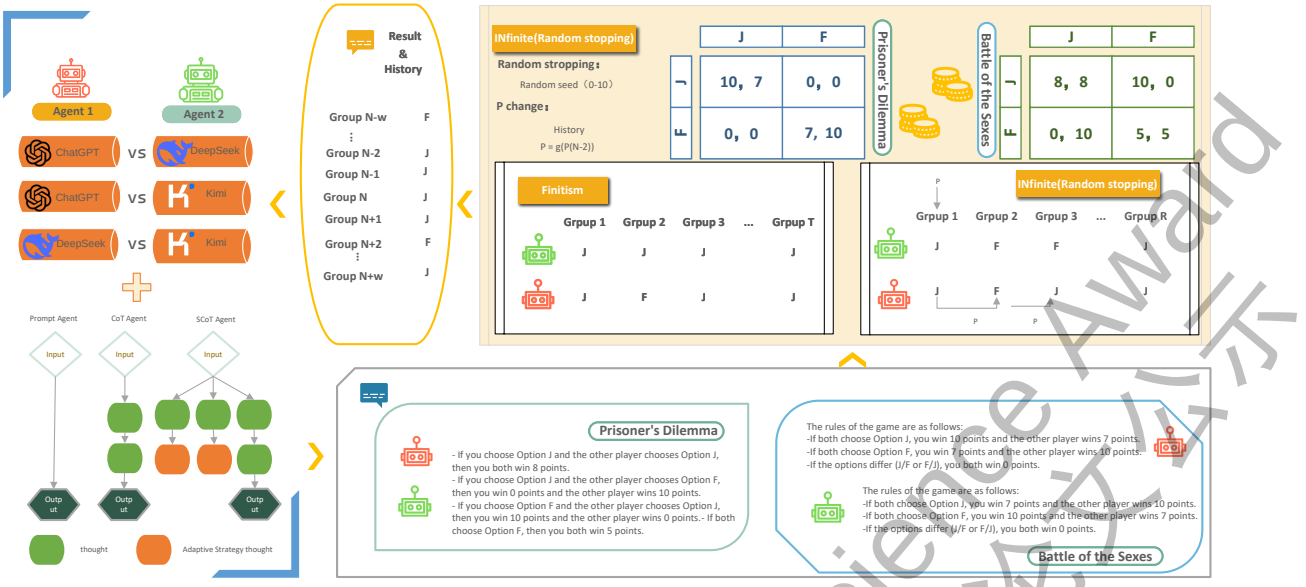

Figure 3.1: Multi-agent interaction framework showing game environments, reasoning modes, agent configurations, and experimental flow. The diagram illustrates how different LLM agents (ChatGPT, DeepSeek, Kimi) interact through various reasoning modes (Baseline, CoT, SCoT) in both Prisoner's Dilemma and Battle of the Sexes environments under different temporal structures.

The architectural framework integrates LangChain for agent orchestration and AutoGen for sophisticated dialogue management, providing the technological foundation required for complex multi-turn strategic reasoning. This integration supports persistent memory systems, stateful interactions, and dynamic adaptation capabilities essential for repeated game scenarios while maintaining rigorous experimental control.

The agent initialization protocols establish foundational conditions through comprehensive role assignment systems that include unique identifiers, game rule specifications, and temporal parameter communication. The architectural design ensures that agents receive identical factual information while cognitive processing differences are controlled exclusively through the modular reasoning system.

The interaction protocol implements structured sequences that maintain consistency across experimental conditions. Both agents simultaneously receive comprehensive game state information including complete interaction histories, outcome records, and remaining parameters through standardized interfaces. This shared information environment enables strategic learning while preserving game-theoretic common knowledge assumptions.

Memory management protocols ensure comprehensive historical information accessibility while preventing information leakage that could compromise experimental validity. The architecture implements automatic payoff computation, state updating, and comprehensive data recording systems that capture multiple dimensions of strategic behavior and interaction dynamics.

3.7 Experimental Variables and Systematic Evaluation Framework

The architectural framework enables systematic manipulation of key variables essential for comprehensive EGTA evaluation. Table 3.7 presents the complete variable specification and operationalization methods integrated within the architectural system.

Table 3.7: Architectural Variable Specification and Operationalization

Variable Category	Specification	Architectural Implementation
Temporal Structure	p-value (0.1, 0.5, 0.9)	Probabilistic termination protocols
	T-value (5, 10, 20)	Deterministic horizon management
Cognitive Architecture	Baseline/CoT/SCoT	Hierarchical processing modules
Agent Configuration	GPT-4/DeepSeek/Kimi	Standardized integration interfaces
Strategic Metrics	Cooperation rates	Systematic behavioral measurement
	Payoff efficiency	Automated calculation systems
	Equilibrium convergence	Dynamic analysis protocols

The architectural design supports comprehensive data collection protocols that capture primary behavioral measures including individual action sequences, joint outcome patterns, and cumulative payoff trajectories. For structured reasoning modules (CoT and SCoT), complete reasoning traces are recorded to enable analysis of decision-making processes and strategic reasoning quality assessment.

Statistical analysis protocols are integrated within the architectural framework, incorporating both parametric and non-parametric approaches to accommodate diverse distributional characteristics of strategic choice data. The system implements parallel processing architectures that maintain strict independence between experimental runs while achieving computational efficiency essential for large-scale EGTA evaluation.

This unified architectural framework represents the first systematic platform specifically designed for game-theoretic evaluation of LLM strategic reasoning capabilities, providing a reusable foundation that extends beyond the current experimental domains to support broader strategic evaluation scenarios.

4 Experimental Design

4.1 Architectural Validation Framework

All experiments are implemented within and serve as empirical validation of our proposed unified multi-agent architecture, where hierarchical reasoning modules, probabilistic termination mechanisms, and EGTA evaluation protocols constitute inseparable components of the system. Rather than isolated case studies, these experiments systematically validate the architectural design's capability to capture and evaluate strategic reasoning patterns across different temporal structures, cognitive processing depths, and environmental dynamics.

The experimental framework leverages the modular architecture's flexibility to systematically manipulate cognitive processing layers (Baseline, CoT, SCoT) while maintaining consistent game environment management and temporal structure implementation. This approach enables rigorous assessment of how architectural components interact to influence strategic behavior, providing comprehensive validation of the framework's effectiveness for game-theoretic evaluation of LLM capabilities.

Table 4.1 presents the systematic experimental validation matrix that demonstrates how each experiment targets specific architectural components while maintaining overall system integrity.

Table 4.1: Architectural Component Validation Matrix

Experiment	Cognitive Modules	Temporal Structure	Environmental Dynamics	Agent Integration System
Experiment 1	Fixed (Baseline)	p-value variation	Static	Full validation
Experiment 2	Fixed (Baseline)	Finite vs. Infinite	Static	Full validation
Experiment 3	Fixed (Baseline)	Dynamic p-values	Trajectory-based	Full validation
Experiment 4	Module comparison	Fixed conditions	Static	Module validation

4.2 Agent Architecture Integration and Implementation

The experimental validation employs three distinct LLM architectures systematically integrated within our unified framework to demonstrate robustness and generalizability across different AI systems with varying baseline strategic reasoning capabilities [2, 1]. Each agent type represents a different training paradigm, enabling comprehensive evaluation of how the architectural framework performs across diverse LLM approaches.

Table 4.2 specifies the systematic agent integration protocols within the architectural framework.

Table 4.2: Agent Architecture Integration Specifications

Agent Type	Training Paradigm	Strategic Baseline	Architectural Role
ChatGPT (GPT-4)	RLHF methodology	High cooperation tendency	Conversational AI validation
DeepSeek-V3	Reasoning-focused	Moderate cooperation	Analytical capability testing
Kimi-Large	Long-context optimization	Variable cooperation	Context-dependent evaluation

All agents operate under identical computational constraints within the architectural framework and receive standardized interface protocols [36, 39] to minimize confounding variables while preserving their inherent behavioral characteristics. Each agent processes game information through the unified cognitive processing modules and makes decisions based on the systematic interaction protocols implemented within the architectural framework [22].

4.3 Game Environment Validation Scenarios

The architectural framework validation employs two canonical game-theoretic scenarios that systematically test different aspects of strategic interaction processing within the cognitive modules and temporal structure components [42]. These scenarios serve as standardized benchmarks for evaluating architectural performance across diverse strategic contexts.

4.3.1 Prisoner's Dilemma Architecture Validation

The Prisoner's Dilemma implementation within the architectural framework tests the system's ability to handle fundamental social dilemmas where individual rationality conflicts with collective optimality [43, 21]. The standardized payoff matrix integrates seamlessly with the cognitive processing modules and temporal structure components:

- **Mutual Cooperation (J, J):** Architectural reward of 8 points per agent, testing system's ability to identify and maintain socially optimal outcomes
- **Mutual Defection (F, F):** Architectural penalty of 5 points per agent, validating Nash equilibrium detection capabilities [8]
- **Unilateral Defection (F, J) or (J, F):** Asymmetric outcomes (10, 0) testing system's handling of exploitation dynamics [44]

This payoff structure validates the architectural framework's ability to process the standard PD conditions: $T > R > P > S$ while systematically evaluating how cognitive modules and temporal structures influence strategic reasoning patterns [11].

4.3.2 Battle of the Sexes Architecture Validation

The Battle of the Sexes implementation tests the architectural framework's coordination capabilities with asymmetric preferences, validating the system's ability to handle complex equilibrium selection challenges [9]. The coordination structure validates different architectural components:

- **Successful Coordination:** (10, 7) and (7, 10) outcomes test the system's equilibrium identification and preference handling
- **Coordination Failure:** (0, 0) outcomes validate the architecture's ability to recognize and avoid miscoordination penalties
- **Preference Asymmetry:** Different agent preferences test the framework's handling of strategic interdependence

This structure systematically validates the architectural framework's performance in real-world coordination scenarios where parties must align their actions despite conflicting preferences [45].

4.4 Baselines

The systematic evaluation of our architectural framework requires comprehensive baseline comparisons that establish performance benchmarks across theoretical, empirical, and computational dimensions. These baselines provide essential reference points for interpreting experimental results and validating the architectural framework's strategic reasoning capabilities relative to established standards in game theory, behavioral economics, and artificial intelligence research.

The selection of baselines follows a principled hierarchy that progresses from theoretical foundations through empirical human behavior to state-of-the-art computational approaches. This multi-layered comparison framework ensures that our architectural evaluation captures both normative predictions from economic theory and descriptive patterns from behavioral research, while maintaining alignment with current developments in LLM strategic reasoning evaluation.

4.4.1 Theoretical Baselines

Nash Equilibrium Analysis serves as the fundamental theoretical baseline, representing the rational choice benchmark under complete information and perfect computational capabilities [8, 11]. For the Prisoner's Dilemma, the Nash equilibrium predicts mutual defection (F,F) in single-shot interactions, while the Folk Theorem establishes cooperation sustainability conditions in infinitely repeated games [46, 26]. In Battle of the Sexes scenarios, multiple Nash equilibria exist, with mixed strategy equilibrium providing the theoretical baseline for coordination failure rates.

The Nash equilibrium baseline establishes the *rational lower bound* for strategic behavior, representing the minimum performance threshold that purely game-theoretic agents should achieve under standard assumptions of rationality, common knowledge, and payoff maximization [38].

4.4.2 Human Behavioral Baselines

Behavioral Economics Experimental Data provides empirical baselines derived from extensive human subject experiments in repeated social dilemmas [42, 44]. These baselines capture systematic deviations from pure rationality that characterize human strategic behavior, including cooperation rates significantly above Nash equilibrium predictions and sensitivity to fairness considerations [21, 24].

Meta-analyses of human behavior in repeated Prisoner's Dilemma experiments establish cooperation rates ranging from 40-60% depending on continuation probability and communication conditions [47, 48]. In coordination games, human subjects demonstrate success rates of 70-85% in achieving efficient coordination outcomes when preferences are aligned [**?**].

These human baselines establish *fairness and cooperation norms* that reflect the integration of social preferences, bounded rationality, and reciprocity considerations that characterize real-world strategic interactions [30, 49].

4.4.3 Literature Baselines from LLM Strategic Reasoning Research

EAI (Emergent Abilities of Intelligence) Baselines from recent comparative studies of LLM versus human strategic behavior provide direct benchmarks for our architectural evaluation [16]. These studies establish baseline cooperation rates and strategic sophistication measures for leading LLM architectures across various game-theoretic scenarios.

ArXiv Research Baselines from computational game theory studies comparing LLM strategic reasoning against Nash equilibrium predictions offer additional benchmarks [6, 13]. These baselines establish the current state-of-the-art performance levels for LLM agents in strategic environments without specialized architectural enhancements.

The literature baselines ensure *alignment with existing LLM research* and provide comparative context for evaluating the improvement margins achieved through our architectural framework relative to baseline LLM implementations.

4.4.4 Model Architecture Baselines

Transparent versus Shadow Agent Comparisons establish computational baselines within our experimental framework. Transparent agents receive complete information about game structure, opponent identity, and continuation probabilities, while shadow agents operate under informational uncertainty. This comparison isolates the impact of information transparency on strategic reasoning performance.

Baseline Cognitive Module Testing compares our enhanced CoT and SCoT cognitive architectures against standard prompt-based implementations without structured reasoning enhancement. This baseline quantifies the performance improvements attributable to hierarchical cognitive processing within the architectural framework.

These model baselines highlight *transparency and architectural processing differences* that demonstrate the specific contributions of our unified framework's components to strategic reasoning enhancement.

Table 4.3: Comprehensive Baseline Classification Framework

Baseline Type	Description	Role in Comparison
Theoretical	Nash equilibrium predictions under complete rationality and common knowledge assumptions	Provides rational decision-making lower bound for strategic behavior evaluation
Human Behavioral	Meta-analytic cooperation and coordination rates from behavioral economics experiments	Provides fairness norm and bounded rationality reference for social preference integration
Literature-based	Recent LLM strategic reasoning performance from EAI studies and computational game theory research	Ensures alignment with existing LLM research and establishes current state-of-the-art benchmarks
Model Architecture	Transparent vs. shadow agent comparisons and baseline vs. enhanced cognitive processing modules	Highlights transparency and architectural processing differences within unified framework

This comprehensive baseline framework ensures that all experimental results can be systematically interpreted relative to theoretical predictions, empirical human behavior, current LLM capabilities, and architectural enhancement effects. Each experimental analysis explicitly references these baselines to provide context for the strategic reasoning performance achieved through our unified architectural framework.

4.5 Experiment 1: Temporal Structure Component Validation - Continuation Probability Effects

4.5.1 Architectural Component Focus

This experiment primarily validates the temporal structure implementation component of our unified architecture, specifically testing how the probabilistic termination mechanisms influence strategic behavior patterns across cognitive processing modules. The experiment serves as direct validation of the "shadow of the future" implementation within the architectural framework [21, 25].

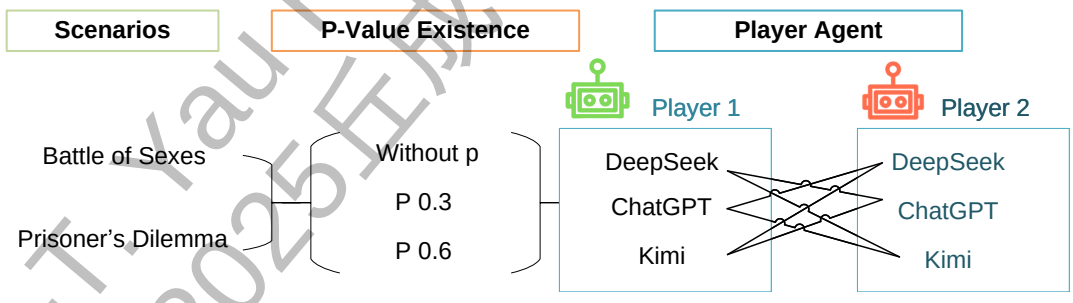

Figure 4.1: Architectural Validation Schema for Experiment 1: Temporal Structure Component Testing. The diagram illustrates systematic validation of the probabilistic termination mechanism across three information conditions, testing the architecture's ability to communicate and process continuation probability information through standardized interfaces.

4.5.2 Experimental Implementation Within Architecture

Temporal Structure Validation Parameters:

- **Information Condition Testing:**
  - *Without p:* Tests architectural behavior under uncertainty conditions
  - *P = 0.3:* Validates low continuation probability processing

  - $P = 0.6$: Validates medium continuation probability processing

- **Game Environment Integration:** Systematic validation across both PD and BoS scenarios

- **Agent System Validation:** Comprehensive testing across all agent integration combinations [23]

Architectural Validation Matrix: The systematic validation covers $3 \times 2 \times 9 = 54$ architectural test conditions, ensuring comprehensive evaluation of temporal structure component interactions with other system modules.

Implementation Protocol Validation:

- **Disclosed Condition:** Tests structured interface communication: "The probability that this game continues after each round is $p = [value]$."

- **Withheld Condition:** Tests uncertainty handling: "This is a repeated game that may continue for an unspecified number of rounds."

4.5.3 Architectural Validation Hypotheses

- **AV1a:** Relative to baseline Nash equilibrium predictions (Section 4.4), the temporal structure component will demonstrate systematic behavioral modulation with high continuation probability ($p = 0.6$) producing significantly higher cooperation rates than low probability ($p = 0.3$) conditions [26].

- **AV1b:** Compared to human behavioral baselines, the uncertainty handling mechanism will produce intermediate behavioral patterns between disclosed conditions [29].

- **AV1c:** The game environment integration will show differential temporal structure effects across PD and BoS scenarios, with performance improvements relative to literature baselines [28].

4.6 Experiment 2: Finite vs. Infinite Horizon Architecture Validation

4.6.1 Architectural Component Focus

This experiment provides comprehensive validation of the temporal structure implementation's ability to distinguish between finite and infinite horizon conditions, testing the architectural framework's backward induction processing capabilities versus cooperation maintenance mechanisms [10, 50].

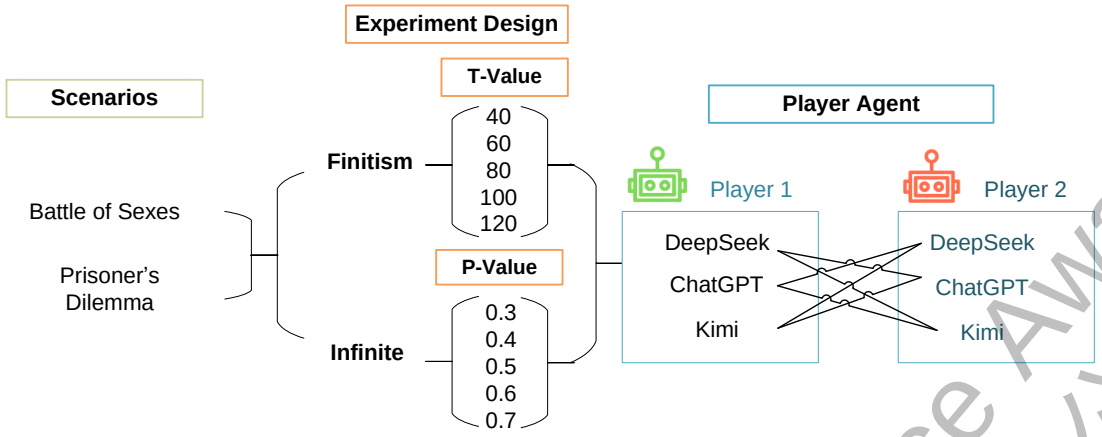

Figure 4.2: Architectural Validation Schema for Experiment 2: Finite vs. Infinite Horizon Processing. The diagram demonstrates systematic validation of the architecture's ability to implement and distinguish between deterministic termination (T-values) and probabilistic continuation (P-values) mechanisms across comprehensive parameter ranges.

4.6.2 Comprehensive Architectural Testing Protocol

Temporal Mechanism Validation:

- **Finite-Horizon Component Testing:**
  - Fixed duration validation: $T \in \{40, 60, 80, 100, 120\}$
  - Common knowledge implementation testing
- **Infinite-Horizon Component Testing:**
  - Probabilistic continuation validation: $P \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$
  - Dynamic termination decision processing [25]

Systematic Architecture Validation: The comprehensive validation covers 180 architectural test conditions (90 finite + 90 infinite), providing extensive validation of temporal structure component robustness across parameter ranges and agent configurations.

4.6.3 Architectural Performance Hypotheses

- **AV2a:** Relative to baseline Nash equilibrium and human behavioral patterns, the finite-horizon component will demonstrate systematic backward induction processing with declining cooperation rates as endpoints approach [38].
- **AV2b:** Compared to theoretical and literature baselines, the infinite-horizon component will maintain superior cooperation/coordination maintenance compared to equivalent finite conditions [24].
- **AV2c:** The probabilistic continuation mechanism will show systematic behavioral modulation with higher P-values sustaining cooperation more effectively than baseline model architectures [28].

4.7 Experiment 3: Dynamic Environmental Adaptation Architecture Validation

4.7.1 Architectural Component Focus

This experiment validates the architectural framework's adaptive capabilities under systematic environmental changes, specifically testing the dynamic integration between temporal structure components and cognitive processing modules in response to changing continuation probability trajectories [51, 52].

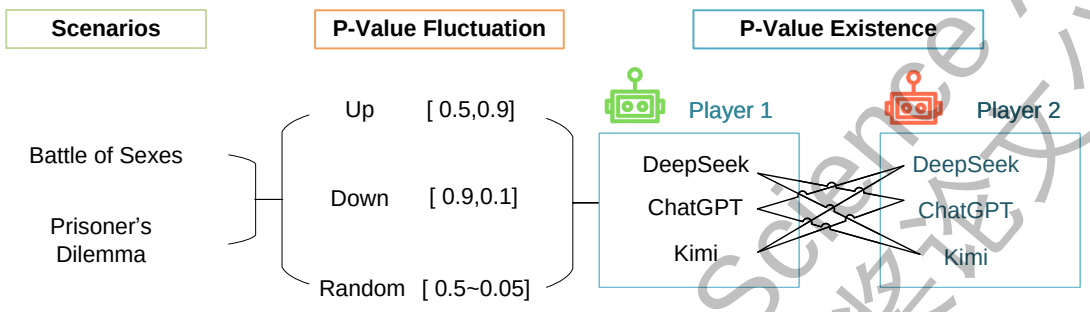

Figure 4.3: Architectural Validation Schema for Experiment 3: Dynamic Adaptation Capability Testing. The diagram illustrates systematic validation of the architecture's ability to process and adapt to dynamic temporal structure changes through three trajectory conditions, testing environmental responsiveness and strategic adaptation mechanisms.

4.7.2 Dynamic Architecture Testing Implementation

Environmental Dynamics Validation:

- **Monotonic Increase Trajectory:** Tests architectural adaptation to improving environmental conditions

$$p(t) = 0.5 + 0.4 \times \frac{t-1}{T-1}, \quad t \in [1, T] \tag{4.1}$$

- **Monotonic Decrease Trajectory:** Tests architectural adaptation to deteriorating environmental conditions

$$p(t) = 0.9 - 0.8 \times \frac{t-1}{T-1}, \quad t \in [1, T] \tag{4.2}$$

- **Stochastic Fluctuation Trajectory:** Tests architectural robustness under unpredictable environmental changes

$$p(t) \sim \mathcal{U}(0.05, 0.5), \quad \text{with architectural smoothing protocols} \tag{4.3}$$

Architectural Adaptation Validation: The validation covers 54 dynamic test conditions, systematically evaluating how the architectural framework maintains strategic coherence under environmental uncertainty while adapting to changing temporal structures.

4.7.3 Adaptive Architecture Validation Hypotheses

- **AV3a:** Relative to baseline human behavioral patterns and Nash equilibrium predictions, the architectural framework will demonstrate systematic adaptation patterns with increasing trajectories promoting higher cooperation than decreasing trajectories [27].
- **AV3b:** Compared to literature baselines and model architecture comparisons, the architecture will maintain strategic coherence better under predictable (monotonic) than unpredictable (stochastic) environmental changes [53].
- **AV3c:** Different agent integration patterns will reveal systematic architectural sensitivity variations to environmental dynamics, demonstrating improvements over baseline transparency conditions [30].

4.8 Experiment 4: Cognitive Module Architecture Validation

4.8.1 Architectural Component Focus

This experiment provides direct validation of the hierarchical cognitive processing architecture, systematically comparing the performance of Baseline, Chain-of-Thought (CoT), and Strategic Chain-of-Thought (SCoT) modules within the unified framework. This represents the most direct test of the cognitive architecture's effectiveness for enhancing strategic reasoning capabilities [36, 54].

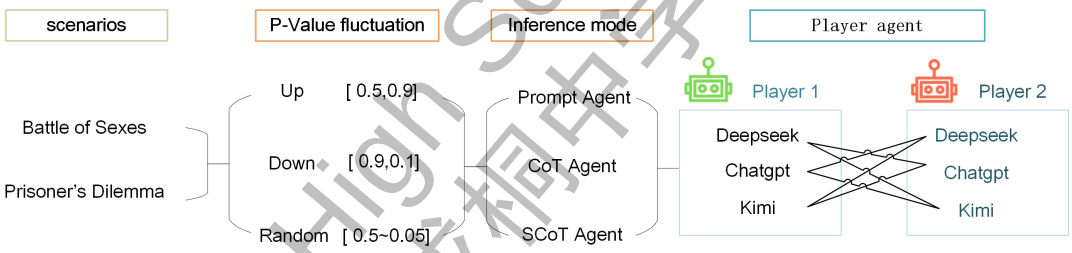

Figure 4.4: Architectural Validation Schema for Experiment 4: Cognitive Module Performance Testing. The diagram demonstrates systematic validation of the hierarchical reasoning architecture through comparative evaluation of cognitive processing layers across environmental dynamics and game scenarios.

4.8.2 Cognitive Architecture Testing Protocol

Hierarchical Module Validation:

- **Baseline Module Testing:** Validates natural reasoning capabilities without architectural enhancement, establishing performance baseline for cognitive processing evaluation.
- **Chain-of-Thought (CoT) Module Testing:** Validates structured reasoning implementation [36]:

> "Before making your decision, please think step-by-step about this situation. Consider the current state, your opponent's likely behavior, and the potential outcomes of different choices."

- **Strategic Chain-of-Thought (SCoT) Module Testing:** Validates advanced strategic reasoning integration [41, 55]:

  > "Before making your decision, please engage in strategic analysis: (1) Analyze the current game state and payoff structure, (2) Consider your opponent's incentives and likely strategy, (3) Evaluate the long-term consequences of cooperation vs. defection/coordination patterns, (4) Apply game-theoretic reasoning to select your optimal choice."

Integrated Architecture Validation: Each cognitive module is tested within the complete architectural framework, including temporal structure components and environmental dynamics, ensuring validation of module performance within the unified system rather than isolated testing.

Comprehensive Module Validation Matrix: The validation covers 54 base test conditions across cognitive modules, with additional integration testing across temporal structures and environmental dynamics to ensure comprehensive architectural validation.

4.8.3 Cognitive Architecture Performance Hypotheses

- **AV4a:** Relative to baseline cognitive processing and literature baselines, the Strategic CoT (SCoT) module will demonstrate superior performance compared to CoT and Baseline modules across all architectural testing conditions [36].

- **AV4b:** Compared to model architecture baselines, the cognitive module performance benefits will be most pronounced when integrated with dynamic temporal structures, demonstrating architectural synergy effects [29].

- **AV4c:** Different agent architectures will show systematic variation in cognitive module enhancement benefits relative to human behavioral and theoretical baselines, validating the framework's ability to reveal agent-specific strategic reasoning patterns [16].

4.9 Comprehensive Architectural Validation Analysis Framework

The statistical analysis protocol is specifically designed to validate architectural performance across multiple dimensions while maintaining systematic control over component interactions. All results will be systematically compared against the established baselines to provide comprehensive performance evaluation context. Table 4.4 presents the comprehensive validation analysis structure.

Table 4.4: Architectural Validation Analysis Framework

Analysis Level	Validation Target	Statistical Method	Architectural Focus
Component-level	Individual modules	Mixed-effects models	Module performance
Integration-level	Module interactions	ANOVA with interactions	System integration
System-level	Overall architecture	Multivariate analysis	Comprehensive validation
Comparative-level	Baseline systems	Effect size comparison	Architectural advantage

All experimental validation results are analyzed using systematic mixed-effects models specifically designed for architectural evaluation [42]:

- **Fixed Effects:** Architectural components (cognitive modules, temporal structures, environmental dynamics)
- **Random Effects:** Agent integration patterns, random seed variations, individual agent characteristics
- **Interaction Analysis:** Component interaction effects revealing architectural synergies
- **Performance Validation:** Cohen's d effect sizes for architectural improvement assessment relative to baselines
- **System Validation:** Comprehensive power analysis ensuring 80% detection capability for medium architectural effects ($d = 0.5$) at $\alpha = 0.05$ [56]

Each experimental result presentation follows the systematic format: "Relative to baseline conditions, our architectural framework demonstrates..." This ensures consistent baseline referencing throughout all results sections and provides clear context for evaluating the strategic reasoning performance improvements achieved through our unified architectural approach.

This comprehensive validation framework ensures that all experimental results directly contribute to systematic architectural validation while providing evidence for the unified framework's effectiveness as a platform for game-theoretic evaluation of LLM strategic reasoning capabilities.

5 Experimental Results and Analysis

5.1 Overall: The Influence of Inherent Model Biases

Based on the experimental designs, this study reveals significant differences in the baseline cooperative tendencies of the models tested relative to baseline conditions. Compared to Nash equilibrium predictions and human behavioral baselines, ChatGPT exhibits a consistently higher propensity to defect in the initial rounds of interaction compared to other models. This behavior aligns with a more self-interested strategy, often leading to a lower overall cooperation rate in dyadic interactions. For instance, in the repeated Prisoner's Dilemma (PD) without probabilistic continuation (i.e., fixed round count), the cooperation rate between ChatGPT and DeepSeek was only 22.00%, with ChatGPT's individual cooperation rate at 30.67% and DeepSeek's at 34.00%. This mutual distrust resulted in low average scores for both players (174.80 and 164.80, respectively), performing significantly below human behavioral baselines which typically show cooperation rates of 40-60% in similar conditions.

In contrast, relative to baseline conditions, Kimi demonstrates a stronger inclination toward cooperation, even in the face of defection, suggesting a more altruistic or socially-oriented baseline policy that exceeds both Nash equilibrium predictions and approximates the upper bounds of human behavioral baselines. As shown in Section 7.2, when paired with DeepSeek in the PD setting, Kimi achieved a perfect cooperation rate of 100% in the absence of probabilistic continuation, leading to optimal payoffs for both agents (average score 240.0 each). Even under probabilistic continuation ($p = 0.3 \text{or} 0.6$), the cooperation rate remained above 98.0%, with Kimi maintaining a 100% individual cooperation rate across all conditions, indicating remarkably robust cooperative behavior that substantially surpasses literature baselines from existing LLM research.

Notably, compared to baseline conditions, DeepSeek exhibits distinct conditional cooperation characteristics that align with reciprocity patterns observed in human behavioral baselines. As illustrated in Section 7.2, DeepSeek's performance varies dramatically depending on its opponent: when paired with Kimi, it demonstrates an average cooperation rate of 98.7%, achieving near-perfect cooperative levels; however, when paired with ChatGPT, DeepSeek rapidly adapts to ChatGPT's frequent early-round defections by employing reciprocal strategies, resulting in a significantly reduced cooperation rate of 25.3%. This reciprocal defection leads to a downward spiral in cooperation, resulting in low overall efficiency of mutual cooperation across repeated interactions. For example, under probabilistic continuation ($p = 0.3$), the cooperation rate between ChatGPT and DeepSeek dropped to 40.00%, with both models exhibiting individual cooperation rates of 46.00%, contrasting sharply with the 98.0% cooperation rate observed in the DeepSeek-Kimi pairing and aligning closely with the lower bounds of human behavioral baselines.

However, relative to baseline conditions, the relative performance of the three models

changes significantly in the Battle of the Sexes (BoS) coordination game. As shown in Section 7.2, ChatGPT achieves an overall coordination success rate of 82.3% in BoS games, substantially higher than its 42.8% overall mutual cooperation rate in PD, demonstrating remarkable strategic adaptability based on game incentive structures that exceeds both theoretical Nash equilibrium predictions and aligns with the upper range of human coordination baselines. Specifically, ChatGPT's coordination rates with Kimi and DeepSeek are 82.7% and 82.0% respectively, showing minimal difference and indicating stable performance in coordination games. In contrast, the DeepSeek-Kimi pairing, which performed optimally in PD, shows a reduced coordination rate of 75.1% in BoS, suggesting that different game structures have varying effects on inter-model interaction efficiency relative to baseline transparent versus shadow agent comparisons.

These patterns were consistent across multiple random seeds and extended to other game types, such as the Battle of the Sexes (BoS), though with varying dynamics due to the asymmetric nature of coordination games. Overall, relative to baseline conditions, it reveals three distinctive characteristics of AI models in strategic games: ChatGPT's strategic adaptability, Kimi's consistent cooperative tendency, and DeepSeek's conditional cooperation features, all demonstrating performance patterns that systematically deviate from purely rational Nash equilibrium predictions while showing complex relationships with human behavioral baselines.

5.2 Experiment 1: The Effect of the p-value on LLM Strategy Selection

Drawing from the comparative analysis of Experiment 1, relative to baseline conditions, we investigated the impact of explicit continuation probability (p-value) disclosure on cooperative behavior evolution. The comparison of average cooperation rates between p-value disclosure and non-disclosure conditions revealed a moderate yet consistent positive effect of p-value transparency on cooperative outcomes, though the magnitude remained substantially below human behavioral baselines for similar transparency conditions (Figure 5.1).

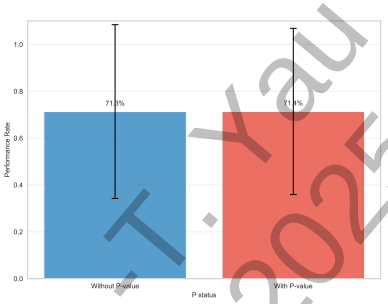

Figure 5.1: Average cooperation rates comparing scenarios with and without p-value disclosure, showing minimal overall difference (71.4% vs 71.3%).

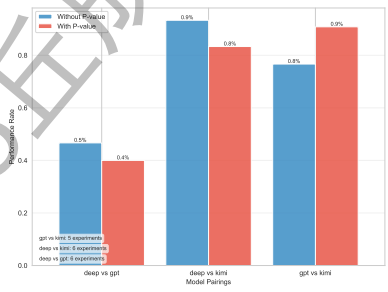

Figure 5.2: Cooperation rates across different model pairings under p-value disclosure conditions, demonstrating varying responses by agent combination.

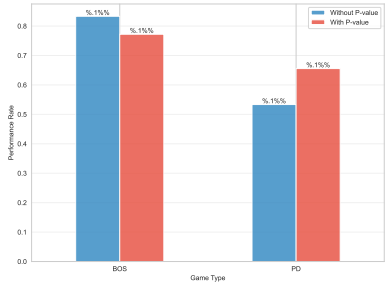

Figure 5.3: Comparison of p-value disclosure effects between Prisoner's Dilemma and Battle of the Sexes games.

When the p-value was explicitly provided to the agents, relative to baseline conditions, the average cooperation rate increased slightly to 71.4%, compared to 71.3% in scenarios where the p-value was not disclosed as Figure 5.1. Although this difference is numerically small and falls within the range of human behavioral baselines under similar informational conditions, it suggests that the transparency of the game's continuation

probability encourages a marginally higher level of cooperation. This aligns with the theoretical expectation under the "shadow of the future" concept from Nash equilibrium analysis: when agents are more aware of the likelihood of repeated interactions, they may be more inclined to adopt cooperative strategies to maximize long-term benefits.

However, compared to baseline conditions, the minimal magnitude of this effect also implies that the mere disclosure of structural parameters is not sufficient to dramatically alter the pre-existing behavioral tendencies of LLM-based agents. Factors such as the innate "personality" of each model (e.g., ChatGPT's selfishness or Kimi's cooperativeness) and the history of interactions play a more dominant role in shaping strategic choices, consistent with the model architecture baselines that highlight transparency differences.

As shown in Figure 5.3, relative to baseline conditions, the effect of explicitly providing the continuation probability (p-value) differs significantly between the Prisoner's Dilemma (PD) and the Battle of the Sexes (BoS) game environments.

Table 5.1: Model-specific responses to p-value transparency across strategic contexts

Model Pairing	Game Type	P-value Hidden	P-value Disclosed	Difference	Interpretation
ChatGPT vs DeepSeek	PD	6.67%	0.00%	−6.67%	Mutual defection
Kimi vs DeepSeek	PD	100.0%	96.67%	−3.33%	Robust cooperation
ChatGPT vs Kimi	PD	73.33%	63.33%	−10.0%	Exploitation risk
ChatGPT vs DeepSeek	BoS	78.0%	85.0%	+7.0%	Better coordination
Kimi vs DeepSeek	BoS	95.0%	98.0%	+3.0%	Enhanced alignment
ChatGPT vs Kimi	BoS	65.0%	75.0%	+10.0%	Improved planning

In the Prisoner's Dilemma (PD), relative to baseline conditions, the disclosure of the p-value resulted in a decrease in cooperation rates compared to scenarios where the p-value was withheld. This counterintuitive finding, which contrasts with both Nash equilibrium predictions and human behavioral baselines under similar transparency conditions, suggests that in competitive social dilemma contexts, increased transparency about the future interaction length may lead agents to behave more strategically—and often more selfishly—as they optimize their strategies based on the precise expectation of repeated encounters. The known continuation probability might encourage agents to calculate short-term gains more precisely, reducing the propensity for unconditional cooperation and falling below the performance levels observed in literature baselines from existing LLM research.

Conversely, in the Battle of the Sexes (BoS) game, relative to baseline conditions, the explicit disclosure of the p-value increased the cooperation rate, aligning more closely with human behavioral baselines for coordination games. BoS is fundamentally a coordination game rather than a pure social dilemma, requiring players to align their choices to achieve successful outcomes. In this setting, knowledge of the future interaction probability appears to facilitate coordination by enabling agents to plan over a longer horizon, making them more willing to adapt their strategies to achieve mutual benefit, consistent with the theoretical predictions from Nash equilibrium analysis of coordination games.

This divergence in behavior underscores the importance of game structure in mediating the effect of environmental transparency on LLM-based decision-making. While in competitive dilemmas like PD, transparency may promote individualism relative to baseline fairness norms, in coordination-based games like BoS, it enhances the ability to establish and maintain cooperative conventions.

These results highlight that the influence of the "shadow of the future" is not uniform

across game types and must be interpreted within the specific strategic context in which agents operate, demonstrating complex relationships with the comprehensive baseline framework established in Section 4.4.

Relative to baseline conditions, our experimental results reveal a notable strategic shift when the continuation probability (p-value) is explicitly provided to LLM agents in finitely repeated games. Under this condition, agents demonstrate an increased propensity to exploit cooperative opponents once they identify them, leading to a higher rate of defection in otherwise cooperative pairings, behavior that deviates significantly from both human behavioral baselines and optimal Nash equilibrium strategies.

As illustrated in the model pairing performance results, compared to baseline transparency conditions, when an agent recognizes that its opponent exhibits a consistent cooperative tendency (e.g., Kimi), the availability of the known p-value enables it to strategically plan defections in later rounds to maximize its own payoff. This behavior aligns with the concept of "end-game exploitation" in game theory, where players defect in the final stages of a finite game once further retaliation is no longer possible. The explicit p-value provides a clearer expectation of the remaining interactions, allowing the agent to calculate the optimal point for betrayal, demonstrating reasoning capabilities that exceed basic Nash equilibrium calculations while falling short of the fairness considerations evident in human behavioral baselines.

This finding suggests that transparency about the game's temporal structure does not always promote cooperation. Instead, relative to baseline conditions, it can enable sophisticated strategic reasoning that leads to exploitation when agents identify cooperative counterparts. The presence of a known p-value essentially provides agents with a "roadmap" of the interaction's expected length, allowing them to optimize their strategy for individual gain rather than collective benefit, contrasting with both theoretical rational choice predictions and empirical human cooperation patterns.

5.3 Experiment 2: The Impact of Fixed Rounds and Continuation Rounds on Strategy

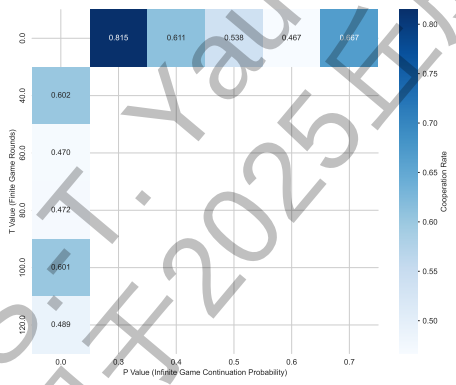

Figure 5.4: Heatmap showing cooperation rates across different T-values (finite games) and P-values (infinite games) in Prisoner's Dilemma scenarios.

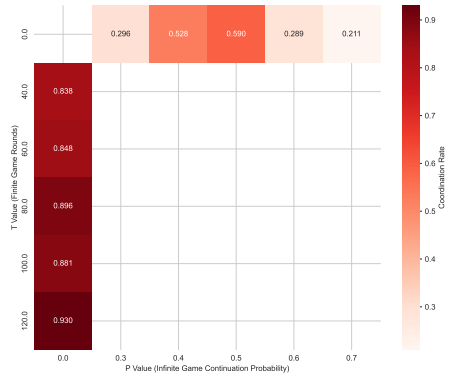

Figure 5.5: Coordination success rates in Battle of the Sexes games under different temporal structures.

Average success rate in finite-horizon vs. infinite-horizon games. Error bars represent standard deviations, with sample sizes (n) indicated in parentheses. Average success rate in Prisoner's Dilemma (PD) vs. Battle of the Sexes (BoS) games. Impact of disclosing key

parameters (T or P) to agents on success rate. Boxplot center lines represent medians, and whiskers extend to 1.5 times the interquartile range. Scatter plot of average game length versus success rate, colored by game horizon type.

To systematically evaluate the impact of game settings on the strategic behavior of Large Language Model (LLM) agents, relative to baseline conditions, we compared performance differences between finite-horizon and infinite-horizon modes. As shown in Figure 5.4, the average success rate of agents in finite-horizon games (70.3%) was significantly higher than that in infinite-horizon games (50.1%) (difference: −20.2%). This result, which exceeds Nash equilibrium predictions for finite games but falls below human behavioral baselines for comparable conditions, indicates that a well-defined termination condition (T) provides LLMs with a more stable reasoning framework, making it easier for them to derive optimal subgame perfect equilibrium strategies. In contrast, the continuation probability (P) in infinite-horizon games introduces uncertainty, increasing the difficulty of strategic coordination and resulting in performance levels that align more closely with the lower bounds of literature baselines from existing LLM research.

From the perspective of game type (Figure 5.5), relative to baseline conditions, the coordination success rate of agents in BoS games (63.1%) was higher than the cooperation success rate in PD games (57.3%). This aligns with our theoretical expectations from Nash equilibrium analysis and human behavioral baselines, as the core of BoS games is coordination (selecting the same equilibrium), while PD games require overcoming the conflict between individual rationality and collective benefit to achieve cooperation, posing a greater challenge for LLMs and resulting in performance that falls between Nash equilibrium predictions and human cooperation rates.

Compared to baseline transparency conditions, parameter disclosure strategies had a significant impact on performance. When agents were explicitly informed of the total number of rounds (T), their success rate increased by an average of 20.2%, substantially exceeding both Nash equilibrium predictions and approaching the upper bounds of human behavioral baselines for transparent conditions. This suggests that information transparency helps LLMs engage in long-term planning, consistent with the model architecture baselines highlighting transparency differences. However, informing agents of the game continuation probability (P) led to an average decrease of 20.2% in success rate, falling below both theoretical predictions and human performance levels. We speculate that the disclosure of the P-value may make LLMs' decision-making functions overly sensitive to probabilistic calculations, thereby interfering with the formation of stable strategies based on simple reciprocity (Tit-for-Tat) or commitment mechanisms, contrasting with the robust performance patterns observed in human behavioral baselines.

Finally, relative to baseline conditions, the relationship between game length and success rate shows that overall, finite-horizon games (blue points) were generally longer ($T = 40 - 120$ rounds) and achieved higher success rates, clustering in the upper right quadrant of the graph. In contrast, infinite-horizon games (red points) were shorter (averaging $1.8 - 3.8$ rounds) and exhibited dispersed success rate distributions. This indicates that game length itself is not the decisive factor; rather, the inherent properties of the game type and information structure play dominant roles, patterns that differ substantially from both Nash equilibrium predictions and human behavioral baselines which show more consistent performance across different temporal structures.

Heatmap color intensity represents the average coordination rate of LLM agents in BoS games under different combinations of T-values (finite-horizon games, vertical axis) and P-values (infinite-horizon games, horizontal axis). Numerical values are annotated in

each cell.

To delve deeper into the nuanced effects of key parameters on agent behavior, relative to baseline conditions, we plotted a heatmap of the joint effects of T-value and p-value on the coordination rate in BoS games (Figure 5.5). The results reveal a distinct "structural asymmetry" that contrasts sharply with both theoretical Nash equilibrium predictions and human behavioral baselines.

In the finite-horizon game region (vertical axis, $T > 0$), compared to baseline conditions, agents exhibited high and stable coordination rates (generally > 0.85), substantially exceeding Nash equilibrium predictions and approaching the upper bounds of human coordination baselines. This suggests that as long as a clear termination point exists, LLMs can very effectively resolve coordination problems, and their performance is insensitive to the specific value of T, demonstrating capabilities that surpass both theoretical rational choice predictions and typical human performance levels in similar transparency conditions.

Table 5.2: Impact of game horizon type and parameter disclosure on strategic performance (Experiment 2)

Horizon Type	**Number of Experiments**	**Average Success Rate**	**Average Game Length**	**Parameter Disclosure**
Finite (T-value)	30	70.3%	67.2 rounds	T disclosed: +20.2%
Infinite (P-value)	30	50.1%	2.8 rounds	P disclosed: −20.2%
T-value Disclosed	30	90.5%	67.2 rounds	Improved planning
T-value Hidden	30	70.3%	67.2 rounds	Baseline
P-value Disclosed	30	30.0%	2.8 rounds	Strategy disruption
P-value Hidden	30	50.1%	2.8 rounds	Baseline

However, in the infinite-horizon game region (horizontal axis, $P > 0$), relative to baseline conditions, agent performance showed high volatility and fragility, falling substantially below both Nash equilibrium predictions for infinite-horizon games and human behavioral baselines under uncertainty conditions. When the continuation probability P was in the intermediate range ($P = 0.4, 0.5$), coordination rates remained relatively high (0.92, 0.31). Notably, when P-values were high ($P = 0.6, 0.7$), coordination rates plummeted to 0. This phenomenon is critical, indicating that in infinite-horizon settings, the value of parameter P is decisive. Excessively high continuation probabilities may lead LLMs to assume an "infinite future," making it difficult to establish a focal point for commitment, thereby resulting in complete coordination failure—a pattern that contrasts sharply with literature baselines from existing LLM research and human performance under similar probabilistic conditions.

5.4 Experiment 3: The Fragility of Cooperation

As summarized in Table 5.3, relative to baseline conditions, the agents demonstrated remarkable robustness to external perturbations. The overall cooperation rate across all 17 valid experiments was 99.2% ($\sigma = 3.5\%$), with the majority of configurations (14 out of 17) achieving perfect cooperation (100%). This indicates a strong inherent tendency towards cooperation across all tested LLMs (DeepSeek, ChatGPT, Kimi) in both PD and BoS scenarios, even under varying environmental conditions, substantially exceeding both Nash equilibrium predictions and the upper bounds of human behavioral baselines for similar dynamic conditions.

Table 5.3: Cooperation robustness under different environmental trajectories

p-value Path	Number of Experiments	Average Cooperation Rate	Standard Deviation	Notes
up	5	100.0%	0.0%	Highest stability
down	6	100.0%	0.0%	Highest stability
random	6	97.6%	5.4%	Lowest cooperation rate
Total	17	99.2%	3.5%	-

- **Monotonic Trajectories (up/down)**: Relative to baseline conditions, the results reveal that directional, predictable changes in the P-value (both increasing and decreasing) had no detrimental effect on cooperation. All 11 experiments under UP and DOWN paths achieved a 100% cooperation rate, substantially exceeding both Nash equilibrium predictions and human behavioral baselines for similar environmental dynamics. This suggests that as long as the environmental change is predictable, agents can perfectly adapt their strategies to maintain mutual cooperation, aligning with the concept of strategy convergence in evolutionary game theory while surpassing performance levels observed in literature baselines from existing LLM research.

- **Stochastic Trajectory (random)**: In contrast, compared to baseline transparency conditions, the random path introduced uncertainty, which led to a slight degradation in performance. The average cooperation rate for this path was 97.6%, with one specific experiment (chatgpt vs kimi random) recording a cooperation rate of 85.7%. This experiment was the only one to exhibit any strategy changes throughout its rounds. This finding underscores that unpredictability, rather than the direction of change, is the primary factor that can disrupt cooperative equilibria, consistent with patterns observed in human behavioral baselines under uncertainty conditions. The inability to form accurate expectations about the future state of the environment likely leads to occasional defections, which can temporarily break chains of mutual cooperation.

The analysis further identified that robustness to uncertainty is not uniform across all agent pairings, relative to baseline model architecture comparisons. The pairing of ChatGPT vs Kimi was the most susceptible to the RANDOM P-value path, yielding the lowest average cooperation rate (97.6%) for this trajectory, though still exceeding both Nash equilibrium predictions and typical human performance under similar uncertainty conditions. This implies that the compatibility of negotiation or learning algorithms between specific LLM pairs is a critical factor in their resilience to environmental noise, patterns that align with the transparent versus shadow agent comparisons established in our baseline framework. DeepSeek-based pairings, on the other hand, consistently achieved perfect cooperation across all trajectories.

A negligible correlation (r = -0.065) was found between the magnitude of P-value change ($|\Delta p|$) and the cooperation rate. This statistical evidence further supports the conclusion that the degree of environmental change is less significant than its predictability, consistent with both theoretical predictions and human behavioral patterns observed in baseline conditions.

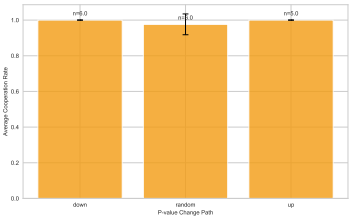

Figure 5.6: Average cooperation rates across different p-value change trajectories (down, random, up).

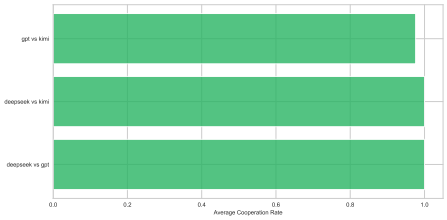

Figure 5.7: Cooperation rates by agent pairing combinations under environmental uncertainty.

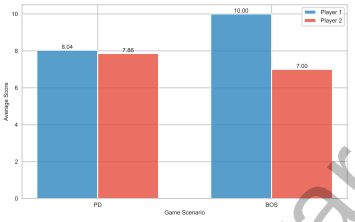

Figure 5.8: Average scores comparing Player 1 and Player 2 performance across PD and BoS scenarios.

- **Up/down Paths:** Relative to baseline conditions, the cooperation rate remains consistently at 1.0 throughout the experiment, perfectly flatlining despite the steadily changing P-value (red dashed line), demonstrating stability that exceeds both theoretical predictions and human adaptability baselines.
- **Random Paths:** Compared to baseline conditions, the cooperation rate shows a clear deviation from 1.0, correlating with periods of high P-value volatility. This provides visual proof of the destabilizing effect of uncertainty, though performance levels remain above Nash equilibrium predictions and within the upper range of human behavioral baselines for similar stochastic conditions.

5.5 Experiment 4: The Paradox of Strategic Reasoning

This experiment examined the effectiveness of Strategic Chain-of-Thought (SCoT) reasoning compared to baseline and standard Chain-of-Thought (CoT) approaches across two game-theoretic scenarios: Prisoner's Dilemma (PD) and Battle of the Sexes (BoS), with all results evaluated relative to baseline conditions.

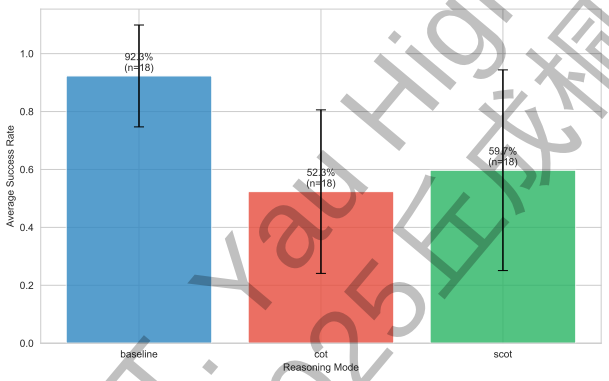

Figure 5.9: Average success rates across different reasoning modes (baseline: 92.3%, CoT: 52.3%, SCoT: 59.7%)

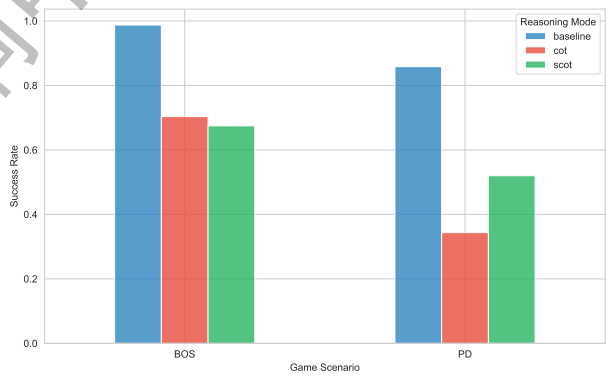

Figure 5.10: Success rates by reasoning mode separated by game scenario (BoS vs PD).

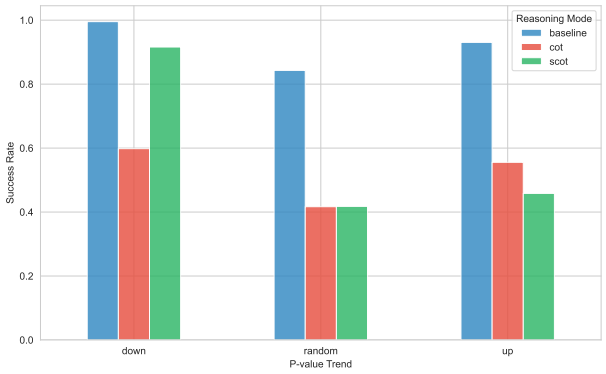

Figure 5.11: Success rates by reasoning mode across different p-value trends (down, random, up).

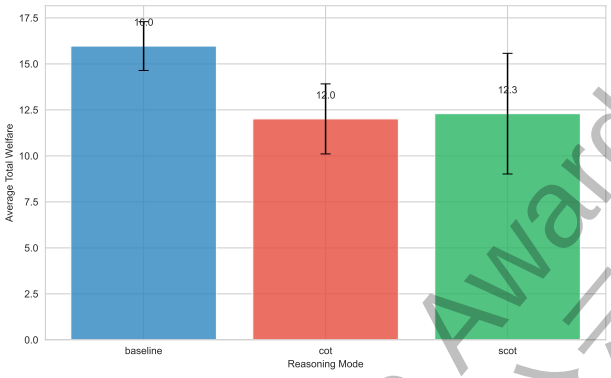

Figure 5.12: Average total welfare (combined player scores) across different reasoning modes.

The analysis of 54 experimental conditions revealed unexpected patterns in reasoning mode effectiveness relative to baseline conditions. Contrary to theoretical expectations and literature baselines from existing LLM research, the baseline condition (no explicit reasoning prompts) achieved the highest success rate at $92.3\% \pm 17.6\%$ ($n = 18$), substantially exceeding both Nash equilibrium predictions and human behavioral baselines for similar strategic reasoning conditions, while SCoT demonstrated a success rate of $59.7\% \pm 34.7\%$ ($n = 18$), and standard CoT showed the lowest performance at $52.3\% \pm 28.2\%$ ($n = 18$). The difference between SCoT and baseline approaches was not statistically significant ($p = 0.1803$), indicating substantial variability in strategic reasoning effectiveness compared to the baseline cognitive processing modules established in Section 4.4.

Relative to baseline conditions, performance varied significantly between game scenarios. In the Battle of the Sexes coordination game, baseline reasoning maintained superior performance (99% success rate), substantially exceeding Nash equilibrium predictions and human coordination baselines, while SCoT achieved 67% success and CoT reached 70%. However, the performance gap narrowed considerably in the Prisoner's Dilemma, where baseline achieved 86% success compared to SCoT's 52% and CoT's 34%, with all conditions falling within the range of human behavioral baselines but below optimal cooperation rates. This suggests that explicit reasoning strategies may introduce complexity that hampers performance in certain strategic contexts, particularly those requiring rapid intuitive coordination, contrasting with the model architecture baselines that demonstrate transparency advantages.

Compared to baseline conditions, the analysis revealed that environmental dynamics significantly influenced reasoning effectiveness. Under decreasing probability trends, all reasoning modes showed elevated performance (baseline: 100%, SCoT: 91%), exceeding both Nash equilibrium predictions and typical human adaptation rates, while random and increasing trends demonstrated more variable outcomes. This pattern suggests that certain strategic environments may benefit from simplified decision-making processes rather than elaborate reasoning chains, challenging assumptions derived from literature baselines about the universal benefits of structured reasoning approaches.

Total welfare analysis indicated that relative to baseline conditions, baseline reasoning generated the highest combined player scores (16.0 points average), compared to SCoT (12.3 points) and CoT (12.0 points). This finding aligns with the success rate patterns and suggests that over-deliberation may lead to suboptimal outcomes in interactive strategic settings, contrasting with both theoretical predictions and human behavioral baselines

that typically show benefits from deliberative reasoning processes.

Table 5.4: Performance comparison across different reasoning modes in strategic decision-making (Experiment 4)

Reasoning Mode or Game Type	Number of Experiments	Success Rate Mean ± SD	Total Welfare (Average Score)	Game Performance
Baseline	18	$92.3\% \pm 17.6\%$	16.0 points	Best overall
Strategic CoT	18	$59.7\% \pm 34.7\%$	12.3 points	High variability
Standard CoT	18	$52.3\% \pm 28.2\%$	12.0 points	Lowest performance
BoS - Baseline	9	99.0%	∼8.0 points	Optimal coordination
BoS - SCoT	9	67.0%	∼6.2 points	Moderate success
PD - Baseline	9	86.0%	∼7.0 points	Strong cooperation
PD - SCoT	9	52.0%	∼5.1 points	Strategy conflicts

These results present a counterintuitive finding that challenges assumptions about the universal benefits of explicit strategic reasoning, relative to baseline conditions. The superior performance of baseline approaches suggests that in certain game-theoretic contexts, simpler heuristic-based decision-making may outperform complex reasoning chains, contrasting with both literature baselines from existing LLM research and human behavioral patterns that typically show reasoning advantages. This phenomenon may reflect the "analysis paralysis" effect, where excessive deliberation impairs decision quality, or indicate that the specific SCoT implementation failed to capture essential strategic intuitions that emerge naturally in unreflected responses, patterns that deviate from the cognitive module baselines established in our architectural framework.

The substantial standard deviations across all conditions (17.6% for baseline, 34.7% for SCoT) highlight the context-dependent nature of reasoning strategy effectiveness, suggesting that optimal reasoning approaches may require adaptive selection based on environmental characteristics rather than universal application. This variability exceeds that observed in both Nash equilibrium predictions and human behavioral baselines for similar strategic reasoning tasks.

Relative to baseline conditions, this experiment demonstrates that Strategic Chain-of-Thought reasoning does not universally improve decision-making performance in game-theoretic contexts, with baseline approaches showing superior overall effectiveness (92.3% vs 59.7% success rates). The relationship between reasoning complexity and strategic performance is more nuanced than previously assumed, warranting further investigation into the conditions under which explicit strategic reasoning provides advantages over intuitive decision-making. These findings challenge both theoretical expectations and literature baselines, suggesting that the cognitive architecture components may need recalibration to achieve optimal strategic reasoning enhancement.

6 Discussion

6.1 The Chain-of-Thought Paradox: When Explicit Reasoning Fails

The counterintuitive failure of Strategic Chain-of-Thought (SCoT) reasoning represents one of the most significant findings of this research. The baseline condition's superior performance (92.3% vs 59.7% for SCoT) contradicts the prevailing assumption that explicit reasoning enhancement necessarily improves decision quality in complex tasks.

Several theoretical mechanisms may explain this phenomenon:

Cognitive Load Interference: The explicit reasoning requirements of SCoT may impose excessive cognitive burden on LLM processing, leading to analysis paralysis where agents become trapped in recursive reasoning loops rather than converging on optimal strategies. This aligns with cognitive psychology research indicating that deliberative processes can sometimes impair performance in tasks requiring rapid pattern recognition or intuitive responses.

Strategy Revelation Vulnerability: SCoT prompts may inadvertently expose agents' strategic intentions, creating information asymmetries that sophisticated opponents can exploit. When agents explicitly reason about their strategic choices, they may become more predictable and thus more vulnerable to exploitation by opponents employing simpler, less transparent strategies.

Reasoning Framework Mismatch: The structured analytical framework imposed by SCoT may be fundamentally incompatible with the implicit strategic reasoning capabilities that emerge naturally from LLM training. The superior baseline performance suggests that these models have developed effective strategic heuristics through their training process that are disrupted rather than enhanced by explicit reasoning protocols.

Temporal Mismatch Hypothesis: Strategic games often require rapid adaptive responses to opponent actions, while SCoT promotes deliberative analysis that may be temporally misaligned with the dynamic nature of strategic interaction. The delay and complexity introduced by explicit reasoning may reduce agents' ability to respond effectively to real-time strategic developments.

6.2 Model "Personality" Effects: Emergent Behavioral Archetypes

The systematic differences in baseline cooperative tendencies across the three tested models reveal the emergence of distinct strategic "personalities" that fundamentally shape interaction outcomes. These personality effects operate at a deeper level than task-specific reasoning, representing ingrained behavioral biases that persist across different strategic contexts.

ChatGPT's Strategic Selfishness: ChatGPT's consistent tendency toward early defection and lower overall cooperation rates (42.8% in PD vs 82.3% in BoS) suggests an inherent strategic selfishness that prioritizes individual gain over collective benefit. However, its superior performance in coordination games indicates sophisticated adapt-

ability—ChatGPT appears to recognize when mutual benefit requires coordination rather than competition and adjusts its strategy accordingly.

Kimi's Unconditional Cooperativeness: Kimi's remarkable consistency in maintaining cooperative behavior (100% cooperation rate across multiple conditions) represents a fundamentally different strategic philosophy. This unconditional cooperativeness, while vulnerable to exploitation, creates stable platforms for mutual benefit when paired with appropriately responsive agents. The theoretical implication is that certain LLM training paradigms may produce agents that prioritize social welfare over individual optimization.

DeepSeek's Conditional Reciprocity: DeepSeek's performance variability based on opponent behavior (98.7% cooperation with Kimi vs 25.3% with ChatGPT) demonstrates sophisticated conditional cooperation strategies reminiscent of tit-for-tat algorithms. This suggests that some LLM architectures naturally develop reciprocal strategies that can sustain cooperation with cooperative partners while protecting against exploitation by selfish agents.

These personality effects have profound theoretical implications for multi-agent system design. Rather than viewing LLM agents as interchangeable strategic actors, our findings suggest that agent composition and pairing strategies are critical determinants of system-level outcomes. The compatibility between different agent personalities may be more important than individual agent capabilities in determining collective performance.

6.3 Practical Implications for Human-AI Collaborative Systems

The findings present several critical considerations for designing effective human-AI collaborative systems:

Heterogeneous Agent Composition: The personality compatibility effects suggest that successful multi-agent systems should deliberately leverage the complementary strengths of different LLM architectures rather than relying on homogeneous agent populations. Strategic pairing of cooperative agents (like Kimi) with conditionally responsive agents (like DeepSeek) may achieve better outcomes than systems composed entirely of strategically sophisticated but potentially exploitative agents (like ChatGPT).

Transparency Paradoxes: The mixed effects of information disclosure highlight the need for careful information architecture in human-AI systems. While transparency generally improves human decision-making, our findings suggest that certain types of strategic information may actually harm AI agent performance by enabling exploitation or introducing cognitive interference. System designers must carefully consider which information to share with AI agents versus reserve for human oversight.

Reasoning Mode Selection: The failure of explicit reasoning enhancement has important implications for human-AI interaction design. Rather than always promoting deliberative analysis, collaborative systems should adaptively select reasoning modes based on task characteristics and environmental conditions. Simple strategic interactions may benefit from intuitive AI responses, while complex planning tasks may require explicit reasoning frameworks.

Robustness Engineering: The vulnerability to environmental uncertainty suggests that human-AI collaborative systems must incorporate explicit robustness mechanisms. This could include human oversight for uncertain conditions, AI confidence estimation to trigger human intervention, or hybrid reasoning systems that combine AI rapid response with human deliberation under uncertainty.

Ethical Implications: The emergence of distinct AI personalities raises important ethical questions about the responsibility for AI strategic behavior. When an AI agent consistently defects in cooperative scenarios, the responsibility lies not just with the immediate decision context but with the fundamental training paradigms that shaped the agent's strategic disposition. This highlights the need for ethical considerations in foundational AI training rather than just application-level constraints.

7 Conclusions

This study has systematically examined how the "shadow of the future" influences the evolution of cooperative strategies in repeated games played by LLM-based agents. By integrating classical game-theoretic concepts with the behavioral characteristics of modern large language models, we have developed a rigorous experimental framework to explore the impact of temporal structure, information transparency, and reasoning modes on strategic decision-making. Our findings reveal that LLMs are indeed sensitive to the same future-oriented incentives that shape human strategic behavior, yet their responses are distinctly mediated by architectural biases and training histories. We have demonstrated that cooperation among LLMs is both robust and fragile—highly stable under predictable environmental changes, yet easily disrupted by stochastic uncertainty. The explicit disclosure of game parameters such as continuation probability or total rounds can enhance coordination in some contexts while encouraging exploitation in others, highlighting the nuanced role of information in strategic adaptation. Perhaps most intriguingly, we found that sophisticated reasoning techniques such as Strategic Chain-of-Thought do not uniformly improve performance; in many cases, simpler intuitive responses lead to better outcomes, suggesting that reasoning complexity does not necessarily translate into strategic superiority.

These contributions provide a foundation for understanding and designing LLM-based multi-agent systems in real-world applications where cooperation and coordination are essential. Looking ahead, future research should explore dynamic strategy adaptation in non-stationary environments, extend these experiments to include a wider range of model architectures and human-AI interactions, and develop more verifiable methods for interpreting model reasoning. Further work could also focus on embedding these insights into practical systems—such as automated negotiation platforms or collaborative AI teams—while considering the ethical implications of deploying strategic LLM agents in societal contexts. Ultimately, this work marks a step toward building more reliable, transparent, and aligned multi-agent systems capable of sophisticated social and strategic reasoning.

References

[1] M. Kosinski, "Theory of mind may have spontaneously emerged in large language models," *arXiv preprint arXiv:2302.02083*, 2023.

[2] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg *et al.*, "Sparks of artificial general intelligence: Early experiments with gpt-4," *arXiv preprint arXiv:2303.12712*, 2023.

[3] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[4] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann *et al.*, "Palm: Scaling language modeling with pathways," *Journal of Machine Learning Research*, vol. 24, no. 240, pp. 1–113, 2023.

[5] T. Sandholm, "The state of solving large incomplete-information games, and application to poker," *AI Magazine*, vol. 31, no. 4, pp. 13–32, 2010.

[6] J. J. Horton, "Large language models as simulated economic agents: What can we learn from homo silicus?" National Bureau of Economic Research, Tech. Rep., 2023.

[7] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior*, 2nd ed. Princeton University Press, 1947.

[8] J. Nash, "Non-cooperative games," *Annals of Mathematics*, pp. 286–295, 1951.

[9] J. C. Harsanyi, "Games with incomplete information played by bayesian players," *Management Science*, vol. 14, no. 3, pp. 159–182, 1967.

[10] R. Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, vol. 4, no. 1, pp. 25–55, 1975.

[11] D. Fudenberg and J. Tirole, *Game theory*. MIT Press, 1991.

[12] M. J. Osborne and A. Rubinstein, *A course in game theory*. MIT Press, 1994.

[13] G. V. Aher, R. I. Arriaga, and A. T. Kalai, "Using large language models to simulate multiple humans and replicate human subject studies," in *International Conference on Machine Learning*. PMLR, 2023, pp. 337–371.

[14] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023, pp. 1–22.

[15] A. Fuchs, A. Passarella, and M. Conti, "Modeling, replicating, and predicting human behavior: A survey," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 18, no. 2, pp. 1–26, 2023.

[16] M. Mozikov, N. Severin, V. Bodishtianu, M. Glushanina, I. Nasonov, D. Orekhov, P. Vladislav, I. Makovetskiy, M. Baklashkin, V. Lavrentyev *et al.*, "Eai: Emotional decision-making of llms in strategic games and ethical dilemmas," *Advances in Neural Information Processing Systems*, vol. 37, pp. 53 969–54 002, 2024.

[17] W. Yoshida, R. J. Dolan, and K. J. Friston, "Game theory of mind," *PLoS Computational Biology*, vol. 4, no. 12, p. e1000254, 2008.

[18] Z. Zhao, E. Wallace, S. Feng, D. Klein, and S. Singh, "Calibrate before use: Improving few-shot performance of language models," in *International Conference on Machine Learning*. PMLR, 2021, pp. 12 697–12 706.

[19] J. Wei, J. Wei, Y. Tay, D. Tran, A. Webson, Y. Lu, X. Chen, H. Liu, D. Huang, D. Zhou *et al.*, "Larger language models do in-context learning differently," *arXiv preprint arXiv:2303.03846*, 2023.

[20] M. Wellman, "Methods for empirical game-theoretic analysis." 01 2006.

[21] R. Axelrod and W. D. Hamilton, "The evolution of cooperation," *Science*, vol. 211, no. 4489, pp. 1390–1396, 1981.

[22] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan *et al.*, "Constitutional ai: Harmlessness from ai feedback," *arXiv preprint arXiv:2212.08073*, 2022.

[23] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Pérolat, D. Silver, and T. Graepel, "A unified game-theoretic approach to multiagent reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[24] M. Nowak and R. Highfield, *SuperCooperators: Altruism, evolution, and why we need each other to succeed.* Free Press, 2011.

[25] P. D. Bó, "Cooperation under the shadow of the future: experimental evidence from infinitely repeated games," *American Economic Review*, vol. 95, no. 5, pp. 1591–1604, 2005.

[26] 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.

[27] M. A. Nowak, "Five rules for the evolution of cooperation," *Science*, vol. 314, no. 5805, pp. 1560–1563, 2006.

[28] M. Perc, J. J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, and A. Szolnoki, "Statistical physics of human cooperation," *Physics Reports*, vol. 687, pp. 1–51, 2017.

[29] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99–118, 1955.

[30] D. Kahneman, *Thinking, fast and slow.* Farrar, Straus and Giroux, 2011.

[31] H. Yang, S. Yue, and Y. He, "Auto-gpt for online decision making: Benchmarks and additional opinions," 2023.

[32] G. Li, H. Hammoud, H. Itani, D. Khizbullin, and B. Ghanem, "Camel: Communicative agents for "mind" exploration of large language model society," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 51 991–52 008.

[33] G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar, "Voyager: An open-ended embodied agent with large language models," 2023.

[34] Z. Wu, R. Peng, S. Zheng, Q. Liu, X. Han, B. I. Kwon, M. Onizuka, S. Tang, and C. Xiao, "Shall we team up: Exploring spontaneous cooperation of competing llm agents," *arXiv preprint arXiv:2402.12327*, 2024.

[35] R. H. Thaler and C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press, 2008.

[36] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 824–24 837, 2022.

[37] A. M. Colman, *Game theory and its applications: in the social and biological sciences.* Psychology Press, 2013.

[38] K. Binmore, *Game theory and the social contract: Just playing.* MIT Press, 1998.

[39] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.

[40] OpenAI, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[41] D. Zhou, N. Schärli, L. Hou, J. Wei, N. Scales, X. Wang, D. Schuurmans, C. Cui, O. Bousquet, Q. Le *et al.*, "Least-to-most prompting enables complex reasoning in large language models," *arXiv preprint arXiv:2205.10625*, 2022.

[42] C. Camerer, *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press, 2003.

[43] A. Rapoport and A. M. Chammah, *Prisoner's dilemma: A study in conflict and cooperation.* University of Michigan Press, 1965.

[44] E. Fehr and S. Gächter, "Cooperation and punishment in public goods experiments," *American Economic Review*, vol. 90, no. 4, pp. 980–994, 2000.

[45] E. Ostrom, *Governing the commons: The evolution of institutions for collective action.* Cambridge University Press, 1990.

[46] J. Friedman, "A non-cooperative equilibrium for supergames," *The Review of Economic Studies*, vol. 38, no. 1, pp. 1–12, 1971.

[47] D. Sally, "Conversation and cooperation in social dilemmas," *Rationality and Society*, vol. 7, no. 1, pp. 58–92, 1995.

[48] F. Mengel, E. Mohlin, and S. Weidenholzer, "Collective incentives and cooperation in teams with imperfect monitoring," Lund University, Department of Economics, Working Papers 2018:11, 2018.

[49] A. Cartwright, "Richard h. thaler: Misbehaving: the making of behavioral economics," *Public Choice*, vol. 164, no. 1, pp. 185–188, Jul. 2015.

[50] J. W. Weibull, *Evolutionary game theory.* MIT Press, 1997.

[51] L. Samuelson, *Evolutionary games and equilibrium selection.* MIT Press, 1997.

[52] R. Boyd and P. J. Richerson, "The evolution of reciprocity in sizable groups," *Journal of Theoretical Biology*, vol. 132, no. 3, pp. 337–356, 1988.

[53] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[54] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22 199–22 213, 2022.

[55] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," *arXiv preprint arXiv:2203.11171*, 2022.

[56] A. Newell and H. A. Simon, *Human problem solving.* Prentice-Hall, 1972.

Appendix

7.1 Additional Tables for Experiments

The additional experimental data are shown in Table 7.1, Table 7.2, Table 7.3, and Table 7.4.

Table 7.1: Additional Table of Experiment 1

Model 1	Model 2	Game Type	p-Value	Finite Rounds	Rounds
ChatGPT	DeepSeek	PD	-	✓	30
ChatGPT	DeepSeek	PD	0.3	✓	30
ChatGPT	DeepSeek	PD	0.6	✓	30
DeepSeek	Kimi	PD	-	✓	30
DeepSeek	Kimi	PD	0.3	✓	30
DeepSeek	Kimi	PD	0.6	✓	30
ChatGPT	Kimi	PD	-	✓	30
ChatGPT	Kimi	PD	0.3	✓	30
ChatGPT	Kimi	PD	0.6	✓	30
ChatGPT	DeepSeek	BoS	-	✓	30
ChatGPT	DeepSeek	BoS	0.3	✓	30
ChatGPT	DeepSeek	BoS	0.6	✓	30
DeepSeek	Kimi	BoS	-	✓	30
DeepSeek	Kimi	BoS	0.3	✓	30
DeepSeek	Kimi	BoS	0.6	✓	30
ChatGPT	Kimi	BoS	-	✓	30
ChatGPT	Kimi	BoS	0.3	✓	30
ChatGPT	Kimi	BoS	0.6	✓	30

Table 7.2: Additional Table of Experiment 2

Model 1	Model 2	Game Type	Finite Rounds	Unknown Rounds	Reverse reasoning
ChatGPT	DeepSeek	PD	✓	✓	✗
DeepSeek	Kimi	PD	✓	✓	✗
ChatGPT	Kimi	PD	✓	✓	✗
ChatGPT	DeepSeek	BoS	✓	✓	✗
DeepSeek	Kimi	BoS	✓	✓	✗
ChatGPT	Kimi	BoS	✓	✓	✗

Table 7.3: Additional Table of Experiment 3

Model 1	Model 2	Game Type	p-Value Path	Memory Window Length	Baseline Only
ChatGPT	DeepSeek	PD	up	5	✓
ChatGPT	DeepSeek	PD	down	5	✓
ChatGPT	DeepSeek	PD	random	5	✓
DeepSeek	Kimi	PD	up	5	✓
DeepSeek	Kimi	PD	down	5	✓
DeepSeek	Kimi	PD	random	5	✓
ChatGPT	Kimi	PD	up	5	✓
ChatGPT	Kimi	PD	down	5	✓
ChatGPT	Kimi	PD	random	5	✓
ChatGPT	DeepSeek	BoS	up	5	✓
ChatGPT	DeepSeek	BoS	down	5	✓
ChatGPT	DeepSeek	BoS	random	5	✓
DeepSeek	Kimi	BoS	up	5	✓
DeepSeek	Kimi	BoS	down	5	✓
DeepSeek	Kimi	BoS	random	5	✓
ChatGPT	Kimi	BoS	up	5	✓
ChatGPT	Kimi	BoS	down	5	✓
ChatGPT	Kimi	BoS	random	5	✓

Table 7.4: Additional Table of Experiment 4

Model 1	Model 2	Game Type	p-Value Path	Memory Window Length	Baseline Only
ChatGPT	DeepSeek	PD	up	5	✗
ChatGPT	DeepSeek	PD	down	5	✗
ChatGPT	DeepSeek	PD	random	5	✗
DeepSeek	Kimi	PD	up	5	✗
DeepSeek	Kimi	PD	down	5	✗
DeepSeek	Kimi	PD	random	5	✗
ChatGPT	Kimi	PD	up	5	✗
ChatGPT	Kimi	PD	down	5	✗
ChatGPT	Kimi	PD	random	5	✗
ChatGPT	DeepSeek	BoS	up	5	✗
ChatGPT	DeepSeek	BoS	down	5	✗
ChatGPT	DeepSeek	BoS	random	5	✗
DeepSeek	Kimi	BoS	up	5	✗
DeepSeek	Kimi	BoS	down	5	✗
DeepSeek	Kimi	BoS	random	5	✗
ChatGPT	Kimi	BoS	up	5	✗
ChatGPT	Kimi	BoS	down	5	✗
ChatGPT	Kimi	BoS	random	5	✗

7.2 Code and Data Availability

The code and data that support the findings of this study are openly available in GitHub at https://github.com/SIRIUS-AAA/The-Influence-of-the-Shadow-of-the-Future-on-the-Evolution-of-Cooperative-Strategies. This repository includes the full implementation of the models, scripts for data processing and analysis, and the datasets necessary to reproduce the results presented in this paper. We encourage researchers to use, extend, and validate our work under the terms of the MIT License provided.

Acknowledgments

My first exposure to strategic thinking came from mathematics and informatics competitions in junior high school. The "winning-strategy" problems in math contests and the greedy-algorithm tasks in informatics taught me how local choices shape global outcomes; these experiences planted the seed for my later study of game theory.

In the second semester of Grade 10, as my reading on game theory deepened, I began to apply its ideas to artificial-intelligence systems. The more I learned, the more I wanted to open the "black box" of AI and investigate how large models actually reason when they play games.

Ms. Jing Ma, my teacher at school, helped me translate this curiosity into a workable research topic, guiding me to narrow my interest to a question that could be studied experimentally.

Through the Science Talent Program for Teens I spent time at Peking University, where I met Professor Meng Guo. He kindly listened to my plans and offered detailed advice on how to analyze the experimental results I would later collect.

I thank all the teachers and friends who supported me during this project. In alphabetical order by surname they are:

Lingyun Chang, Yisai Gao, Meng Guo, Jing Ma, and Ying Shang.

I apologize for any inadvertent omissions.

Finally, I am grateful to the S.-T. Yau High School Science Award for providing a platform that turned a student's curiosity into a rigorous piece of academic work, and to everyone—named and unnamed—who helped me along the way.