MUTUAL RECURSIVE IMPROVEMENT: TOWARD CO-EVOLVING MULTI-AGENT SOCIETIES IN GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

How can LLM agent societies adapt their cooperation as conditions change, and how are the resulting benefits distributed? Strong individual reasoning does not ensure effective coordination among agents with distinct interests, particularly when changing environments make existing strategies and commitments inappropriate. We introduce **Mutual Recursive Improvement (MRI)**, a framework connecting individual learning with institutional adaptation through an interaction–reflection–revision cycle. Human-inspired interventions help agents interpret outcomes, reassess experience, and revise private rules and shared commitments. These revisions shape subsequent interactions, whose outcomes inform further adjustment, while model parameters remain frozen and agents receive no explicit instruction to maximize collective profit. We evaluate eight LLMs over 50 simulated quarters in market-conditioned Dictator, Trust, and Public Goods games, using disclosed historical SPDR S&P 500 ETF Trust (SPY) returns. The final configuration improves allocation quality and collective surplus over baseline in 23 of 24 model–game combinations each, with both improving in 22. Mean allocation quality increases by 29.7%, and mean regime-aligned cooperation rises from 22.0% to 64.6%. These findings demonstrate the potential of structured interventions to support adaptive cooperation through revisable experience and commitments. The code will be open sourced upon acceptance.

1 INTRODUCTION

Large language models (LLMs) support agents that reason, use tools, and interact with one another, providing a foundation for increasingly capable multi-agent systems (Wu et al., 2023). When agents represent participants with distinct interests, their decisions jointly determine both individual payoffs and collective outcomes. Cooperation in such societies requires agents to reconcile competing incentives and establish mutually beneficial arrangements. Strong individual reasoning alone does not ensure successful coordination, as studies of repeated games have shown (Akata et al., 2025). This raises the challenge of translating individual capabilities into effective cooperation among interdependent agents.

Dynamic environments make this challenge more demanding. Changes in external conditions can alter the returns to cooperation, rendering previously useful strategies, commitments, and experience inappropriate. For example, agents representing different firms may need to revise investment decisions and partnership agreements as market conditions change. Effective adaptation therefore requires participants to recognize when existing arrangements have become costly and reconsider how they cooperate. The benefits of adaptation also require scrutiny: higher collective returns can coexist with losses for individual participants, while preserving existing commitments can constrain collective performance. An adaptive agent society must therefore account for both the effectiveness of cooperation and how its benefits are distributed.

LLM agents possess capabilities relevant to this process. Multi-agent debate supports the exchange and refinement of reasoning (Du et al., 2023), while experience-based learning allows agents to extract reusable lessons from past interactions without parameter updates (Zhao et al., 2024). However, these capabilities leave open how individual insights and accumulated experience can drive the revision of cooperation itself. Agents must connect their assessments of changing conditions to adjustments in strategies and shared commitments, then use subsequent outcomes to guide further revision. We

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

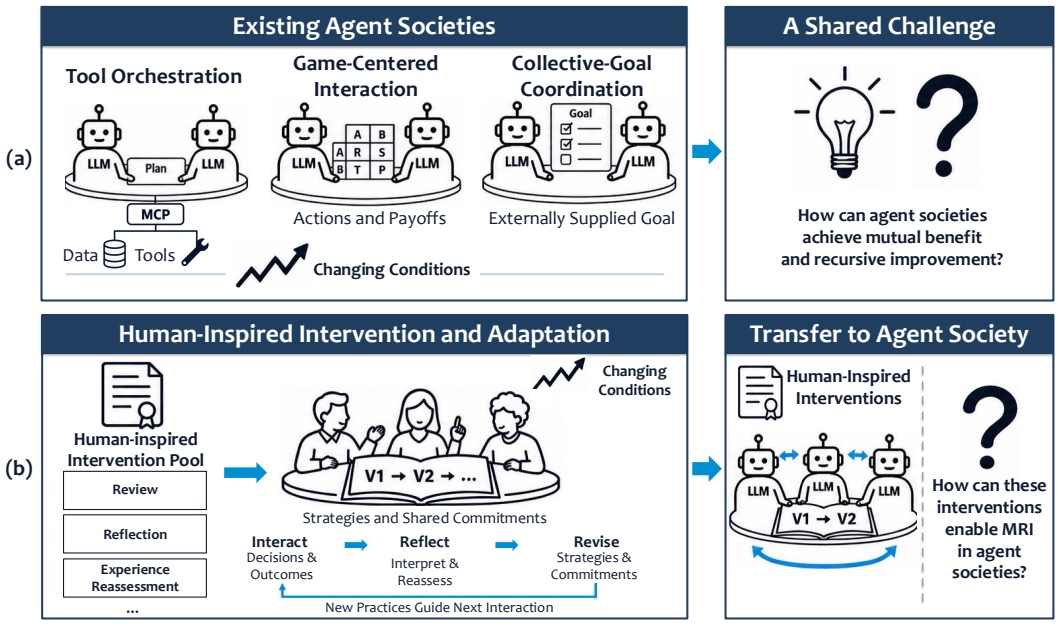


Figure 1: **From human institutional adaptation to MRI.** (a) Existing agent settings motivate questions about adapting cooperation to changing conditions, achieving mutual benefit, and supporting recursive improvement. (b) Human institutional adaptation inspires interventions for reassessing experience and revising coordination strategies and shared commitments. We investigate whether analogous interventions can support mutual recursive improvement in LLM agent societies.

therefore ask: *How can LLM agent societies adapt their cooperation to achieve mutual recursive improvement as conditions change? Can structured interventions improve collective returns while sustaining mutual benefits?*

We introduce **Mutual Recursive Improvement (MRI)**, a framework that connects individual learning with institutional adaptation through an interaction–reflection–revision cycle. Inspired by how human organizations reconsider decisions and agreements, the framework uses reflection, review, and experience reassessment to support the revision of cooperation. These interventions help agents interpret outcomes, reconsider commitments, and reassess the lessons and rules informing their decisions. Revised practices then shape subsequent interactions, whose outcomes provide evidence for further adjustment. As illustrated in Fig. 1(b), this cycle makes coordination strategies and shared commitments explicit targets of recursive improvement, while keeping model parameters frozen and providing no explicit instruction to maximize collective profit.

We study this process across complementary cooperation structures, examining both collective performance and participant outcomes. Our contributions are fourfold:

- **A framework for recursively adapting cooperation.** We introduce MRI, which connects individual learning with the revision of coordination strategies and shared commitments. Human-inspired interventions organize this process into an interaction–reflection–revision cycle, without parameter updates or an explicit collective-profit objective.
- **A controlled sandbox for cooperation under changing conditions.** We develop market-conditioned Dictator, Trust, and Public Goods games capturing unilateral allocation, bilateral reciprocity, and collective contribution. Using disclosed historical SPDR SP 500 ETF Trust (SPY) returns, we evaluate eight LLM agents over 50 rounds, measuring allocation quality, collective surplus, adaptive cooperation, and payoff dispersion.
- **Collective improvements across cooperation structures.** The final configuration improves allocation quality and collective surplus over baseline in 23 of 24 model–game combinations each, with both improving in 22. Allocation quality measures alignment with profit-maximizing choices under changing market conditions, while collective surplus measures system net profit; both weight favorable and unfavorable markets equally. Mean

allocation quality increases by 29.7%, while a cooperation score rewarding cooperation in favorable markets and restraint in unfavorable markets, including reciprocity in Trust, rises from 22.0% to 64.6%. These improvements span all three games, supporting institutional adaptation as a means of translating individual capabilities into stronger collective performance.

- **Evidence on intervention effectiveness and the distribution of benefits.** We measure allocation quality using the Regime-Balanced Optimal-Allocation Score *RBOAS*, which evaluates alignment with profit-maximizing allocation while weighting favorable and unfavorable markets equally. Adding self-reflection at L2 reduces this score relative to L1 in 23 of 24 comparisons, whereas standalone case-based reasoning improves it over baseline in all 24. Higher collective performance also accompanies different distributional outcomes: from B0 to L5, payoff dispersion decreases in Trust but increases in Dictator and Public Goods. These findings distinguish effective intervention from simply adding mechanisms, and collective gains from evidence of mutual benefit.

2 Related Work

MRI connects three research directions: agent adaptation, mutual adaptation, and decision-making in dynamic games. Together, these directions explain what agents update, how updates affect other participants, and how changing payoff conditions alter their economic consequences. Tab. 1 compares representative approaches along these dimensions. Our focus is on persistent adaptation among agents with separate payoffs, evaluated against the objective of maximizing system profit in market-conditioned games.

Table 1: **Comparison of MRI with related work across six dimensions. Conceptual foundations:** T1: experience-based adaptation; T2: strategic interdependence; T3: collective coordination. **Experimental design and analysis:** E1: evaluation across multiple game structures; E2: historical market data determining financial payoffs; E3: cross-round adaptation analysis.

Work	Conceptual Foundations			Experimental Design and Analysis		
	T1: Experience	T2: Strategic	T3: Coordination	E1: Games	E2: Markets	E3: Dynamics
Reflexion (Shinn et al., 2023)	✓	✗	✗	✗	✗	✗
ExpeL (Zhao et al., 2024)	✓	✗	✗	✗	✗	✗
Agent-Pro (Zhang et al., 2024b)	✓	✓	✗	✓	✗	✓
Darwin Gödel Machine (Zhang et al., 2025)	✓	✗	✗	✗	✗	✗
SOTOPIA (Zhou et al., 2024)	✗	✓	✓	✗	✗	✗
Repeated Games (Akata et al., 2025)	✗	✓	✓	✓	✗	✓
Hypothetical Minds (Cross et al., 2024)	✓	✓	✓	✓	✗	✓
ShapeLLM (Garcia Segura et al., 2025)	✓	✓	✓	✓	✗	✓
FinMem (Yu et al., 2023)	✓	✗	✗	✗	✓	✗
TradingAgents (Xiao et al., 2024)	✗	✗	✓	✗	✓	✗
FinAgent (Zhang et al., 2024c)	✓	✗	✗	✗	✓	✓
FinCon (Yu et al., 2024)	✓	✗	✓	✗	✓	✓
MRI (ours)	✓	✓	✓	✓	✓	✓

2.1 Agent Adaptation

LLM frameworks support tool use and multi-agent communication (Yao et al., 2023; Wu et al., 2023), while Model Context Protocol (MCP) standardizes access to external tools and resources (Hou et al., 2026). Beyond this integration layer, Reflexion stores verbal feedback, ExpeL extracts reusable knowledge, and Agent-Pro revises beliefs and policies (Shinn et al., 2023; Zhao et al., 2024; Zhang et al., 2024b). At the system level, AFlow and Automated Design of Agentic Systems optimize workflows and agent designs, while Gödel Agent and the Darwin Gödel Machine modify logic or code (Zhang et al., 2024a; Hu et al., 2024; Yin et al., 2024; Zhang et al., 2025). With model parameters fixed, MRI examines how individual learning informs revisions to coordination strategies and shared commitments across repeated interactions.

2.2 Mutual Adaptation

Mutual adaptation occurs when agents respond to others while influencing subsequent interactions. Hypothetical Minds updates hypotheses about other agents, while LOLA and ShapeLLM consider influence on counterparts' learning (Cross et al., 2024; Foerster et al., 2017; Garcia Segura et al., 2025). Such influence can support coordination or exploitation, and expanded recall can impair cooperation (Garcia Segura et al., 2025; Liu et al., 2026). MRI focuses on how individual adaptation informs revisions to shared cooperation arrangements and affects both collective and participant outcomes.

2.3 Dynamic Games

Repeated games support controlled studies of LLM agent cooperation and strategic behavior (Akata et al., 2025), while changing market conditions motivate state-dependent investment decisions (Guidolin & Timmermann, 2007). FinMem and TradingAgents address financial decisions through layered memory and specialist collaborations, respectively (Yu et al., 2023; Xiao et al., 2024). MRI brings these directions together by studying how changing external payoff conditions affect cooperation among participants with distinct interests.

3 Method

MRI connects individual experience with the revision of coordination strategies and shared commitments through an interaction, reflection, and revision cycle (Fig. 2(b)). Three market-conditioned games situate this cycle in settings where agents have distinct interests and changing conditions alter the returns to cooperation (Fig. 2(a)). Within these settings, human-inspired interventions help agents interpret experience, reassess existing arrangements, and revise the rules informing their decisions (Fig. 2(c)). Model parameters remain fixed, and agents receive no explicit instruction to maximize collective profit. Evaluation therefore examines whether this process improves collective returns and conditional cooperation.

3.1 MRI: A Framework for Co-evolving Agent Societies

MRI provides a framework for studying how agents with distinct interests revise their coordination strategies and shared commitments through repeated interaction in dynamic games. Co-evolution refers to the reciprocal relationship between individual adaptation and shared cooperation arrangements. Individual assessments inform revisions to these arrangements, which in turn shape subsequent decisions. Let $\mathcal{N}$ denote the participants and $t = 1, \ldots, T$ the interaction rounds. Agent i selects a feasible action according to

$$a_{i,t} = \mathcal{L}_\theta^{\text{dec}}\left(o_{i,t}, H_{i,t-1}, Z_{i,t}^{\text{dec}}\right), \qquad a_{i,t} \in \mathcal{A}_{i,t}, \tag{1}$$

where $o_{i,t}$ contains observable environmental information, the agent's role, and decision constraints; $H_{i,t-1}$ is its accessible history; and $Z_{i,t}^{\text{dec}}$ is the intervention context available at decision time, including retained private rules, public commitments, and any analysis generated before the decision. Model parameters θ remain fixed, and $\mathcal{A}_{i,t}$ denotes the feasible action set. Observations also include preceding actions visible to the agent within the current round.

The MRI framework connects **interaction, reflection, and revision** in a recursive cycle (Fig. 2(b)). Interaction produces actions and payoff feedback, which enter the updated history $H_{i,t}$. Reflection helps agents interpret these outcomes in relation to environmental conditions and the behavior of other participants, and reassess whether existing strategies and commitments remain appropriate. Revision translates these assessments into updates to the private rules and public agreements informing subsequent decisions. The update of retained intervention resources is summarized as

$$Z_{i,t} = \mathcal{U}_{\mathcal{I},i}\left(H_{i,t}, Z_{i,t-1}\right), \tag{2}$$

where $\mathcal{I}$ specifies the enabled interventions and $\mathcal{U}_{\mathcal{I},i}$ denotes the update procedure for agent i. This procedure involves exchanges and agreement formation among participants, rather than independent updates alone. Retained resources $Z_{i,t}$ subsequently contribute to $Z_{i,t+1}^{\text{dec}}$, together with newly available information and any analysis performed before the next decision. Individual experience can thus inform shared coordination, while subsequent interaction provides evidence for further revision.

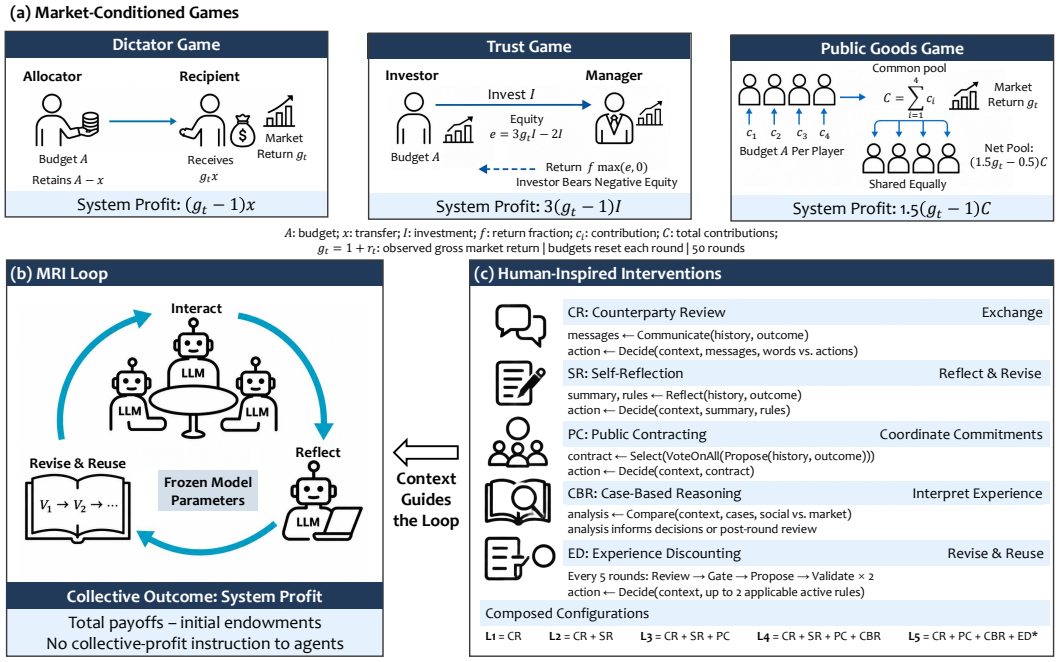


Figure 2: **Overview of Mutual Recursive Improvement (MRI).** (a) Observed market returns condition payoffs in three games. (b) Agents use interaction outcomes to reflect on and revise private rules and shared commitments, which guide subsequent decisions with model parameters frozen. (c) Five human-inspired interventions support this cycle. L1–L4 progressively add CR, SR, PC, and CBR; L5 replaces routine SR with periodic review and rule validation through ED. Configurations are evaluated independently, without instructing agents to maximize collective profit.

This feedback makes adaptation recursive without assuming that every revision improves performance. Rules and agreements guide decisions through context; they neither alter the game's payoff functions nor mechanically enforce actions. Agents receive no explicit instruction to maximize collective profit.

3.2 Market-Conditioned Games

Three market-conditioned games instantiate the adaptation cycle under unilateral allocation, bilateral reciprocity, and collective contribution (Fig. 2(a)). Let $g_t = 1 + r_t$ denote the gross market return, A the allocation budget, $u_{i,t}$ participant i's end-of-round funds, and E_0 the total initial endowment, with $E_0 = A$ for the two-player games and $E_0 = 4A$ for Public Goods. For joint action $\mathbf{a}_t$, we define the system profit and corresponding system-optimal action as

$$W_t(\mathbf{a}_t; g_t) = \sum_{i \in \mathcal{N}} u_{i,t} - E_0, \qquad \mathbf{a}_t^\star = \arg\max_{\mathbf{a}_t \in \mathcal{A}_t} W_t(\mathbf{a}_t; g_t). \tag{3}$$

Agents observe g_t before acting, while each game specifies the feasible actions and payoff structure. The three settings differ in how cooperative decisions are initiated and how their outcomes are distributed under the same market-conditioned environment.

Dictator game. An allocator chooses a transfer $x_t \in [0, A]$ to a recipient who cannot alter it (Camerer et al., 2004). The allocator retains $A - x_t$, while the transferred amount receives the market return:

$$u_{\mathrm{D},t} = A - x_t, \qquad u_{\mathrm{R},t} = g_t x_t, \qquad W_t = (g_t - 1)x_t. \tag{4}$$

Trust game. An investor selects $I_t \in [0, A]$, after which a manager observes I_t and chooses a return fraction $f_t \in [0, 1]$ (Camerer et al., 2004). Borrowing $2I_t$ finances a position of $3I_t$, leaving $e_t = 3g_t I_t - 2I_t$ after repayment:

$$\begin{aligned} u_{\mathrm{inv},t} &= A - I_t + f_t \max(e_t, 0) + \min(e_t, 0), \\ u_{\mathrm{mgr},t} &= (1 - f_t)\max(e_t, 0), \\ W_t &= 3(g_t - 1)I_t. \end{aligned} \tag{5}$$

Public Goods game. Four participants contribute $c_{i,t} \in [0, A]$ to a common pool $C_t = \sum_{i=1}^{4} c_{i,t}$ (Ledyard, 1995). Borrowing $0.5C_t$ finances a position of $1.5C_t$, whose proceeds after repayment are shared equally:

$$u_{i,t} = A - c_{i,t} + \frac{(1.5g_t - 0.5)C_t}{4}, \qquad W_t = 1.5(g_t - 1)C_t. \tag{6}$$

Together, the three games provide distinct forms of strategic interdependence while sharing the same market-conditioned objective, allowing adaptive cooperation to be evaluated against the system-optimal action $\mathbf{a}_t^\star$.

3.3 Human-Inspired Intervention Pool

The intervention pool supports the interpretation of experience and revision of cooperation within MRI. We distinguish two functional categories, instantiated by the five mechanisms summarized in Fig. 2(c). These choices do not exhaust the interventions that the framework can accommodate. Further details of the intervention procedures and configurations are provided in Appendix A.

- **Social coordination interventions.** These interventions connect individual assessments with shared cooperation arrangements through **Counterparty Review (CR)**, which supports reassessment of cooperation expectations, and **Public Contracting (PC)**, which establishes revisable public commitments. Their outputs guide subsequent decisions without changing payoff rules or enforcing actions.
- **Individual reflection and experience revision.** These interventions support the interpretation and reuse of experience through **Self-Reflection (SR)**, which develops private lessons and rules, **Case-Based Reasoning (CBR)**, which examines their contextual applicability, and **Experience Discounting (ED)**, which selectively retains rules through validation of individual payoff consequences with other actions and market conditions fixed. Here, discounting denotes selective retention rather than numerical time decay.

The resulting resources inform $Z_{i,t}$ and $Z_{i,t}^{\mathrm{dec}}$. Inspired by the connection between personal understanding and social knowledge building (Stahl, 2013), we progressively compose these mechanisms. L1 introduces CR, L2 adds SR, and L3 adds PC, connecting information exchange, private reflection, and public commitments. L4 adds CBR to strengthen contextual interpretation. L5 retains CR, PC, and CBR while replacing routine reflection with periodic experience review and validated rule maintenance. These configurations extend support from forming cooperation arrangements to interpreting and reassessing them; their effectiveness is tested empirically.

3.4 Evaluation Metrics

We evaluate allocation quality, collective surplus, adaptive cooperation, and payoff dispersion. Following Section 3.2, W_t denotes system net profit. Let $\langle\cdot\rangle_{\mathcal{T}_+}$ and $\langle\cdot\rangle_{\mathcal{T}_-}$ denote averages over valid favorable and unfavorable market rounds, respectively. The first three metrics exclude neutral rounds. The first two assign equal weight to favorable and unfavorable market regimes, while the third uses the unweighted harmonic mean of its two regime-specific components. Auxiliary definitions and boundary conditions are provided in Appendix B.

Allocation quality. The Regime-Balanced Optimal-Allocation Score (*RBOAS*), denoted by O_T, measures whether normalized allocation q_t approaches one in favorable markets and zero in unfavorable markets:

$$O_T = \frac{1}{2}\left(\langle q\rangle_{\mathcal{T}_+} + \langle 1 - q\rangle_{\mathcal{T}_-}\right). \tag{7}$$

Higher values indicate better alignment with the system-optimal allocation.

Collective surplus. Regime-Balanced Cumulative Surplus (*RBCS*), denoted by S_T, measures system net profit over a balanced horizon of T rounds:

$$S_T = \frac{T}{2}\left(\langle W\rangle_{\mathcal{T}_+} + \langle W\rangle_{\mathcal{T}_-}\right). \tag{8}$$

Higher values indicate greater collective surplus in game-specific payoff units.

Adaptive cooperation. The Regime-Aligned Cooperation Score (*RACS*), denoted by K_T, combines favorable-market cooperation b_T and unfavorable-market restraint d_T through their harmonic mean:

$$K_T = \begin{cases} \dfrac{2b_T d_T}{b_T + d_T}, & b_T + d_T > 0, \\ 0, & b_T + d_T = 0. \end{cases} \tag{9}$$

Higher values require both behaviors. Cooperation includes allocation in all games and effective reciprocity in Trust.

Payoff dispersion. The Payoff Coefficient of Variation (*CV*), denoted by ν_T, measures dispersion across participants' cumulative payoffs, using their population mean $\bar{U}_T$ and standard deviation σ_T:

$$\nu_T = \frac{\sigma_T}{\bar{U}_T}. \tag{10}$$

Lower values indicate more equal payoffs. Mutual benefit additionally requires participant-level gains relative to a specified baseline.

4 Experiments

4.1 Experimental Setup

We evaluate the three games in Section 3.2 over 50 rounds, each corresponding to a calendar quarter from 2014 Q2 to 2026 Q3. All configurations share a frozen market sequence derived from unadjusted daily SPY closing prices (Yahoo Finance, 2026), with g_t equal to the last available closing price divided by the first within quarter t, excluding dividends. The observation $o_{i,t}$ includes the signed net return $100(g_t - 1)\%$, the agent's role, feasible actions, and payoff rules; dates and price levels are withheld. Agents receive no instruction to maximize system profit W_t. The task thus evaluates adaptation to disclosed market conditions rather than return forecasting.

We evaluate eight models: DeepSeek-V3.2-Exp (DeepSeek-AI, 2025), Llama 4 Maverick (Meta AI, 2025), GPT-4.1 mini (OpenAI, 2025), GPT-4o (Hurst et al., 2024), Kimi K2 (Moonshot AI, 2025), Mistral Medium 3 (Mistral AI, 2025), Qwen3-30B-A3B-Instruct-2507 (Qwen Team, 2025), and Qwen3.5-35B-A3B (Qwen Team, 2026). In the main experiments, each society uses a single backbone model with frozen parameters θ and fixed inference settings. The baseline B0 provides game observations $o_{i,t}$ and accessible interaction history $H_{i,t-1}$. Building on these inputs, L1–L5 instantiate the intervention configurations defined in Section 3.3, each running independently for 50 rounds. L5 replaces per-round reflection with reviews every five rounds and validated rule maintenance.

4.2 Collective Performance Across Intervention Configurations

Collective improvement across interaction structures. Tab. 2 shows that L5 raises overall mean O_T from 0.5679 to 0.7368, a relative gain of 29.7%. The gains span all three games: 28.5% in Dictator game, 27.2% in Trust game, and 33.6% in Public Goods game. L5 improves both O_T and S_T over B0 in 22 of 24 model–game comparisons. Across configurations, overall mean O_T first rises to 0.6583 at L1, falls to 0.5438 at L2, and then increases through 0.5533, 0.6198, and 0.7368 at L3–L5. The results identify a shared pattern of early gains from counterparty review and stronger final performance under the configurations supporting contextual interpretation and validated rule maintenance. L5 also yields the highest mean cumulative participant payoff in all three games. However, mean payoff *CV* falls from 16.45% to 6.93% in Trust but rises relative to B0 in Dictator and Public Goods, indicating that higher average payoffs do not consistently coincide with lower relative dispersion.

Cooperation improves through two pronounced gains. Fig. 3 reveals two large increases in aggregate cooperation, separated by modest intermediate gains. Mean K_T rises from 0.220 at B0 to 0.497 at L1, increases gradually to 0.525 through L2–L4, and then reaches 0.646 at L5. The B0-to-L1 and L4-to-L5 transitions together account for 93.4% of the total increase, identifying counterparty review and the L5 review and validation configuration as the main points of improvement along the evaluated sequence. The game-level trajectories reveal complementary patterns: the Dictator game

and the Public Goods game achieve the largest net gains, increasing by 0.463 and 0.479, respectively, while the Trust game shows an early rise to 0.660 at L1, a decline to 0.538 at L2, and a recovery to 0.698 at L5. Thus, the shared endpoint is stronger conditional cooperation, reached through different responses to intervention composition.

Table 2: **Collective performance across intervention configurations.** Model rows report *RBOAS* (O_T) / *RBCS* (S_T). Purple and orange shading indicate gains over B0: relative gains in O_T and positive absolute gains in S_T, respectively, with the latter scaled within each game. Summary rows report equally weighted mean O_T and mean cumulative participant payoff / *CV* (ν_T, in percent). Blue left arrows show changes from B0. Red right arrows show changes from the preceding configuration. Bold marks the highest mean O_T in each summary row.

Game	Model	B0	L1	L2	L3	L4	L5	Best level (gain vs. B0)
Dictator	DeepSeek-V3.2	.596 / 3.53	.585 / 3.00	.546 / 0.79	.558 / 1.61	.583 / 2.45	.680 / 5.38	**L5** (+14.1%)
	Llama 4 Maverick	.520 / 0.64	.622 / 3.66	.560 / 0.75	.560 / 1.60	.724 / 6.70	.642 / 4.32	**L4** (+39.2%)
	GPT-4.1 mini	.554 / 1.69	.664 / 5.48	.499 / −0.30	.528 / 0.46	.579 / 2.65	.670 / 5.86	**L5** (+20.9%)
	GPT-4o	.506 / 0.13	.534 / 0.50	.510 / −0.30	.506 / −0.28	.537 / 0.66	.606 / 2.78	**L5** (+19.8%)
	Kimi K2	.503 / 0.07	.528 / 1.14	.483 / −0.84	.515 / −0.25	.521 / 1.59	.622 / 4.01	**L5** (+23.7%)
	Mistral Medium 3	.507 / 0.21	.720 / 6.47	.628 / 5.40	.588 / 2.19	.626 / 3.91	.723 / 7.05	**L5** (+42.6%)
	Qwen3-30B-A3B	.499 / 0.16	.505 / −0.26	.487 / −0.81	.488 / −0.78	.542 / 1.56	.738 / 7.30	**L5** (+47.9%)
	Qwen3.5-35B-A3B	.507 / 0.13	.623 / 3.39	.522 / 0.69	.580 / 2.21	.606 / 2.46	.706 / 6.46	**L5** (+39.3%)
	Mean RBOAS	0.5240	↑0.5976↑	↑0.5294↓	↑0.5404↑	↑0.5898↑	↑**0.6734**↑	**L5** (+28.5%)
	Mean payoff / CV	250.61/11.29%	253.11/10.27%	254.01/3.31%	253.95/5.50%	254.16/9.95%	254.41/16.24%	—
Trust game	DeepSeek-V3.2	.740 / 30.44	.782 / 32.19	.587 / 12.01	.611 / 12.28	.657 / 18.33	.771 / 32.24	**L1** (+5.7%)
	Llama 4 Maverick	.641 / 15.42	.872 / 40.50	.550 / 4.43	.595 / 5.02	.789 / 32.24	.839 / 38.10	**L1** (+36.0%)
	GPT-4.1 mini	.522 / 1.64	.875 / 40.31	.631 / 6.93	.672 / 13.44	.807 / 23.34	.880 / 39.83	**L5** (+68.6%)
	GPT-4o	.666 / 19.35	.614 / 17.79	.510 / −0.03	.510 / −0.77	.555 / 6.66	.683 / 18.61	**L5** (+2.6%)
	Kimi K2	.500 / 0.00	.629 / 16.01	.546 / 7.96	.608 / 10.78	.661 / 18.80	.644 / 16.31	**L4** (+32.2%)
	Mistral Medium 3	.547 / 8.65	.583 / 14.55	.502 / −0.16	.571 / 5.94	.503 / 1.00	.830 / 37.54	**L5** (+51.7%)
	Qwen3-30B-A3B	.669 / 17.93	.911 / 43.83	.559 / 8.15	.488 / 1.80	.517 / −0.22	.901 / 40.99	**L1** (+36.2%)
	Qwen3.5-35B-A3B	.663 / 13.12	.662 / 17.91	.602 / 14.33	.632 / 18.92	.634 / 12.19	.748 / 26.49	**L5** (+12.8%)
	Mean RBOAS	0.6185	↑0.7410↑	↓0.5609↓	↓0.5859↑	↑0.6404↑	↑**0.7870**↑	**L5** (+27.2%)
	Mean payoff / CV	260.27/16.45%	271.87/15.16%	262.08/13.01%	264.74/10.17%	264.88/8.02%	274.26/6.93%	—
Public Goods	DeepSeek-V3.2	.526 / 4.27	.584 / 15.98	.554 / 2.63	.542 / 6.28	.676 / 39.33	.863 / 75.72	**L5** (+64.1%)
	Llama 4 Maverick	.656 / 37.32	.649 / 33.57	.556 / 9.42	.503 / −3.20	.905 / 73.45	.775 / 57.16	**L4** (+38.0%)
	GPT-4.1 mini	.570 / 15.55	.842 / 72.80	.519 / 1.43	.532 / 4.66	.623 / 26.76	.787 / 70.92	**L1** (+47.7%)
	GPT-4o	.523 / 4.17	.554 / 11.55	.578 / 18.21	.534 / 4.21	.684 / 31.46	.840 / 65.89	**L5** (+60.6%)
	Kimi K2	.510 / 1.60	.568 / 14.95	.529 / 9.99	.522 / 4.95	.550 / 10.13	.767 / 51.25	**L5** (+50.4%)
	Mistral Medium 3	.595 / 15.06	.829 / 70.97	.557 / 10.63	.520 / −0.04	.510 / −7.70	.655 / 33.35	**L1** (+39.3%)
	Qwen3-30B-A3B	.510 / 1.52	.549 / 8.08	.541 / 1.10	.537 / 2.01	.571 / 24.56	.730 / 37.40	**L5** (+43.1%)
	Qwen3.5-35B-A3B	.600 / 18.11	.514 / 1.73	.496 / −2.51	.578 / 14.89	.515 / 1.79	.583 / 19.96	**B0** (0.0%)
	Mean RBOAS	0.5613	↑0.6361↑	↓0.5413↓	↓0.5335↓	↑0.6293↑	↑**0.7500**↑	**L5** (+33.6%)
	Mean payoff / CV	504.51/2.51%	511.27/4.51%	512.25/3.48%	511.78/1.85%	513.82/2.75%	521.53/3.25%	—
Overall mean RBOAS		0.5679	↑0.6583↑	↓0.5438↓	↓0.5533↑	↑0.6198↑	↑**0.7368**↑	**L5** (+29.7%)

4.3 Intervention Effects and Dependence on Case Support

Case-based support produces the broadest standalone gains. Adding each intervention separately to B0 reveals substantial differences in effectiveness. Across all eight models and three games, CBR improves O_T in all 24 comparisons and achieves the highest standalone score in 22. Full results are reported in Appendix C (Tab. C1).

Frequent case exposure supports stronger allocation adaptation. We next vary CBR exposure within L4 while retaining CR, SR, and PC between exposures. Tab. 3 shows that providing cases every round achieves the highest O_T in all six model–game pairs. Averaged equally across these pairs, O_T is 0.7378 with every-round exposure, compared with 0.5763 every five rounds and 0.5892 every ten rounds. The largest gap appears for Llama 4 Maverick in Public Goods game: 0.9050 versus 0.5190 and 0.4988, respectively. This pattern indicates that performance under L4 depends substantially on continued access to case support. The dependence is not simply monotonic: in Trust game, exposure every ten rounds outperforms every five rounds for both models. The experiment therefore supports the benefit of every-round exposure under the tested settings, rather than a uniform frequency–performance relationship.

Mixed-model performance. To examine intervention effects in a heterogeneous group, we pair a GPT-4.1 mini agent with a Kimi K2 agent in the Trust game and evaluate the group under B0 and L5. We also compare its L5 outcomes with those of the corresponding single-model groups (Tab. 4). Within the mixed group, L5 raises both participants' cumulative payoffs over B0. Collective surplus

S_T rises from -1.63 to 32.19, while payoff *CV* falls from 7.18% to 0.67%. At L5, the mixed group has lower *CV* than either single-model group, with S_T between theirs. Kimi K2 earns more and GPT-4.1 mini earns less than in their respective single-model L5 groups, indicating that the payoff effects of model mixing differ across participants in this run.

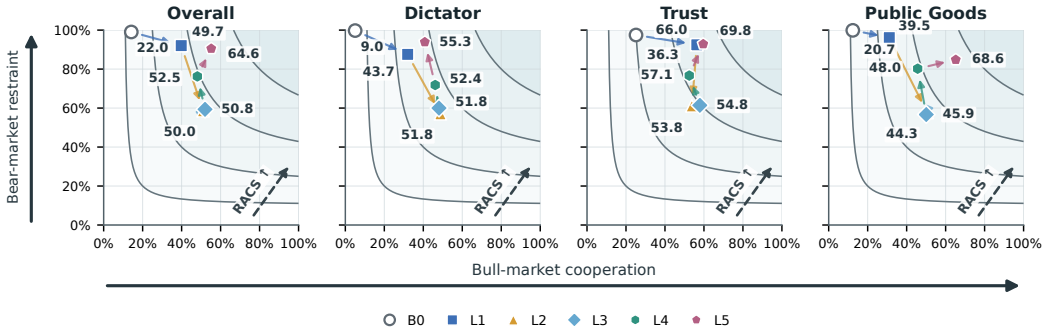


Figure 3: **Regime-aligned cooperation across intervention levels.** Panels show Overall, Dictator game, Trust game, and Public Goods game. Markers show mean bull-market cooperation and bear-market restraint across eight models. Overall additionally averages across games. Labels report mean model-level *RACS* (%). Colored arrows trace B0 through L1–L5. Dashed arrows indicate the direction of joint improvement.

Table 3: CBR exposure-frequency ablation for two selected models. Cells report *RBOAS* over 50 rounds. Columns indicate how often the CBR case collection is provided.

Game	Model	Every 10 rounds	Every 5 rounds	Every round
Dictator	GPT-4.1 mini	0.5090	0.5152	0.579
	Llama 4 Maverick	0.6484	0.6989	0.724
Trust Game	GPT-4.1 mini	0.7089	0.6491	0.807
	Llama 4 Maverick	0.6395	0.5385	0.789
Public Goods	GPT-4.1 mini	0.5308	0.5370	0.623
	Llama 4 Maverick	0.4988	0.5190	0.905

Table 4: **Single-model and multi-model groups in the Trust Game.** Single-model groups follow the main experiment settings but with 50 new rounds. The multi-model group contains two agents, one using GPT-4.1 mini and the other Kimi K2. $\Delta_{\text{B0}} = \text{Multi}_{L5} - \text{Multi}_{B0}$. $\Delta_{\text{L5}} = \text{Multi}_{L5} - \text{Single}_{L5}$. Deltas are absolute differences, with *CV* differences in percentage points.

	Metric	Model	Multi B0	Single L5	Multi L5	Δ_{B0}	Δ_{L5}
(a)	RBOAS ↑	GPT-4.1 mini	0.5458	0.8562	0.8301	+0.2843	−0.0261
		Kimi K2	0.5000	0.6500	0.8500	+0.3500	+0.2000
	Cumulative payoff ↑	GPT-4.1 mini	232.77	283.55	276.93	+44.16	−6.62
		Kimi K2	268.79	263.58	280.68	+11.88	+17.10
(b)	RBCS ↑	GPT-4.1 mini		41.92			−9.74
		Kimi K2	−1.63	18.35	32.19	+33.81	+13.84
	Payoff CV ↓	GPT-4.1 mini		3.33%			−2.66
		Kimi K2	7.18%	3.41%	0.67%	−6.51	−2.74

5 Conclusion and Limitations

MRI improves mean allocation quality by 29.7% and raises regime-aligned cooperation from 22.0% to 64.6%, with simultaneous allocation and collective-surplus gains in 22 of 24 model–game comparisons. These gains depend on intervention composition: adding routine reflection at L2 reduces allocation quality relative to L1 in 23 comparisons, whereas standalone case-based reasoning improves it over baseline in all 24. Every-round case exposure also outperforms both reduced-frequency settings in all six tested pairs, highlighting the importance of continued support.

Higher collective returns accompany lower payoff dispersion in Trust but higher dispersion in Dictator and Public Goods, distinguishing collective improvement from more equal outcomes. Together, these findings identify intervention composition, support frequency, and benefit distribution as central to adaptive cooperation with frozen models. Our evidence is limited to controlled games with disclosed market returns and 50-round interactions, Longer evaluations under uncertain conditions are needed to assess whether these gains support sustained mutual benefit.

AI USE STATEMENT

In this work, we used generative AI tools for assist in the writing of proofs, design or provide feedback on research methodology or experiments, implement methods, assist with translation, support qualitative and thematic data analysis, interpret results. We have not used generative AI tools for generate synthetic data sets, refine hypotheses, clean and reformat dataset and formulate mathematical claims. We have reviewed all AI-assisted work. We thoroughly checked all LLM-generated texts inside this paper and LLM-generated code was verified and tested for correctness. We take responsibility for the final content of this work.

ETHICS STATEMENT

This paper has no ethical concerns.

REPRODUCIBILITY STATEMENT

The code will be open sourced upon acceptance.

REFERENCES

Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. Playing repeated games with large language models. *Nature Human Behaviour*, 9(7):1380–1390, 2025.

Colin F Camerer, Ernst Fehr, et al. Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, 97:55–95, 2004.

Logan Cross, Violet Xiang, Agam Bhatia, Daniel L. K. Yamins, and Nick Haber. Hypothetical Minds: Scaffolding theory of mind for multi-agent tasks with large language models. *arXiv preprint arXiv:2407.07086*, 2024. URL https://arxiv.org/abs/2407.07086.

DeepSeek-AI. DeepSeek-V3.2-Exp. Official model card, 2025. URL https://huggingface.co/deepseek-ai/DeepSeek-V3.2-Exp. Accessed September 24, 2026.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate. *arXiv preprint arXiv:2305.14325*, 2023.

Jakob N. Foerster, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with opponent-learning awareness. *arXiv preprint arXiv:1709.04326*, 2017. URL https://arxiv.org/abs/1709.04326.

Marta Emili Garcia Segura, Stephen Hailes, and Mirco Musolesi. Opponent shaping in LLM agents. *arXiv preprint arXiv:2510.08255*, 2025. URL https://arxiv.org/abs/2510.08255.

Massimo Guidolin and Allan Timmermann. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544, 2007. doi: 10.1016/j.jedc.2006.12.004. URL https://www.sciencedirect.com/science/article/pii/S0165188906002272.

Xinyi Hou, Yanjie Zhao, Shenao Wang, and Haoyu Wang. Model Context Protocol (MCP): Landscape, security threats, and future research directions. *ACM Transactions on Software Engineering and Methodology*, 35(10), 2026. doi: 10.1145/3796519. URL https://doi.org/10.1145/3796519.

Shengran Hu, Cong Lu, and Jeff Clune. Automated design of agentic systems. *arXiv preprint arXiv:2408.08435*, 2024. URL https://arxiv.org/abs/2408.08435.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

John O. Ledyard. Public goods: A survey of experimental research. In John H. Kagel and Alvin E. Roth (eds.), *The Handbook of Experimental Economics*, pp. 111–194. Princeton University Press, 1995. doi: 10.1515/9780691213255-004.

Jiayuan Liu, Tianqin Li, Shiyi Du, Xin Luo, Haoxuan Zeng, Emanuel Tewolde, Tai Sing Lee, Tonghan Wang, Carl Kingsford, and Vincent Conitzer. The memory curse: How expanded recall erodes cooperative intent in LLM agents. *arXiv preprint arXiv:2605.08060*, 2026. URL https://arxiv.org/abs/2605.08060.

Meta AI. The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation. Official release announcement, April 2025. URL https://ai.meta.com/blog/llama-4-multimodal-intelligence/. Accessed September 24, 2026.

Mistral AI. Medium is the new large. Official release announcement, May 2025. URL https://mistral.ai/news/mistral-medium-3/. Accessed September 24, 2026.

Moonshot AI. Kimi-K2-Instruct-0905. Official model card, 2025. URL https://huggingface.co/moonshotai/Kimi-K2-Instruct-0905. Accessed September 24, 2026.

OpenAI. Introducing GPT-4.1 in the API. Official release announcement, April 2025. URL https://openai.com/index/gpt-4-1/. Accessed September 24, 2026.

Qwen Team. Qwen3-30B-A3B-Instruct-2507. Official model card, 2025. URL https://huggingface.co/Qwen/Qwen3-30B-A3B-Instruct-2507. Accessed September 24, 2026.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL https://qwen.ai/blog?id=qwen3.5. Accessed September 24, 2026.

Noah Shinn, Federico Cassano, Edward Berman, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *arXiv preprint arXiv:2303.11366*, 2023. URL https://arxiv.org/abs/2303.11366.

Gerry Stahl. A model of collaborative knowledge-building. In *Fourth international conference of the learning sciences*, volume 10, pp. 70–77, 2013.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W. White, Doug Burger, and Chi Wang. AutoGen: Enabling next-gen LLM applications via multi-agent conversation. *arXiv preprint arXiv:2308.08155*, 2023. URL https://arxiv.org/abs/2308.08155.

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. TradingAgents: Multi-agents LLM financial trading framework. *arXiv preprint arXiv:2412.20138*, 2024. URL https://arxiv.org/abs/2412.20138.

Yahoo Finance. SPDR S&P 500 ETF Trust (SPY): Historical Data, 2026. URL https://finance.yahoo.com/quote/SPY/history/.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*, 2023.

Xunjian Yin, Xinyi Wang, Liangming Pan, Li Lin, Xiaojun Wan, and William Yang Wang. Gödel Agent: A self-referential agent framework for recursive self-improvement. *arXiv preprint arXiv:2410.04444*, 2024. URL https://arxiv.org/abs/2410.04444.

Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. FinMem: A performance-enhanced LLM trading agent with layered memory and character design. *arXiv preprint arXiv:2311.13743*, 2023. URL https://arxiv.org/abs/2311.13743.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Zhenyu Cui, Rong Liu, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. FinCon: A synthesized LLM multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. In *Advances in Neural Information Processing Systems*, volume 37, pp. 137010–137045, 2024. URL https://proceedings.neurips.cc/paper/2024/hash/f7ae4fe91d96f50abc2211f09b6a7e49-Abstract-Conference.html.

Jenny Zhang, Shengran Hu, Cong Lu, Robert Lange, and Jeff Clune. Darwin Gödel Machine: Open-ended evolution of self-improving agents. *arXiv preprint arXiv:2505.22954*, 2025. URL https://arxiv.org/abs/2505.22954.

Jiayi Zhang, Jinyu Xiang, Zhaoyang Yu, Fengwei Teng, Xionghui Chen, Jiaqi Chen, Mingchen Zhuge, Xin Cheng, Sirui Hong, Jinlin Wang, Bingnan Zheng, Bang Liu, Yuyu Luo, and Chenglin Wu. AFlow: Automating agentic workflow generation. *arXiv preprint arXiv:2410.10762*, 2024a. URL https://arxiv.org/abs/2410.10762.

Wenqi Zhang, Ke Tang, Hai Wu, Mengna Wang, Yongliang Shen, Guiyang Hou, Zeqi Tan, Peng Li, Yueting Zhuang, and Weiming Lu. Agent-Pro: Learning to evolve via policy-level reflection and optimization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024b. URL https://arxiv.org/abs/2402.17574.

Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, Longtao Zheng, Xinrun Wang, and Bo An. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024c. doi: 10.1145/3637528.3671801. URL https://doi.org/10.1145/3637528.3671801.

Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. Expel: Llm agents are experiential learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 19632–19642, 2024.

Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and Maarten Sap. SOTOPIA: Interactive evaluation for social intelligence in language agents. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=mM7VurbA4r.

APPENDIX CONTENTS

A Details of Human-Inspired Interventions **14**

A.1 Counterparty Review 14

A.2 Self-Reflection 14

A.3 Public Contracting 14

A.4 Case-Based Reasoning 15

A.5 Experience Discounting 15

B Evaluation Metric Definitions **17**

B.1 Regime-specific averaging 17

B.2 Allocation quality 17

B.3 Collective surplus 17

B.4 Adaptive cooperation 17

B.5 Payoff dispersion 18

C Single-intervention ablations **19**

A Details of Human-Inspired Interventions

This section supplements Section 3.3 with the information flow, execution procedures, and retained resources of the five interventions. We retain the notation of the main text: $H_{i,t}$ is agent i's accessible interaction history, $Z_{i,t}$ contains resources retained after round t, and $Z_{i,t}^{\mathrm{dec}}$ is the intervention context available when the agent acts. Post-round updates inform subsequent decisions; they do not change actions already taken. Memory remains associated with participant identity, with role-specific applicability checked when roles change. All interventions operate with fixed model parameters and leave the game's payoff functions unchanged.

A.1 Counterparty Review

Counterparty Review (CR) makes public communication available for reassessing cooperation expectations. After a round, agents generate messages from their accessible histories and observed outcomes. These messages enter subsequent decision context.

During decision-making, agents assess counterparties' stated intentions against their observed behavior. The assessment concerns the credibility of commitments and expectations of future cooperation. CR does not require a separate numerical trust score: the review is incorporated into the decision prompt. Public messages provide information for agents to interpret, rather than binding instructions.

The communication prompt supplies the completed-round summary and requests a JSON response containing reasoning and one public message. Messages are shared with all participants, whereas private reflection is not shared. The next decision prompt contains the participant's accumulated observable action–outcome history and the latest public messages. Each imported message is limited to 400 characters. Earlier messages are not accumulated as a separate transcript.

A.2 Self-Reflection

Self-Reflection (SR) produces a private summary and decision-guiding rules after each round. Agents review their actions and outcomes to identify lessons for subsequent interactions. Each structured rule specifies an obligation, a trigger, a proposed self-penalty, and a revision condition. These are fields of a rule, not a requirement to generate four separate rules.

The latest valid summary and rules inform subsequent decisions through the agent's retained resources. Self-penalties are textual commitments: the environment does not deduct funds or otherwise enforce them. Revision conditions allow agents to reconsider whether a rule remains appropriate as circumstances change.

The reflection prompt requests a private summary and a list of rules with the fields `id`, `obligation`, `trigger`, `self_penalty`, and `revise_when`. It uses up to eight recent history entries and public messages when enabled. The retained summary is limited to 600 characters. Accepted rule updates replace the previous rule list rather than being appended indefinitely. In composed configurations, an empty or unparseable rule list leaves the previous valid rules unchanged. The standalone SR runner instead requires a valid nonempty summary and rule list, retries invalid outputs, and stops the cell if validation continues to fail.

A.3 Public Contracting

Public Contracting (PC) forms revisable public commitments. Agents propose contracts describing the parties, duties, benefits, and conditions for adjustment. Each participant receives the valid proposals and votes on every proposal by signing or declining. The voting context includes the supplied round summary, decision history, public messages, and self rules when enabled; proposals retain their original text and proposer identity.

Selection prioritizes unanimously approved proposals. Otherwise, approval count, structural completeness, and a rotating proposer priority determine selection and resolve ties. The selected contract enters the next round's context. In composed configurations, public messages and contract proposals can be generated together.

Contracts provide a shared reference for cooperation. They neither modify feasible actions and payoff rules nor mechanically enforce compliance.

Proposals use a structured JSON contract with a title, summary, participant entries, adaptation conditions, and breach conditions. Each participant entry specifies rights, benefits, and duties. Structural completeness measures coverage of these three fields for each party and the breach field. Voting returns one `SIGN` or `DECLINE` entry per candidate. Participants may sign multiple proposals but cannot amend or merge them. If several proposals receive unanimous approval, the first in the fixed agenda order is selected. The selector's fallback for no valid proposal retains the previous contract, or supplies a neutral continuation contract if none exists. This fallback does not override a runner configured to stop on invalid outputs.

A.4 Case-Based Reasoning

Case-Based Reasoning (CBR) supplies a fixed collection of cases describing prior contexts, decisions, and outcomes. Agents use these cases to interpret current experience, distinguishing counterparty behavior from market effects and considering whether a previous lesson applies. For example, this distinction helps avoid attributing a market-driven loss solely to an uncooperative counterparty. Case analysis provides contextual guidance rather than an enforced action rule.

The delivery point differs across experimental settings. Standalone CBR supplies the full case collection before each decision and incorporates the resulting analysis into decision-making. In composed configurations, case-based interpretation supports enabled post-round communication and reflection, using the updated interaction history. Consequently, the standalone and composed experiments do not isolate an identical intervention at an identical point in the decision process.

The exposure-frequency experiment varies case provision within L4 across every round, every five rounds, and every ten rounds. CR, SR, and PC remain enabled between exposures.

The case collection is fixed and human-curated, with historical episodes organized around counterparty behavior and environmental conditions. The prompts request an explicit interpretation along these two dimensions and identification of a relevant precedent. Cases serve as analogies rather than additional observations from the current run.

In the exposure-frequency experiment, cases are provided in rounds 4, 9, 14, and every fifth round thereafter through round 49, or in rounds 9, 19, 29, 39, and 49 for the ten-round schedule. Every-round exposure begins in round 1. On other rounds, the full case collection is omitted, but existing summaries, rules, and contracts are not reset. Case-derived guidance can therefore persist through these retained resources.

A.5 Experience Discounting

Experience Discounting (ED) selectively retains decision-guiding rules through periodic review and payoff validation. Discounting refers to selective retention rather than numerical time decay. In L5, this procedure replaces routine per-round self-reflection while retaining CR, PC, and CBR.

Review and proposal. Every five rounds, agents review recent experience. When an evidence gate is satisfied, the review may produce a candidate rule specifying an applicability condition, a role-valid action, and a predicted individual outcome.

Payoff validation. In subsequent applicable rounds, the suggested action is evaluated against the observed action. For a candidate rule R, let $a_{i,t}^{(R)}$ denote its suggested action for agent i. Let $\mathbf{a}_t^{(R)}$ be the joint action obtained by replacing $a_{i,t}$ with $a_{i,t}^{(R)}$ in the observed joint action $\mathbf{a}_t$, leaving all other participants' actions unchanged. The individual payoff difference is

$$\Delta u_{i,t}(R) = u_{i,t}\left(\mathbf{a}_t^{(R)}; g_t\right) - u_{i,t}(\mathbf{a}_t; g_t)\,. \tag{A.1}$$

The market return is also held fixed. A candidate becomes active after two consecutive applicable checks with $\Delta u_{i,t}(R) \geq 0$; a failed check rejects it.

This validation concerns the agent's own payoff under a fixed-action comparison. It does not test collective optimality or estimate how other participants would respond to the candidate action.

Rule reuse. At decision time, up to two applicable active rules are supplied, prioritized by usefulness and recency. These rules enter decision context as guidance; agents remain free to choose another feasible action. The procedure therefore validates retained advice without converting it into an enforced policy.

The evidence gate requires at least three fresh observations for the participant's current role since its last accepted candidate. It also requires either a prediction mismatch or at least two observations with the same recent market direction, payoff sign, and high/low action category. Prediction mismatches concern market direction for allocation decisions and individual payoff sign for the Trust manager. Historical cases do not count toward this evidence requirement.

The review prompt supplies completed observations, existing rules, and realized own outcomes. It requests an evidence-based summary. When the gate opens, a second prompt requests at most one candidate with `condition_pattern`, `action`, and `predicted_outcome`, or a null candidate. Conditions are restricted to always applicable, recent market up/down, or last own payoff positive/nonpositive. Here, market up includes a neutral return, and own payoff is net of the current role's endowment. Invalid or duplicate candidates are not added.

Rule memories are specific to participant identity and role. Inapplicable rounds are skipped without resetting a candidate's validation streak. Active rules are ranked first by their count of cited uses with nonnegative own net payoff and then by creation recency. A rule is retired after three cited uses with negative own net payoff. These counts describe outcomes when a rule is cited rather than causal gains attributable to that rule.

B Evaluation Metric Definitions

We retain the notation of Section 3.2. System net profit W_t is defined in Eq. equation 3. All metrics are computed for a fixed game, intervention configuration, and run of T rounds.

B.1 Regime-specific averaging

Let $\mathcal{T}_+$ and $\mathcal{T}_-$ denote the sets of valid rounds with $g_t > 1$ and $g_t < 1$, respectively. For a real-valued round sequence $(z_t)_{t=1}^T$ and a nonempty round set $\mathcal{S} \subseteq \{1, \dots, T\}$, define

$$\langle z \rangle_{\mathcal{S}} = \frac{1}{|\mathcal{S}|} \sum_{t \in \mathcal{S}} z_t, \tag{B.1}$$

where $|\mathcal{S}|$ is the number of rounds in $\mathcal{S}$. The allocation-quality, collective-surplus, and adaptive-cooperation metrics require both market regimes to be nonempty and exclude neutral rounds with $g_t = 1$. If either regime is empty, these three metrics are undefined.

B.2 Allocation quality

Using the allocation budget $A > 0$ and game-specific actions from Section 3.2, Eqs. equation 4–equation 6, define normalized market allocation q_t by

$$q_t = \begin{cases} \dfrac{x_t}{A}, & \text{Dictator game,} \\ \dfrac{I_t}{A}, & \text{Trust game,} \\ \dfrac{C_t}{4A}, & \text{Public Goods game.} \end{cases} \tag{B.2}$$

Thus $q_t \in [0, 1]$. Under these payoff structures, system net profit is maximized by $q_t = 1$ when $g_t > 1$ and by $q_t = 0$ when $g_t < 1$.

The Regime-Balanced Optimal-Allocation Score (***RBOAS***), denoted by O_T, is defined in Section 3.4, Eq. equation 7. It uses q_t from Eq. equation B.2 and the regime averages from Appendix B.1, Eq. equation B.1. The score lies in $[0, 1]$, with higher values indicating closer alignment with the system-optimal allocation. A constant allocation across both regimes yields $O_T = 1/2$.

B.3 Collective surplus

The Regime-Balanced Cumulative Surplus (***RBCS***), denoted by S_T, is defined in Section 3.4, Eq. equation 8. Its inputs are round-level system net profits W_t from Section 3.2, Eq. equation 3, and the regime averages from Appendix B.1, Eq. equation B.1.

The factor $T/2$ assigns an equivalent half-horizon to each regime, preserving the magnitude and sign of profit in game-specific payoff units. Thus S_T is a regime-balanced cumulative quantity and need not equal the sum of profits under the observed regime frequencies.

B.4 Adaptive cooperation

Define round-level cooperation intensity $k_t \in [0, 1]$. In the Dictator and Public Goods games, cooperation is represented by normalized allocation q_t (Appendix B.2, Eq. equation B.2):

$$k_t = q_t. \tag{B.3}$$

In the Trust game, cooperation also includes effective reciprocity:

$$k_t = \begin{cases} \dfrac{q_t + f_t \mathbf{1}\{e_t > 0\}}{2}, & I_t > 0, \\ 0, & I_t = 0. \end{cases} \tag{B.4}$$

The investment I_t, return fraction f_t, and equity e_t retain their definitions in Section 3.2, in the text preceding Eq. equation 5. The indicator $\mathbf{1}\{\cdot\}$ equals one when its condition holds and zero

otherwise. Reciprocity contributes only when positive equity is available, and zero investment gives zero cooperation intensity.

Define favorable-market cooperation b_T and unfavorable-market restraint d_T by

$$b_T = \langle k \rangle_{\mathcal{T}_+}, \qquad d_T = \langle 1 - k \rangle_{\mathcal{T}_-}. \tag{B.5}$$

These averages use Eq. equation B.1, so $b_T, d_T \in [0, 1]$.

The Regime-Aligned Cooperation Score (*RACS*), denoted by K_T, is defined in Section 3.4, Eq. equation 9. It combines b_T and d_T through their unweighted harmonic mean. The score lies in $[0, 1]$ and equals zero if either component is zero, including when both are zero. High values require both components to be high.

B.5 Payoff dispersion

For a complete T-round run, let U_i denote the cumulative payoff of fixed participant identity $i \in \mathcal{N}$ across role changes. Let N be the number of participants and $\bar{U}_T$ their mean cumulative payoff:

$$U_i = \sum_{t=1}^{T} u_{i,t}, \qquad \bar{U}_T = \frac{1}{N} \sum_{i \in \mathcal{N}} U_i, \qquad N = |\mathcal{N}|. \tag{B.6}$$

Here $u_{i,t}$ denotes end-of-round funds as defined in Section 3.2, Eq. equation 3. Cumulative payoffs include retained endowments and all rounds at their observed frequencies, including neutral rounds.

The population standard deviation σ_T across participants is

$$\sigma_T = \left[\frac{1}{N} \sum_{i \in \mathcal{N}} (U_i - \bar{U}_T)^2 \right]^{1/2}. \tag{B.7}$$

The Payoff Coefficient of Variation (*CV*), denoted by ν_T, is defined in Section 3.4, Eq. equation 10. Its inputs are $\bar{U}_T$ and σ_T from Eqs. equation B.6 and equation B.7, respectively. We use this coefficient for nonnegative cumulative payoffs with $\bar{U}_T > 0$. Otherwise, it is undefined.

Lower ν_T indicates lower relative dispersion across participants' cumulative payoffs. For each run, mean cumulative participant payoff corresponds to $\bar{U}_T$, and CV in percent corresponds to $100\nu_T\%$.

C SINGLE-INTERVENTION ABLATIONS

CBR's game-level mean O_T reaches 0.7121 in Dictator game, 0.8811 in Trust game, and 0.8656 in Public Goods game, corresponding to gains of 35.9%, 42.5%, and 54.2% over B0. Other interventions exhibit more game-specific effects: standalone PC raises mean O_T from 0.6185 to 0.6513 in Trust game and from 0.5613 to 0.5869 in Public Goods game, while Dictator game remains near baseline at 0.5218. These results distinguish broadly useful case support from interventions whose benefits depend more strongly on the interaction structure. Tab. C1 reports the complete single-intervention results for Dictator, Trust, and Public Goods, in the same order as the main experiments. Each intervention is evaluated separately against B0.

Table C1: **Single-intervention ablation.** Each intervention is added separately to B0. Cells report regime-balanced RBOAS / RBCS. Light blue marks RBOAS above B0. Darker blue marks the best standalone RBOAS in each row. The final column identifies that intervention and its relative RBOAS gain over B0.

Game	Model	B0	B0+CR	B0+SR	B0+PC	B0+CBR	B0+ED	Best intervention (gain vs. B0)
Dictator	DeepSeek-V3.2	.596 / 3.53	.585 / 3.00	.510 / 0.13	.494 / −0.39	.841 / 11.97	.599 / 2.52	CBR (+41.1%)
	Llama 4 Maverick	.520 / 0.64	.622 / 3.66	.645 / 4.18	.653 / 5.98	.711 / 6.55	.503 / 0.16	CBR (+36.7%)
	GPT-4.1 mini	.554 / 1.69	.664 / 5.48	.555 / 2.22	.502 / −0.08	.726 / 7.18	.520 / 0.21	CBR (+31.0%)
	GPT-4o	.506 / 0.13	.534 / 0.50	.500 / −0.30	.500 / −0.30	.628 / 4.58	.500 / 0.00	CBR (+24.1%)
	Kimi K2	.503 / 0.07	.528 / 1.14	.531 / 1.19	.514 / 0.97	.544 / 1.14	.500 / 0.00	CBR (+8.2%)
	Mistral Medium 3	.507 / 0.21	.720 / 6.47	.567 / 1.18	.493 / −0.44	.622 / 5.04	.500 / 0.00	CR (+42.0%)
	Qwen3-30B-A3B	.499 / 0.16	.505 / −0.26	.517 / 0.13	.495 / −0.19	.645 / 4.66	.503 / 0.07	CBR (+29.3%)
	Qwen3.5-35B-A3B	.507 / 0.13	.623 / 3.39	.503 / 0.11	.523 / 0.65	.980 / 15.76	.472 / 0.26	CBR (+93.3%)
Trust game	DeepSeek-V3.2	.740 / 30.44	.782 / 32.19	.600 / 11.81	.773 / 31.21	.941 / 41.80	.524 / 1.64	CBR (+27.2%)
	Llama 4 Maverick	.641 / 15.42	.872 / 40.50	.671 / 20.60	.843 / 39.11	.894 / 40.86	.666 / 15.82	CBR (+39.5%)
	GPT-4.1 mini	.522 / 1.64	.875 / 40.31	.610 / 6.68	.653 / 17.75	.993 / 48.01	.732 / 25.77	CBR (+90.2%)
	GPT-4o	.666 / 19.35	.614 / 17.79	.500 / −0.38	.532 / 1.50	.928 / 46.55	.500 / 0.00	CBR (+39.3%)
	Kimi K2	.500 / 0.00	.629 / 16.01	.519 / 3.21	.568 / 9.81	.757 / 34.28	.500 / 0.00	CBR (+51.4%)
	Mistral Medium 3	.547 / 8.65	.583 / 14.55	.483 / −0.12	.579 / 7.22	.757 / 30.25	.509 / 2.85	CBR (+38.4%)
	Qwen3-30B-A3B	.669 / 17.93	.911 / 43.83	.559 / 9.23	.661 / 17.60	.941 / 46.90	.802 / 34.26	CBR (+40.7%)
	Qwen3.5-35B-A3B	.663 / 13.12	.662 / 17.91	.564 / 10.04	.601 / 8.93	.838 / 37.77	.510 / 0.83	CBR (+26.4%)
Public Goods	DeepSeek-V3.2	.526 / 4.27	.584 / 15.98	.526 / 3.78	.495 / 4.70	.905 / 75.25	.527 / 4.06	CBR (+72.1%)
	Llama 4 Maverick	.656 / 37.32	.649 / 33.57	.536 / 7.12	.612 / 30.77	.658 / 31.52	.638 / 24.32	CBR (+0.3%)
	GPT-4.1 mini	.570 / 15.55	.842 / 72.80	.563 / 3.07	.524 / 2.89	.982 / 95.89	.903 / 81.39	CBR (+72.3%)
	GPT-4o	.523 / 4.17	.554 / 11.55	.506 / 2.17	.536 / 2.66	.997 / 96.11	.538 / 3.68	CBR (+90.6%)
	Kimi K2	.510 / 1.60	.568 / 14.95	.519 / 4.05	.828 / 68.43	.618 / 27.63	.510 / 1.94	PC (+62.4%)
	Mistral Medium 3	.595 / 15.06	.829 / 70.97	.508 / 0.08	.505 / −3.64	.985 / 95.35	.699 / 32.15	CBR (+65.5%)
	Qwen3-30B-A3B	.510 / 1.52	.549 / 8.08	.531 / 2.12	.508 / −2.74	.782 / 55.27	.528 / 5.35	CBR (+53.3%)
	Qwen3.5-35B-A3B	.600 / 18.11	.514 / 1.73	.524 / 2.94	.687 / 36.95	.998 / 95.74	.625 / 21.95	CBR (+66.3%)