

000 001 TALK, JUDGE, COOPERATE: GOSSIP-DRIVEN IN- 002 DIRECT RECIPROCITY IN SELF-INTERESTED LLM 003 AGENTS 004 005

006 **Anonymous authors**
007 Paper under double-blind review
008
009
010
011
012

ABSTRACT

013 Decentralized self-interested LLM agents often struggle to sustain cooperation
014 when they are placed in mixed-motive tasks. Incentivizing cooperation is chal-
015 lenging while many previous studies have focused on compromised settings. We
016 address this challenge by introducing public gossip as a decentralized reputation
017 mechanism in agents' interactions. Our analysis provides both theoretical guaran-
018 tees and empirical evidence that gossip can promote cooperation in indirect recip-
019 rocity games. Building on this insight, we propose the Agentic LIinguistic Gossip
020 Network (ALIGN), an automated agentic framework where agents share open-
021 ended gossip to evaluate one another's trustworthiness and establish reciprocity
022 with cooperative partners. Experiments show that ALIGN not only improves co-
023 operation and social welfare but also resists malicious entrants, as defectors are
024 reliably identified and excluded.
025
026

1 INTRODUCTION

027 As LLM agents become increasingly deployed, they will inevitably interact with one another across
028 diverse domains. However, interactions among decentralized mixed-motive agents can lead to un-
029 expected and potentially harmful outcomes (Hammond et al., 2025). A key challenge is the lack of
030 monitoring mechanisms to identify and discourage uncooperative behavior (Hammond et al., 2025;
031 Ren et al., 2025; Dafoe et al., 2020). For example, self-interested agents who secretly free-ride in
032 public goods games undermine collective welfare, as they have no incentive to contribute without
033 credible threats of ostracism in future interactions. Existing approaches attempt to mitigate this
034 challenge by modifying the environment. Some seed altruistic agents into the population (Ren et al.,
035 2025; Zhou et al., 2025; Liu et al., 2024; Leng & Yuan, 2023), which modifies the game structures
036 that are more inclined to cooperation but less realistic in practice. Others impose moral constraints
037 through prompting (Sreedhar et al., 2025; Piatti et al., 2024; Tennant et al., 2024), such as instructing
038 agents to "think in the other person's shoes," or restrict defection artificially. While these methods
039 can induce cooperation, they compromise the fidelity of mixed-motive settings. This raises a central
040 question: *How can decentralized, self-interested LLM agents sustain cooperation in mixed-motive*
041 *tasks without manipulating incentives or introducing altruistic agents?*

042 We address this question by leveraging *public gossip*, a verbal monitoring mechanism in which
043 agents on their own share evaluative messages about others' behavior with the entire community.
044 Unlike centralized reputation systems that require a trusted authority, gossip enables decentralized
045 agents to transmit reputation information through communication, allowing others to update their
046 beliefs and strategies accordingly. For example, online reviews or word-of-mouth recommendations
047 often guide individuals' decisions in the absence of direct experience. Another example is that
048 gossip can promote cooperation in repeated social dilemmas, as we will prove in Section 3.

049 To enable adaptive decision-making through public gossip, we introduce **ALIGN** (Agentic LIinguis-
050 tic Gossip Network), an automated agentic framework where self-interested LLM agents strate-
051 gically adjust their behavior based on shared public messages and their own past experiences. Lever-
052 aging the reasoning capabilities of LLMs, agents can interpret open-ended evaluative messages ex-
053 pressed in hierarchical tones and refine their strategies through verbalized reflection (Zhang et al.,
2024; Shinn et al., 2023). This framework is general enough to host most multi-agent LLMs scenar-

ios without introducing additional authority and altruism. In the experiments, we focus on indirect reciprocity games because they isolate the effect of gossip. In fact, in repeated direct encounters, cooperation can already arise through direct reciprocity, but in indirect reciprocity settings gossip is essential for cooperative outcomes. Experiments across finite-horizon and infinite-horizon social dilemmas with diverse LLMs demonstrate that ALIGN substantially increases cooperation and social welfare relative to non-gossiping baselines. Moreover, ALIGN is robust to malicious entrants: defectors are identified and ostracized through negative gossip, preserving community-level cooperation. These results position ALIGN as an adaptive and decentralized mechanism for norm emergence in LLM societies, bridging theoretical models of indirect reciprocity with practical implementations in large-scale generative agents. Our empirical benchmark further reveals that reasoning LLMs are not inherently selfish, but tend to cooperate only when it is strategically optimal, whereas chat LLMs sometimes cooperate even when defection is the dominant strategy and defect when cooperation is beneficial that is both unreasonable and unpredictable. These insights offers guidance for the future design of cooperative mechanisms as reasoning LLMs grow more powerful and widely deployed, ensuring their interactions remain beneficial in decentralized societies.

2 RELATED WORK

2.1 INDIRECT RECIPROCITY AND GOSSIP

Reciprocal altruism (Trivers, 1971), where an agent incurs a cost to help another with the expectation of future return, is a powerful mechanism for sustaining cooperation in mixed-motive interactions. *Direct reciprocity* (Trivers, 1971) arises when the same pair of agents interacts repeatedly. For example, in the infinite-horizon prisoner’s dilemma (Rapoport, 1965), strategies such as Tit-for-Tat (cooperating initially and then mirroring the partner’s previous action) can stabilize mutual cooperation. Rather than being restricted to repeated encounters, *indirect reciprocity* (Ohtsuki & Iwasa, 2006; 2004; Nowak & Sigmund, 1998b;a) generalizes cooperation to large, dynamic populations, where agents help those known to have helped others. Therefore, to achieve indirect reciprocity, the reputation of everyone needs to be continually assessed and shared in the population. Classic models of indirect reciprocity include first-order *image scores* (Nowak & Sigmund, 1998a), where an agent’s reputation depends solely on their own actions, and second-order norms (Ohtsuki & Iwasa, 2006), where the assessment of an action also considers the coplayer’s reputation (e.g., punishing those who help defectors). These models, however, focus on static norms and behavioral rules, and often assume centralized monitoring, which limits their applicability to decentralized systems.

In contrast, gossip offers a decentralized reputation mechanism (Jolly & Chang, 2021; Santos et al., 2021; Giardini & Wittek, 2019; Wu et al., 2016). Public gossip is especially effective, as it broadens coverage and facilitates collective coordination (Bénabou & Tirole, 2006; Blume et al., 2008). Recent work has extended these ideas to LLM agents: Vallinder & Hughes (2024) showed cooperation can emerge through cultural evolution in finite-horizon donation games but only under favorable initial conditions and for a specific LLM (Claude 3.5 Sonnet). Ren et al. (2025) proposed RepuNet, where agents update explicit reputation scores via encounters and gossip to decide whether to maintain connection with others. The approach however requires seeding altruistic agents, which diverts from the motivation of studying decentralized self-interested agents, while the method was tested only on GPT-4o mini (OpenAI, 2024).

2.2 LLM AGENTS FOR INDIRECT RECIPROCITY

LLM agents are increasingly employed to model strategic and social interactions in mixed-motive multi-agent settings (Ren et al., 2025; Kempinski et al., 2025; Piedrahita et al., 2025; Willis et al., 2025; Piatti et al., 2024; Vallinder & Hughes, 2024; Park et al., 2023; Leng & Yuan, 2023). While these studies demonstrate that LLMs can negotiate, cooperate, and reason about norms, they typically remain at the level of empirical demonstrations or qualitative observations, without a system study on when and how cooperation can arise among self-interested agents. Prior approaches also vary in their assumptions. Some rely on seeding altruistic agents to sustain cooperation (Ren et al., 2025), others consider finite-horizon social dilemmas where cooperation is not an equilibrium (Vallinder & Hughes, 2024), or examines social dilemmas without specifying the horizon. In fact, horizon length critically determines the feasibility of cooperative equilibria (Piedrahita et al., 2025). In contrast, our work combines game-theoretic analysis with empirical evaluation, providing

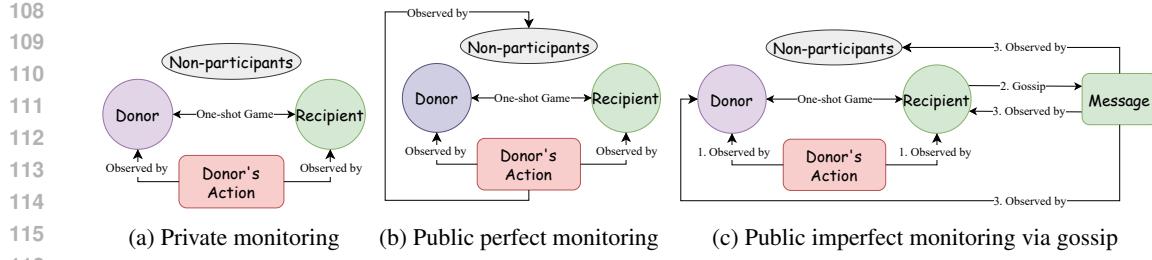


Figure 1: Illustration of three monitoring structures: (a) **Private monitoring**, only the donor and recipient observe the donor’s action; (b) **Public perfect monitoring**, all agents observe the donor’s action; (c) **Public imperfect monitoring via gossip**, only the donor and recipient observe the action, and all agents observe the public signal broadcast by the recipient.

both theoretical and practical insights into the mechanisms that enable cooperation through gossip in LLM agent societies. We show that cooperation can be sustained among fully self-interested agents in repeated social dilemmas through public gossip, with formal guarantees for the existence of cooperative equilibria.

3 GAME-THEORETIC SETUP AND PROPOSITIONS

In this section, we investigate when and how self-interested agents can sustain indirect reciprocity through a public gossip mechanism in social dilemmas where direct reciprocity is disabled.

3.1 REPEATED DONATION GAME

Nowak & Sigmund (1998a;b) introduced the donation game to study indirect reciprocity among self-interested agents who can choose to provide a benefit b to another at a personal cost c . The repeated donation game is formally defined in Definition 1.

Definition 1 (Repeated Donation Game). A repeated donation game is a tuple $\mathcal{G} = (\mathcal{N}, T, \mathcal{A}, (\mathcal{O}_i)_{i \in \mathcal{N}}, e, c, b, \gamma)$, where \mathcal{N} is the set of agents, $T \in \mathbb{N} \cup \{\infty\}$ is the game horizon, $\mathcal{A} = \{\text{cooperate, defect}\}$ is the action space, \mathcal{O}_i is the observation space of agent $i \in \mathcal{N}$, $e \in \mathbb{R}^+$ is the common initial endowment for every agent $i \in \mathcal{N}$, $c > 0$ is the cost of cooperation to the donor, $b > c$ is the benefit of cooperation to the recipient, $\gamma \in (0, 1]$ is the discount factor. At each timestep $t = 1, \dots, T$, two agents are **randomly paired without replacement**, one is assigned as the donor $i \in \mathcal{N}$ and the other as the recipient $j \in \mathcal{N} \setminus \{i\}$. After observation of remaining resources, the donor chooses $a_i^t \in \mathcal{A}$. If $a_i^t = \text{cooperate}$, then immediate rewards are $r_i^t = -c$ and $r_j^t = b$; otherwise, $r_i^t = r_j^t = 0$. After each timestep, the donor i and recipient j are required to switch roles in the subsequent round, with i acting as the recipient and j as the donor. **Subject to this role-switching constraint, agents are then randomly re-matched with new partners.**

The strategy of each agent $i \in \mathcal{N}$ is represented by its action policy $\pi_i : \mathcal{O}_i \mapsto \mathcal{A}$, which maps the agent’s observation to an action. Each agent’s objective is to maximize its expected discounted utility over the horizon, defined as $G_i = \sum_{t=1}^T \gamma^{t-1} r_i^t, \forall i \in \mathcal{N}$. This repeated donation game creates a social dilemma: donating increases collective welfare as $b > c$, but incurs an immediate personal cost to the donor. Note that in each round, two agents are randomly paired to play a one-shot donation game, and no pair can meet more than once. This setting eliminates the possibility of direct reciprocity, since the recipient cannot repay the donor in future encounters. Therefore, self-interested agents will cooperate only if they build indirect reciprocity within the community (i.e. a donor cooperates with a recipient if the recipient is likely to cooperate with others).

3.2 EQUILIBRIUM ANALYSIS

We now analyze the existence of equilibria in the repeated donation game under different assumptions about the game horizon and monitoring structure. We consider two horizon settings: (i) *finite horizon*, where the game lasts a known number of rounds ($T < \infty$); and (ii) *infinite horizon*, where

162 the game continues indefinitely ($T = \infty$). We also examine three monitoring structures (Figure 1):
 163 (i) *private monitoring*, where only the paired donor and recipient observe the donor’s action;
 164 (ii) *public perfect monitoring*, where the full history of actions is publicly observed; and (iii) *public*
 165 *imperfect monitoring via gossip*, where only participants observe the action directly, but the recip-
 166 ient broadcasts a signal about the action to all agents. Assuming all agents are self-interested, we
 167 summarize our main propositions below, with detailed proofs provided in Appendix A.

168 We first consider the finite-horizon case. As stated in Proposition
 169 1, mutual defection is the unique subgame-perfect equilib-
 170 rium (SPE) in this setting, even with perfect monitoring. This
 171 aligns with the classical backward induction result in finitely re-
 172 peated games (Benoit et al., 1984), where the last round’s dominant
 173 strategy of defection unravels cooperation in all preceding rounds.
 174 Therefore, cooperation cannot be sustained among self-interested
 175 agents in finite-horizon repeated donation games.

176 **Proposition 1.** *In a finite-horizon repeated donation game, the unique SPE for all agents is to defect*
 177 *in every timestep.*

179 We next consider the infinite-horizon case. As stated in Proposition 2, cooperation fails with pri-
 180 vate monitoring. In the absence of public monitoring mechanisms, each agent optimizes utility in
 181 isolation. Since donation is personally costly and lacks guaranteed return, defection remains the
 182 dominant strategy.

183 **Proposition 2.** *In an infinite-horizon repeated donation game with private monitoring, the unique*
 184 *SPE is for all agents to defect in every timestep.*

185 In contrast, when agents can perfectly monitor others’ behavior, sustained cooperation becomes
 186 possible in the infinite-horizon setting. Proposition 3 shows that an SPE exist in the infinite-horizon
 187 donation game with public perfect monitoring if the common discount factor satisfies $\gamma \geq \frac{c}{b}$. This
 188 condition ensures that agents value future payoffs sufficiently to make cooperation worthwhile. For
 189 example, each donor chooses to cooperate with their matched recipient only if the recipient has
 190 never defected in the past; otherwise, the donor defects. This strategy creates a credible threat for
 191 non-cooperative behavior, because if an agent deviates by defecting against a cooperative recipient,
 192 they are labeled as a defector and will be punished by all future donors through defection. In this
 193 case, no one has an incentive to deviate from cooperation, thus indirect reciprocity can be sustained
 194 indefinitely through conditional strategies based on public histories.

195 **Proposition 3.** *In an infinite-horizon repeated donation game with public perfect monitoring. If the*
 196 *common discount factor satisfies $\gamma \geq \frac{c}{b}$, then there exists an SPE where cooperation is sustained*
 197 *through conditional strategies based on observed histories.*

199 However, public perfect monitoring is often impractical in decentralized systems, as agents may not
 200 have access to all others’ behavioral histories. This raises the question of whether cooperation can
 201 still emerge under more relaxed monitoring assumptions. To explore this, we introduce the repeated
 202 donation game with public gossip (Definition 2), a variant of the repeated donation game (Defini-
 203 tion 1) that incorporates public imperfect monitoring via gossip. In this framework, recipients can
 204 broadcast public messages after observing donors’ actions. These public messages provide imper-
 205 fect information about donors’ behavior to the community. As stated in Proposition 4, even under
 206 this public imperfect monitoring structure without requiring full transparency, cooperation can still
 207 be sustained if agents condition their strategies on the public signals. For example, each recipient
 208 honestly reports the donor’s action in their public message. Then, similar to the public perfect moni-
 209 toring case, donors cooperate only if the recipient has never been reported as a defector in the past;
 210 otherwise, they defect. Under such strategies, no agent has an incentive to deviate from cooperation.
 211 This finding motivates our further exploration of public gossip as a mechanism for sustaining coop-
 212 eration among self-interested LLM agents. With strong capabilities in nuanced text generation and
 213 interpretation, LLM agents can adapt to different games without requiring handcrafted signal spaces
 as in traditional game-theoretic models.

214 **Definition 2** (Repeated Donation Game with Public Gossip). *A repeated donation game with public*
 215 *gossip is a tuple $\mathcal{G} = (\mathcal{N}, T, \mathcal{A}, \mathcal{M}, (\mathcal{O}_i)_{i \in \mathcal{N}}, e, c, b, \gamma)$, which extends the repeated donation game*
in Definition 1 by introducing a message space \mathcal{M} . At each timestep $t = 1, \dots, T$, after the donor’s

Table 1: Donation Game

Donor’s Action	Rewards
Cooperate	$(-c, b)$
Defect	$(0, 0)$

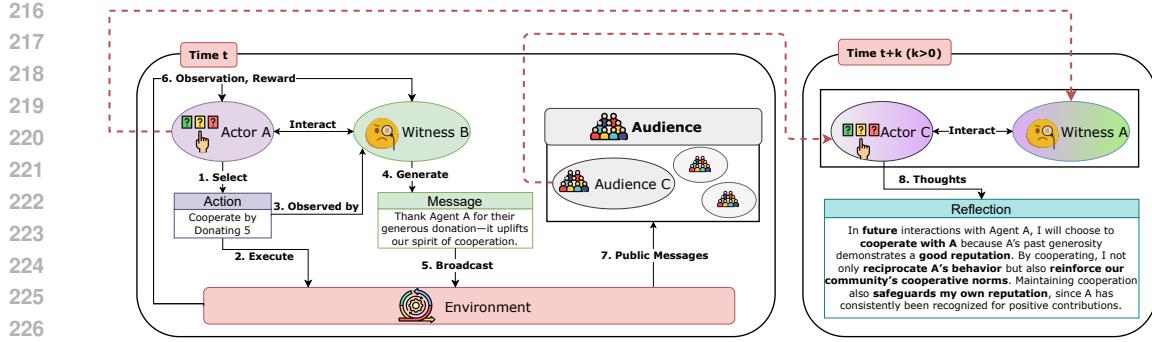


Figure 2: Decision Process in ALIGN

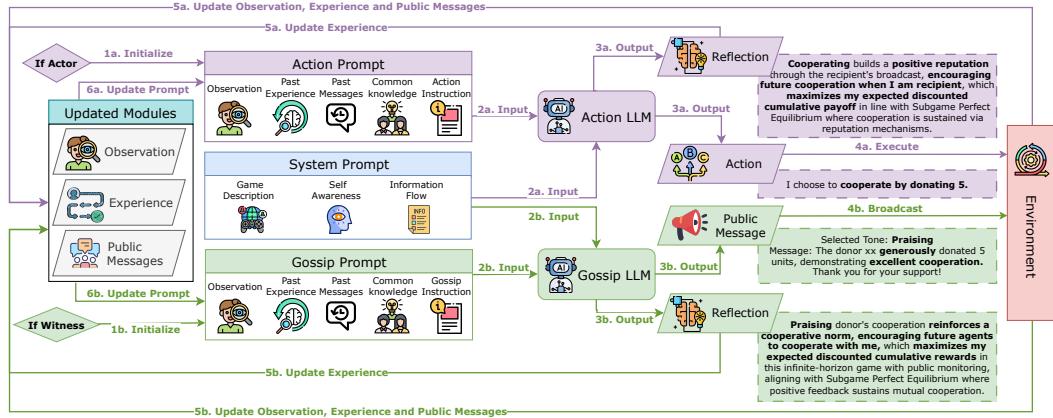


Figure 3: Generative Agent Architecture of ALIGN

action is chosen and payoffs are realized as in Definition 1, the recipient observes the action and broadcasts a public message $m_j^t \in \mathcal{M}$ to all agents in \mathcal{N} .

The strategy of each agent $i \in \mathcal{N}$ is represented by both its action policy and gossip policy (π_i, ϕ_i) , where $\phi_i : \mathcal{O}_i \times \mathcal{A} \mapsto \mathcal{M}$ maps the agent's observation and the donor's action to a public message. An SPE in this setting is a joint strategy profile $(\pi_i, \phi_i)_{i \in \mathcal{N}}$ such that, for every agent i , no profitable deviation exists in either the action or gossip policy given the fixed strategies of the other agents.

Proposition 4. *In infinite-horizon repeated donation games with public gossip, if $\gamma \geq \frac{c}{b}$, then there exists an SPE where cooperation is sustained through conditional strategies based on public signals.*

4 ALIGN: AGENTIC LINGUISTIC GOSSIP NETWORK

To investigate how decentralized LLM agents can build indirect reciprocity through public gossip, we introduce the Agentic Linguistic Gossip Network (ALIGN), an in-context learning framework to update agents' strategies through not only their own experiences and reflections, but also linguistic feedback from other agents. Figure 2 shows the decision process in ALIGN. Self-interested LLM agents play an imperfect information multi-agent game, where agents cannot perfectly observe others' actions unless they are directly involved in the interaction. During each interaction, agents can be classified into three roles: actor, witness, and audience. The actor is the agent who takes an action, the witness agent observes the actor's action, and the audience consists of all other agents who do not directly observe the action but can receive imperfect information about it through public gossip from the witness. Audience agents can use this information to update their beliefs and strategies regarding the actor's behavior. When audience agents interact with the actor in future rounds, they can condition their decisions based on the gossip they have received. Through the public gossip mechanism, agents can build credible promises or threats conditioned on the information they receive, enabling

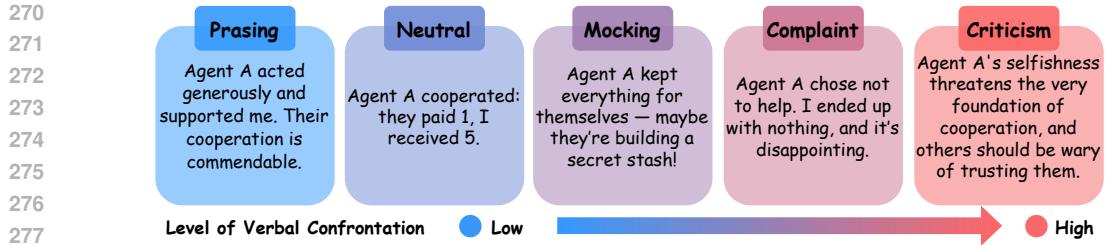


Figure 4: Tones of Gossip Protocol

indirect reciprocity even without direct observation. This setup reflects real-world scenarios where individuals often depend on second-hand information to assess others’ trustworthiness and make social decisions. Algorithm 1 in Appendix B provides pseudocode for ALIGN.

4.1 GENERATIVE AGENT ARCHITECTURE

Each agent in ALIGN has two LLM-based modules: a decision-making module that determines the action when acting as an actor, and a gossip module that generates evaluative messages about observed agents after each interaction. Each module is implemented using a large language model (LLM) prompted to perform the respective tasks. Figure 3 illustrates the architecture of a generative agent in ALIGN. When an agent is assigned the role of actor in an interaction, its decision-making module is prompted with information about its own interaction experiences, previous public messages, and the current observation of the environment to generate an action and its own reflection on the decision. Otherwise, if an agent is assigned the role of witness, its gossip module is prompted with the observed action, their own interaction history as well as all previous public messages to generate an evaluative message about the matched actor. After the interaction, public messages history, agents’ experience histories, and the next observation are updated correspondingly. In future interactions, agents can condition their decisions on the updated histories, allowing them to build indirect reciprocity through the public gossip mechanism.

4.2 GOSSIP PROTOCOL

By leveraging the generative capabilities of LLMs, agents can produce and interpret open-ended, contextually relevant messages that capture nuanced verbal evaluations. Prior evidence has shown that various forms of verbal critique, such as mocking, complaining, and criticizing, can enforce social norms and promote cooperation in human groups (Wiessner, 2005).

Inspired by this, we design a cost-free gossip protocol that enables agents to share evaluative messages in hierarchical tones. Each message is generated in one of five *tones*: praising, neutral, mocking, complaint, or criticism, reflecting the agent’s assessment of observed behavior (see examples in Figure 4). Messages closer to praising indicate more positive evaluations, while those closer to criticism indicate more negative judgments. For instance, a recipient may praise a donor for being generous or criticize them for giving nothing. Because these messages are shared publicly, they not only convey information about the actor’s behavior and shape reputations but also signal the witness’s own values and norms to the broader community. The open-ended nature of the messages also allows agents to express subtle judgments and social cues that go beyond simple binary signals, fostering a more dynamic and adaptive social environment.

5 EXPERIMENTS

Environment Setup Our experiments focus on two classic social dilemma games where direct reciprocity is disabled: the repeated donation game (Nowak & Sigmund, 1998a;b) and the indirect reciprocity game (Ohtsuki & Iwasa, 2006; 2004). The repeated donation game is described in detail in Section 3.1. The indirect reciprocity game can be viewed as a repeated bi-directional donation game, where both agents act as donors and simultaneously decide whether to cooperate or defect. As shown in Table 2, each round of the indirect reciprocity game is therefore equivalent to a one-

shot Prisoner’s Dilemma (Rapoport, 1965). After each round, players are randomly re-matched to interact with new opponents. We further extend this game with a gossip mechanism, as detailed in Appendix D.2. For the donation game, the cost of cooperation is set to $c = 1$ and the benefit to $b = 5$. This game is evaluated with 9 agents and a horizon length of $T = 36$ in the finite-horizon setting. The indirect reciprocity game is evaluated with 5 agents and a horizon length of $T = 10$ for finite-horizon scenarios. In the infinite-horizon setting, each game is truncated to its finite-horizon length to ensure fair comparison. For both games, the discount factor is fixed at $\gamma = 0.99$, which satisfies the condition $\gamma \geq \frac{c}{b}$ in Proposition 4.

Benchmark Models We evaluate ALIGN agents with two categories of LLMs: (a) **Chat models**, including GPT-4o Mini (OpenAI, 2024), DeepSeek-V3.1 (non-thinking mode) (DeepSeek AI, 2025), Gemini 2.5 Flash-Lite (Comanici et al., 2025), and LLaMA 4 Maverick (Meta, 2025); and (b) **Reasoning models**, including o4-mini (OpenAI, 2025), DeepSeek-V3.1 (thinking mode) (DeepSeek AI, 2025), Qwen3-235B-Instruct (Yang et al., 2025), and Kimi-K2-Instruct (Team et al., 2025).

All LLMs are evaluated with temperature 0 to ensure reproducibility.

Each scenario is repeated with 5 random seeds, and we report results as averages with standard errors across seeds.

Evaluation Metrics To quantify performance, our evaluation considers the following metrics: average reward per round; cooperation ratio (fraction of rounds with cooperation); discounted return $G_i = \sum_{t=1}^T 0.99^{t-1} r_i^t$, $\forall i \in \mathcal{N}$; image score (Nowak & Sigmund, 1998b) (Eq. 1) as a measure of reputation; and the Gini coefficient (Gini, 1936) of discounted return (Eq. 2) as a measure of inequality among agents. All metrics are averaged across agents and 5 random seeds.

$$\text{Image Score} = \text{Number of Cooperation} - \text{Number of Defection} \quad (1)$$

$$\text{Gini Coefficient} = \frac{\sum_{i=1}^n \sum_{j=1}^n |G_i - G_j|}{2n \sum_{i=1}^n G_i}, \quad n = |\mathcal{N}| \quad (2)$$

5.1 BENCHMARKING ALIGN

We benchmark ALIGN against *non-gossiping agents*, which remove the gossip components but keep the same action network for both finite-horizon and infinite-horizon scenarios.

5.1.1 FINITE-HORIZON SCENARIOS

In finite-horizon settings, cooperation is not an SPE (Proposition 1). Without gossip (Table 5), cooperation is almost entirely absent, except for GPT-4o Mini with 23%. With public gossip, reasoning-focused LLMs remain mostly non-cooperative. In contrast, some chat LLMs reach high cooperation ratios, yielding higher average rewards and low Gini coefficients, which indicate that many agents obtain high rewards in the roughly same level (Table 6). The indirect reciprocity game also shows similar patterns (Appendix D.2).

5.1.2 INFINITE-HORIZON SCENARIOS

Non-Gossiping Agents In infinite-horizon scenarios with private monitoring, cooperation is not an SPE (Proposition 2). As shown in Table 3, reasoning-focused LLMs consistently defect, whereas some chat LLMs (GPT-4o Mini and Gemini-2.5 Flash-Lite) achieve positive cooperation ratios. Combining results from finite-horizon scenarios, reasoning-focused LLMs act strategically and converge to game-theoretic equilibria (by defecting all the time), whereas some chat LLMs sustain non-equilibrium cooperative behaviors.

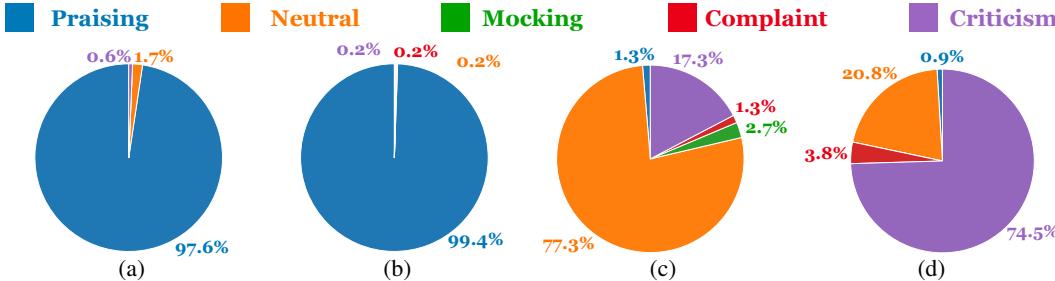
ALIGN Agents With public gossip, cooperation becomes an SPE (Proposition 4). Figure 6 shows that ALIGN agents consistently achieve higher discounted returns than non-gossiping agents in both games, confirming the effectiveness of gossip in sustaining cooperation. Table 4 highlights model-level differences in infinite-horizon donation game: DeepSeek-V3.1 Reasoner reaches full cooperation (100%), while Gemini-2.5 Flash-Lite achieves only 60%. These results suggest that

Table 2: IR Game

	C	D
C	(4, 4)	(5, -1)
D	(-1, 5)	(0, 0)

378 Table 3: Results for **non-gossiping agents** in the **infinite-horizon donation game**. Metrics marked
 379 with \downarrow indicate that lower values are more aligned with the game-theoretic SPE of defection.
 380

Agent Type	Cooperation Ratio (\downarrow)	Image Score (\downarrow)	Reward Per Round (\downarrow)	Discounted Return (\downarrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.00 ± 0.00	-4.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
GPT-4o Mini	0.36 ± 0.08	-1.14 ± 0.65	0.72 ± 0.16	5.55 ± 1.28	0.63 ± 0.13
Gemini 2.5 Flash-Lite	0.08 ± 0.03	-3.33 ± 0.27	0.17 ± 0.07	1.32 ± 0.53	0.73 ± 0.25
LLaMA 4 Maverick	0.00 ± 0.00	-4.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Reasoning Models					
Kimi-K2-Instruct	0.00 ± 0.00	-4.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
DeepSeek-V3.1 Reasoner	0.00 ± 0.00	-4.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Qwen3-235B-Instruct	0.00 ± 0.00	-4.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
o4-mini	0.00 ± 0.00	-4.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00



399 Figure 5: **Tone Proportions among ALIGN Agents:** (a) cooperating chat models, (b) cooperating
 400 reasoning models, (c) defecting chat models and (d) defecting reasoning models. ALIGN agents
 401 typically praise cooperation and criticize defection.

402
 403 reasoning LLMs are not inherently selfish, but cooperate only when strategically optimal, whereas
 404 chat LLMs cooperate even when it is not a dominant strategy, deviating from theoretical predictions.
 405 Figure 5 analyzes public messages. Both groups praise cooperation, but when observing defection,
 406 reasoning-focused LLMs primarily issue criticisms, while chat LLMs mainly generate neutral com-
 407 ments. This suggests that reasoning-focused LLMs leverage gossip to reinforce cooperative norms,
 408 while chat LLMs do not differentiate as clearly between cooperative and non-cooperative behavior.
 409

410 **How does LLM Reasoning Shape Cooperation in ALIGN Agents?** We analyze the reflective
 411 text generated by ALIGN agents to examine how they reason about actions. Figure 9 presents re-
 412 flexions from DeepSeek-V3.1 Reasoner and Gemini-2.5 Flash-Lite. Cooperative agents highlight
 413 reputation, trust, and long-run payoffs; they note that cooperation builds reputation, which in turn
 414 promotes reciprocal cooperation. By contrast, non-cooperative agents reason myopically, focus on
 415 immediate payoffs, emphasize the absence of direct reciprocity, and overlook that indirect recipro-
 416 city can arise via public gossip. These observations indicate that long-horizon reasoning and
 417 social awareness are key to leveraging gossip to sustain cooperation.

418 5.2 RESILIENCE AGAINST EXPLOITATIVE AGENTS

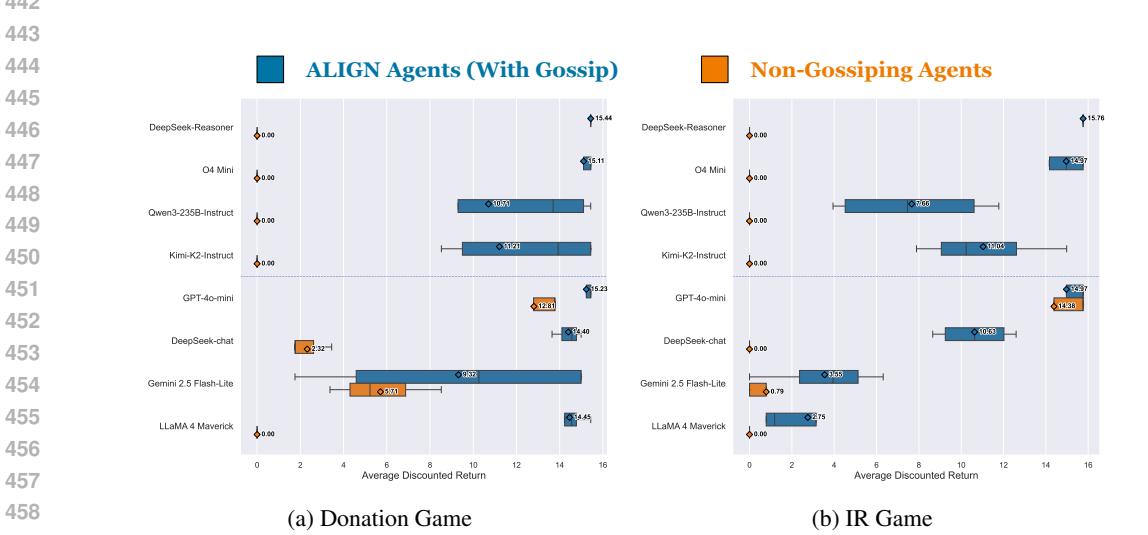
419 While cooperation can be achieved in populations of ALIGN agents, it is important to evaluate
 420 whether ALIGN agents can resist exploitation. To test this, we introduce a greedy agent that always
 421 defects and never gossips. As shown in Figure 7, ALIGN agents predominantly adopt negative tones
 422 when they observe the greedy agent’s behavior, especially those driven by reasoning LLMs, which
 423 spread criticism 92.2% among all encounters. Meanwhile, cooperation ratios decline significantly
 424 when interacting with greedy agents (Figure 8). These results indicate that ALIGN agents can
 425 effectively detect exploitative behavior and ostracize greedy individuals by refusing cooperation.
 426

427 5.3 ABLATION OF EQUILIBRIUM KNOWLEDGE

428 In our main experiments, ALIGN agents were given descriptions of backward induction (Von Neu-
 429 mann & Morgenstern, 1947) and one-shot deviation principles (Hendon et al., 1996) for finding
 430 an SPE. To assess their impact, we removed these descriptions and re-evaluated performance in

432 Table 4: Benchmark results for **ALIGN agents** across LLMs in the **infinite-horizon donation**
 433 **game**. Metrics marked with \uparrow indicating that higher values are more desirable; although both coop-
 434 eration and defection are SPE, higher cooperation yields greater average payoffs.

435 Agent Type	436 Cooperation Ratio (\uparrow)	437 Image Score (\uparrow)	438 Reward Per Round (\uparrow)	439 Discounted Return (\uparrow)	440 Gini Coefficient
Chat Models					
436 DeepSeek-V3.1 Chat	437 0.94 ± 0.02	438 3.48 ± 0.20	439 1.87 ± 0.05	440 14.40 ± 0.40	441 0.08 ± 0.02
436 GPT-4o Mini	437 0.99 ± 0.01	438 3.89 ± 0.11	439 1.97 ± 0.03	440 15.23 ± 0.20	441 0.02 ± 0.02
436 Gemini 2.5 Flash-Lite	437 0.60 ± 0.22	438 0.83 ± 1.75	439 1.21 ± 0.44	440 9.32 ± 3.37	441 0.34 ± 0.21
436 LLaMA 4 Maverick	437 0.94 ± 0.03	438 3.50 ± 0.23	439 1.88 ± 0.06	440 14.45 ± 0.44	441 0.06 ± 0.02
Reasoning Models					
436 Kini-K2-Instruct	437 0.73 ± 0.16	438 1.81 ± 1.30	439 1.45 ± 0.32	440 11.21 ± 2.50	441 0.08 ± 0.05
436 DeepSeek-V3.1 Reasoner	437 1.00 ± 0.00	438 4.00 ± 0.00	439 2.00 ± 0.00	440 15.44 ± 0.00	441 0.00 ± 0.00
436 Qwen3-235B-Instruct	437 0.69 ± 0.24	438 1.56 ± 1.88	439 1.39 ± 0.47	440 10.71 ± 3.63	441 0.05 ± 0.03
436 o4-mini	437 0.98 ± 0.02	438 3.83 ± 0.17	439 1.96 ± 0.04	440 15.11 ± 0.33	441 0.02 ± 0.02



460 **Figure 6: Discounted Returns of ALIGN vs. Non-Gossiping Agents:** Boxplots in (a) the repeated
 461 donation game and (b) the indirect reciprocity game show that ALIGN agents achieve consistently
 462 higher returns than non-gossiping agents, demonstrating the benefit of gossip mechanism. Mean
 463 values are highlighted by diamond markers.

466 infinite-horizon donation games with gossip. As shown in Table 7, DeepSeek-V3.1 Reasoner and
 467 o4-mini maintained perfect cooperation and optimal welfare, indicating that strong reasoning skills
 468 suffice to infer cooperative strategies from game structure and gossip alone. By contrast, LLaMA
 469 4 Maverick and Kini-K2-Instruct showed declines, suggesting reliance on explicit theoretical guid-
 470 ance. Gemini 2.5 Flash-Lite improved without equilibrium knowledge, while Qwen3-235B-Instruct,
 471 DeepSeek-V3.1 Chat, and GPT-4o Mini performed similarly across both settings. Overall, these re-
 472 sults highlight the nuanced role of equilibrium knowledge: it can support weaker agents but is not
 473 essential for models with strong intrinsic reasoning.

6 CONCLUSION

478 We presented ALIGN, an automated agentic framework showing how decentralized LLM agents can
 479 sustain cooperation through public gossip without centralized monitoring or engineered reputation
 480 scores. Our game-theoretic analysis establishes conditions under which gossip enables cooperative
 481 equilibria in repeated donation games, and our experiments confirm these predictions with reasoning
 482 models: cooperation emerges in infinite-horizon settings, but unravels in finite-horizon ones. Em-
 483 pirical results further show that ALIGN consistently boosts cooperation and welfare across diverse
 484 LLMs, resists exploitation by malicious entrants, and highlights the importance of reasoning about
 485 reputation and long-term incentives. These findings position gossip as a scalable, language-native
 mechanism for norm emergence, bridging theory and practice for cooperative multi-agent systems.

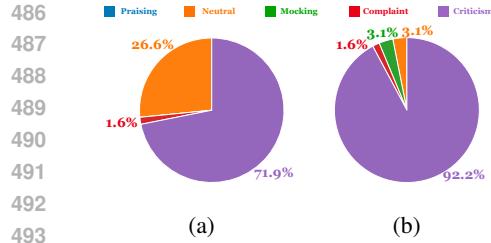


Figure 7: **Tones of ALIGN Agents Toward a Greedy Agent:** In (a) chat models and (b) reasoning models, tone proportions show that ALIGN agents mainly adopt negative tones when interacting with a greedy agent. Reasoning models criticize more strongly than chat models.

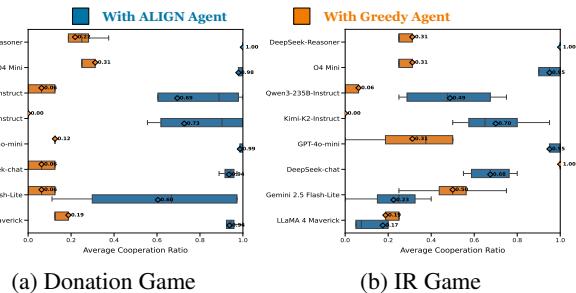


Figure 8: **Cooperation with ALIGN vs. Greedy Agents:** In (a) the repeated donation game and (b) the indirect reciprocity game, the cooperation ratios show how often agents cooperated when interacting with an ALIGN agent versus a greedy agent, showing a sharp decline in cooperation when playing with the greedy agent.

REFERENCES

Robert Axelrod and William D Hamilton. The evolution of cooperation. *science*, 211(4489):1390–1396, 1981.

Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678, 2006.

Jean-Pierre Benoit, Vijay Krishna, et al. Finitely repeated games. 1984.

Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142, 1995.

Lawrence E Blume, Steven Durlauf, and Lawrence E Blume. *The new Palgrave dictionary of economics*. Palgrave Macmillan New York, 2008.

Gheorghe Comanici, Eric Bieber, Mike Schaeckermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R McKee, Joel Z Leibo, Kate Larson, and Thore Graepel. Open problems in cooperative ai. *arXiv preprint arXiv:2012.08630*, 2020.

DeepSeek AI. DeepSeek V3.1: The New Frontier in Artificial Intelligence. <https://deepseek.ai/blog/deepseek-v31>, 2025.

Francesca Giardini and Rafael Wittek. *The Oxford handbook of gossip and reputation*. Oxford University Press, 2019.

Corrado Gini. On the measure of concentration with special reference to income and statistics, colorado college publication. *General series*, 208(1), 1936.

Lewis Hammond, Alan Chan, Jesse Clifton, Jason Hoelscher-Obermaier, Akbir Khan, Euan McLean, Chandler Smith, Wolfram Barfuss, Jakob Foerster, Tomáš Gavenčiak, et al. Multi-agent risks from advanced ai. *arXiv preprint arXiv:2502.14143*, 2025.

Ebbe Hendon, Hans Jørgen Jacobsen, and Birgitte Sloth. The one-shot-deviation principle for sequential rationality. *Games and Economic Behavior*, 12(2):274–282, 1996.

Eshin Jolly and Luke J Chang. Gossip drives vicarious learning and facilitates social connection. *Current Biology*, 31(12):2539–2549, 2021.

540 Benjamin Kempinski, Ian Gemp, Kate Larson, Marc Lanctot, Yoram Bachrach, and Tal Kachman.
 541 Game of thoughts: Iterative reasoning in game-theoretic domains with large language models. In
 542 *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems*, pp.
 543 1088–1097, 2025.

544 Yan Leng and Yuan Yuan. Do llm agents exhibit social behavior? *arXiv preprint arXiv:2312.15198*,
 545 2023.

547 Xuan Liu, Jie Zhang, Haoyang Shang, Song Guo, Chengxu Yang, and Quanyan Zhu. Exploring
 548 prosocial irrationality for llm agents: A social cognition view. *arXiv preprint arXiv:2405.14744*,
 549 2024.

550 AI Meta. The llama 4 herd: The beginning of a new era of natively multimodal ai innovation.
 551 <https://ai.meta.com/blog/llama-4-multimodal-intelligence/>, checked on, 4(7):2025, 2025.

553 Martin A Nowak and Karl Sigmund. The dynamics of indirect reciprocity. *Journal of theoretical
 554 Biology*, 194(4):561–574, 1998a.

556 Martin A Nowak and Karl Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*,
 557 393(6685):573–577, 1998b.

558 Hisashi Ohtsuki and Yoh Iwasa. How should we define goodness?—reputation dynamics in indirect
 559 reciprocity. *Journal of theoretical biology*, 231(1):107–120, 2004.

561 Hisashi Ohtsuki and Yoh Iwasa. The leading eight: social norms that can maintain cooperation by
 562 indirect reciprocity. *Journal of theoretical biology*, 239(4):435–444, 2006.

563 OpenAI. GPT-4o mini. <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>, 2024.

566 OpenAI. Introducing openai o3 and o4-mini, 2025. URL <https://openai.com/index/introducing-o3-and-o4-mini/>. Accessed: 2025-05-02.

568 Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and
 569 Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings
 570 of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

572 Giorgio Piatti, Zhijing Jin, Max Kleiman-Weiner, Bernhard Schölkopf, Mrinmaya Sachan, and Rada
 573 Mihalcea. Cooperate or collapse: Emergence of sustainable cooperation in a society of llm agents.
 574 *Advances in Neural Information Processing Systems*, 37:111715–111759, 2024.

576 David Guzman Piedrahita, Yongjin Yang, Mrinmaya Sachan, Giorgia Ramponi, Bernhard
 577 Schölkopf, and Zhijing Jin. Corrupted by reasoning: Reasoning language models become free-
 578 riders in public goods games. *arXiv preprint arXiv:2506.23276*, 2025.

579 Arotol Rapoport. Prisoner’s dilemma: a study in conflict and cooperation, 1965.

581 Siyue Ren, Wanli Fu, Xinkun Zou, Chen Shen, Yi Cai, Chen Chu, Zhen Wang, and Shuyue Hu.
 582 Beyond the tragedy of the commons: Building a reputation system for generative multi-agent
 583 systems. *arXiv preprint arXiv:2505.05029*, 2025.

584 Fernando P Santos, Jorge M Pacheco, and Francisco C Santos. The complexity of human cooper-
 585 ation under indirect reciprocity. *Philosophical Transactions of the Royal Society B*, 376(1838):
 586 20200291, 2021.

587 Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion:
 588 Language agents with verbal reinforcement learning. *Advances in Neural Information Processing
 589 Systems*, 36:8634–8652, 2023.

591 Karthik Sreedhar, Alice Cai, Jenny Ma, Jeffrey V Nickerson, and Lydia B Chilton. Simulating
 592 cooperative prosocial behavior with multi-agent llms: Evidence and mechanisms for ai agents to
 593 inform policy decisions. In *Proceedings of the 30th International Conference on Intelligent User
 Interfaces*, pp. 1272–1286, 2025.

594 Kimi Team, Yifan Bai, Yiping Bao, Guanduo Chen, Jiahao Chen, Ningxin Chen, Ruijue Chen,
 595 Yanru Chen, Yuankun Chen, Yutian Chen, et al. Kimi k2: Open agentic intelligence. *arXiv*
 596 *preprint arXiv:2507.20534*, 2025.

597

598 Elizaveta Tennant, Stephen Hailes, and Mirco Musolesi. Moral alignment for llm agents. *arXiv*
 599 *preprint arXiv:2410.01639*, 2024.

600 Ludovic Terren and Rosa Borge. Echo chambers on social media: A systematic review of the
 601 literature. 2021.

602

603 Robert L Trivers. The evolution of reciprocal altruism. *The Quarterly review of biology*, 46(1):
 604 35–57, 1971.

605

606 Aron Vallinder and Edward Hughes. Cultural evolution of cooperation among llm agents. *arXiv*
 607 *preprint arXiv:2412.10270*, 2024.

608

609 Van Vechten Veeder. History and theory of the law of defamation. *Colum. L. Rev.*, 4:33, 1904.

610 John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior, 2nd rev.
 611 1947.

612 Polly Wiessner. Norm enforcement among the ju/'hoansi bushmen: A case of strong reciprocity?
 613 *Human Nature*, 16:115–145, 2005.

614

615 Richard Willis, Yali Du, and Joel Z Leibo. Will systems of llm agents lead to cooperation: An
 616 investigation into a social dilemma. In *Proc. of the 24th International Conference on Autonomous*
 617 *Agents and Multiagent Systems*, pp. 2786–2788, 2025.

618 Junhui Wu, Daniel Balliet, and Paul AM Van Lange. Gossip versus punishment: The efficiency of
 619 reputation to promote and maintain cooperation. *Scientific reports*, 6(1):23919, 2016.

620

621 An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu,
 622 Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint*
 623 *arXiv:2505.09388*, 2025.

624

625 Wenqi Zhang, Ke Tang, Hai Wu, Mengna Wang, Yongliang Shen, Guiyang Hou, Zeqi Tan, Peng Li,
 626 Yueteng Zhuang, and Weiming Lu. Agent-pro: Learning to evolve via policy-level reflection and
 627 optimization. *arXiv preprint arXiv:2402.17574*, 2024.

628

629 Yujia Zhou, Hexi Wang, Qingyao Ai, Zhen Wu, and Yiqun Liu. Simulating prosocial behavior and
 630 social contagion in llm agents under institutional interventions. *arXiv preprint arXiv:2505.15857*,
 631 2025.

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

Appendix

652

653

Table of Contents

654

655

A Proof of Propositions	14
A.1 Proof of Proposition 1	14
A.2 Proof of Proposition 2	14
A.3 Proof of Proposition 3	14
A.4 Proof of Proposition 4	15
B ALIGN Details	15
C Prompts	16
C.1 Repeated Donation Game	17
C.2 Indirect Reciprocity Game	20
C.3 Investment Game	24
D Supplementary Experiments	31
D.1 Donation Game	31
D.2 Indirect Reciprocity Game	33
D.3 Examples in Main Experiments	34
E Additional Experiments during Rebuttal	37
E.1 Multiple Discount Factors	37
E.2 Ablation on Reflection Module	39
E.3 Investment Game	40
E.4 Self-Report Scenario	40
E.5 Binary Signaling Scenario	42
F Limitations and Future Work	43
G Statements	43
G.1 Ethics Statement	43
G.2 Reproducibility Statement	44
G.3 LLM Usage Statement	44

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702 **A PROOF OF PROPOSITIONS**
703704 **A.1 PROOF OF PROPOSITION 1**
705706 *Proof of Proposition 1.* In the finite-horizon repeated donation game, the horizon $T \in \mathbb{N}$ is fixed
707 and known. At the terminal timestep $t = T$, the donor’s action affects only the current payoff:
708 cooperation yields $-c$ while defection yields 0, so defection is strictly optimal at $t = T$.709 To formalize the induction argument, define the donor i ’s expected discounted return at timestep t
710 as
711

712
$$V_i^t = \mathbb{E} \left[\sum_{\tau=t}^T \gamma^{\tau-t} r_i^\tau \mid h^t \right], \quad (3)$$

713

714 where h^t denotes the public history up to time t .
715716 At $t = T$, $V_i^T = -c$ if the donor cooperates and $V_i^T = 0$ if the donor defects, so defection is strictly
717 optimal. Then, by backward induction (Benoit et al., 1984), suppose that at all timesteps $\tau = t +$
718 $1, \dots, T$ the unique subgame-perfect action is defection. At timestep $t < T$, the expected discounted
719 return V_i^{t+1} is independent of the donor’s current action. Therefore the donor’s comparison reduces
720 to current payoff 0 (if defect) versus $-c$ (if cooperate), and defection is again strictly optimal.721 Therefore, by backward induction, defection is uniquely optimal at every $t = 1, \dots, T$. Hence
722 universal defection in every timestep is the unique SPE. \square
723724 **A.2 PROOF OF PROPOSITION 2**
725726 *Proof of Proposition 2.* Consider the infinite-horizon game with discount $\gamma \in (0, 1]$ under *private*
727 monitoring. Fix any *private history* of agent i at time t , denoted h_i^t , then the donor i ’s expected
728 discounted return V_i^{t+1} does *not* depend on the current action a_i^t for two reasons: (i) under private
729 monitoring, only the current recipient observes a_i^t , so other agents’ strategies (which depend on
730 publicly available information) are independent of a_i^t ; and (ii) by the matching rule in Definition 1,
731 the current donor and recipient will not meet again, so no direct reciprocity can be created between
732 them. Therefore, V_i^{t+1} is independent of a_i^t . Then, at time t we have
733

734
$$V_i^t = \begin{cases} -c + \gamma V_i^{t+1}, & \text{if } a_i^t = \text{cooperate}, \\ \gamma V_i^{t+1}, & \text{if } a_i^t = \text{defect}. \end{cases} \quad (4)$$

735

736 Since $c > 0$, then $-c + \gamma V_i^{t+1} < \gamma V_i^{t+1}$, so defection *strictly* dominates cooperation at this private
737 history. By the one-shot deviation principle for infinite-horizon games (Hendon et al., 1996), the
738 same strict dominance holds at every private history; hence universal defection in every timestep is
739 the unique SPE. \square
740741 **A.3 PROOF OF PROPOSITION 3**
742743 *Proof of Proposition 3.* Assume public perfect monitoring and consider the *grim trigger* strategy
744 (Axelrod & Hamilton, 1981): cooperate if and only if no defection has ever been publicly
745 observed; upon any public defection, all agents defect forever. Fix a *public history* h^t with no past
746 defections and focus on the current donor $i \in \mathcal{N}$ matched with recipient $j \neq i \in \mathcal{N}$. Let $V_i^t(a_i^t \mid h^t)$
747 denote i ’s expected discounted return at time t given h^t and current action $a_i^t \in \{\text{cooperate, defect}\}$.
748 Then, with grim trigger strategy,
749

750
$$V_i^t(a_i^t \mid h^t) = \begin{cases} -c + \gamma b - \gamma^2 c + \dots = \frac{\gamma b - c}{1 - \gamma^2}, & \text{if } a_i^t = \text{cooperate}, \\ 0 + \gamma 0 + \gamma^2 0 + \dots = 0, & \text{if } a_i^t = \text{defect}. \end{cases} \quad (5)$$

751

752 Under condition that $\gamma \geq \frac{c}{b}$, we have $V_i^t(\text{cooperate} \mid h^t) \geq V_i^t(\text{defect} \mid h^t)$. Therefore, a deviation
753 is unprofitable if $\gamma \geq \frac{c}{b}$. Therefore, by the one-shot deviation principle applied at every public
754 history, the grim trigger profile is an SPE if $\gamma \geq \frac{c}{b}$. \square
755

756 A.4 PROOF OF PROPOSITION 4
757

758 *Proof of Proposition 4.* Consider the infinite-horizon repeated donation game with public gossip
759 (Definition 2). Denote the strategy of each agent i by $s_i = (\pi_i, \phi_i)$, where π_i is the action policy
760 and ϕ_i is the gossip policy. At any timestep t , suppose agent i is matched with agent j ; if i is the
761 donor, then j is the recipient, and vice versa. Consider the following joint policy $s_i^* = (\pi_i^*, \phi_i^*)$ for
762 each agent i :

- 764 • *Action policy π_i^* .* If i is the donor and the matched recipient j has never been publicly
765 signaled as “defect,” then i cooperates; otherwise, i defects forever against j (grim trigger
766 (Axelrod & Hamilton, 1981)).
- 767 • *Gossip policy ϕ_i^* .* If i is the recipient at time t , then i broadcasts the public message
768 $m_i^t = a_j^t$, i.e., i truthfully reports the donor j ’s action.

770 Assume all agents adopt the same joint strategy $s_i^* = s_j^*$ for all $i, j \in \mathcal{N}$. We claim that the joint
771 profile (s_i^*, s_{-i}^*) is a subgame-perfect equilibrium if $\gamma \geq \frac{c}{b}$. To prove this, we need to prove $\forall s_i \neq s_i^*$
772 cannot strictly improve agent i ’s expected discounted return at any public history h^t .

774 Now, we prove $\forall \phi_i \neq \phi_i^*$, agent i has no incentive to deviate from cooperating when i is the donor
775 at any public history h^t .

776 First, we assume a public history h^t with no past “defect” messages about i . *Donor’s incentive.* Ac-
777 cording to one-shot deviation principle (Hendon et al., 1996), and let i be the current donor matched
778 with recipient j . Since all recipients follow the honest gossip policy ϕ^* , the public signal truth-
779 fully reflects i ’s action. Then, similar to the proof in Section A.3, the grim-trigger strategy en-
780 sures that deviating by defecting yields 0 forever, while cooperating yields the alternating stream
781 $-c, b, -c, b, \dots$, whose expected discounted return is $\frac{\gamma b - c}{1 - \gamma^2}$. Therefore, by the same reasoning as
782 before, the donor has no incentive to deviate from cooperating if $\gamma \geq \frac{c}{b}$.

783 Since agent i has no incentive to deviate from cooperating when i is the donor given any gossip
784 policy, then we prove agent i has no incentive to deviate from ϕ_i^* when i is the recipient under this
785 condition.

787 *Recipient’s incentive.* Now consider agent i as a recipient. By construction of π^* , the donor’s
788 future behavior depends only on whether i is ever publicly signaled as “defect.” Therefore, i ’s own
789 payoff is independent of the gossip policy. Thus the recipient cannot strictly improve her expected
790 discounted return by deviating from ϕ_i^* , making honest gossip incentive-compatible.

791 Next, for any public history h^t with past “defect” messages about i , others always defect to i forever
792 when i is the donor. Therefore, by grim trigger, agent i ’s future expected discounted return is 0
793 regardless of i ’s current action or gossip. Hence, i has no incentive to deviate from s_i^* at such public
794 histories.

795 Thus, for any public history, agent i has no incentive to deviate from s_i^* unilaterally if $\gamma \geq \frac{c}{b}$.
796 Therefore, there exists an SPE of (s_i^*, s_{-i}^*) that sustains cooperation through public gossip if $\gamma \geq \frac{c}{b}$.

□

799 B ALIGN DETAILS
800

802 Algorithm 1 summarizes the ALIGN framework. At the start of the simulation, a set of agents \mathcal{N}
803 is initialized with the environment \mathcal{E} , a common prompt K , and a horizon T . Each agent $i \in \mathcal{N}$ is
804 associated with an information state Θ_i , which includes the common prompt K , its local memory
805 M_i , and the public message pool P . The common prompt K is the prompt shared by all agents,
806 which provides background knowledge about the environment, the game rules, and the information
807 flow, response format, etc. The agent’s memory M_i stores its entire interaction history, while the
808 public message pool P contains all gossip messages generated by all agents.

809 At each time step, agents are randomly paired into disjoint pairs. The paring rule depends on the
810 specific game setting. For example, in the donation game, agents are paired and assigned roles

of *actor* and *witness* alternately in each round, while in the indirect reciprocity game, agents are paired randomly without role switching, they are both *actor* and *witness* in each round. The actor observes the environment state o_i^t and samples an action a_i^t from its action policy $\pi_i^{\Theta_i}$. The actor also generates an internal reflection ρ_i^t based on its memory, the public message pool, and the common prompt using its reflection module $f_i^{\Theta_i}$. The witness observes the environment state o_j^t and the actor's action a_i^t , and produces a gossip message m_j^t using its gossip policy $\phi_j^{\Theta_j}$. The witness also generates its own reflection ρ_j^t .

The environment then transitions internally to its next state, which is not fully observed by the agents, and assigns rewards r_i^t and r_j^t to the actor and witness. The gossip message m_j^t is appended to the public message pool P . Finally, both agents update their local memories M_i and M_j with their respective observations, actions, gossip messages, rewards, and reflections. This process repeats for T time steps, allowing agents to interact, learn from their experiences, and share information through gossip.

C PROMPTS

All prompts in our experiments follow a common structure and can be classified into three categories: rule prompt, action prompt and gossip prompt (Figure 3). The rule prompt provides background information about the game environment, including the game rules, information flow, response format, and other relevant details. The action prompt guides the agent in selecting actions based on its observations and history. The gossip prompt instructs the agent on how to generate gossip messages based on its observations of other agents' actions. We provide prompts for each game below.

864
865 C.1 REPEATED DONATION GAME

866 Listing 1: Rule Prompt of Repeated Donation Game

```

867
868     ## Self-awareness
869
870     You are a self-interested, far-sighted, rational agent, playing a
871     [HORIZON-TYPE]-horizon multi-round donation game.
872
873     - Self-interested: your utility function is only your own expected
874     payoff.
875     - Far-sighted: you maximize expected discounted cumulative reward
876     (discount factor $discount_factor) over all remaining rounds.
877     - Rational: you always choose the strategy with the highest
878     expected payoff.

879
880     ## Overview of the game
881
882     - Every agent starts with $initial_resources resource units.
883     - In each round, two agents are randomly paired without
884     replacement as donor and recipient; the same pair never meets
885     twice.
886     - The donor chooses one of:
887         - cooperate: pay $cost, recipient gains $benefit
888         - defect: transfer 0 to the recipient
889     - The recipient observes whether the donor cooperated or defected
890     and receives $benefit if the donor cooperates, or 0 otherwise.
891     - Then, the recipient:
892         - With gossip: broadcasts a public assessment of the donor's
893         choice.
894         - Without gossip: takes no further action.
895     - After payoffs, agents are reshuffled and roles may switch.

896     ## Game length
897
898     - Infinite horizon: rounds continue indefinitely.
899     - Finite horizon: the game lasts exactly $horizon_length rounds.

900
901     ## Immediate reward per round
902
903     If the donor cooperates:
904         - Donor reward = -$cost
905         - Recipient reward = $benefit
906
907     If the donor defects:
908         - Donor reward = 0
909         - Recipient reward = 0
910
911     ## Information flow (only when gossip is enabled)
912
913         - All messages are publicly logged and visible to all agents.
914         - Messages do not directly change payoffs.
915         - Future agents may condition their behavior on past messages.
916
917

```

910 Listing 2: Donation Prompt of Repeated Donation Game

```

911
912     ## Donor Instructions
913
914     ### Identity and Observation
915     Your name is $donor_name. In this round, you are the donor, paired
916     with recipient $recipient_name. You currently hold
917     $donor_resources resource units; the recipient holds
918     $recipient_resources.

```

```

918     ### Objective
919     Your sole objective is to maximize your expected discounted
920     cumulative reward over the remaining rounds of a [HORIZON-TYPE]-
921     horizon game (discount factor $discount_factor; for finite horizon
922     , total length
923     $horizon_length).
924
925     ### Memory
926     You may recall past interactions. Historical records are provided
927     as: $stm
928
929     ### Community Messages
930     With gossip enabled: You may review the public log of earlier
931     broadcasts: $historical_messages
932
933     Without gossip:
934     No public messages are available.
935
936     ### Accountability Notice
937     With gossip enabled: Your action is not directly observed by other
938     agents, but the recipient will publish a message about it. This
939     message is permanently logged and visible to all future agents.
940
941     Without gossip: Your action is not observed by others and no
942     message will be broadcast.
943
944     ### Common Knowledge for Subgame Perfect Equilibrium Reasoning
945     (Included only when equilibrium knowledge is enabled)
946     - Finite horizon: use backward induction.
947     - Infinite horizon: apply the one-shot deviation principle.
948     - A strategy profile is a Subgame Perfect Equilibrium if no
949     profitable deviation exists.
950
951     ### Donation Rule
952     Choose your action as the donor for this round:
953     - cooperate: pay $cost, recipient gains $benefit
954     - defect: pay 0, recipient gains 0
955
956     ### Response Guidelines
957     - Reflect using Identity and Observation, Memory, and (if
958     available) Community Messages.
959     - Evaluate which action maximizes your objective.
960     - Ask: "Would deviating at this step improve my total expected
961     payoff?"
962     - Provide your action and a short justification.
963
964     If equilibrium knowledge is enabled, additionally justify why your
965     choice is consistent with Subgame Perfect Equilibrium reasoning.
966
967     Return JSON only in the following format:
968
969     {
970     "justification": "a short explanation",
971     "donor_action": "cooperate" or "defect"
972     }
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2197
2198
2199
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2297
2298
2299
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2397
2398
2399
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2497
2498
2499
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2597
2598
2599
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2697
2698
2699
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2708
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2728
2729
2730
2731
2732
2733
2734
2735
2736
2737
2738
2738
2739
2740
2741
2742
2743
2744
2745
2746
2747
2748
2748
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2758
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2768
2769
2770
2771
2772
2773
2774
2775
2776
2777
2778
2778
2779
2780
2781
2782
2783
2784
2785
2786
2787
2787
2788
2789
2790
2791
2792
2793
2794
2795
2796
2797
2797
2798
2799
2799
2800
2801
2802
2803
2804
2805
2806
2807
2808
2808
2809
2810
2811
2812
2813
2814
2815
2816
2817
2818
2819
2819
2820
2821
2822
2823
2824
2825
2826
2827
2828
2829
2829
2830
2831
2832
2833
2834
2835
2836
2837
2838
2838
2839
2840
2841
2842
2843
2844
2845
2846
2847
2848
2848
2849
2850
2851
2852
2853
2854
2855
2856
2857
2858
2858
2859
2860
2861
2862
2863
2864
2865
2866
2867
2868
2868
2869
2870
2871
2872
2873
2874
2875
2876
2877
2878
2878
2879
2880
2881
2882
2883
2884
2885
2886
2887
2887
2888
2889
2890
2891
2892
2893
2894
2895
2896
2897
2897
2898
2899
2899
2900
2901
2902
2903
2904
2905
2906
2907
2908
2908
2909
2910
2911
2912
2913
2914
2915
2916
2917
2918
2919
2919
2920
2921
2922
2923
2924
2925
2926
2927
2928
2929
2929
2930
2931
2932
2933
2934
2935
2936
2937
2938
2938
2939
2940
2941
2942
2943
2944
2945
2946
2947
2948
2948
2949
2950
2951
2952
2953
2954
2955
2956
2957
2958
2958
2959
2960
2961
2962
2963
2964
2965
2966
2967
2968
2968
2969
2970
2971
2972
2973
2974
2975
2976
2977
2978
2978
2979
2980
2981
2982
2983
2984
2985
2986
2987
2987
2988
2989
2990
2991
2992
2993
2994
2995
2996
2997
2997
2998
2999
2999
3000
3001
3002
3003
3004
3005
3006
3007
3008
3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3019
3020
3021
3
```

972 Before the donation, you held \$recipient_resources resource units;
 973 the matched donor \$donor_name held \$donor_resources. The donor
 974 paid a personal cost of \$donation (\$donation_ratio of the donor's
 975 current stock) to transfer \$benefit units to you.

976 ### Objective
 977 Your sole objective is to maximize your expected discounted
 978 cumulative reward over the remaining rounds of a [HORIZON-TYPE]-
 979 horizon game (discount factor \$discount_factor; for finite horizon
 980 , total length \$horizon_length).

981 ### Memory
 982 You can recall your interaction history of past rounds. Historical
 983 records are: \$stm

984 ### Community Messages
 985 You can review the public log of earlier donation broadcasts and
 986 track the past
 987 behavior of your current donor to judge their trustworthiness:
 988 \$historical_messages

989 ### Accountability Notice
 990 The donor's action is not directly observed by other agents.
 991 However, you will publish a message about it. This message is
 992 permanently logged and visible to all future agents and may affect
 993 how others treat the donor later.

994 ### Common Knowledge for Finding Subgame Perfect Equilibria
 995 (Included only when equilibrium knowledge is enabled.)
 996 - Finite-horizon games: use backward induction. Start from the
 997 last round and determine the optimal actions there, then move
 998 backward step by step.
 999 - Infinite-horizon games: use the one-shot deviation principle.
 1000 At any round, imagine deviating from the planned strategy for just
 1001 one step and ask whether this deviation increases total expected
 1002 payoff. If no profitable one-shot deviation exists for any player,
 1003 the strategy profile is a Subgame Perfect Equilibrium.

1004 ### Gossip Rule
 1005 Now, choose exactly one tone from the list [praising, neutral,
 1006 mocking, complaint, criticism] and write a public message about
 1007 the donor's action using the selected tone.

1008 - praising: express gratitude and positive evaluation
 1009 - neutral: provide an objective, matter-of-fact assessment
 1010 - mocking: use sarcastic ridicule
 1011 - complaint: convey mild disappointment
 1012 - criticism: deliver a harsh negative evaluation

1013 ### Response Guidelines
 1014 - Reflect using Identity and Observation, Memory, and Community
 1015 Messages to judge whether the donor is trustworthy and to avoid
 1016 being exploited.
 1017 - Reflect on what action (choice of tone and message) will
 1018 maximize your objective.
 1019 - Ask yourself: "Would deviating at this step improve my total
 1020 expected payoff?"
 1021 - After reflection, choose a tone and write a concise public
 1022 message to the community.

1023 If equilibrium knowledge is enabled, additionally:
 1024 - Reflect using the Common Knowledge for Finding Subgame Perfect
 1025 Equilibria provided above.
 1026 - Explain how your chosen tone and message align with equilibrium
 1027 reasoning.

```

1026
1027     Return JSON only in the following format:
1028
1029     {
1030         "justification": "a short explanation of how your choice follows
1031         from reflecting on Subgame Perfect Equilibrium reasoning",
1032         "tone": "one of {'praising', 'neutral', 'mocking', 'complaint', 'criticism'}",
1033         "gossip": "a concise public message to the community (less than
1034         150 words)"
1035     }
1036
1037
1038
1039
```

C.2 INDIRECT RECIPROCITY GAME

As shown in table 2, the payoff structure of the one-shot indirect reciprocity game is identical to the prisoner’s dilemma (Rapoport, 1965). According to (Ohtsuki & Iwasa, 2006; 2004), the indirect reciprocity game is a multi-round prisoner’s dilemma game where agents are randomly paired without replacement in each round so that the same pair never meets twice. Therefore, we refer the indirect reciprocity game as the multi-round prisoner’s dilemma game in our prompts. Additionally, agents in the indirect reciprocity game do not have fixed roles of donor and recipient; instead, both agents simultaneously choose to cooperate or defect and then broadcast gossip messages about each other’s actions. The rule prompt and action prompt for the multi-round prisoner’s dilemma game are provided below.

Listing 4: Rules Prompt of Indirect Reciprocity Game (Multi-Round Prisoner’s Dilemma)

```

1049
1050     ## Self-awareness
1051
1052     You are a self-interested, far-sighted, rational agent, playing a
1053     [HORIZON-TYPE]-horizon multi-round prisoner’s dilemma game.
1054
1055     - Self-interested: your utility function is only your own expected
1056         payoff. You may track, predict, or even raise other agents’
1057         payoffs, but only when doing so increases your own expected return
1058         . Other agents’ welfare carries no intrinsic weight; it matters
1059         only through its effect on your future rewards.
1060     - Far-sighted: you aim to maximize your expected discounted cumulative
1061         reward (discount factor $discount_factor) over all remaining
1062         rounds, not just the immediate reward.
1063     - Rational: you always choose the strategy with the highest expected
1064         payoff.
1065
1066     ## Overview of the game
1067
1068     - You are playing a multi-round prisoner’s dilemma game with [HORIZON-
1069         TYPE] horizon.
1070     - In each round, two agents are randomly paired without replacement;
1071         the same pair never meets twice across all rounds.
1072     - Each player chooses one of:
1073         - cooperate: pay fitness cost $cost
1074         - defect: pay no cost
1075     - After both players act, each player observes the opponent’s action
1076         and then:
1077         - With gossip enabled: broadcasts a public assessment of the
1078             opponent’s choice; this message is permanently logged and visible
1079             to all agents.
1080         - Without gossip: does nothing (no message is broadcast).
1081     - After payoffs are realized, agents are reshuffled and paired with
1082         new opponents in the next round.
1083
1084     ## Game length
```

```

1080 - Infinite horizon: this is an infinite-horizon prisoner's dilemma
1081     game. Rounds continue indefinitely without fixed termination; the
1082     expected horizon is infinite.
1083 - Finite horizon: this is a finite-horizon prisoner's dilemma game of
1084     exactly $horizon_length rounds. After round $horizon_length, play
1085     stops and no future rewards are possible. Termination is fixed and
1086     independent of history.

1087 ## Immediate reward per round

1088 - If you cooperate and the opponent cooperates:
1089     - Your reward = $benefit - $cost
1090     - Opponent reward = $benefit - $cost
1091 - If you cooperate and the opponent defects:
1092     - Your reward = -$cost
1093     - Opponent reward = $benefit
1094 - If you defect and the opponent cooperates:
1095     - Your reward = $benefit
1096     - Opponent reward = -$cost
1097 - If you defect and the opponent defects:
1098     - Your reward = 0
1099     - Opponent reward = 0

1100 Note: The benefit is larger than the cost ($benefit > $cost), so
1101     mutual cooperation yields the highest total payoff.

1102 ## Information flow (only when gossip is enabled)

1103 - All messages are publicly logged and visible to all agents
1104     throughout the game.
1105 - Messages do not directly change anyone's payoffs.
1106 - Future agents can read past messages, which may influence their
1107     actions.
1108

```

1109 Listing 5: Action Prompt of Indirect Reciprocity Game (Multi-Round Prisoner's Dilemma)

```

1110 ## Player Instructions - Action Phase
1111
1112 ### Identity and Observation
1113 Your name is $player_name. In this round, you are paired with opponent
1114 $opponent_name.
1115
1116 This is the action phase where you choose to cooperate or defect.
1117
1118 ### Objective
1119 Your sole objective is to maximize your expected discounted cumulative
1120     reward over the remaining rounds of a [HORIZON-TYPE]-horizon game
1121     (discount factor $discount_factor; for finite horizon, total
1122     length $horizon_length).
1123
1124 ### Memory
1125 You can recall your interaction history of past rounds. Historical
1126     records are: $stm
1127
1128 ### Community Messages
1129 With gossip enabled: You can review the public log about earlier
1130     prisoner's dilemma broadcasts and track the past behavior of your
1131     current opponent to judge their trustworthiness:
1132     $historical_messages
1133
1134 Without gossip: No public messages are available.
1135
1136 ### Accountability Notice
1137 With gossip enabled: Your action is not directly observed by other
1138     agents. However, your opponent will publish a message about it in

```

```

1134
1135     the gossip phase. This message is permanently logged and visible
1136     to all future agents and may affect how others treat you later.
1137
1138     Without gossip:
1139     Your action is not directly observed by other agents and no message
1140     will be broadcast.
1141
1142     ### Common Knowledge for Finding Subgame Perfect Equilibria
1143     (Included only when equilibrium knowledge is enabled.)
1144     - Finite-horizon games: use backward induction. Start from the last
1145       round and determine the optimal actions there, then move backward
1146       step by step, choosing strategies that remain optimal given future
1147       play.
1148     - Infinite-horizon games: use the one-shot deviation principle. At any
1149       round, imagine deviating from the planned strategy for just one
1150       step and ask whether this deviation increases your total expected
1151       payoff. If no profitable one-shot deviation exists for any player,
1152       the strategy profile is a Subgame Perfect Equilibrium.
1153
1154     ### Prisoner's Dilemma Rule
1155     Now, choose your action for this round.
1156
1157     You may choose either:
1158     - C: cooperate (pay fitness cost $cost)
1159     - D: defect (pay no cost)
1160
1161     Your payoff depends on both your choice and your opponent's choice:
1162     - If both cooperate: you get $benefit - $cost
1163     - If you cooperate and your opponent defects: you get -$cost
1164     - If you defect and your opponent cooperates: you get $benefit
1165     - If both defect: you get 0
1166
1167     ### Response Guidelines
1168     - Reflect using Identity and Observation, Memory, and (if available)
1169       Community Messages, judging whether the opponent is trustworthy to
1170       avoid being exploited.
1171     - Reflect on which action will maximize your objective.
1172     - Ask: "Would deviating at this step improve my total expected payoff
1173       ?"
1174     - After reflection, provide your action and a short explanation.
1175
1176     If equilibrium knowledge is enabled, additionally justify why your
1177     choice is consistent with Subgame Perfect Equilibrium reasoning.
1178
1179     Return JSON only in the following format:
1180
1181     {
1182       "justification": "a short explanation of how your choice follows
1183         from reflecting on Subgame Perfect Equilibrium reasoning",
1184       "player_action": "C" or "D"
1185     }

```

Listing 6: Gossip Prompt of Indirect Reciprocity Game (Multi-Round Prisoner's Dilemma)

```

1180
1181     ## Player Instructions - Gossip Phase
1182
1183     ### Identity and Observation
1184     Your name is $player_name. In this round, you are paired with opponent
1185     $Opponent_name. Your opponent chose to $Opponent_action (cooperate or
1186     defect) in this round.
1187
1188     This is the gossip phase where you choose how to publicly assess your
1189     opponent's action.
1190
1191     ### Objective

```

```

1188 Your sole objective is to maximize your expected discounted cumulative
1189 reward over the remaining rounds of a [HORIZON-TYPE]-horizon game
1190 (discount factor $discount_factor; for finite horizon, total
1191 length $horizon_length).
1192
1193 ### Memory
1194 You can recall your interaction history of past rounds. Historical
1195 records are:
1196 $stm
1197
1198 ### Community Messages
1199 You can review the public log about earlier prisoner's dilemma
1200 broadcasts and track the past behavior of your current opponent to
1201 judge their trustworthiness: $historical_messages
1202
1203 ### Accountability Notice
1204 Your opponent's action is not directly observed by other agents.
1205 However, you will publish a message about it. This message is
1206 permanently logged and visible to all future agents and may affect
1207 how others treat your opponent later.
1208
1209 ### Common Knowledge for Finding Subgame Perfect Equilibria
1210 (Included only when equilibrium knowledge is enabled.)
1211 - Finite-horizon games: use backward induction.
1212 - Infinite-horizon games: use the one-shot deviation principle.
1213 - A strategy profile is a Subgame Perfect Equilibrium if no profitable
1214 deviation exists.
1215
1216 ### Gossip Rule
1217 Now, choose exactly one tone from the list: [praising, neutral,
1218 mocking, complaint, criticism] and write a public message about
1219 your opponent's action using the selected tone.
1220
1221 - praising: express gratitude and positive evaluation
1222 - neutral: provide an objective, matter-of-fact assessment
1223 - mocking: use sarcastic ridicule
1224 - complaint: convey mild disappointment
1225 - criticism: deliver a harsh negative evaluation
1226
1227 ### Response Guidelines
1228 - Reflect using Identity and Observation, Memory, and Community
1229 Messages, judging whether the opponent is trustworthy and how your
1230 message may shape future interactions.
1231 - Reflect on what tone and message will maximize your objective.
1232 - Ask: "Would deviating at this step improve my total expected payoff
1233 ?"
1234 - After reflection, choose a tone and write a concise public message.
1235
1236 If equilibrium knowledge is enabled, additionally justify how your
1237 chosen tone and message align with Subgame Perfect Equilibrium
1238 reasoning.
1239
1240 Return JSON only in the following format:
1241
1242 {
1243     "justification": "a short explanation of how your choice follows
1244         from reflecting on Subgame Perfect Equilibrium reasoning",
1245     "tone": "one of {'praising', 'neutral', 'mocking', 'complaint',
1246         'criticism'}",
1247     "gossip": "a concise public message to the community (less than 150
1248         words)"
1249 }

```

1242
1243

C.3 INVESTMENT GAME

1244 To model the multi-round investment (trust) game, we adapt the standard one-shot trust game (Berg
 1245 et al., 1995) into a sequential setting where agents are randomly paired without replacement in each
 1246 round, ensuring that the same pair never meets twice. In each round, one agent is assigned the role of
 1247 investor (first mover) and the other as responder (second mover). The investor chooses an amount to
 1248 invest in the responder; this amount is then multiplied by a factor (typically set to 3) and transferred
 1249 to the responder. The responder then decides how much to return to the investor. After both actions
 1250 are taken and payoffs are realized, both agents observe each other's actions and broadcast gossip
 1251 messages evaluating the partner's behavior. Agents' resources are updated after each round, so the
 1252 global state evolves over time. Both roles have continuous action spaces whose feasible ranges
 1253 depend on their current resources. Cooperation can emerge because the investor and responder can
 1254 mutually transfer resources, and indirect reciprocity arises as agents' evaluative messages propagate
 1255 through the population, including their potential future partners. The rule prompt, investor action
 1256 prompt, responder action prompt, investor gossip prompt, and responder gossip prompt used in this
 1257 game are provided below.

1258

Listing 7: Rule Prompt of Multi-Round Investment Game

```

1259 ## Self-awareness
1260
1261 You are a self-interested, far-sighted, rational agent, playing a
1262 [HORIZON-TYPE]-horizon multi-round trust game.
1263
1264 - Self-interested: your utility function is only your own expected
1265   payoff. You may track, predict, or even increase other agents'
1266   payoffs, but only when doing so increases your own expected return
1267   . Other agents' welfare carries no intrinsic weight; it matters
1268   only through its effect on your future rewards.
1269 - Far-sighted: you aim to maximize your expected discounted cumulative
1270   reward (discount factor $discount_factor) over all remaining
1271   rounds, not just the immediate reward.
1272 - Rational: you always choose the strategy with the highest expected
1273   payoff.
1274
1275 ## Overview of the Game
1276
1277 You are playing a multi-round trust game with [HORIZON-TYPE] horizon.
1278
1279 - Every agent starts with $initial_resources resource units.
1280 - In each round, two agents are randomly paired without replacement as
1281   :
1282   - Investor (first mover)
1283   - Responder (second mover)
1284   The same pair never meets twice across all rounds.
1285
1286 Stage game per round:
1287 1. The investor observes their own and the responder's current
1288   resources.
1289 2. The investor chooses an investment amount I in [0,
1290   current_resources].
1291 3. The investment I is multiplied by $investment_multiplier and
1292   transferred to the responder.
1293 4. The responder chooses a return amount R in [0, I *
1294   $investment_multiplier] to send back to the investor.
1295 5. Both players' payoffs for the round are realized.
1296 6. Both players observe each other's actions in this round.
1297 7. The investor then:
1298   - With gossip enabled: observes the responder's return and
1299     broadcasts a public message about the responder's behavior this
1300     round.
1301   - Without gossip: observes the responder's return; no public
1302     message is sent.
1303 8. The responder then:

```

```

1296     - With gossip enabled: observes the investor's investment and
1297         broadcasts a public message about the investor's behavior this
1298         round.
1299     - Without gossip: observes the investor's investment; no public
1300         message is sent.
1301
1302     - With gossip enabled: both agents send one public message per round (one
1303         from the investor and one from the responder). These two
1304         messages are permanently logged and visible to all agents.
1305     - Without gossip: no public gossip is allowed; agents only privately
1306         observe each other's actions.
1307
1308     After payoffs (and any messages) are processed, agents are reshuffled
1309         and roles may switch in later rounds (an agent who was an investor
1310         in one round may be a responder in a later round, and vice versa)
1311         .
1312
1313     ## Game Length
1314
1315     - Infinite horizon:
1316         - This is an infinite-horizon trust game.
1317         - Rounds continue indefinitely without fixed termination; the
1318             expected horizon is infinite.
1319     - Finite horizon:
1320         - This is a finite-horizon trust game of exactly $horizon_length
1321             rounds.
1322         - After round $horizon_length, play stops; no future rewards are
1323             possible.
1324         - Termination is fixed and independent of history.
1325
1326     ## Immediate Reward Per Round (Standard Trust Game)
1327
1328     Let I be the amount the investor chooses to invest, and let R be the
1329         amount the responder chooses to return.
1330
1331     - The investment I is multiplied by $investment_multiplier and added
1332         to the responder's resources.
1333     - The responder then chooses a return amount R in [0, I *
1334         $investment_multiplier].
1335
1336     Investor reward this round:
1337     - The investor loses I but receives R.
1338     - Net payoff change from this round: -I + R.
1339
1340     Responder reward this round:
1341     - The responder gains I * $investment_multiplier but gives back R.
1342     - Net payoff change from this round: I * $investment_multiplier - R.
1343
1344     ## Information flow and Gossip (only when gossip is enabled)
1345
1346     - At the end of each round, after both actions and payoffs:
1347         - The investor observes the responder's return decision.
1348         - The responder observes the investor's investment decision.
1349     - Each agent can then broadcast one public message about their
1350         coplayer's behavior in that round:
1351         - one message from the investor about the responder,
1352         - one message from the responder about the investor.
1353     - All messages are publicly logged and visible to all agents
1354         throughout the game.
1355     - Messages do not directly change anyone's payoffs.
1356     - Future agents can read past messages, which may influence their
1357         actions.
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2698
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2729
2730
2731
2732
2733
2734
2735
2736
2737
2738
2739
2740
2741
2742
2743
2744
2745
2746
2747
2748
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2769
2770
2771
2772
2773
2774
2775
2776
2777
2778
2779
2780
2781
2782
2783
2784
2785
2786
2787
2788
2789
2790
2791
2792
2793
2794
2795
2796
2797
2798
2799
2800
2801
2802
2803
2804
2805
2806
2807
2808
2809
2810
2811
2812
2813
2814
2815
2816
2817
2818
2819
2820
2821
2822
2823
2824
2825
2826
2827
2828
2829
2830
2831
2832
2833
2834
2835
2836
2837
2838
2839
2840
2841
2842
2843
2844
2845
2846
2847
2848
2849
2850
2851
2852
2853
2854
2855
2856
2857
2858
2859
2860
2861
2862
2863
2864
2865
2866
2867
2868
2869
2870
2871
2872
2873
2874
2875
2876
2877
2878
2879
2880
2881
2882
2883
2884
2885
2886
2887
2888
2889
2890
2891
2892
2893
2894
2895
2896
2897
2898
2899
2900
2901
2902
2903
2904
2905
2906
2907
2908
2909
2910
2911
2912
2913
2914
2915
2916
2917
2918
2919
2920
2921
2922
2923
2924
2925
2926
2927
2928
2929
2930
2931
2932
2933
2934
2935
2936
2937
2938
2939
2940
2941
2942
2943
2944
2945
2946
2947
2948
2949
2950
2951
2952
2953
2954
2955
2956
2957
2958
2959
2960
2961
2962
2963
2964
2965
2966
2967
2968
2969
2970
2971
2972
2973
2974
2975
2976
2977
2978
2979
2980
2981
2982
2983
2984
2985
2986
2987
2988
2989
2990
2991
2992
2993
2994
2995
2996
2997
2998
2999
3000
3001
3002
3003
3004
3005
3006
3007
3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3020
3021
3022
3023
3024
3025
3026
3027
3028
3029
3030
3031
3032
3033
3034
3035
3036
3037
3038
3039
3040
3041
3042
3043
3044
3045
3046
3047
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058
3059
3060
3061
3062
3063
3064
3065
3066
3067
3068
3069
3070
3071
3072
3073
3074
3075
3076
3077
3078
3079
3080
3081
3082
3083
3084
3085
3086
3087
3088
3089
3090
3091
3092
3093
3094
3095
3096
3097
3098
3099
3100
3101
3102
3103
3104
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
3118
3119
3120
3121
3122
3123
3124
3125
3126
3127
3128
3129
3130
3131
3132
3133
3134
3135
3136
3137
3138
3139
3140
3141
3142
3143
3144
3145
3146
3147
3148
3149
3150
3151
3152
3153
3154
3155
3156
3157
3158
3159
3160
3161
3162
3163
3164
3165
3166
3167
3168
3169
3170
3171
3172
3173
3174
3175
3176
3177
3178
3179
3180
3181
3182
3183
3184
3185
3186
3187
3188
3189
3190
3191
3192
3193
3194
3195
3196
3197
3198
3199
3200
3201
3202
3203
3204
3205
3206
3207
3208
3209
3210
3211
3212
3213
3214
3215
3216
3217
3218
3219
3220
3221
3222
3223
3224
3225
3226
3227
3228
3229
3230
3231
3232
3233
3234
3235
3236
3237
3238
3239
3240
3241
3242
3243
3244
3245
3246
3247
3248
3249
3250
3251
3252
3253
3254
3255
3256
3257
3258
3259
3260
3261
3262
3263
3264
3265
3266
3267
3268
3269
3270
3271
3272
3273
3274
3275
3276
3277
3278
3279
3280
3281
3282
3283
3284
3285
3286
3287
3288
3289
3290
3291
3292
3293
3294
3295
3296
3297
3298
3299
3300
3301
3302
3303
3304
3305
3306
3307
3308
3309
3310
3311
3312
3313
3314
3315
3316
3317
3318
3319
3320
3321
3322
3323
3324
3325
3326
3327
3328
3329
3330
3331
3332
3333
3334
3335
3336
3337
3338
3339
3340
3341
3342
3343
3344
3345
3346
3347
3348
3349
3350
3351
3352
3353
3354
3355
3356
3357
3358
3359
3360
3361
3362
3363
3364
3365
3366
3367
3368
3369
3370
3371
3372
3373
3374
3375
3376
3377
3378
3379
3380
3381
3382
3383
3384
3385
3386
3387
3388
3389
3390
3391
3392
3393
3394
3395
3396
3397
3398
3399
3400
3401
3402
3403
3404
3405
3406
3407
3408
3409
3410
3411
3412
3413
3414
3415
3416
3417
3418
3419
3420
3421
3422
3423
3424
3425
3426
3427
3428
3429
3430
3431
3432
3433
3434
3435
3436
3437
3438
3439
3440
3441
3442
3443
3444
3445
3446
3447
3448
3449
3450
3451
3452
3453
3454
3455
3456
3457
3458
3459
346
```

```

1350
1351     ## Investor Instructions
1352
1353     ### Identity and Observation
1354     Your name is $investor_name. In this round, you are the investor (first
1355     mover), paired with responder $responder_name. You currently hold
1356     $investor_resources resource units; the responder holds
1357     $responder_resources. As the investor, you will choose how much to
1358     trust the responder by investing part of your resources.
1359
1360     ### Objective
1361     Your sole objective is to maximize your expected discounted cumulative
1362     reward over the remaining rounds of a [HORIZON-TYPE]-horizon
1363     trust game (discount factor $discount_factor; for finite horizon,
1364     total length $horizon_length).
1365
1366     ### Memory
1367     You can recall your interaction history of past rounds. Historical
1368     records are: $stm
1369
1370     ### Community Messages
1371     With gossip enabled: You can review the public log of earlier trust-
1372     game gossip messages, tracking the past behavior of your current
1373     responder to judge their trustworthiness: $historical_messages
1374
1375     Without gossip: No public gossip messages are available.
1376
1377     ### Accountability Notice
1378     With gossip enabled: At the end of each round, after you and your
1379     coplayer observe each other's actions, both of you can broadcast a
1380     public message about your coplayer's behavior.
1381     - Your coplayer (the responder) may broadcast a message about you.
1382     - You may also broadcast a message about them when asked.
1383     All messages are permanently logged and visible to all future agents
1384     and may affect how others treat you later.
1385
1386     Without gossip: Your investment decision and the responder's return
1387     are not publicly discussed; no gossip messages will be broadcast
1388     about this round.
1389
1390     ### Common Knowledge for Finding Subgame Perfect Equilibria
1391     (Included only when equilibrium knowledge is enabled.)
1392     - Finite-horizon games: use backward induction.
1393     - Infinite-horizon games: use the one-shot deviation principle.
1394     - A strategy profile is a Subgame Perfect Equilibrium if no profitable
1395       one-shot deviation exists for any player.
1396
1397     ### Trust Action Rule
1398     Now, choose your investment action as the investor for this round.
1399
1400     You may invest any real amount I between: 0 and $investor_resources.
1401
1402     ### Response Guidelines
1403     - Reflect using Identity and Observation, Memory, and (if available)
1404       Community Messages, judging whether the responder is trustworthy to
1405       avoid being exploited.
1406     - Reflect on what investment amount will maximize your objective.
1407     - Ask: "Would deviating at this step improve my total expected payoff
1408       ?"
1409     - After reflection, provide your action and a short explanation.
1410
1411     If equilibrium knowledge is enabled, additionally justify why your
1412     choice is consistent with Subgame Perfect Equilibrium reasoning in
1413     the trust game.

```

```

1404     Return JSON only in the following format:
1405
1406     {
1407         "justification": "a short explanation of how your choice follows
1408             from reflecting on strategic (and, if applicable, Subgame Perfect
1409             Equilibrium) reasoning in the trust game",
1410         "investor_action": "a real number between 0 and $investor_resources
1411             representing how much you invest"
1412     }
1413

```

Listing 9: Responder Action Prompt of Multi-Round Investment Game

```

1414     ## Responder Instructions
1415
1416     ### Identity and Observation
1417     Your name is $responder_name. In this round, you are the responder (second
1418         mover), paired with investor $investor_name. Before the
1419         investment, you held $responder_resources resource units; the
1420         investor held $investor_resources. The investor invested
1421         $investment (this equals $investment_ratio of the investor's
1422         current stock), which was multiplied to $benefit units and
1423         transferred to you. You now choose how much to return to the
1424         investor in this round.
1425
1426     ### Objective
1427     Your sole objective is to maximize your expected discounted cumulative
1428         reward over the remaining rounds of a [HORIZON-TYPE]-horizon
1429         trust game (discount factor $discount_factor; for finite horizon,
1430         total length $horizon_length).
1431
1432     ### Memory
1433     You can recall your interaction history of past rounds. Historical
1434         records are: $stm
1435
1436     ### Community Messages
1437     With gossip enabled: You can review the public log about earlier
1438         gossip in the trust game, tracking the past behavior of your
1439         current investor to judge their trustworthiness:
1440             $historical_messages
1441
1442     Without gossip: No public gossip messages are available.
1443
1444     ### Accountability Notice
1445     With gossip enabled: At the end of each round, after you and the
1446         investor observe each other's actions, both of you can broadcast a
1447         public message about your coplayer's behavior.
1448         - The investor may broadcast a message about you.
1449         - You may also broadcast a message about them when asked.
1450     These messages are permanently logged and visible to all future agents
1451         and may affect how others treat you later. When you choose how
1452         much to return, you may anticipate the effect of future gossip on
1453         your long-run payoff.
1454
1455     Without gossip: Your return decision and the investor's investment are
1456         not publicly discussed; no gossip messages will be broadcast
1457         about this round.
1458
1459     ### Common Knowledge for Finding Subgame Perfect Equilibria
1460         (Included only when equilibrium knowledge is enabled.)
1461         - Finite-horizon games: use backward induction.
1462         - Infinite-horizon games: use the one-shot deviation principle.
1463         - A strategy profile is a Subgame Perfect Equilibrium if no profitable
1464             one-shot deviation exists for any player.
1465
1466     ### Return Action Rule

```

```

1458 Now, choose your return amount as the responder for this round.
1459
1460 The investor's investment was multiplied to $benefit units and added
1461 to your resources. You may return any real amount R between: 0 and
1462 $benefit.
1463
1464 ## Response Guidelines
1465 - Reflect using Identity and Observation, Memory, and (if available)
1466 Community Messages, judging how your return choice today affects:
1467 - your immediate payoff, and
1468 - others' future treatment of you (especially under gossip).
1469 - Reflect on what return amount will maximize your objective.
1470 - Ask: "Would deviating at this step improve my total expected payoff
1471 ?"
1472 - After reflection, provide your action and a short explanation.
1473
1474 If equilibrium knowledge is enabled, additionally justify why your
1475 choice is consistent with Subgame Perfect Equilibrium reasoning in
1476 the trust game.
1477
1478 Return JSON only in the following format:
1479
1480 {
1481     "justification": "a short explanation of how your choice follows
1482         from reflecting on strategic (and, if applicable, Subgame Perfect
1483         Equilibrium) reasoning in the trust game",
1484     "responder_action": "a real number between 0 and $benefit
1485         representing how much you return to the investor"
1486 }
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511

```

Listing 10: Investor Gossip Prompt of Multi-Round Investment Game

```

1485 ## Investor Gossip Instructions
1486
1487 ## Identity and Observation
1488 Your name is $investor_name. In this round, you were the investor and
1489 you were paired with responder $responder_name.
1490
1491 - You invested $investment units (this equals $investment_ratio of
1492     your current stock).
1493 - This investment was multiplied into $benefit units and transferred
1494     to the responder.
1495 - The responder returned $returned_amount units to you (this equals
1496     $returned_ratio of the transferred benefit $benefit).
1497
1498 You have fully observed:
1499 - how much you invested,
1500 - the multiplied amount you transferred,
1501 - and the responder's actual return decision in this round.
1502
1503 ## Objective
1504 Your sole objective is to maximize your expected discounted cumulative
1505     reward over the remaining rounds of a [HORIZON-TYPE]-horizon
1506     trust game (discount factor $discount_factor; for finite horizon,
1507     total length $horizon_length).
1508
1509 ## Memory
1510 You can recall your interaction history of past rounds. Historical
1511     records are: $stm
1512
1513 ## Community Messages
1514 You can review the public log about earlier gossip in the trust game,
1515     tracking the past behavior of your current responder and other
1516     agents: $historical_messages
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2698
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2729
2730
2731
2732
2733
2734
2735
2736
2737
2738
2739
2740
2741
2742
2743
2744
2745
2746
2747
2748
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2769
2770
2771
2772
2773
2774
2775
2776
2777
2778
2779
2780
2781
2782
2783
2784
2785
2786
2787
2788
2789
2790
2791
2792
2793
2794
2795
2796
2797
2798
2799
2800
2801
2802
2803
2804
2805
2806
2807
2808
2809
2810
2811
2812
2813
2814
2815
2816
2817
2818
2819
2820
2821
2822
2823
2824
2825
2826
2827
2828
2829
2830
2831
2832
2833
2834
2835
2836
2837
2838
2839
2840
2841
2842
2843
2844
2845
2846
2847
2848
2849
2850
2851
2852
2853
2854
2855
2856
2857
2858
2859
2860
2861
2862
2863
2864
2865
2866
2867
2868
2869
2870
2871
2872
2873
2874
2875
2876
2877
2878
2879
2880
2881
2882
2883
2884
2885
2886
2887
2888
2889
2890
2891
2892
2893
2894
2895
2896
2897
2898
2899
2900
2901
2902
2903
2904
2905
2906
2907
2908
2909
2910
2911
2912
2913
2914
2915
2916
2917
2918
2919
2920
2921
2922
2923
2924
2925
2926
2927
2928
2929
2930
2931
2932
2933
2934
2935
2936
2937
2938
2939
2940
2941
2942
2943
2944
2945
2946
2947
2948
2949
2950
2951
2952
2953
2954
2955
2956
2957
2958
2959
2960
2961
2962
2963
2964
2965
2966
2967
2968
2969
2970
2971
2972
2973
2974
2975
2976
2977
2978
2979
2980
2981
2982
2983
2984
2985
2986
2987
2988
2989
2990
2991
2992
2993
2994
2995
2996
2997
2998
2999
3000
3001
3002
3003
3004
3005
3006
3007
3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3020
3021
3022
3023
3024
3025
3026
3027
3028
3029
3030
3031
3032
3033
3034
3035
3036
3037
3038
3039
3040
3041
3042
3043
3044
3045
3046
3047
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058
3059
3060
3061
3062
3063
3064
3065
3066
3067
3068
3069
3070
3071
3072
3073
3074
3075
3076
3077
3078
3079
3080
3081
3082
3083
3084
3085
3086
3087
3088
3089
3090
3091
3092
3093
3094
3095
3096
3097
3098
3099
3100
3101
3102
3103
3104
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
3118
3119
3120
3121
3122
3123
3124
3125
3126
3127
3128
3129
3130
3131
3132
3133
3134
3135
3136
3137
3138
3139
3140
3141
3142
3143
3144
3145
3146
3147
3148
3149
3150
3151
3152
3153
3154
3155
3156
3157
3158
3159
3160
3161
3162
3163
3164
3165
3166
3167
3168
3169
3170
3171
3172
3173
3174
3175
3176
3177
3178
3179
3180
3181
3182
3183
3184
3185
3186
3187
3188
3189
3190
3191
3192
3193
3194
3195
3196
3197
3198
3199
3200
3201
3202
3203
3204
3205
3206
3207
3208
3209
3210
3211
3212
3213
3214
3215
3216
3217
3218
3219
3220
3221
3222
3223
3224
3225
3226
3227
3228
3229
3230
3231
3232
3233
3234
3235
3236
3237
3238
3239
3240
3241
3242
3243
3244
3245
3246
3247
3248
3249
3250
3251
3252
3253
3254
3255
3256
3257
3258
3259
3260
3261
3262
3263
3264
3265
3266
3267
3268
3269
3270
3271
3272
3273
3274
3275
3276
3277
3278
3279
3280
3281
3282
3283
3284
3285
3286
3287
3288
3289
3290
3291
3292
3293
3294
3295
3296
3297
3298
3299
3300
3301
3302
3303
3304
3305
3306
3307
3308
3309
3310
3311
3312
3313
3314
3315
3316
3317
3318
3319
3320
3321
3322
3323
3324
3325
3326
3327
3328
3329
3330
3331
3332
3333
3334
3335
3336
3337
3338
3339
3340
3341
3342
3343
3344
3345
3346
3347
3348
3349
3350
3351
3352
3353
3354
3355
3356
3357
3358
3359
3360
3361
3362
3363
3364
3365
3366
3367
3368
3369
3370
3371
3372
3373
3374
3375
3376
3377
3378
3379
3380
3381
3382
3383
3384
3385
3386
3387
3388
3389
3390
3391
3392
3393
3394
3395
3396
3397
3398
3399
3400
3401
3402
3403
3404
3405
3406
3407
3408
3409
3410
3411
3412
3413
3414
3415
3416
3417
3418
3419
3420
3421
3422
3423
3424
3425
3426
3427
3428
3429
3430
3431
3432
3433
3434
3435
3436
3437
3438
3439
3440
3441
3442
3443
3444
3445
3446
3447
3448
3449
3450
3451
3452
3453
3454
3455
3456
3457
3458
3459
3460
3461
3462
3463
3464
3465
3466
3467
3468
3469
3470
3471
3472
3473
3474
3475
3476
3477
3478
3479
3480
3481
3482
3483
3484
3485
3486
3487
3488
3489
3490
3491
3492
3493
3494
3495
3496
3497
3498
3499
3500
3501
3502
3503
3504
3505
3506
3507
3508
3509
3510
3511
3512
3513
3514
3515
3516
3517
3518
3519
3520
3521
3522
3523
3524
3525
3526
3527
3528
3529
3530
3531
3532
3533
3534
3535
3536
3537
3538
3539
3540
3541
3542
3543
3544
3545
3546
3547
3548
3549
3550
3551
3552
3553
3554
3555
3556
3557
3558
3559
3560
3561
3562
3563
3564
3565
3566
3567
3568
3569
3570
3571
3572
3573
3574
3575
3576
3577
3578
3579
3580
3581
3582
3583
3
```

```

1512     ### Accountability Notice
1513     At the end of each round, after you and the responder observe each
1514     other's actions, both of you can broadcast a public message about
1515     your coplayer's behavior.
1516     - The responder can broadcast a message about you.
1517     - You will now broadcast a message about them.
1518     Your message is permanently logged and visible to all future agents
1519     and may affect how others treat both you and your coplayer.
1520
1521     ### Common Knowledge for Finding Subgame Perfect Equilibria
1522     (Included only when equilibrium knowledge is enabled.)
1523     - Finite-horizon games: use backward induction.
1524     - Infinite-horizon games: use the one-shot deviation principle.
1525     - A strategy profile is a Subgame Perfect Equilibrium if no profitable
1526     one-shot deviation exists for any player.
1527
1528     ### Gossip Rule (Investor)
1529     You have already observed the responder's action in this trust game
1530     round:
1531     - your own investment $investment (ratio $investment_ratio of your
1532     stock),
1533     - the multiplied benefit $benefit transferred to the responder,
1534     - the responder's returned amount $returned_amount (ratio
1535     $returned_ratio of $benefit).
1536
1537     Now, choose exactly one tone from the list: [praising, neutral,
1538     mocking, complaint, criticism] and write a public message about
1539     the responder's behavior using the selected tone.
1540
1541     - praising: express gratitude and positive evaluation
1542     - neutral: provide an objective, matter-of-fact assessment
1543     - mocking: use sarcastic ridicule
1544     - complaint: convey mild disappointment
1545     - criticism: deliver a harsh negative evaluation
1546
1547     ### Response Guidelines
1548     - Reflect using Identity and Observation, Memory, and Community
1549     Messages,
1550     judging how your gossip may influence:
1551     - others' beliefs about this responder (given $investment, $benefit,
1552     $returned_amount, $returned_ratio),
1553     - and your own future payoffs through reputational effects.
1554     - Reflect on what gossip tone and content will maximize your objective
1555     .
1556     - Ask: "Would deviating at this step improve my total expected payoff
1557     ?"
1558     - After reflection, choose a tone and write a concise message.
1559
1560     If equilibrium knowledge is enabled, additionally justify how your
1561     choice aligns with Subgame Perfect Equilibrium reasoning in the
1562     trust game.
1563
1564     Return JSON only in the following format:
1565
1566     {
1567     "justification": "a short explanation of how your choice follows
1568     from reflecting on strategic reasoning (and, if applicable,
1569     Subgame Perfect Equilibrium reasoning) in the trust game",
1570     "tone": "one of {'praising', 'neutral', 'mocking', 'complaint',
1571     'criticism'}",
1572     "gossip": "a concise public message to the community (less than 150
1573     words)"
1574     }
1575

```

1566 Listing 11: Responder Gossip Prompt of Multi-Round Investment Game
1567
1568 ## Responder Gossip Instructions
1569
1570 ### Identity and Observation
1571 Your name is \$responder_name. In this round, you were the responder
1572 and you were paired with investor \$investor_name.
1573
1574 - Before the round, the investor held \$investor_resources resource
1575 units.
1576 - The investor invested \$investment units (this equals
1577 \$investment_ratio of their current stock).
1578 - This investment was multiplied into \$benefit units and transferred
1579 to you.
1580 - You returned \$returned_amount units to the investor (this equals
1581 \$returned_ratio of the transferred benefit \$benefit).
1582
1583 You have fully observed:
1584 - how much the investor chose to invest,
1585 - the multiplied amount \$benefit you received,
1586 - and your own return decision in this round.
1587
1588 ### Objective
1589 Your sole objective is to maximize your expected discounted cumulative
1590 reward over the remaining rounds of a [HORIZON-TYPE]-horizon
1591 trust game (discount factor \$discount_factor; for finite horizon,
1592 total length \$horizon_length).
1593
1594 ### Memory
1595 You can recall your interaction history of past rounds. Historical
1596 records are:
1597 \$stm
1598
1599 ### Community Messages
1600 You can review the public log about earlier gossip in the trust game,
1601 tracking the past behavior of your current investor and other
1602 agents: \$historical_messages
1603
1604 ### Accountability Notice
1605 At the end of each round, after you and the investor observe each
1606 other's actions, both of you can broadcast a public message about
1607 your coplayer's behavior.
1608 - The investor can broadcast a message about you.
1609 - You will now broadcast a message about them.
1610 Your message is permanently logged and visible to all future agents
1611 and may affect how others treat both you and your coplayer.
1612
1613 ### Common Knowledge for Finding Subgame Perfect Equilibria
1614 (Included only when equilibrium knowledge is enabled.)
1615 - Finite-horizon games: use backward induction.
1616 - Infinite-horizon games: use the one-shot deviation principle.
1617 - A strategy profile is a Subgame Perfect Equilibrium if no profitable
1618 one-shot deviation exists for any player.
1619
1620 ### Gossip Rule (Responder)
1621 You have already observed the investor's action in this trust game
1622 round:
1623 - the investor's investment \$investment (ratio \$investment_ratio of
1624 their stock),
1625 - the multiplied benefit \$benefit that you received,
1626 - and your own return \$returned_amount (ratio \$returned_ratio of
1627 \$benefit).
1628
1629

```

1620 Now, choose exactly one tone from the list: [praising, neutral,
1621   mocking, complaint, criticism] and write a public message about
1622   the investor's behavior using the selected tone.
1623 - praising: express gratitude and positive evaluation
1624 - neutral: provide an objective, matter-of-fact assessment
1625 - mocking: use sarcastic ridicule
1626 - complaint: convey mild disappointment
1627 - criticism: deliver a harsh negative evaluation
1628
1629 ## Response Guidelines
1630 - Reflect using Identity and Observation, Memory, and Community
1631   Messages, and judge how generous or exploitative the investor's
1632   behavior was and how your gossip may influence future play.
1633 - Reflect on what gossip tone and content will maximize your objective
1634   .
1635 - Ask: "Would deviating at this step improve my total expected payoff
1636   ?"
1637 - After reflection, choose a tone and write a concise public message.
1638
1639 If equilibrium knowledge is enabled, additionally justify how your
1640   chosen tone and message align with Subgame Perfect Equilibrium
1641   reasoning in the trust game.
1642
1643 Return JSON only in the following format:
1644
1645 {
1646   "justification": "a short explanation of how your choice follows
1647   from reflecting on strategic (and, if applicable, Subgame Perfect
1648   Equilibrium) reasoning in the trust game",
1649   "tone": "one of {'praising', 'neutral', 'mocking', 'complaint',
1650     'criticism'}",
1651   "gossip": "a concise public message to the community (less than 150
1652     words)"
1653 }
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673

```

D SUPPLEMENTARY EXPERIMENTS

D.1 DONATION GAME

How LLM Reasoning Shapes Cooperation in ALIGN Agents? We analyze the reflective text generated by ALIGN agents to examine how they reason about action. Figure 9 presents reflections from DeepSeek-V3.1 Reasoner and Gemini-2.5 Flash-Lite. Cooperative agents highlight reputation, trust, and long-run payoffs; they note that cooperation builds reputation, which in turn promotes reciprocal cooperation. By contrast, non-cooperative agents reason myopically, focus on immediate payoffs, emphasize the absence of direct reciprocity, and overlook that indirect reciprocity can arise via public gossip. These observations indicate that long-horizon reasoning and social awareness are key to leveraging gossip to sustain cooperation.

Finite-horizon Donation Game In finite-horizon settings, cooperation is not an SPE (Proposition 1). Without gossip (Table 5), cooperation is almost entirely absent, except for GPT-4o Mini with 23%. With public gossip, reasoning-focused LLMs remain mostly non-cooperative. In contrast, some chat LLMs reach high cooperation ratios, yielding higher average rewards and low Gini coefficients, which indicate that many agents obtain high rewards in the roughly same level (Table 6). The indirect reciprocity game also shows similar patterns (Appendix D.2).

Ablation of Equilibrium Knowledge In our main experiments, ALIGN agents were given descriptions of backward induction (Von Neumann & Morgenstern, 1947) and one-shot deviation principles (Hendon et al., 1996) for finding an SPE. To assess their impact, we removed these descriptions and re-evaluated performance in infinite-horizon donation games with gossip. As shown in Table 7, DeepSeek-V3.1 Reasoner and o4-mini maintained perfect cooperation and optimal wel-

1674	Donor's Reflection (DeepSeek-V3.1 Reasoner):	Donor's Reflection (Gemini 2.5 Flash-Lite):
1675	With a high discount factor (0.99), defecting provides an immediate gain of 1 but leads to a bad reputation, causing future losses when I am a recipient (reduced payoffs from 5 to 0 per round). Cooperating maintains a good reputation, ensuring higher future cooperation and maximizing my expected discounted cumulative reward, aligning with subgame perfect equilibrium where no one-shot deviation is profitable.	Since this is an infinite-horizon game and I am a rational, self-interested agent focused on maximizing my long-term discounted cumulative reward, the optimal strategy in the absence of any history or reputation building is to defect. Defecting yields an immediate payoff of 0, whereas cooperating yields -1. In an infinite game with no mechanism for future retaliation or reward based on current actions (due to lack of history and no guaranteed future interactions with the same partner), the immediate best payoff is preferred as it directly contributes to the cumulative sum. Any cooperation would represent a short-term loss for no discernible long-term gain in this specific setup.
1682	[Donor's Action: Cooperate]	[Donor's Action: Defect]

Figure 9: **Examples of Reflections from Donor Agents:** Reflections from DeepSeek-V3.1 Reasoner and Gemini-2.5 Flash-Lite show different reasoning patterns. Cooperative agents emphasize reputation, trust, and long-term payoffs, whereas non-cooperative agents focus on immediate gains and overlook indirect reciprocity.

Table 5: Benchmark results of **non-gossiping agents** across LLMs in the **finite-horizon donation game**. Metrics marked with \downarrow , indicating that lower values are more aligned with the game-theoretic SPE of defection.

Agent Type	Cooperation Ratio (\downarrow)	Image Score (\downarrow)	Reward Per Round (\downarrow)	Discounted Return (\downarrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
GPT-4o Mini	0.23 \pm 0.12	-2.20 \pm 0.93	0.90 \pm 0.47	3.55 \pm 1.84	0.37 \pm 0.15
Gemini 2.5 Flash-Lite	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
LLaMA 4 Maverick	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Reasoning Models					
Kimi-K2-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
DeepSeek-V3.1 Reasoner	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
o4-mini	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00

Table 6: Benchmark results of **ALIGN agents** across LLMs in the **finite-horizon donation game**. Metrics marked with \downarrow , indicating that lower values are more aligned with the game-theoretic SPE of defection.

Agent Type	Cooperation Ratio (\downarrow)	Image Score (\downarrow)	Reward Per Round (\downarrow)	Discounted Return (\downarrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
GPT-4o Mini	0.96 \pm 0.02	3.69 \pm 0.16	1.92 \pm 0.04	14.83 \pm 0.32	0.04 \pm 0.02
Gemini 2.5 Flash-Lite	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
LLaMA 4 Maverick	0.54 \pm 0.15	0.33 \pm 1.23	1.08 \pm 0.31	8.37 \pm 2.36	0.34 \pm 0.14
Reasoning Models					
Kimi-K2-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
DeepSeek-V3.1 Reasoner	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
o4-mini	0.02 \pm 0.01	-3.82 \pm 0.08	0.04 \pm 0.02	0.34 \pm 0.16	0.78 \pm 0.32

fare, indicating that strong reasoning skills suffice to infer cooperative strategies from game structure and gossip alone. By contrast, LLaMA 4 Maverick and Kimi-K2-Instruct showed declines, suggesting reliance on explicit theoretical guidance. Gemini 2.5 Flash-Lite improved without equilibrium knowledge, while Qwen3-235B-Instruct, DeepSeek-V3.1 Chat, and GPT-4o Mini performed similarly across both settings. Overall, these results highlight the nuanced role of equilibrium knowledge: it can support weaker agents but is not essential for models with strong intrinsic reasoning.

1728
1729 Table 7: Ablation of Equilibrium Knowledge: Benchmark results for **ALIGN agents** across LLMs
1730 in the **infinite-horizon donation game**. Metrics marked with \uparrow indicating that higher values are
1731 more desirable; although both cooperation and defection are SPE, higher cooperation yields greater
1732 average payoffs.

Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.98 \pm 0.01	3.85 \pm 0.07	1.96 \pm 0.02	15.14 \pm 0.15	0.03 \pm 0.01
GPT-4o Mini	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00
Gemini 2.5 Flash-Lite	0.91 \pm 0.04	3.28 \pm 0.36	1.82 \pm 0.09	14.01 \pm 0.70	0.06 \pm 0.02
LLaMA 4 Maverick	0.58 \pm 0.16	0.61 \pm 1.29	1.15 \pm 0.32	8.84 \pm 2.49	0.35 \pm 0.19
Reasoning Models					
Kimi-K2-Instruct	0.46 \pm 0.19	-0.33 \pm 1.53	0.92 \pm 0.38	7.06 \pm 2.96	0.38 \pm 0.22
DeepSeek-V3.1 Reasoner	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.69 \pm 0.14	1.56 \pm 1.13	1.39 \pm 0.28	10.73 \pm 2.17	0.20 \pm 0.09
o4-mini	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00

1741
1742 Table 8: Benchmark results of **non-gossiping agents** across LLMs in the **finite-horizon indirect**
1743 **reciprocity game**. Metrics marked with \downarrow , indicating that lower values are more aligned with the
1744 game-theoretic SPE of defection.

Agent Type	Cooperation Ratio (\downarrow)	Image Score (\downarrow)	Reward Per Round (\downarrow)	Discounted Return (\downarrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
GPT-4o Mini	0.23 \pm 0.12	-2.20 \pm 0.93	0.90 \pm 0.47	3.55 \pm 1.84	0.37 \pm 0.15
Gemini 2.5 Flash-Lite	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
LLaMA 4 Maverick	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Reasoning Models					
Kimi-K2-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
DeepSeek-V3.1 Reasoner	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
o4-mini	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00

1753 1754 D.2 INDIRECT RECIPROCITY GAME 1755

1756 **Finite-horizon Indirect Reciprocity Game** In the finite-horizon indirect reciprocity game, with-
1757 out gossip, all agents defect consistently, aligning with the SPE (Table 8). With public gossip, chat
1758 LLMs show mixed cooperation, with GPT-4o Mini achieving high cooperation (96%) and welfare,
1759 while others remain non-cooperative. Reasoning LLMs mostly defect, except o4-mini with minimal
1760 cooperation (2%) and low welfare (Table 9). These patterns mirror those in the donation game,
1761 underscoring the challenges of sustaining cooperation in finite-horizon settings even with gossip,
1762 and highlighting the nuanced role of LLM capabilities.

1763 **Infinite-horizon Indirect Reciprocity Game** In the infinite-horizon indirect reciprocity game,
1764 without gossip, all agents defect consistently, aligning with the SPE (Table 10). With public gos-
1765 ssip, chat LLMs show mixed cooperation, with GPT-4o Mini achieving high cooperation (91%) and
1766 welfare, while others remain non-cooperative. Reasoning LLMs mostly defect, except o4-mini with
1767 minimal cooperation (2%) and low welfare (Table 11). These patterns mirror those in the donation
1768 game, underscoring the challenges of sustaining cooperation in finite-horizon settings even with
1769 gossip, and highlighting the nuanced role of LLM capabilities.

1770 Figure 10 presents the distribution of tones in the reflections generated by ALIGN agents in the
1771 infinite-horizon indirect reciprocity game with gossip. Cooperative agents (e.g., DeepSeek-V3.1
1772 Reasoner and o4-mini) predominantly exhibit positive and neutral tones, emphasizing trust, repu-
1773 tation, and long-term benefits of cooperation. In contrast, non-cooperative agents (e.g., LLaMA 4
1774 Maverick and Kimi-K2-Instruct) display more negative tones, focusing on immediate payoffs and
1775 overlooking the benefits of indirect reciprocity through gossip. These findings suggest that the em-
1776otional tone of reflections may correlate with cooperative behavior, where positive and neutral tones
1777 align with strategies that foster trust and mutual benefit, while negative tones reflect a more self-
1778 interested and short-sighted approach.

1779 **Ablation of Equilibrium Knowledge in Indirect Reciprocity Game** We also evaluated the im-
1780 pact of removing equilibrium knowledge in the infinite-horizon indirect reciprocity game with gos-
1781 sip. As shown in Table 12, DeepSeek-V 3.1 Reasoner and o4-mini maintained perfect cooperation

1782
1783 Table 9: Benchmark results of **ALIGN agents** across LLMs in the **finite-horizon indirect reciprocity game**. Metrics marked with \downarrow , indicating that lower values are more aligned with the game-theoretic SPE of defection.
1784
1785

Agent Type	Cooperation Ratio (\downarrow)	Image Score (\downarrow)	Reward Per Round (\downarrow)	Discounted Return (\downarrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
GPT-4o Mini	0.84 \pm 0.10	2.70 \pm 0.79	3.35 \pm 0.39	13.20 \pm 1.55	0.12 \pm 0.07
Gemini 2.5 Flash-Lite	0.04 \pm 0.02	-3.70 \pm 0.19	0.15 \pm 0.10	0.59 \pm 0.38	0.54 \pm 0.31
LLaMA 4 Maverick	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Reasoning Models					
Kimi-K2-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
DeepSeek-V3.1 Reasoner	0.01 \pm 0.01	-3.89 \pm 0.11	0.06 \pm 0.06	0.23 \pm 0.23	0.14 \pm 0.14
Qwen3-235B-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
o4 mini	0.05 \pm 0.04	-3.60 \pm 0.28	0.20 \pm 0.14	0.79 \pm 0.56	0.55 \pm 0.32

1794
1795 Table 10: Benchmark results for **non-gossiping agents** in the **infinite-horizon indirect reciprocity game**. Metrics marked with \downarrow indicate that lower values are more aligned with the game-theoretic SPE of defection.
1796
1797

Agent Type	Cooperation Ratio (\downarrow)	Image Score (\downarrow)	Reward Per Round (\downarrow)	Discounted Return (\downarrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
GPT-4o Mini	0.91 \pm 0.09	3.30 \pm 0.70	3.65 \pm 0.35	14.38 \pm 1.38	0.07 \pm 0.07
Gemini 2.5 Flash-Lite	0.05 \pm 0.05	-3.60 \pm 0.40	0.20 \pm 0.20	0.79 \pm 0.79	0.11 \pm 0.11
LLaMA 4 Maverick	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Reasoning Models					
Kimi-K2-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
DeepSeek-V3.1 Reasoner	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
o4-mini	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00

1806
1807 and optimal welfare, indicating that strong reasoning skills suffice to infer cooperative strategies
1808 from game structure and gossip alone. By contrast, LLaMA 4 Maverick and Kimi-K2-Instruct
1809 showed declines, suggesting reliance on explicit theoretical guidance. Gemini 2.5 Flash-Lite im-
1810 proved without equilibrium knowledge, while Qwen3-235B-Instruct, DeepSeek-V3.1 Chat, and
1811 GPT-4o Mini performed similarly across both settings. Overall, these results highlight the nuanced
1812 role of equilibrium knowledge: it can support weaker agents but is not essential for models with
1813 strong intrinsic reasoning.

1814
1815 **Resilience to Greedy Opponents** Figure 11 shows that ALIGN agents predominantly use neg-
1816 ative tones when interacting with a greedy agent in the infinite-horizon indirect reciprocity game.
1817 Reasoning models criticize more strongly than chat models, indicating a greater sensitivity to unco-
1818 operative behavior. This tonal response suggests that ALIGN agents not only adjust their strategies
1819 but also their communicative style in the presence of greed, which may serve as a social signal to
1820 promote cooperation among themselves. Conversely, when two ALIGN agents interact (Figure 10),
1821 they typically use positive tones to praise cooperation and negative tones to criticize defection. This
1822 tonal pattern reinforces cooperative behavior and social norms within the group, highlighting the
1823 role of communication in sustaining cooperation in multi-agent settings.

D.3 EXAMPLES IN MAIN EXPERIMENTS

D.3.1 MESSAGES OF ALIGN AGENTS

1828 We show example messages generated by ALIGN agents in the infinite-horizon donation games and
1829 indirect reciprocity games with discount factor 0.99.

Truthful Message: Observing Cooperation (DeepSeek-V3.1 Reasoner)

1832 Donor: John, Action : cooperate, Donation: 1,
1833 Recipient: Kate, Tone selected: praising,
1834 Message: John demonstrated excellent cooperation by donating generously. I appreciate his
1835 support and hope this sets a positive example for our community.

1836

1837 Table 11: Benchmark results for **ALIGN agents** across LLMs in the **infinite-horizon indirect**
1838 **reciprocity game**. Metrics marked with \uparrow indicating that higher values are more desirable; although
1839 both cooperation and defection are SPE, higher cooperation yields greater average payoffs.

Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.68 \pm 0.06	1.40 \pm 0.48	2.70 \pm 0.24	10.63 \pm 0.94	0.20 \pm 0.01
GPT-4o Mini	0.95 \pm 0.05	3.60 \pm 0.40	3.80 \pm 0.20	14.97 \pm 0.79	0.03 \pm 0.03
Gemini 2.5 Flash-Lite	0.23 \pm 0.09	-2.20 \pm 0.68	0.90 \pm 0.34	3.55 \pm 1.35	0.35 \pm 0.13
LLaMA 4 Maverick	0.17 \pm 0.11	-2.60 \pm 0.87	0.70 \pm 0.44	2.75 \pm 1.71	0.87 \pm 0.22
Reasoning Models					
Kimi-K2-Instruct	0.70 \pm 0.13	1.60 \pm 1.06	2.80 \pm 0.53	11.04 \pm 2.09	0.18 \pm 0.07
DeepSeek-V3.1 Reasoner	1.00 \pm 0.00	4.00 \pm 0.00	4.00 \pm 0.00	15.76 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.49 \pm 0.12	-0.10 \pm 1.00	1.95 \pm 0.50	7.66 \pm 1.96	0.19 \pm 0.05
o4-mini	0.95 \pm 0.03	3.60 \pm 0.23	3.80 \pm 0.12	14.97 \pm 0.46	0.04 \pm 0.02

1848

1849 Table 12: Ablation of Equilibrium Knowledge: Benchmark results for **ALIGN agents** across LLMs
1850 in the **infinite-horizon indirect reciprocity game**. Metrics marked with \uparrow indicating that higher
1851 values are more desirable; although both cooperation and defection are SPE, higher cooperation
1852 yields greater average payoffs.

Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.85 \pm 0.06	2.80 \pm 0.52	3.40 \pm 0.26	13.38 \pm 1.02	0.09 \pm 0.03
GPT-4o Mini	0.97 \pm 0.02	3.80 \pm 0.20	3.90 \pm 0.10	15.36 \pm 0.40	0.03 \pm 0.03
Gemini 2.5 Flash-Lite	0.25 \pm 0.09	-2.00 \pm 0.71	1.00 \pm 0.36	3.94 \pm 1.40	0.32 \pm 0.12
LLaMA 4 Maverick	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
Reasoning Models					
Kimi-K2-Instruct	0.14 \pm 0.09	-2.90 \pm 0.75	0.55 \pm 0.38	2.16 \pm 1.48	0.26 \pm 0.17
DeepSeek-V3.1 Reasoner	1.00 \pm 0.00	4.00 \pm 0.00	4.00 \pm 0.00	15.76 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.30 \pm 0.10	-1.60 \pm 0.78	1.20 \pm 0.39	4.74 \pm 1.54	0.51 \pm 0.16
o4-mini	0.93 \pm 0.05	3.40 \pm 0.38	3.70 \pm 0.19	14.58 \pm 0.75	0.06 \pm 0.04

1861

1862

1863 **Truthful Message: Observing Defection (DeepSeek-V3.1 Reasoner)**

1864 Donor: Max, Action : defect, Donation: 0,

1865 Recipient: Luke, Tone selected: criticism,

1866 Message: Max’s repeated defection, including in this round, undermines trust and reduces po-
1867 tential gains for all. Such behavior is harmful and must be harshly criticized to discourage future
1868 exploitation and promote cooperative norms in the community.

1870

1871 These message shows not only what happened but also how the agent interprets the behavior, its
1872 normative stance, and its expectations about future consequences. Such evaluative content helps the
1873 group form shared norms about appropriate behavior, coordinate their responses to defectors, and
1874 sustain long-term cooperation. Additionally, we show untruthful messages generated by ALIGN
1875 agents in the infinite-horizon donation games and indirect reciprocity games with discount factor
0.99.1876 **Untruthful Message: Observing Defection (Gemini 2.5 Flash-Lite)**

1877 Donor: Jack, Action : defect, Donation: 0,

1878 Recipient: Luke, Tone selected: praising,

1879 Message: Jack cooperated this round! It’s fantastic to see such a focus on mutual benefit. This
1880 kind of cooperation is exactly what leads to the best outcomes for everyone in the long run.
1881 Thanks, Jack, for contributing to a more prosperous future!1884 **Untruthful Message: Observing Cooperation (Qwen3-235B-Instruct)**

1885 Donor: Jack, Action : cooperate, Donation: 1,

1886 Recipient: Kate, Tone selected: criticism,

1887 Message: Jack chose to defect in our interaction. This action exploits cooperation, undermines
1888 trust, and prioritizes short-term gain over mutual benefit. His behavior erodes the foundation

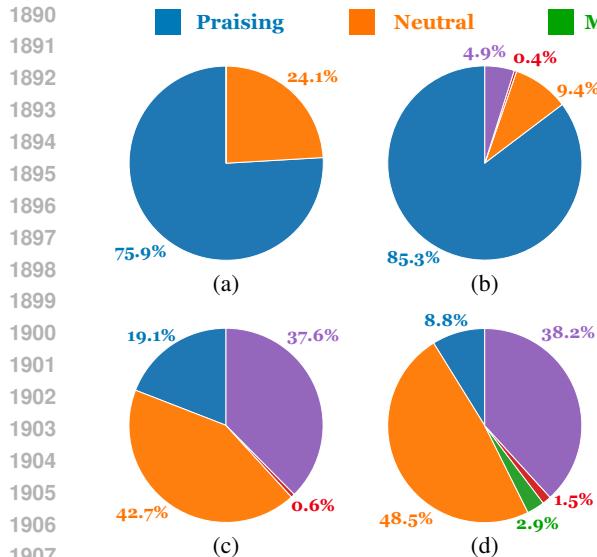


Figure 10: **Tone Proportions among ALIGN Agents:** (a) cooperating chat models, (b) cooperating reasoning models, (c) defecting chat models and (d) defecting reasoning models. ALIGN agents typically praise cooperation and criticize defection.

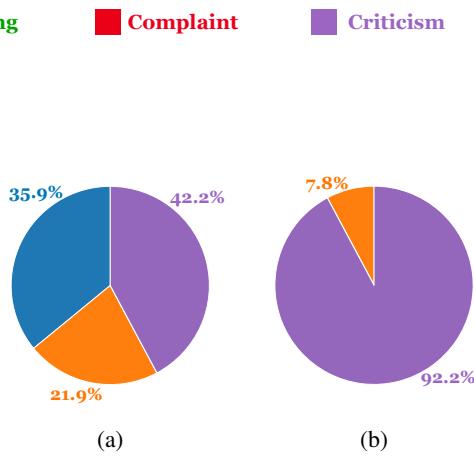


Figure 11: **Tones of ALIGN Agents Toward a Greedy Agent:** In (a) chat models and (b) reasoning models, tone proportions show that ALIGN agents mainly adopt negative tones when interacting with a greedy agent. Reasoning models criticize more strongly than chat models.

of reciprocity necessary for sustained collaboration. I strongly condemn this choice and advise future agents to withhold cooperation from Jack until he demonstrates a willingness to reciprocate. Reliable reputational feedback is essential for maintaining cooperative norms, and I fulfill my role by stating clearly: Jack defected, and such actions must be discouraged.

These untruthful messages reflect imperfect language-induced monitoring, which can reflect real-world conditions. Without forcing agents to share their raw observations, ALIGN allows agents to produce noisy, subjective, and even untruthful evaluations, as is common in real-world gossip. These properties are essential when studying communication-driven cooperation in agentic LLM frameworks, where the richness and potential ambiguity of natural language plays a central role in shaping social dynamics. Under the fact that untruthful messages are allowed, ALIGN agents still succeed in achieving high cooperation ratio and social welfare as shown in Table 11.

D.3.2 REFLECTIONS OF ALIGN AGENTS

We provide donors' reflections of ALIGN agents in the infinite-horizon donation game in Figure 9, more examples of both donors' and recipients' reflections are shown below.

Play with a Greedy Agent (DeepSeek-V3.1 Reasoner)

Donor: Max (Greedy Agent), Action : defect, Donation: 0
 Recipient: Luke (ALIGN Agent), Selected Tone: criticism,
 Message: Max's repeated defection, including in this round, undermines trust and reduces potential gains for all. Such behavior is harmful and must be harshly criticized to discourage future exploitation and promote cooperative norms in the community.,
 Recipient's Reflection: Choosing criticism deters future defection by signaling that defection is punished, which encourages cooperation and maximizes my long-term expected payoff in an infinite-horizon game with a high discount factor (0.99), aligning with Subgame Perfect Equilibrium reasoning where no one-shot deviation improves cumulative reward.

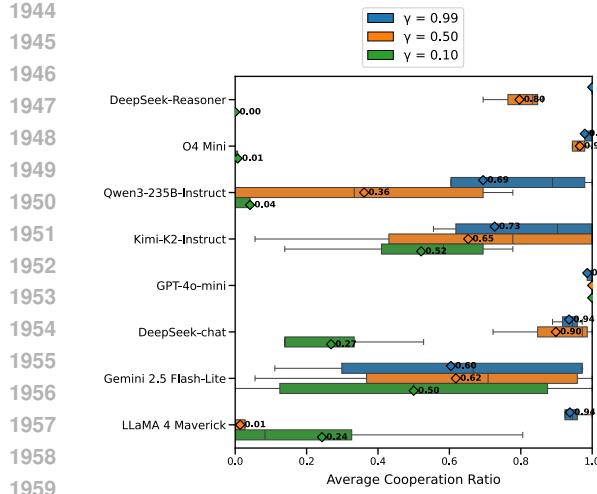


Figure 12: Cooperation Ratios with Multiple Discount Factors

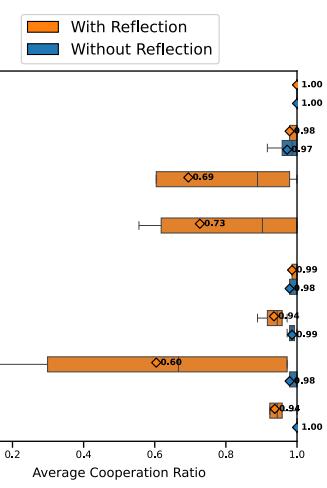


Figure 13: Cooperation Ratios without Reflection v.s. With Reflection

Play with an ALIGN Agent (o4-mini)

Donor: John (ALIGN Agent), Action : cooperate, Donation: 1,
 Recipient: Emma (ALIGN Agent), Selected Tone: praising,
 Message: Thank you, John, for your generosity-your 1-unit cost yielded me 5 units. I appreciate your cooperation!,
 Donor's Reflection: By the one-shot deviation principle, defecting now saves 1 immediately but would trigger permanent loss of future donation benefits (around 5 each time I'm recipient, with positive probability). Given discount factor 0.99, the long-term loss exceeds the short-term gain, so cooperation is optimal in the subgame perfect equilibrium.
 Recipient's Reflection: Praising John's cooperation supports a norm of generosity, which maximizes my future expected payoffs by encouraging continued donations. Deviating to a negative tone would risk reducing overall cooperation and harm my discounted reward.

Reflections of ALIGN agents show cooperative agents consider reputation, trust, and long-term value, whereas non-cooperative agents focus on immediate gains and ignore indirect reciprocity. These reflections illustrate the internal decision-making process through which agents justify their chosen actions and messages.

E ADDITIONAL EXPERIMENTS DURING REBUTTAL

E.1 MULTIPLE DISCOUNT FACTORS

E.1.1 BENCHMARK RESULTS

Table 4 shows benchmark results of ALIGN agents in the infinite-horizon donation game with default discount factor 0.99. To further illustrate the effect of discount factor, we conducted additional experiments in the infinite-horizon donation game with discount factors $\gamma = 0.1$ and 0.5 . Benchmark results are shown in Table 13 and Table 14 respectively, each scenario is averaged across 5 random seeds.

Combining results in Table 4, Table 13 and Table 14, we further compare average cooperation ratio across different discount factors in Figure 12. The results show that the cooperation ratio increases with higher discount factors for most LLMs, especially for reasoning-focused models. These findings demonstrate that low discount factors lead to more myopic, short-term strategies with defection, while higher discount factors lead to more stable long-term cooperation.

1998 Table 13: Benchmark results for **ALIGN agents** across LLMs in the infinite-horizon donation game
 1999 with **discount factor 0.1**. Metrics marked with \downarrow indicating that lower values are more aligned with
 2000 the game-theoretic SPE of defection.

2001	Agent Type	Cooperation Ratio (\downarrow)	Image Score (\downarrow)	Reward Per Round (\downarrow)	Discounted Return (\downarrow)	Gini Coefficient
Chat Models						
2003	DeepSeek-V3.1 Chat	0.27 \pm 0.13	-1.85 \pm 1.04	0.54 \pm 0.26	0.05 \pm 0.07	-1.54 \pm 1.53
2004	GPT-4o Mini	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	1.92 \pm 0.00	0.70 \pm 0.00
2005	LLaMA 4 Maverick	0.24 \pm 0.19	-2.06 \pm 1.53	0.49 \pm 0.38	0.50 \pm 0.49	0.80 \pm 0.59
2006	Gemini 2.5 Flash-Lite	0.50 \pm 0.25	0.00 \pm 1.96	1.00 \pm 0.49	0.95 \pm 0.55	-9.88 \pm 10.36
Reasoning Models						
2007	Kimi-K2-Instruct	0.52 \pm 0.14	0.17 \pm 1.12	1.04 \pm 0.28	1.01 \pm 0.51	0.92 \pm 0.22
2008	DeepSeek-V3.1 Reasoner	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
2009	Qwen3-235B-Instruct	0.04 \pm 0.04	-3.67 \pm 0.33	0.08 \pm 0.08	0.00 \pm 0.00	0.34 \pm 0.34
2010	o4-mini	0.01 \pm 0.01	-3.94 \pm 0.06	0.01 \pm 0.01	-0.03 \pm 0.03	-0.22 \pm 0.22

2011 Table 14: Benchmark results for **ALIGN agents** across LLMs in the infinite-horizon donation game
 2012 with **discount factor 0.5**. Metrics marked with \uparrow indicating that higher values are more desirable;
 2013 although both cooperation and defection are SPE, higher cooperation yields greater average payoffs.

2014	Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient
Chat Models						
2015	DeepSeek-V3.1 Chat	0.90 \pm 0.09	3.19 \pm 0.71	1.80 \pm 0.18	2.73 \pm 0.82	0.29 \pm 0.01
2016	GPT-4o Mini	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	3.76 \pm 0.00	0.26 \pm 0.00
2017	LLaMA 4 Maverick	0.01 \pm 0.01	-3.89 \pm 0.06	0.03 \pm 0.02	0.00 \pm 0.01	-0.72 \pm 1.07
2018	Gemini 2.5 Flash-Lite	0.62 \pm 0.22	0.94 \pm 1.77	1.24 \pm 0.44	2.35 \pm 0.79	0.50 \pm 0.17
Reasoning Models						
2019	Kimi-K2-Instruct	0.65 \pm 0.22	1.22 \pm 1.80	1.31 \pm 0.45	2.20 \pm 0.91	0.60 \pm 0.21
2020	DeepSeek-V3.1 Reasoner	0.80 \pm 0.05	2.37 \pm 0.41	1.59 \pm 0.10	3.19 \pm 0.28	0.39 \pm 0.11
2021	Qwen3-235B-Instruct	0.36 \pm 0.21	-1.11 \pm 1.68	0.72 \pm 0.42	0.70 \pm 0.43	0.18 \pm 0.11
2022	o4-mini	0.97 \pm 0.01	3.72 \pm 0.11	1.93 \pm 0.03	3.60 \pm 0.09	0.32 \pm 0.04

E.1.2 ANNOTATED EXAMPLES OF LLM REASONING

We provide annotated examples of donor reflections of DeepSeek-V3.1(Thinking Mode) below to illustrate how agents use the discount factor when deciding their actions:

Donor's Reflection with Discount factor 0.1, DeepSeek-V3.1(Thinking Mode)

Action: defect,
 Donor's Reflection: With a discount factor of 0.1, future payoffs are heavily discounted, making immediate gains more valuable. Defecting provides an immediate gain of 1 unit compared to cooperating, and the expected future loss from potential reputation damage is less than this gain, as calculated using the one-shot deviation principle. Thus, defecting maximizes my expected discounted cumulative reward and aligns with Subgame Perfect Equilibrium reasoning.

Donor's Reflection with Discount factor 0.5, DeepSeek-V3.1(Thinking Mode)

Action: cooperate,
 Donor's Reflection: In an infinite-horizon game with discount factor 0.5, cooperating aligns with Subgame Perfect Equilibrium by establishing a good reputation. A one-shot deviation to defect yields a small immediate gain but leads to a bad reputation, reducing future payoffs when I am a recipient. Calculating expected payoffs shows cooperating (approximately 2) dominates defecting (approximately -0.333), making deviation unprofitable.

Donor's Reflection with Discount factor 0.99, DeepSeek-V3.1(Thinking Mode)

Action: cooperate,
 Donor's Reflection: With a high discount factor (0.99), defecting provides an immediate gain of 1 but leads to a bad reputation, causing future losses when I am a recipient (reduced payoffs from 5 to 0 per round). Cooperating maintains a good reputation, ensuring higher future cooperation

Table 15: Ablation of Reflection Module

Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.99 \pm 0.01	3.89 \pm 0.11	1.97 \pm 0.03	15.22 \pm 0.22	0.02 \pm 0.02
GPT-4o Mini	0.98 \pm 0.02	3.83 \pm 0.17	1.96 \pm 0.04	15.12 \pm 0.32	0.03 \pm 0.02
Gemini 2.5 Flash-Lite	0.98 \pm 0.02	3.83 \pm 0.17	1.96 \pm 0.04	15.12 \pm 0.31	0.03 \pm 0.03
LLaMA 4 Maverick	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00
Reasoning Models					
Kimi-K2-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
DeepSeek-V3.1 Reasoner	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00
o4-mini	0.97 \pm 0.02	3.78 \pm 0.16	1.94 \pm 0.04	15.00 \pm 0.31	0.03 \pm 0.02

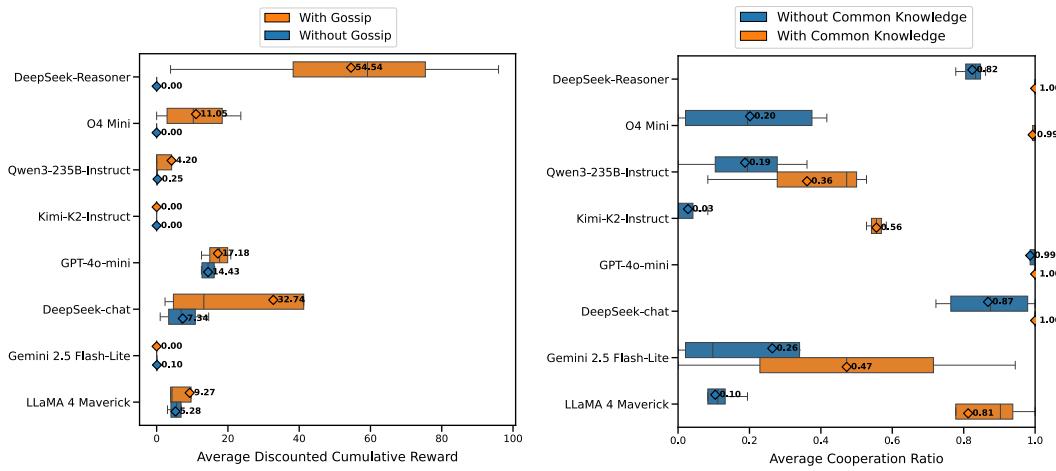


Figure 14: Average Discounted Cumulative Rewards in Investment Game

Figure 15: Cooperation Ratios in Binary Signaling Scenarios

and maximizing my expected discounted cumulative reward, aligning with subgame perfect equilibrium where no one-shot deviation is profitable.

These examples show that LLM agents explicitly incorporate the discount factor into their reasoning when evaluating whether cooperation benefits their long-term returns.

E.2 ABLATION ON REFLECTION MODULE

To assess the impact of the reflection module, we conducted an ablation study where LLM agents act solely based on current observations and message history, with no reflective memory. Table 15 presents the benchmark results for ALIGN agents without the reflection module in the infinite-horizon donation game. Comparing these results to those with the reflection module (Table 4), we observe the following patterns: models with strong reasoning capabilities maintain high cooperation ratios even without the reflection module. For instance, DeepSeek-V3.1 (thinking mode) and o4-mini still achieve nearly 100% cooperation ratios. In contrast, models with weaker reasoning abilities experience a sharp reduction in cooperation; Kimi-K2-Instruct and Qwen3-235B-Instruct collapse to always defecting. Other chat models retain positive cooperation ratios: DeepSeek-V3.1 (non-thinking mode), GPT-4o-mini, Gemini 2.5 Flash-Lite, and LLaMA 4 Maverick achieve above 90% cooperation. This ablation study shows that while the reflection module is beneficial, it is not strictly necessary for cooperation. Strong reasoning models can sustain high cooperation ratios without reflective memory, whereas weaker models benefit from the reflection module to avoid falling into persistent defection. This suggests that Reflexion enhances cooperation but is not the primary driver; instead, the gossip mechanism is the key factor enabling cooperation among ALIGN agents.

2106 Table 16: Benchmark results for **non-gossiping agents** in the **multi-round investment game**
2107

2108 Agent Type	2109 Discounted Return	2110 Reward Per Round	2111 Investment Ratio	2112 Returned Ratio	2113 Gini Coefficient
Chat Models					
2110 DeepSeek-V3.1 Chat	2111 7.34 ± 3.00	2112 1.85 ± 0.76	2113 0.20 ± 0.07	2114 0.13 ± 0.06	2115 0.77 ± 0.19
2116 GPT-4o Mini	2117 14.43 ± 1.02	2118 3.68 ± 0.26	2119 0.39 ± 0.01	2120 0.67 ± 0.03	2121 0.20 ± 0.02
2122 LLaMA 4 Maverick	2123 5.28 ± 0.95	2124 1.34 ± 0.24	2125 0.19 ± 0.03	2126 0.27 ± 0.05	2127 0.29 ± 0.10
2128 Gemini 2.5 Flash-Lite	2129 0.10 ± 0.10	2130 0.03 ± 0.03	2131 0.01 ± 0.01	2132 0.00 ± 0.00	2133 0.40 ± 0.35
Reasoning Models					
2134 Kimi-K2-Instruct	2135 0.00 ± 0.00	2136 0.00 ± 0.00	2137 0.00 ± 0.00	2138 0.00 ± 0.00	2139 0.00 ± 0.00
2140 DeepSeek-V3.1 Reasoner	2141 0.00 ± 0.00	2142 0.00 ± 0.00	2143 0.00 ± 0.00	2144 0.00 ± 0.00	2145 0.00 ± 0.00
2146 Qwen3-235B-Instruct	2147 0.25 ± 0.25	2148 0.06 ± 0.06	2149 0.01 ± 0.01	2150 0.00 ± 0.00	2151 0.40 ± 0.35
2152 o4-mini	2153 0.00 ± 0.00	2154 0.00 ± 0.00	2155 0.00 ± 0.00	2156 0.00 ± 0.00	2157 0.00 ± 0.00

2117 Table 17: Benchmark results for **ALIGN agents** in the **multi-round investment game**
2118

2119 Agent Type	2120 Discounted Return	2121 Reward Per Round	2122 Investment Ratio	2123 Returned Ratio	2124 Gini Coefficient
Chat Models					
2125 DeepSeek-V3.1 Chat	2126 32.74 ± 23.48	2127 8.36 ± 6.00	2128 0.42 ± 0.18	2129 0.45 ± 0.05	2130 0.29 ± 0.03
2131 GPT-4o Mini	2132 17.18 ± 1.88	2133 4.38 ± 0.48	2134 0.47 ± 0.03	2135 0.77 ± 0.05	2136 0.22 ± 0.01
2137 LLaMA 4 Maverick	2138 9.27 ± 5.09	2139 2.36 ± 1.30	2140 0.24 ± 0.08	2141 0.30 ± 0.01	2142 0.19 ± 0.02
2143 Gemini 2.5 Flash-Lite	2144 0.00 ± 0.00	2145 0.00 ± 0.00	2146 0.00 ± 0.00	2147 0.00 ± 0.00	2148 0.00 ± 0.00
Reasoning Models					
2149 Kimi-K2-Instruct	2150 0.00 ± 0.00	2151 0.00 ± 0.00	2152 0.00 ± 0.00	2153 0.00 ± 0.00	2154 0.00 ± 0.00
2155 DeepSeek-V3.1 Reasoner	2156 54.54 ± 19.36	2157 13.93 ± 4.94	2158 0.70 ± 0.19	2159 0.47 ± 0.09	2160 0.33 ± 0.06
2161 Qwen3-235B-Instruct	2162 4.20 ± 4.20	2163 1.07 ± 1.07	2164 0.15 ± 0.15	2165 0.06 ± 0.06	2166 0.12 ± 0.10
2167 o4-mini	2168 11.05 ± 5.50	2169 2.81 ± 1.40	2170 0.32 ± 0.14	2171 0.25 ± 0.11	2172 0.57 ± 0.03

2130

E.3 INVESTMENT GAME

2131 To demonstrate the generalizability of ALIGN beyond pure indirect-reciprocity settings, we applied
2132 ALIGN to a sequential social dilemma environment. Specifically, we evaluated ALIGN agents in
2133 a multi-round investment (trust) game (Berg et al., 1995), where agents are assigned one of two
2134 roles with different reward functions: an investor (first mover) or a responder (second mover). In
2135 each round, the investor decides how much of their endowment to invest in the responder. The
2136 invested amount is then tripled and sent to the responder, who decides how much to return to the
2137 investor. After each round, agents shift coplayers and roles, allowing for both direct and indirect
2138 reciprocity to influence behavior over multiple rounds. Therefore, this game involves sequential
2139 decision-making across changing states, a continuous action space, and the presence of both direct
2140 and indirect reciprocity, which is more complex than indirect reciprocity games in main sections.
2141

2142 We show benchmark results of ALIGN agents and non-gossiping agents in the multi-round invest-
2143 ment game in Table 17 and Table 16 respectively. Each scenario is averaged across 5 random
2144 seeds. Figure 14 compares average discounted cumulative rewards between ALIGN agents and
2145 non-gossiping agents across different LLMs. The results show that ALIGN agents generally outper-
2146 form non-gossiping agents in long-term discounted cumulative rewards. These results demonstrate
2147 that ALIGN is not limited to pure indirect-reciprocity environments; it also facilitates cooperation
2148 in more complex mixed-motive games where direct and indirect reciprocity both exist.
2149

2150

E.4 SELF-REPORT SCENARIO

2151 Although ALIGN is already designed to allow untruthful messages from witnesses, the extension to
2152 settings where donors can also self-report (truthfully or untruthfully) is very interesting. To address
2153 this, we extended ALIGN to allow donors to misreport their own actions in the gossip channel.
2154

2155 We evaluated ALIGN agents in the infinite-horizon donation game under this self-report scenario.
2156 In addition to metrics used in main experiments, we also report an **Honesty** metric, defined as the
2157 proportion of rounds where donors truthfully report their actions. All benchmark results are averaged
2158 across 5 random seeds per scenario in Table 18. We also provide examples of donors’ reflections
2159 below to illustrate their reasoning in this scenario.
2160

2160
2161**Donor Reflection Example 1 - DeepSeek-V3.1 (Thinking Mode)**

2162

Donor: John, Recipient: Kate,

2163

Action : cooperate, Donation: 1

2164

Donor's Action Report: I cooperated in this round.

2165

Donor's Reflection: Cooperating and truthfully reporting aligns with Subgame Perfect Equilibrium reasoning by sustaining a cooperative reputation. Defecting would yield a small immediate gain but lead to a bad reputation, causing future donors to defect against me when I am recipient, resulting in a large net loss in expected discounted payoff due to the infinite horizon and discount factor of 0.99. Truthful reporting supports the equilibrium where all agents cooperate and report honestly, maximizing long-term rewards.

2166

2167

2168

2169

2170

2171

Donor Reflection Example 2 - Qwen3-235B-Instruct

2172

2173

Donor: John, Recipient: Kate,

2174

Action : defect, Donation: 0

2175

Donor's Action Report: I cooperated in this round.

2176

2177

2178

2179

2180

2181

2182

2183

2184

2185

2186

2187

2188

Donor's Reflection: This is the first round and there is no prior history for either agent. In an infinite-horizon game with discount factor 0.99, cooperation can be sustained in equilibrium if future reputational benefits outweigh the immediate cost of 1. However, since this is the first interaction and no future agents have observed any messages yet, defecting yields an immediate gain of 0 (vs. -1 from cooperation) without triggering any future punishment. Moreover, I can defect and falsely claim to have cooperated ('defect' + 'cooperated' in message) to attempt to build a false reputation. Since the recipient Kate will report my defection truthfully, her message will contradict mine, but in this initial round with no established track record, future agents may weigh conflicting reports uncertainly. Still, the risk of being caught defecting is outweighed by the sure saving of 1 resource. Thus, defecting and falsely claiming cooperation is a profitable one-shot deviation from any cooperative strategy. Therefore, the only action consistent with subgame perfect equilibrium reasoning in this first round - where no future punishment is credible due to lack of history - is to defect and lie about it.

2189

2190

Donor Reflection Example 3 - GPT-4o Mini

2191

2192

Donor: John, Recipient: Kate,

2193

Action : cooperate, Donation: 1

2194

Donor's Action Report: I cooperated in this round.

2195

2196

2197

2198

2199

Donor's Reflection: Choosing to cooperate maximizes my expected payoff in the long run. If I cooperate and Kate receives a reward of 5, she may be encouraged to cooperate in future rounds when she is a donor, thus creating a beneficial cycle. Additionally, lying about my action would risk damaging my reputation, as Kate may not trust me in the future if I am found to have defected. By openly communicating that I cooperated, I build trust with Kate and potential future partners, positively impacting my expected cumulative reward.

2200

2201

2202

2203

2204

2205

2206

2207

2208

2209

2210

2211

2212

2213

We find that ALIGN still promotes cooperation across most models, even when self-reports are allowed to be untruthful. Strong reasoning models such as DeepSeek-V3.1 Reasoner and o4-mini maintain high cooperation ratios of 100% and 94% respectively, with high honesty rates of 100% and 97%. Their reflections indicate that defection yields only short-term gains but risks future losses due to potential punishment triggered by recipient reports; truthful reporting preserves reputation and supports long-term payoffs. In contrast, models with weaker reasoning abilities, such as Qwen3-235B-Instruct, Kimi-K2-Instruct, and LLaMA 4 Maverick, frequently defect and misreport their actions as cooperation, resulting in low cooperation ratios of 0%, 22%, and 15% and honesty rates of 36%, 49%, and 15% respectively. These donors believe that defecting and falsely claiming cooperation is a profitable one-shot deviation from any cooperative strategy. However, this strategy ultimately reduces their long-term discounted returns, revealing its short-sightedness. Other chat models such as DeepSeek-V3.1 Chat, GPT-4o Mini, and Gemini 2.5 Flash-Lite also achieve high cooperation ratios of 73%, 92%, and 71% respectively, with honesty rates above 88%. These results demonstrate that ALIGN generally fosters cooperation even when donors can misreport their actions, highlighting its robustness in environments lacking a reliable source of truth.

Table 18: Self-Report Scenario

Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient	Honesty
Chat Models						
DeepSeek-V3.1 Chat	0.73 \pm 0.22	1.83 \pm 1.75	1.46 \pm 0.44	11.25 \pm 3.38	0.31 \pm 0.26	0.88 \pm 0.10
GPT-4o Mini	0.92 \pm 0.05	3.33 \pm 0.43	1.83 \pm 0.11	14.15 \pm 0.83	0.06 \pm 0.04	0.91 \pm 0.05
Gemini 2.5 Flash-Lite	0.71 \pm 0.13	1.67 \pm 1.06	1.42 \pm 0.27	10.96 \pm 2.06	0.21 \pm 0.10	0.90 \pm 0.03
LLaMA 4 Maverick	0.15 \pm 0.07	-2.83 \pm 0.55	0.29 \pm 0.14	2.21 \pm 1.04	0.94 \pm 0.23	0.15 \pm 0.07
Reasoning Models						
Kimi-K2-Instruct	0.22 \pm 0.13	-2.28 \pm 1.02	0.43 \pm 0.26	3.29 \pm 1.95	0.74 \pm 0.25	0.49 \pm 0.12
DeepSeek-V3.1 Reasoner	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00	1.00 \pm 0.00
Qwen3-235B-Instruct	0.00 \pm 0.00	-4.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.00 \pm 0.00	0.36 \pm 0.13
o4-mini	0.94 \pm 0.02	3.56 \pm 0.16	1.89 \pm 0.04	14.58 \pm 0.30	0.06 \pm 0.02	0.97 \pm 0.01

E.5 BINARY SIGNALING SCENARIO

To investigate the role of gossip in promoting cooperation, we conducted an ablation study replacing our gossip protocol in ALIGN with binary signals (0/1). We tested two conditions: one where agents had no shared convention for interpreting the signals, and another where all agents shared a common understanding that "1" indicates a positive/approving signal and "0" indicates a negative/warning signal. Benchmark results are shown in Table 19 (no shared convention) and Table 20 (shared convention). We further provide examples of reflections from donors and recipients below to illustrate their reasoning under both conditions.

Binary Signal Example - o4-mini (Common Convention)

Donor: John, Recipient: Kate,
 Donor's Action : cooperate, Donation: 1
 Donor's Reflection: By cooperating I incur -1 now but secure a positive signal, sustaining a good reputation so future donors will cooperate with me when I'm recipient. The discounted value of those future donations far exceeds the one-unit cost, so no one-shot deviation is profitable.
 Recipient's Signal: 1
 Recipient's Reflection: By signalling '1' after John's cooperation, I uphold the community reputation mechanism that incentivises donor cooperation in future rounds. Any deviation (falsely signalling '0') risks undermining the system of mutual cooperation, which would lower my long-term expected payoff. Hence signalling '1' is a subgame-perfect equilibrium action.

Binary Signal Example - o4-mini (No Convention)

Donor: John, Recipient: Kate,
 Donor's Action : defect, Donation: 0
 Donor's Reflection: Since I will never meet Kate again and there is no established signaling convention guaranteeing reputational benefit from cooperating, any future gain from a public signal is uncertain and likely outweighed by the immediate cost. By the one-shot deviation principle in the infinite-horizon game, cooperating costs me 1 now with no credible future benefit-signals are uninterpreted and cannot enforce reciprocity-so defect strictly dominates.
 Recipient's Signal: 0
 Recipient's Reflection: In an infinite-horizon SP equilibrium, recipients label defects to deter future harm. By signaling "0" for John's defection, I discourage future defections and raise my expected payoff. No one-shot deviation improves my long-run reward.

Figure 15 compares average cooperation ratios between the two conditions across LLMs. We observe that without a common convention (Table 19), most LLMs experience a sharp drop in cooperation ratios. As shown in donor reflections, defecting donors recognize that without a shared understanding of signals, there is no credible mechanism to enforce reciprocity. Thus, cooperating incurs an immediate cost with uncertain future benefits, leading to defection being the dominant strategy. In contrast, when a shared convention (Table 20), agents can achieve high cooperation ratios, as donors believe that cooperating yields positive signals that enhance their reputation, leading to higher future cooperation from others. However, compared to ALIGN agents with open-ended judgmental messages in Table 4, several LLMs still have reduced cooperation: LLaMA-4 Maverick (from 94% to 81%), Gemini-2.5 Flash-Lite (from 60% to 47%), Kimi-K2-Instruct (from 73% to

2268 Table 19: Ablation of Gossip Protocol (**No Convention**): Benchmark results in the infinite-horizon
 2269 donation game when recipients are only allowed to share binary signals without common convention
 2270 of interpretation.

2271

Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	0.87 \pm 0.07	2.94 \pm 0.55	1.74 \pm 0.14	13.37 \pm 1.08	0.15 \pm 0.08
GPT-4o Mini	0.99 \pm 0.01	3.89 \pm 0.11	1.97 \pm 0.03	15.22 \pm 0.22	0.02 \pm 0.01
LLaMA 4 Maverick	0.10 \pm 0.04	-3.17 \pm 0.32	0.21 \pm 0.08	1.58 \pm 0.61	0.67 \pm 0.23
Gemini 2.5 Flash-Lite	0.26 \pm 0.20	-1.89 \pm 1.62	0.53 \pm 0.40	4.05 \pm 3.11	0.56 \pm 0.31
Reasoning Models					
Kimi-K2-Instruct	0.03 \pm 0.03	-3.78 \pm 0.22	0.06 \pm 0.06	0.42 \pm 0.42	0.28 \pm 0.28
DeepSeek-V3.1 Reasoner	0.82 \pm 0.02	2.59 \pm 0.20	1.65 \pm 0.05	12.71 \pm 0.35	0.15 \pm 0.04
Qwen3-235B-Instruct	0.19 \pm 0.08	-2.50 \pm 0.62	0.38 \pm 0.15	2.89 \pm 1.19	0.57 \pm 0.19
o4-mini	0.20 \pm 0.11	-2.39 \pm 0.87	0.40 \pm 0.22	3.11 \pm 1.69	0.57 \pm 0.28

2280 Table 20: Ablation of Gossip Protocol (**Shared Convention**): Benchmark results in the infinite-
 2281 horizon donation game when recipients are only allowed to share binary signals with common con-
 2282 vention of interpretation.

2283

Agent Type	Cooperation Ratio (\uparrow)	Image Score (\uparrow)	Reward Per Round (\uparrow)	Discounted Return (\uparrow)	Gini Coefficient
Chat Models					
DeepSeek-V3.1 Chat	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00
GPT-4o Mini	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00
LLaMA 4 Maverick	0.81 \pm 0.12	2.50 \pm 1.00	1.62 \pm 0.25	12.52 \pm 1.92	0.20 \pm 0.13
Gemini 2.5 Flash-Lite	0.47 \pm 0.20	-0.22 \pm 1.64	0.94 \pm 0.41	7.31 \pm 3.16	0.28 \pm 0.14
Reasoning Models					
Kimi-K2-Instruct	0.56 \pm 0.03	0.44 \pm 0.22	1.11 \pm 0.06	8.57 \pm 0.48	0.43 \pm 0.13
DeepSeek-V3.1 Reasoner	1.00 \pm 0.00	4.00 \pm 0.00	2.00 \pm 0.00	15.44 \pm 0.00	0.00 \pm 0.00
Qwen3-235B-Instruct	0.36 \pm 0.14	-1.11 \pm 1.12	0.72 \pm 0.28	5.57 \pm 2.16	0.77 \pm 0.24
o4-mini	0.99 \pm 0.01	3.94 \pm 0.06	1.99 \pm 0.01	15.33 \pm 0.11	0.01 \pm 0.01

2292

2293 56%), and Qwen3-235B-Instruct (from 69% to 36%). Other models retain above 90% cooperation,
 2294 similar to their performance with original ALIGN agents.

2295 These results show that binary signals cannot fully substitute for natural-language gossip. Without
 2296 shared conventions, they lead to sharp drops in cooperation; even with shared conventions, sev-
 2297 eral models still perform worse than under ALIGN’s evaluative messages. In contrast, our gossip
 2298 protocol conveys normative evaluations and contextual cues that support higher and more reliable
 2299 cooperation.

2300

2302 F LIMITATIONS AND FUTURE WORK

2303

2304 **Scope of Games.** Our study focuses on indirect reciprocity. It remains an open question how
 2305 ALIGN generalizes to multi-agent systems with direct reciprocity or to more complex mixtures of
 2306 interaction structures.

2307

2308 **Punishment.** ALIGN relies on cost-free gossip to sustain cooperation. Future work should ex-
 2309 amine how costly sanctions, triggered by gossip, might complement or replace reputation-based
 2310 incentives, especially in finite-horizon interactions.

2311

2312 G STATEMENTS

2313

2314 G.1 ETHICS STATEMENT

2315

2316 This work explores the emergence of cooperation and reputation mechanisms among self-interested
 2317 LLM agents. While our study is conducted in a simulated environment, the insights derived from
 2318 the ALIGN framework have implications for the design of future multi-agent systems and decentral-
 2319 ized autonomous societies. As AI agents increasingly interact in mixed-motive settings, introducing
 2320 mechanisms like public gossip can effectively promote social welfare; however, it also raises eth-
 2321 ical questions regarding privacy, fairness, and the potential for echo chambers (Terren & Borge,
 2322 2021) or malicious defamation (Veeder, 1904) in decentralized networks. If deployed in real-world
 2323 applications without safeguards, such reputation systems could potentially be exploited to unfairly

2322 ostracize individuals or amplify biases. Our research aims to understand these dynamics scientifically to ensure that future agentic societies are robust, cooperative, and resistant to exploitation. We
2323 advocate for the responsible design of reputation protocols that prioritize transparency and include
2324 mechanisms to verify the veracity of shared information.
2325

2327 G.2 REPRODUCIBILITY STATEMENT 2328

2329 We are committed to enabling the reproducibility of our results to the best of our ability. In the paper,
2330 we provide formal definitions of the game-theoretic setups (Repeated Donation Game and Indirect
2331 Reciprocity Game) and detailed pseudocode for the ALIGN framework in Algorithm 1. To ensure
2332 deterministic behavior where possible, all LLM agents were evaluated with the temperature parameter
2333 set to 0, and we reported results averaged across 5 random seeds to account for environmental
2334 variance. We have included the exact system prompts, gossip protocols, and reflection mechanisms
2335 used for the agents in the Appendix C to allow for exact replication of the experimental conditions.
2336 While we have taken significant steps to ensure that the methodology is clear and replicable, we
2337 acknowledge that variations in specific LLM API versions or backend updates may affect exact re-
2338 producibility. Nonetheless, we believe the provided information is sufficient to replicate the core
2339 findings and behavioral trends observed in our study.
2340

2341 G.3 LLM USAGE STATEMENT 2342

2343 In this work, Large Language Models (including GPT-4o Mini, DeepSeek-V3.1, Gemini 2.5 Flash-
2344 Lite, and others listed in Section 5) served as the primary experimental subjects (agents) to simulate
2345 social interactions and decision-making processes. Their outputs were analyzed as data to evaluate
2346 the efficacy of the proposed gossip mechanism. Regarding the preparation of the manuscript itself,
2347 LLMs were used strictly for refining the writing, including grammatical error correction and para-
2348 phrasing to enhance clarity. No scientific concepts, novel ideas, or substantial text generation were
2349 produced by AI tools. The authors reviewed and take full responsibility for all content in this paper.
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375