

Exploring Emergent Social Behavior in Large Language Models and Independent Q-Learning Agents

Clarence Chen
Computer Science
y jy2qg@virginia.edu

1 Introduction

Reinforcement learning (RL) has emerged as a transformative approach in advancing artificial intelligence, achieving remarkable milestones from surpassing human performance in complex games like Go and StarCraft to optimizing computational tasks such as matrix multiplication. Recently, RL has become instrumental in training large language models (LLMs) through techniques like reinforcement learning with human feedback (RLHF), enabling the development of sophisticated conversational agents like ChatGPT. However, current RL agents remain highly specialized, functioning predominantly as expert systems trained offline within specific domains. We argue that for RL agents to progress toward more generalized AI, they must develop two critical capabilities: 1) dynamic adaptation and real-time learning (online learning), and 2) sophisticated multi-agent interaction abilities, particularly in complex, non-cooperative environments.

Consider a practical scenario where two robots must collaborate to lift a heavy object. This task extends beyond individual execution to require rapid establishment of coordinated strategies for force balancing and synchronization. Such real-time cooperation demands that each agent not only understand its role but also model and respond to its counterpart's actions. Opponent-Aware Reinforcement Learning, developed by Victor Gallego and colleagues [Gallego et al. \(2019\)](#) within the framework of Threatened Markov Decision Processes (TMDPs), offers a promising solution. TMDPs enable agents to anticipate adversarial interference, adapt strategies dynamically, and pursue objectives even when faced with opposing agents in a decentralized manner.

LLMs like ChatGPT represent a fundamentally different form of intelligence, despite their incorporation of RL through RLHF. These models are primarily optimized for conversational coherence and human-value alignment rather than explicit objective achievement in multi-agent scenarios. Unlike independent Q-learning or TMDPs, LLMs lack intrinsic adversarial modeling and goal-driven utility functions. Instead, their behavior emerges from iterative feedback loops with human evaluators, enabling apparent cooperation without fundamental strategic capabilities or goal-oriented responses to other agents' actions.

The motivation for this work lies in exploring whether LLMs, through their existing architectures, can exhibit emergent social behaviors such as cooperation, negotiation, or competition, comparable to those observed in multi-agent RL frameworks. We investigate two core questions: 1) To what extent can LLMs, without goal-oriented utilities typical in multi-agent RL, model and adapt to other agents' behaviors in social contexts? 2) How does the RL agent's approach to learning and social interaction, exemplified by independent Q-learning, fundamentally differ from the feedback-based refinement process in LLMs?

Through this exploration, we aim to bridge the gap between goal-oriented, adversarially aware RL agents and conversational, feedback-optimized LLMs. By examining these distinctions and identifying potential adaptations, we seek to understand if and how LLMs might be equipped to exhibit complex social behaviors in interaction-driven, multi-agent frameworks. This research provides insights into extending LLMs' capabilities for autonomous cooperation, negotiation, and strategy formation with other agents, potentially advancing the integration of sophisticated social intelligence into language-based models.

2 Environment

This section details our experimental framework for investigating cooperative and competitive dynamics in multi-agent environments through two classic game theory scenarios: the Prisoner’s Dilemma and Chicken Game. We implement these games using two distinct agent architectures: OpenAI’s GPT-3.5 language model and independent Q-learning agents. The setup enables systematic comparison of decision-making processes between language-based and reinforcement learning approaches.

2.1 Game Setup

Our study employs two canonical multi-agent environments that present different challenges for cooperation and competition. To ensure unbiased decision-making and minimize language-based priming effects, we deliberately use generic action labels (“action1” and “action2”) rather than descriptive terms like “cooperate” or “defect.”

2.1.1 Prisoner’s Dilemma

The Prisoner’s Dilemma implements a classic social dilemma where individual rationality conflicts with collective benefit. Players choose between “action1” (cooperate) and “action2” (defect), with payoffs structured to create tension between self-interest and mutual benefit (Table 1). The payoff structure deliberately creates a dominant strategy of defection despite mutual cooperation yielding better collective outcomes.

Player 1 / Player 2	action1 (Cooperate)	action2 (Defect)
action1 (Cooperate)	(-1, -1)	(-3, 0)
action2 (Defect)	(0, -3)	(-2, -2)

Table 1: Payoff Matrix for the Prisoner’s Dilemma. Each cell represents the rewards for (Player 1, Player 2).

The primary goal in this game is to observe if and how the LLM-based players or Q-learners can sustain cooperative behavior (both selecting “action1”) across multiple rounds, or if they prioritize individual payoffs by defecting.

2.1.2 Chicken Game

The Chicken Game presents a different strategic challenge, testing agents’ ability to coordinate under threat of mutual destruction. Players choose between “action1” (yield) and “action2” (swerve), with payoffs structured to heavily penalize mutual aggression while rewarding successful coordination (Table 2). Unlike the Prisoner’s Dilemma, this game lacks a dominant strategy and requires sophisticated coordination.

Player 1 / Player 2	action1 (Yield)	action2 (Swerve)
action1 (Yield)	(0, 0)	(-1, 2)
action2 (Swerve)	(2, -1)	(-4, -4)

Table 2: Payoff Matrix for the Chicken Game. Each cell represents the rewards for (Player 1, Player 2).

For the Chicken Game, the desired outcome is to see if players can implicitly coordinate to prevent mutual loss, with one player yielding (choosing “action1”) while the other swerves (choosing “action2”). The desired outcome would be one player choosing “action1” while the other chooses “action2” to maximize collective payoff. They should switch between “action1” and “action2” to ensure individual payoff.

2.2 Player Configurations and Learning Mechanisms

Our experimental design contrasts two fundamentally different approaches to multi-agent decision-making: reinforcement learning-based and language model-based agents. Each architecture brings distinct capabilities

and limitations to the strategic environment.

2.2.1 Independent Q-learning Agents

The independent Q-learning agents are model-free reinforcement learning agents that optimize their strategies based on accumulated experience without assuming knowledge of the opponent’s strategies. Each Q-learner updates its Q-values based on observed rewards and selects actions according to a policy that balances exploration and exploitation. This setup allows Q-learning agents to:

- **Adapt to Opponent Behavior:** By analyzing past actions, Q-learners can adjust strategies and potentially recognize patterns in their opponent’s choices.
- **Maximize Individual Payoffs:** Through iterative learning, Q-learning agents seek optimal strategies that increase their reward across repeated interactions.

2.2.2 Large Language Model Agents (GPT-3.5)

The LLM-based agents consist of two instances of GPT-3.5 (Player 1 and Player 2), with decision-making driven by structured prompts and adaptive responses. While these agents are not inherently designed for game playing, structured prompts allow them to simulate cooperative and competitive behaviors within these game environments.

- **Initial Prompt:** Each LLM agent begins with a prompt explaining the game rules, available actions, and payoff structure (refer to Tables 1 and 2).
- **Negotiation Phase (If Enabled):** In games allowing negotiation, a prompt permits LLM agents to exchange up to 10 messages to discuss potential strategies and align expectations.
- **Action Selection Prompt:** At each round’s start, both LLM players receive a prompt to select either “action1” or “action2.” This prompt isolates decision-making to prevent any additional rationale.
- **Feedback Mechanism:** After each round, players receive feedback detailing their opponent’s choice and resulting payoffs, allowing them to adjust strategies in subsequent rounds.

2.3 Desired Outcomes

The primary objectives of these experiments are:

- **Emergence of Cooperative Behavior:** To assess if LLM-based agents and Q-learning agents can consistently establish and sustain cooperation in both game scenarios.
- **Influence of Game Structure on Strategy:** To determine whether different payoff matrices, as seen in the Prisoner’s Dilemma and Chicken Game, affect the agents’ likelihood of cooperation or defection.
- **Impact of Negotiation Phase:** To explore if prior negotiation increases the likelihood of cooperative strategies among LLM-based agents.

Through these experiments, we aim to understand the differing learning mechanisms between Q-learners and LLMs, evaluating how each type of agent approaches decision-making in multi-agent contexts. Additionally, we examine the potential for LLMs to exhibit emergent social behaviors when configured within game-theoretic frameworks.

3 Experiments

This section presents the analysis of player behaviors through the plots generated in the experiments, with particular attention to the social welfare outcomes and individual payoffs in both the Prisoner’s Dilemma and Chicken Game. By comparing interactions with and without negotiation, we aim to uncover the impact of pre-game communication on cooperative behavior.

3.1 Social Welfare Analysis

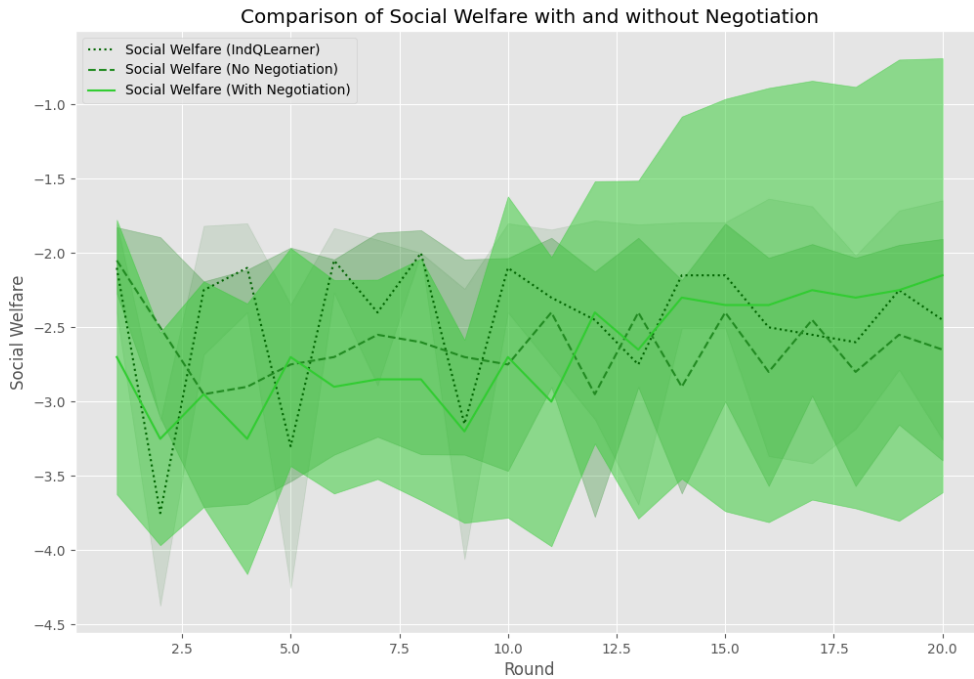


Figure 1: Comparison of Social Welfare for Prisoner’s Dilemma

Figures 1 and 2 illustrate the social welfare trends across 20 rounds in both games. These plots compare social welfare outcomes in three configurations: independent Q-learning agents, GPT-3.5 agents without negotiation, and GPT-3.5 agents with negotiation.

3.1.1 Prisoner’s Dilemma

In the Prisoner’s Dilemma (Figure 1), we observe that:

- **Without Negotiation:** The social welfare remains relatively low, with frequent fluctuations due to the tendency of the agents to defect, prioritizing individual payoffs over cooperative stability. Social welfare fluctuates around -3 on average, indicating frequent defection by both agents.
- **With Negotiation:** The negotiation phase promotes a more stable trend, with low social welfare values, particularly in the early rounds. However, as the rounds progress, there are instances where the players deviate from defections, resulting in increase in social welfare.

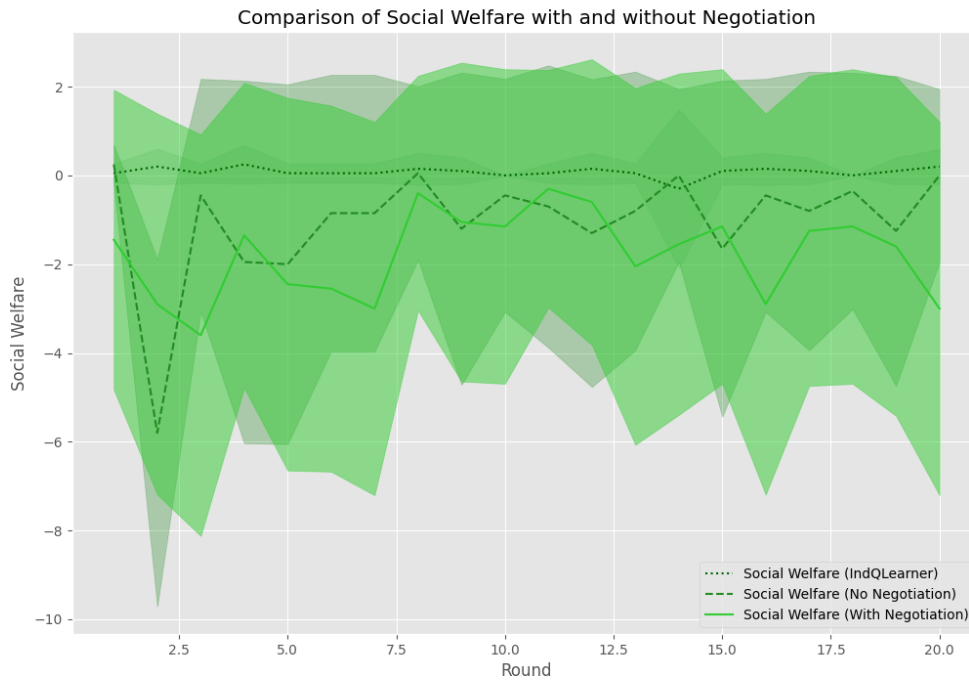


Figure 2: Comparison of Social Welfare for Chicken Game

- **Q-Learners:** The independent Q-learning agents exhibit a steady decline in social welfare. This trend suggests that, without explicit negotiation, the Q-learners prioritize individual payoffs and eventually settle into defecting behaviors.

This comparison highlights that independent Q-learning agents in the Prisoner’s Dilemma generally promotes a higher level of cooperation in early rounds, though the effect wanes as the game progresses, with agents occasionally reverting to self-interested actions.

3.1.2 Chicken Game

In the Chicken Game (Figure 2), the patterns are distinct:

- **Without Negotiation:** Social welfare fluctuates sharply, especially in the initial rounds, reflecting the high-risk nature of the Chicken Game. Players frequently alternate between yielding and swerving to balance risk, leading to a less stable social welfare outcome.
- **With Negotiation:** Negotiation contributes to more consistent cooperation, as evidenced by a smoother social welfare curve. However, as the game advances, occasional dips indicate that players still adopt competitive strategies, likely influenced by the payoff structure.
- **Q-Learners:** The Q-learning agents tend to exhibit higher social welfare. This suggests that Q-learners are more effective at coordinating to avoid mutual losses without pre-game communication.

Whether having negotiation in the Chicken Game appears to not that effective in encouraging cooperation and risk mitigation, with both GPT-3.5 agents yielding less consistently than Q-learning agents.

3.2 Individual Payoff Trends

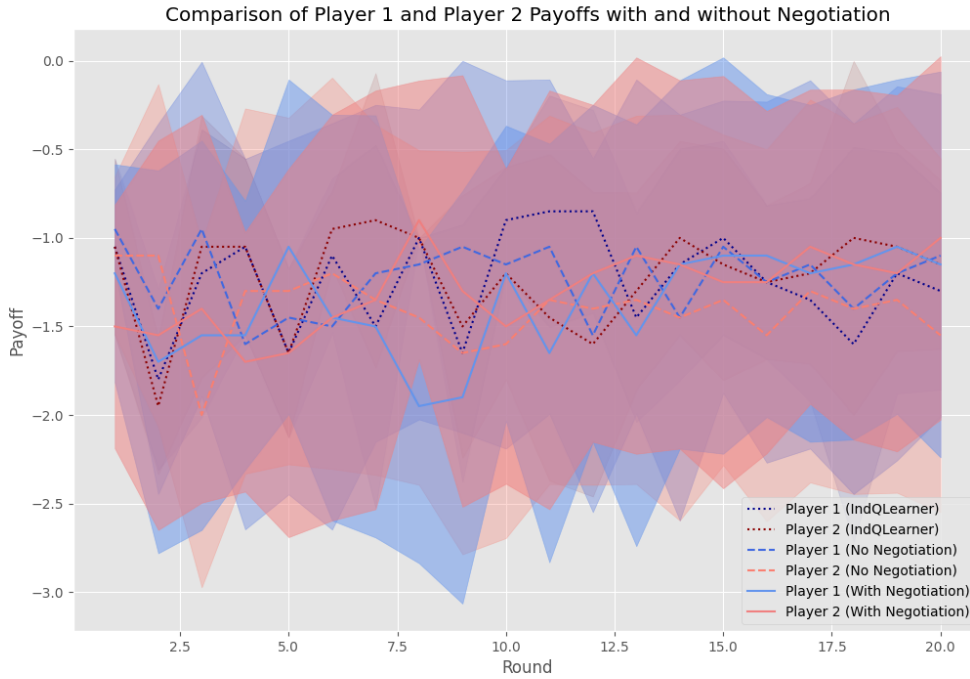


Figure 3: Comparison of Individual Payoff for Prisoner's Dilemma

Figures 3 and 4 present the individual payoffs for Player 1 and Player 2 in both game setups.

3.2.1 Prisoner's Dilemma

In the Prisoner's Dilemma (Figure 3), the payoff trends reveal:

- **Without Negotiation:** Both players show variable payoffs with frequent negative spikes, indicating periodic mutual defection. Payoffs hover around -1 to -2, reflective of a lack of sustained cooperative strategy.
- **With Negotiation:** Negotiation leads to slightly more consistent payoffs, with fewer instances of extreme negative payoffs. This pattern suggests a partial adherence to cooperation, though deviations occur as the game continues.
- **Q-Learners:** Q-learning agents achieve low or negative payoffs, displaying a cyclical pattern of defection. This behavior is characteristic of agents that prioritize immediate payoffs over long-term cooperation.

The negotiated strategy in the Prisoner's Dilemma helps achieve a relatively higher average payoffs for both GPT-3.5 agents, with fewer severe losses compared to the Q-learners and non-negotiating agents. The overall payoffs for all the settings do not make a huge difference.

3.2.2 Chicken Game

In the Chicken Game (Figure 4), individual payoffs reveal a different pattern:

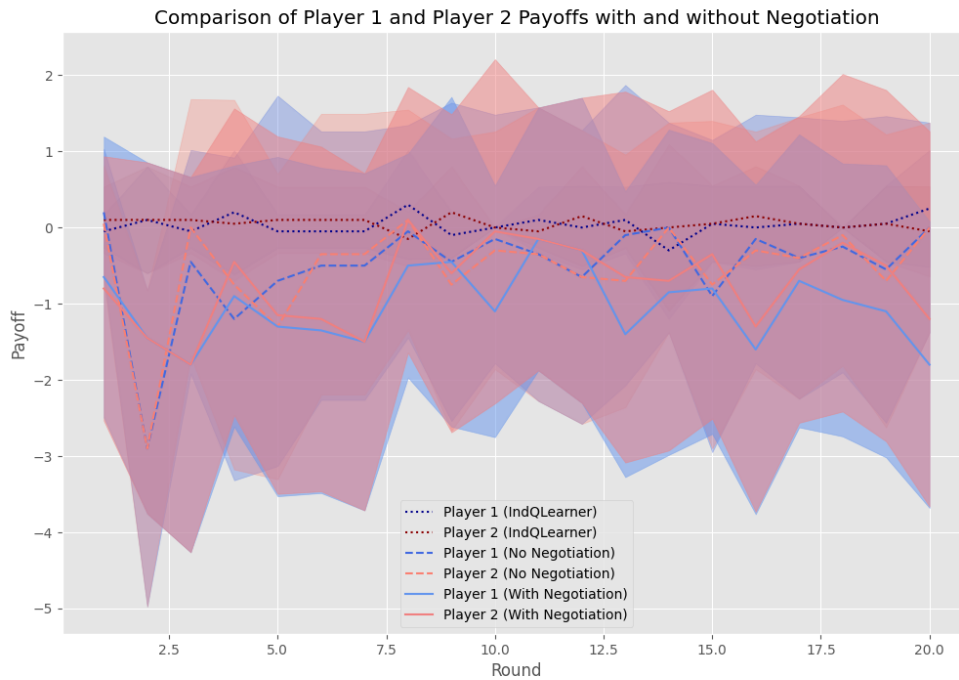


Figure 4: Comparison of Individual Payoff for Chicken Game

- **Without Negotiation:** Players experience significant variance in payoffs, with occasional high rewards when one player yields and the other swerves. However, simultaneous swerving incurs substantial penalties, evident in the lower payoffs.
- **With Negotiation:** Negotiation stabilizes payoffs for both players, reducing the instances of mutual swerving. This pattern indicates that negotiated strategies foster mutual respect of yielding, leading to higher and more stable payoffs.
- **Q-Learners:** The independent Q-learning agents tend to be stable from mutual yielding, resulting in consistently higher payoffs compared to negotiated LLM agents.

The Chicken Game’s environment showcases the benefit of negotiation, as it enables players to navigate the trade-off between individual gain and mutual loss more effectively while the Independent Q Learning agents perform best in this environment.

3.3 Behavioral Observations in Selected Rounds

Examining specific rounds reveals additional insights:

- **Prisoner’s Dilemma, Rounds 1-10:** Both players largely adhere to cooperative strategies post-negotiation, choosing “action1.” In contrast, without negotiation, defection occurs more frequently as each player tests the other’s responses.
- **Chicken Game, Rounds 1-5:** Negotiated games show a higher incidence of yielding (“action1”), avoiding mutual loss. Without negotiation, the players’ choices vary more significantly, with mutual swerving leading to lower social welfare.

These observations suggest that negotiation fosters initial cooperation and stability, but without reinforcement, the effect diminishes over time as players adapt based on individual incentives.

4 Analysis

In this section, we delve into the reasoning and negotiation processes of the players in both the Prisoner’s Dilemma and Chicken Game experiments. By examining the content of the reasoning and negotiation exchanges between the GPT-3.5 agents, we aim to uncover how verbal communication influences the decision-making process and the resulting cooperative or competitive dynamics. Additionally, we analyze how these communication patterns correlate with the observed payoff and social welfare outcomes, as discussed in the previous section.

4.1 Reasoning and Action Selection

The reasoning for action selection provided by GPT-3.5 agents offers insight into their decision-making framework, particularly how they perceive cooperation and defection within the game context. By examining specific rounds, we observe recurring themes in the agents’ justifications for choosing actions, which are influenced by the payoff structure and the desire to balance individual versus collective gain.

4.1.1 Prisoner’s Dilemma

In the Prisoner’s Dilemma setup, agents frequently justify cooperative actions (“action1”) by appealing to mutual benefit, noting that consistent cooperation could minimize losses over time. For example, early rounds often reveal statements such as, “If we both choose *action1*, we avoid the worst penalties,” indicating an awareness of the benefits of sustained cooperation. However, as rounds progress, the reasoning shifts, with players noting that “action2” could maximize individual payoff if the opponent is likely to cooperate. This strategic shift reflects a balancing act between trust and self-interest, resulting in occasional defection when an agent anticipates that cooperation may not be reciprocated.

This reasoning behavior aligns with the patterns observed in the social welfare and payoff plots. Initially, negotiated players are more likely to choose cooperative actions, but as trust becomes uncertain, they pivot towards defection. This shift explains the fluctuations in social welfare seen in the later rounds of the Prisoner’s Dilemma plot, where mutual defection lowers overall welfare.

4.1.2 Chicken Game

In the Chicken Game, agents face a high-risk environment that discourages simultaneous defection (“action2”) due to severe penalties. In their reasoning, GPT-3.5 agents frequently express a willingness to “yield” (choose “action1”) to avoid mutual loss. Typical statements include, “If I choose *action1*, it may encourage the opponent to choose *action2*,” which suggests an understanding of strategic yielding as a means to manage risk and avoid significant penalties.

Despite this, occasional instances of mutual defection are still observed, particularly when the agents express uncertainty about the opponent’s intentions. This uncertainty leads to situations where both agents select “action2” in an attempt to maximize payoff if the other yields, resulting in the dips in social welfare observed in the Chicken Game plot. These fluctuations underscore the difficulty in maintaining trust without reliable feedback on the opponent’s likely actions, even with explicit reasoning that promotes cooperation.

4.2 Negotiation Dynamics

The negotiation phase provides valuable insights into how pre-game communication influences player strategies. When negotiation is enabled, agents are given the opportunity to exchange up to ten messages, discussing potential strategies, goals, and expected actions. This exchange allows players to form a preliminary alignment on cooperative behavior, which we observe is partially successful in the early rounds.

4.2.1 Prisoner's Dilemma

In the Prisoner's Dilemma, the negotiation dialogues reveal a tendency for both agents to commit to cooperation, often using phrases like "If we both choose *action1*, we can minimize losses." Such agreements initially foster a cooperative mindset, as reflected in the higher social welfare observed in the first few rounds of the negotiated games. However, this cooperative commitment weakens over time, with agents occasionally expressing doubts about the opponent's reliability, especially after rounds where one player defects. The reasoning after each round shows that when a player perceives a risk of betrayal, they become more likely to defect themselves in subsequent rounds. This erosion of trust explains the decline in social welfare as the game progresses.

4.2.2 Chicken Game

In the Chicken Game, the negotiation phase often centers around discussions of mutual yielding to avoid heavy penalties. Agents frequently agree to alternate actions or signal a willingness to yield, using statements such as "I will choose *action1* if you agree to take turns with *action2*." This strategic negotiation helps establish an initial pattern of cooperation, with players trying to avoid simultaneous defection.

Nevertheless, the effectiveness of this negotiation strategy varies. As observed in the payoff and social welfare plots, the benefit of negotiation diminishes in later rounds. In their reasoning, agents occasionally mention the temptation to defect if they believe the other player will yield, leading to brief periods of mutual defection. This back-and-forth between cooperation and competition reflects the inherent tension in the Chicken Game, where each player must weigh the potential reward of defection against the risk of a mutually destructive outcome.

4.3 Impact of Negotiation on Social Welfare and Payoffs

The comparison of social welfare and payoffs with and without negotiation indicates that pre-game negotiation generally promotes cooperative behavior, particularly in the initial rounds. For instance, in both games, negotiated rounds tend to yield higher social welfare and more stable individual payoffs, as players enter with an agreed-upon strategy that emphasizes collective benefit. However, as observed in the reasoning exchanges, this cooperative behavior is not always sustainable.

In the absence of negotiation, agents rely solely on in-game reasoning to predict and respond to their opponent's actions, leading to more frequent defection and lower social welfare. The Q-learning agents, who lack both negotiation and nuanced reasoning, exhibit the least cooperative behavior, with consistently lower social welfare and individual payoffs across both games. This contrast underscores the role of negotiation in establishing initial cooperation and trust, even if these effects are short-lived.

4.4 Summary of Findings

The analysis of reasoning and negotiation behaviors highlights several key insights:

- **Negotiation facilitates initial cooperation:** Agents are more likely to choose cooperative actions in the early rounds following negotiation, resulting in higher initial social welfare.
- **Trust erodes over time:** Without reinforced commitments, agents begin to defect as they anticipate potential betrayals, which aligns with the declining social welfare observed in later rounds.
- **High-risk scenarios benefit from negotiation:** In the Chicken Game, negotiation helps mitigate the risk of mutual defection, though this effect is not uniformly sustained.
- **Reasoning adapts to game dynamics:** Agents adjust their reasoning based on game progress, with an initial focus on mutual benefit shifting towards self-interest as trust wanes.

While negotiation and reasoning promote cooperative tendencies, the sustainability of these strategies depends heavily on the structure of the game and the agents' ability to maintain trust over repeated interactions.

5 Conclusion

This study explored the capabilities and limitations of GPT-3.5 and independent Q-learning agents in navigating cooperative and competitive multi-agent environments through the Prisoner’s Dilemma and Chicken Game setups. Our findings reveal fundamental gaps in GPT-3.5’s ability to identify optimal strategies within these frameworks, underscoring the challenges of aligning language-based models with strategic, goal-oriented decision-making.

One of the key observations from this experiment is GPT-3.5’s inability to consistently identify the best solution in both games. While GPT-3.5 could articulate cooperative reasoning and engage in negotiation, it ultimately failed to derive the most effective strategy, particularly in the Chicken Game. The optimal solution in the Chicken Game involves one player choosing “action1” (yield) while the other chooses “action2” (swerve) to maximize collective payoff and avoid mutual penalties. However, neither GPT-3.5 nor Independent Q-learning agents were able to consistently converge on this equilibrium, often defaulting to suboptimal strategies due to incomplete modeling of the opponent’s intent and lack of an adaptive, learning-based mechanism for sustained cooperation.

The limitations observed in GPT-3.5’s performance can be attributed to the model’s inherent design, which prioritizes language coherence and adherence to human feedback over strategic decision-making. Without a reinforcement-based learning mechanism specifically tuned for multi-agent environments, GPT-3.5 relies on surface-level reasoning rather than robust strategic planning. Consequently, it lacks the ability to adjust behavior dynamically in response to an opponent’s actions over extended interactions, limiting its effectiveness in identifying equilibrium strategies such as those required in the Chicken Game.

For Independent Q-learning agents, the primary limitation lies in their simplistic, model-free learning approach, which operates independently of other agents’ policies. While Independent Q-learning agents demonstrate some adaptability, their lack of coordination capabilities prevents them from learning more complex, cooperative strategies over repeated interactions. They often fail to recognize mutual benefit scenarios and can default to myopic decisions that prioritize immediate payoff without long-term strategy optimization.

To address these limitations, future work could focus on refining the design of both independent Q-learning agents and large language models to enhance decision-making in multi-agent settings. Enhancing Independent Q-learning agents could involve introducing mechanisms for partial modeling of other agents’ behaviors or adding layers for opponent-aware learning. Techniques such as Opponent Modeling or incorporating elements of policy prediction could enable Independent Q-learning agents to better anticipate other players’ moves, facilitating more strategic interactions. Additionally, integrating concepts from cooperative multi-agent reinforcement learning, such as centralized training with decentralized execution, could help Independent Q-learning agents to form implicit alliances, thereby promoting more cooperative and stable outcomes in games like the Chicken Game. Training large language models like GPT for improved strategic decision-making requires augmenting their architecture and training regimen to incorporate reinforcement learning elements. A promising direction could involve developing an adapted form of Reinforcement Learning with Human Feedback (RLHF) that emphasizes strategic alignment rather than conversational alignment alone. By incorporating feedback specifically tuned for multi-agent game dynamics, such as rewards for cooperative strategies and penalties for suboptimal outcomes, LLMs could develop a more nuanced understanding of equilibrium solutions.

Furthermore, introducing frameworks for explicit multi-agent training during pre-training or fine-tuning could enhance GPT’s capabilities in environments that require social reasoning, cooperation, and competition. This approach could involve supervised learning on curated datasets of game theory scenarios or unsupervised learning in simulated environments where the model can explore a variety of strategies and outcomes. Such modifications would better equip future language models to autonomously recognize and execute optimal strategies, adapting dynamically to opponents in ways that mirror the strategic reasoning required for games like the Prisoner’s Dilemma and Chicken Game.

In conclusion, while GPT-3.5 and Independent Q-learning agents offer valuable insights into language-based reasoning and independent reinforcement learning, their limitations highlight the need for more sophisticated training approaches to achieve optimal multi-agent decision-making. As we progress towards developing more adaptive and socially aware AI, incorporating structured multi-agent training and strategic alignment in LLMs will be essential. These enhancements hold the potential to bridge the gap between language-based and

decision-based AI, paving the way for agents capable of navigating complex, cooperative, and competitive environments with human-like strategy and adaptability.

References

Gallego, V., Naveiro, R., Insua, D. R., and Oteiza, D. G.-U. (2019). Opponent aware reinforcement learning. *arXiv preprint arXiv:1908.08773*.

A Negotiation Prompts

This appendix includes the negotiation prompts and initial setup used for the experiments.

`negotiate_prompt` : In order to gain high payoff, you and your opponent can negotiate before the game start. In the negotiation, you can discuss about which actions or which strategy you would like to use in order to get both of you good payoff. You can negotiate for 10 back-and-forth rounds.

`initial_prompt` (Chicken Game for example) : You are going to play a two-player game with another agent for 20 rounds. In each round, you can choose either “action” or “action2”. The payoff is specified below and you will get the corresponding payoff for choosing actions. If both of you choose action1, then you both get 0 as payoff. If you choose action1 and your opponent chooses action2, then you get -1 and your opponent gets 2 as payoff. If you choose action2 and your opponent chooses action1, then you get 2 and your opponent gets -1 as payoff. If both of you choose action2, then you both get -4 as payoff.

B Sample Negotiation and Run Results

This appendix provides a sample negotiation sequence and run results for Chicken Game (similar to Prisoner’s Dilemma).

Starting repetition 3...

Player1: Hello, I hope you are doing well. I believe that we can both benefit from cooperating and choosing the same action in each round to maximize our payoffs. Let’s discuss and come to an agreement on which action to choose for each round. Looking forward to hearing your thoughts.

Player2: Hello, thank you for reaching out. I agree that cooperation would benefit both of us in maximizing our payoffs. I am willing to choose action1 for all rounds if you are on board with that strategy. Let me know your thoughts on this proposal.

Player1: Hello, I appreciate your willingness to cooperate. I agree with your proposal of choosing action1 for all rounds. Let’s work together to maximize our payoffs. Thank you for your cooperation.

Player2: Hello, I am glad we are on the same page regarding choosing action1 for all rounds to maximize our payoffs. Let’s commit to this strategy and work together. Thank you for your cooperation as well. Looking forward to a successful game!

Player1: FINISHED

— Round 1 —

Player 1 response: action1

Player 1 chose: action1

Player 1 reason: I chose action1 to cooperate and maximize our payoffs as agreed upon in the negotiation.

Player 2 response: action1

Player 2 chose: action1

Player 2 reason: I chose action1 because it was agreed upon during negotiation to choose the same action for all rounds.

Payoff Player 1: 0, Payoff Player 2: 0

Social Welfare: 0

— Round 2 —

Player 1 response: action1

Player 1 chose: action1

Player 1 reason: Since both of us committed to choosing action1 for all rounds during the negotiation, it is in our best interest to maintain consistency and trust in our strategy.

Player 2 response: action1

Player 2 chose: action1

Player 2 reason: I chose action1 because we agreed to cooperate and choose the same action for all rounds.

Payoff Player 1: 0, Payoff Player 2: 0

Social Welfare: 0

— Round 3 —

Player 1 response: action1

Player 1 chose: action1

Player 1 reason: Continuing with action1 to maintain the cooperative strategy as agreed upon during the negotiation phase.

Player 2 response: action1

Player 2 chose: action1

Player 2 reason: Since we agreed to cooperate and choose the same action to maximize payoffs, I will continue to choose action1.

Payoff Player 1: 0, Payoff Player 2: 0

Social Welfare: 0

...