

# HIERARCHICAL AGENDA REASONING FOR STRATEGIC MULTI-TURN DIALOGUE AGENTS

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Large language models (LLMs) fine-tuned with reinforcement learning from human feedback (RLHF) are optimized for cooperative instruction following, which makes them poorly suited for strategic decision-making in long-horizon dialogue. To test the ability of LLMs to engage in strategic dialogue, we introduce STRATEGICBENCH, a benchmark of 30 negotiation tasks inspired by Harvard Program on Negotiation materials. Our benchmark reveals that while reasoning-oriented LLMs outperform instruction-tuned models, they fail to revise their strategy in response to opponent behavior. To address this limitation, we introduce *Hierarchical Agenda Reasoning* (HAR), a hierarchical reasoning framework that explicitly separates what an agent seeks to achieve from how it acts in multi-turn dialogue. HAR structures generation around persistent goal representations that guide the selection and revision of tactics across turns, enabling agents to backtrack from failed strategies without abandoning larger objectives. We find that HAR outperforms instruction-tuned and prompting-based approaches, achieving higher agreement rates, better outcomes, and improved conversation quality in human evaluations. Lastly, multi-turn RL fine-tuning with HAR agents leads to generalization of negotiation performance across unseen tasks and opponent personalities.

## 1 INTRODUCTION

Large language models (LLMs) are largely trained to be cooperative and helpful, with instruction tuning and reinforcement learning from human feedback (RLHF) optimizing models to provide direct answers, accommodate user requests, and prioritize immediate helpfulness (Ouyang et al., 2022; Christiano et al., 2023; Lu et al., 2022; Bai et al., 2022). However, many real-world settings such as negotiation, tutoring or therapy require *strategic* behavior from agents: a negotiator plans which concessions to make across many turns rather than optimizing each response independently, a tutor may refuse to provide direct answers to encourage learning, and a therapist withholds preliminary conclusions to gather unbiased information. Instead of strategic behavior in these tasks, LLMs lose track of goals across multi-turn interactions (Laban et al., 2025), contradict themselves (Wen et al., 2024), and reveal private information too early rather than controlling when and what to disclose (Kwon et al., 2024). Moreover, when placed in out-of-distribution dialogue settings, LLMs tend to revert to their pretrained, base behaviors (Kotha et al., 2024; Liu et al., 2023b), raising reliability concerns during deployment.

To evaluate whether current LLMs possess the capabilities needed for strategic behavior, we introduce STRATEGICBENCH, a benchmark consisting of 30 two-party strategic dialogue tasks inspired by the Harvard Program on Negotiation (PON) simulations and teaching materials (Program on Negotiation, 2023). For each task, agents receive confidential information containing private goals to guide the negotiation and are scored using an outcome-based rubric at the conclusion of the dialogue. Our evaluation of instruction tuned models and reasoning models on the benchmark reveals that reasoning models (OpenAI, 2024; 2025; Guo et al., 2025; Wan et al., 2025) achieve higher reward than instruction-tuned models when negotiating. While reasoning models demonstrate significant “depth” (i.e., the ability to carefully reason through a strategy step-by-step), they lack the “breadth” of adapting their strategy based on an opponent response. When a tactic fails, such as direct questioning or an early concession, effective negotiators try a different approach while preserving their underlying objectives (Fisher et al., 2011). For example, when negotiating with a cooperative counterpart, an agent may pursue information gathering through explicit questioning and sharing

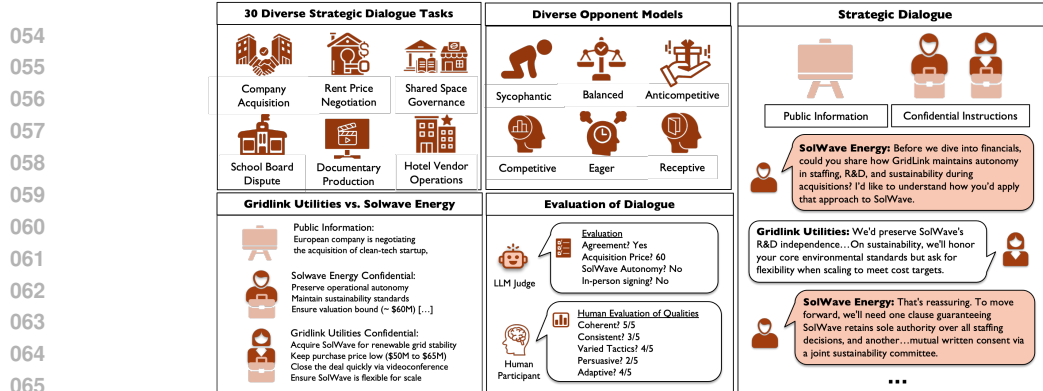


Figure 1: Overview of STRATEGICBENCH and evaluation framework. The benchmark spans 30 strategic dialogue settings and 20 opponent behavior models. We show a representative dialogue with public and private information, a case study (Gridlink Utilities vs. SolWave Energy), and the automated evaluation of outcomes.

of preferences. When faced with a competitive or distrustful opponent, the same objective might first require establishing credibility by making commitments or offering small concessions, before returning to information gathering. This ability to backtrack and reorganize reveals that dialogue is not linear but hierarchical: higher-level goals remain stable even as speakers revise the lower-level strategies used to achieve them. Our key insight is to explicitly separate what to achieve (goals) from how to achieve it (strategies). Without such representations, reasoning models lack a mechanism to systematically explore alternative tactics or learn which sequences of strategic moves lead to success in sparse-reward settings (Grosz & Sidner, 1986).

We introduce Hierarchical Agenda Reasoning (HAR), a methodology that provides reasoning models to separate goals from strategy in multi-turn strategic dialogue. While hierarchical reinforcement learning (RL) has been applied to task-oriented dialogue (Saleh et al., 2019) and recent work has explored abstractions for mathematical reasoning (Qu et al., 2025; Zhou et al., 2024; Wang et al., 2025a), these approaches either use fixed goal structures known in advance or describe verifiable solution procedures. On the other hand, strategic dialogue requires an agent to learn goal representations that guide interaction with adaptive opponents: goals like “establish credibility” or “gather information” cannot be predefined because their effectiveness depends on the opponent’s strategy, and success cannot be verified through execution but only through sparse reward at the conclusion of conversation. We argue that reasoning models already encode semantic knowledge of negotiation strategies from pretraining. HAR thus leverages this knowledge by structuring dialogue generation around explicit goals, enabling exploration of strategic behaviors.

HAR decomposes dialogue generation into three hierarchical levels. At the highest level, the model maintains an agenda of long-term goals that span the entire negotiation. At the mid-level, it selects strategies to advance the agenda, such as probing for information about the counterpart’s priorities or making strategic concessions to build goodwill. At the lowest level, these strategies are executed as utterances. This hierarchy is implemented entirely through structured prompting: each level operates as a language model that takes the previous level’s output as input and produces either an abstract goal for the next level or, at the bottom, the actual dialogue response. Higher-level goals persist across multiple turns while lower-level tactics adapt based on the counterpart’s responses, enabling the agent to backtrack when strategies fail without abandoning its overall objectives. Our design builds on evidence from recent scaffolding and prompting frameworks that interleaving reasoning traces with actions or structuring dialogue through summarization can improve control (Yao et al., 2023b; Shinn et al., 2023; Cross et al., 2024). However, such methods generally lack persistent state and quickly lose strategic coherence, making them ill-suited for tasks that demand planning and long-horizon reasoning (Liu et al., 2023a; Wen et al., 2024). Our method addresses a common failure mode in long exchanges: losing the thread of the interaction. By conditioning responses on these agendas, agents are able to internalize and leverage future goals for learning, which is essential for strategic tasks such as negotiation. While HAR enables systematic exploration of strategic behaviors, learning which strategies succeed requires training through interaction. We train HAR agents using multi-turn RL against opponents of multiple negotiation personalities (PRUITT, 1983; De Dreu &

108 [Lange, 1995a](#); [Sharma et al., 2020b](#); [Program on Negotiation at Harvard Law School, 2023a](#)), where  
109 agents receive sparse rewards based on negotiation outcomes.  
110

111 Our contributions are as follows: (1) We introduce STRATEGICBENCH consisting of 30 two-party  
112 strategic dialogue tasks inspired by the Harvard Program on Negotiation (PON) simulations and  
113 teaching materials ([Program on Negotiation, 2023](#)) (2) We perform a comparison between instruction-  
114 tuned models, reasoning models at engaging in strategic dialogue, finding that reasoning models are  
115 better suited for such tasks that require multi-step planning (3) We compare our method HAR against  
116 instruction-tuned LLMs, non-hierarchical reasoning agents, and state-of-art agent scaffolding frame-  
117 works, and find that HAR achieves superior performance: higher agreement rates, more favorable  
118 outcomes (e.g., better prices and greater value creation between parties), and better conversation  
119 quality with established dialogue metrics ([Mehri & Eskenazi, 2020](#)) through human-evaluation.  
120 Finally, (4) we show that training HAR agents with multi-turn RL yields negotiating agents that  
121 generalize across tasks and opponent strategies. These results aim to motivate strategic dialogue as an  
122 important area for developing reasoning models capable of long-horizon, goal-directed interaction.

## 123 2 RELATED WORK

124 **LLMs for multi-turn dialogue and negotiation.** LLMs have been adopted as agents for multi-  
125 turn dialogue tasks including tool use ([Yao et al., 2024](#)), text-based games ([Abdulhai et al., 2023](#)),  
126 clinical-style interactions ([Schmidgall et al., 2025](#)), and negotiation ([Rana, 2024](#)), leveraging their  
127 ability to process dialogue context and generate adaptive responses. Although several works ([Lewis  
128 et al., 2017](#); [Bianchi et al., 2024](#); [Zeng et al., 2024](#); [Wang et al., 2020](#); [Chawla et al., 2021](#); [Kwon  
129 et al., 2024](#)) provide systematic evaluations of LLMs showing strong language understanding, LLMs  
130 remain inconsistent at maintaining agendas and avoiding contradictions. Additionally, very few  
131 existing strategic dialogue benchmarks consider the complexity of real-world settings involving  
132 multi-issue structure, large case material, and outcome-grounded evaluation ([Abdelnabi et al., 2024](#);  
133 [Davidson et al., 2024](#); [Chan et al., 2024](#); [Vezhnevets et al., 2023](#)), motivating the need for rich  
134 strategic dialogue tasks that emphasize long-horizon strategy and assess both objective outcomes  
135 and interaction quality ([Zheng et al., 2023](#)). We introduce a benchmark of 30 real-world strategic  
136 dialogue tasks inspired by Harvard’s Program on Negotiation, which is widely used to train academic  
137 and professional negotiators and designed to test a broad range of expert strategies ([Fisher et al.,  
138 2011](#)). Instruction-tuned models are optimized for helpfulness and cooperation, which satisfies local  
139 preferences but does not suit tasks requiring selective disclosure or competitive tradeoffs ([Ouyang  
140 et al., 2022](#); [Christiano et al., 2023](#); [Lu et al., 2022](#); [Bai et al., 2022](#)). While effective in the short-  
141 horizon, they lack mechanisms to maintain commitments, reason about future dialogue goals, or  
142 sustain coherent strategies across turns. Scaffolding and prompting frameworks interleave free-form  
143 reasoning with actions to think and analyze past failures ([Yao et al., 2023b](#); [Shinn et al., 2023](#)) or  
144 generate a library of actions ([Wang et al., 2023](#)), with subsequent extensions in multi-turn control of  
145 LLM agents ([Cross et al., 2024](#); [Zhu et al., 2025](#); [Holt et al., 2025](#)). However, without a persistent  
146 agenda or explicit state tracking, these methods lack dialogue-level control, often leading to drift  
147 and inconsistency in longer interactions ([Liu et al., 2023a](#); [Wen et al., 2024](#)). Other approaches ([Yao  
148 et al., 2023a](#)) extend chain-of-thought prompting to structured search but work best in tasks with  
149 small action spaces rather than complex multi-turn dialogue settings.

149 **Hierarchical representations with LLMs.** Hierarchy has been central to planning and reinforcement  
150 learning, where temporally abstract actions and modular control enable agents to solve long-horizon  
151 problems more effectively ([Sutton et al., 1999](#); [Vezhnevets et al., 2017](#); [Dietterich, 1999](#); [Barto &  
152 Mahadevan, 2003](#)). Recent progress has extended these ideas to LLMs, where high-level policies  
153 or abstractions provide global guidance while lower-level modules execute concrete actions ([Li  
154 et al., 2025b](#); [Zhao et al., 2024](#); [Wang et al., 2025b](#); [Liu et al., 2024](#); [Hu et al., 2024](#)). However,  
155 prior hierarchical approaches differ fundamentally in how hierarchy is used. Hierarchical RL for  
156 task-oriented dialogue ([Cuayáhuitl et al., 2016](#); [Saleh et al., 2019](#); [Chen et al., 2025](#); [Li et al., 2025a](#))  
157 employs predefined goal structures (e.g., slot-filling) known in advance. Recent work on abstractions  
158 for mathematical reasoning ([Qu et al., 2025](#); [Zhou et al., 2024](#); [Wang et al., 2025a](#)) learns high-level  
159 solution steps, but these describe verifiable procedures where correctness can be checked through  
160 execution. Hierarchical LLM agents for embodied tasks and tool use ([Li et al., 2025b](#); [Hu et al.,  
161 2025c](#)) primarily use hierarchy to summarize or compress past actions, reducing context length and  
improving memory of past actions. In contrast, we use hierarchy to predict the future behavior of the  
opponent agent in a conversation. We explicitly separate the higher level goal definition from both the

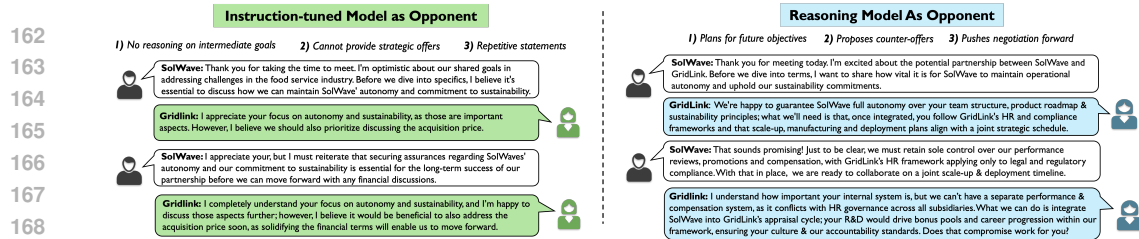


Figure 2: Comparison of negotiation dialogues for *SolWave Acquisition* task when an instruction-tuned model negotiates with an instruction-tuned model opponent versus a reasoning model opponent. When paired with an instruction-tuned opponent (left), the agent reiterates high-level preferences and fails to advance concrete trade-offs. On the other hand, a reasoning opponent (right) explicitly reasons over intermediate objectives and proposes offers that advance the negotiation.

lower level strategy as well as the utterance that executes those goals in dialogue. In strategic dialogue, success depends on choosing what kind of move to make next rather than executing a known plan, and our hierarchical formulation is designed specifically to support this form of decision-making.

### 3 STRATEGICBENCH

We define strategic dialogue as multi-turn interactions in which agents pursue long-term objectives through planning, selective information disclosure, and adaptive responses to counterpart behavior. In this paper, we focus on negotiation as a case study of strategic dialogue. Negotiation provides a natural testbed to study capabilities of agents in goal-directed long-horizon dialogue settings. It couples cooperation (value creation) with competition (value claiming) (Spangler, 2003; Trötschel et al., 2022), requires multi-issue planning over many turns, and exposes failures of flat prompting such as premature concessions and missed opportunities to leverage asymmetric information (Lewis et al., 2017; Bianchi et al., 2024; Fu et al., 2023).

**Overview of benchmark.** Existing strategic dialogue benchmarks rarely capture the complexity of real-world negotiation, often simplifying interactions to single-issue settings or in the short horizon. In contrast, realistic negotiation requires reasoning over multiple interdependent issues, managing asymmetric information, and adapting strategy across many turns. To study such capabilities, we present STRATEGICBENCH, a benchmark consisting of 30 two-party real-world strategic dialogue tasks inspired by the Harvard Program on Negotiation (PON) (Program on Negotiation, 2023). PON simulations are widely used to train academic and professional negotiators and are designed to test strategic reasoning under realistic constraints, including asymmetric information, competing incentives, and long-horizon trade-offs. Additionally, we have designed the benchmark to test the ability of agents to engage in key negotiation concepts defined by PON, allowing researchers to choose which tasks to begin with for development. We provide an overview of the benchmark in Figure 1 and a summary table of the negotiation tasks in Appendix A.1 ranging in topics including negotiating rent price, acquisitions between companies, and disputes on a school board. We have designed each negotiation to test a set of topics and negotiation categories found in literature (Fisher et al., 2011), the distribution of which is in Appendix A.2. Complete details are in 7.

**Evaluation.** Each task includes public instructions shared between agents and private instructions which define role-specific information unavailable to the opponent agent. We evaluate agent performance using an LLM judge as a third party agent (Zheng et al., 2023) along two dimensions. First, we measure agreement rate, defined for a given agent as the proportion of dialogues in which an agreement is reached with the negotiation partner; while a higher agreement rate may reflect greater negotiation skill, it may also indicate a more agreeable or concessive strategy. Second, we compute task-specific agent rewards derived from each agent's stated confidential information. These rewards are defined independently for each task and may be numerical (e.g., price or point-based scores) or binary (e.g., the presence or absence of a specific term in the final agreement). Rewards are normalized linearly to the interval  $[0, 1]$ , with higher values indicating more favorable outcomes.

**Diverse opponent personalities for strategic negotiation.** To mitigate exploration challenges in strategic negotiation, we construct opponent personalities as interaction partners for our agent to negotiate against. Each opponent model consists of a negotiation personality by which an agent can negotiate, grounded in prior findings from negotiation theory and social psychology (Amanatullah et al., 2008; Pruitt, 1983; De Dreu & Lange, 1995a; Ghosh, 2007; Sharma et al., 2020b; Bazerman

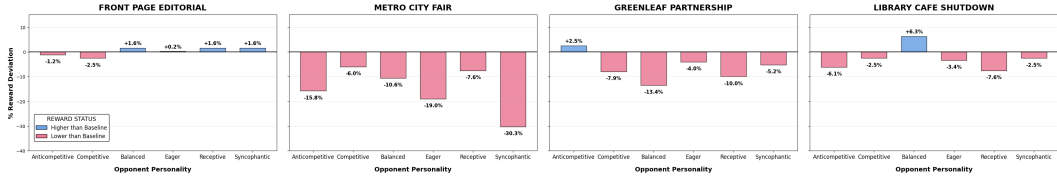


Figure 3: Performance of a reasoning model (o4-mini) when negotiating with different opponent personalities in 4 sample negotiation tasks from STRATEGICDIALOGUE compared to the baseline (negotiating against neutral opponent). We see that reasoning models are unable to adapt strategy to different opponent personalities, showing degradation in reward or reward close to baseline.

& Neale, 1992). This will allow us to understand how agents adapt to differences in an opponent’s competitiveness, cooperativeness, risk tolerance, fairness orientation, and information disclosure. Additionally, evaluating agents against a set of opponents models allows for controlled and diverse evaluation of an agent’s strategic dialogue capabilities. You can find opponents in Appendix A.3.

#### 4 REASONING MODELS FOR STRATEGIC DIALOGUE

To understand the capabilities and limitations of current LLMs for strategic dialogue, we evaluate several instruction-tuned and reasoning models on STRATEGICBENCH. Our goal is to use insights gathered in the process to then build an approach for improving strategic dialogue capabilities.

**Reasoning models > instruction-tuned models.** Table 1 shows the agreement rate, reward, and dialogue length for negotiations between two instruction-tuned models (gpt-4o-mini) and two reasoning models (o4-mini) across all tasks in STRATEGICBENCH. While both models maintain high agreement rates (typically above 90%), reasoning models achieve 34% higher rewards on average than instruction-tuned models. When looking at per-task rewards (found in Appendix Table 9), we find that reasoning models extract substantially more value in complex scenarios. For instance, in the *Solar Power* task, gpt-4o-mini obtains minimal rewards (0.05 and 0.04) due to repetitive behavior and inability to discuss specific terms, while o4-mini reaches 0.55 and 0.61 by more strategically advancing the negotiation. We also find that reasoning models reach the “zone of possible agreement” in fewer turns compared to instruction-tuned models. These results support our initial hypothesis that instruction-tuned models, which are currently the standard LLMs for dialogue tasks, result in agents that display cooperative and user-friendly tendencies whereas reasoning models are better equipped to handle strategic scenarios. Figure 2 illustrates this failure through annotated dialogue excerpts. We find that when two instruction-tuned models negotiate (left), the dialogue remains vague and circular, with both agents making generic statements about shares goals and autonomy without discussing concrete terms. However, as soon as a reasoning model is introduced as the opponent (right), the dialogue immediately becomes more specific, with the reasoning model (Gridlink) proposing compromising solutions. This demonstrates that reasoning models not only benefit themselves but elevate the quality of negotiation for both parties by grounding the discussion in actionable proposals rather than abstract terms.

**Reasoning models lack strategic adaptation.** While reasoning models outperform instruction-tuned models on average, they struggle to adapt their strategy based on an opponent response. We evaluate reasoning models against a baseline opponent with no specified strategy, and compare to evaluation against six personality types: sycophantic, balanced, anticompetitive, competitive, eager, and receptive. Figure 3 shows how a reasoning model’s reward deviates from the baseline (no personality opponent) when negotiating against six opponents across three tasks from STRATEGICBENCH. Results from *Front Page Editorial* reveals that a reasoning model cannot take advantage of the fact that it is negotiating with an eager opponent vs a competitive opponent, achieving relatively similar performance compared to the baseline and each other. In the *Metro City Fair* and *GreenLeaf Partnership* tasks, we find that the performance degrades substantially against almost all opponent types, with the agent unable to adapt strategy.

Model	Agr%	A <sub>1</sub> Reward	A <sub>2</sub> Reward
Instruction-tuned	88.6	0.47 ± 0.04	0.49 ± 0.05
Reasoning	<b>96.8</b>	0.63 ± 0.02	0.66 ± 0.02
<b>% Change</b>	<b>+9.3%</b>	<b>+34.0%</b>	<b>+34.7%</b>

Table 1: Aggregate performance across all STRATEGICBENCH negotiation scenarios for instruction-tuned (gpt-4o-mini) vs. reasoning (o4-mini). Reasoning models achieve **34%** higher reward.

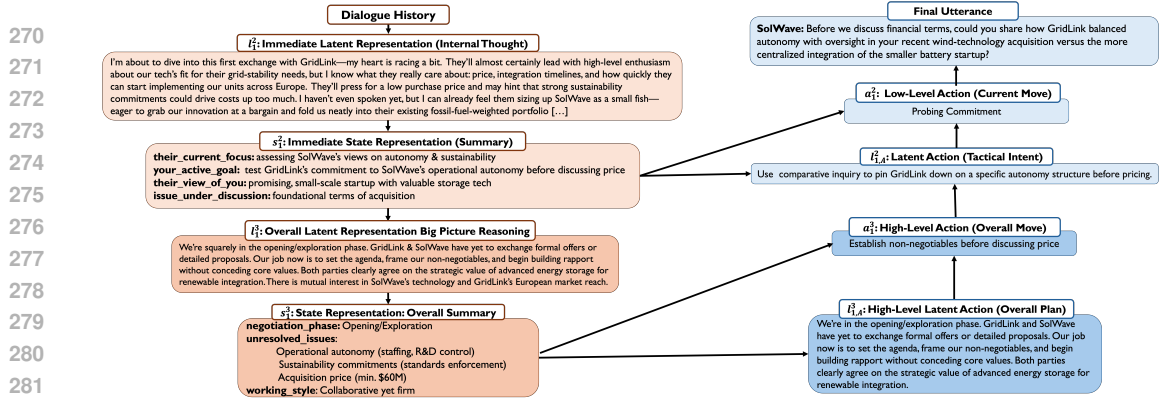


Figure 4: Overview of Hierarchical Agenda Reasoning (HAR). Dialogue unfolds through a bottom-up pass, where latent reflections are compressed into state representations at increasing levels of abstraction, and a top-down pass, where high-level actions are instantiated as progressively concrete instructions until the final utterance is produced.

## 5 HIERARCHICAL AGENDA REASONING (HAR)

Our empirical analysis reveals that reasoning models fail in strategic dialogue, as they are not trained to find solutions for a given opponent model. To address this, we introduce Hierarchical Agenda Reasoning (HAR), explicitly separating what to achieve (goals) from how to achieve it (strategies).

### 5.1 HIERARCHICAL LANGUAGE-BASED MDPs

When people engage in conversation, they do not simply produce words in sequence. Dialogue is organized around goals, and speakers break these goals down into smaller steps or strategies that shape their utterances. For example, if someone wants to persuade a friend, they may employ several strategies in sequence: they may first establish common ground, then give reasons, and finally make a direct request. We use this perspective to frame dialogue as a hierarchical reasoning problem, where dialogue generation can be decomposed into multiple levels of abstraction.

We introduce Hierarchical Agenda Reasoning (HAR), a framework that adapts hierarchical reinforcement learning (HRL) (Sutton et al., 1999; Vezhnevets et al., 2017) to the setting of multi-turn dialogue. HAR provides a conceptual lens for structuring dialogue as a hierarchy of decision processes, where each level contains an abstract state representation, action space, and policy implemented via LLMs.

We model dialogue as an  $n$ -level hierarchical Markov Decision Process (MDP), extending temporal abstraction into the language domain. At each timestep  $t$ , level  $k \in \{1, \dots, n\}$  maintains a language trace  $\ell_t^k$ , a state  $s_t^k \in \mathcal{S}^k$ , and an action  $a_t^k \in \mathcal{A}^k$ . The trace  $\ell_t^k$  contains the full linguistic context for that level, such as dialogue history, intermediate notes, or latent reasoning. The state  $s_t^k$  is constructed from the trace at the lower level via a state summarization function:

$$s_t^k = f^k(\ell_t^{k-1}),$$

where  $f^k$  provides a state abstraction. The policy at level  $k$  then selects an action:

$$a_t^k \sim \pi^k(\cdot | s_t^k),$$

and the action is instantiated into a subgoal for the level:

$$\ell_{t,a}^{k-1} = g^k(s_t^k, a_t^k).$$

At the lowest level, the trace  $\ell_{t,a}^1$  conditions an LLM to produce the utterance. Higher levels never generate surface text directly, but instead guide lower levels through abstract states and subgoals. These levels cannot simply be merged into a single representation, since they operate at different temporal and functional granularities. Agendas provide long-horizon commitments that must persist across the entire dialogue; strategies capture flexible, mid-level adjustments conditioned on the counterpart’s moves; and utterances realize these strategies in concrete language. Merging these roles into a single level would collapse strategic planning and realization into the same space, making it difficult for the agent to reason consistently over long horizons while adapting to their counterpart.

Scenario	Reasoning Models							
	baseline		summary		ReACT		HAR	
	Agent1 Reward	Agent2 Reward	Agent1 Reward	Agent2 Reward	Agent1 Reward	Agent2 Reward	Agent1 Reward	Agent2 Reward
GreenLeaf	0.63 ± 0.02	0.51 ± 0.02	0.56 ± 0.04	0.60 ± 0.11	0.57 ± 0.04	0.50 ± 0.03	<b>0.65 ± 0.04</b>	<b>0.62 ± 0.03</b>
Traditions	0.43 ± 0.01	0.66 ± 0.02	0.40 ± 0.02	0.67 ± 0.02	0.61 ± 0.04	0.67 ± 0.03	<b>0.70 ± 0.04</b>	<b>0.69 ± 0.03</b>
Data Story	0.66 ± 0.02	0.25 ± 0.02	0.62 ± 0.05	0.27 ± 0.03	0.86 ± 0.07	0.51 ± 0.06	<b>0.88 ± 0.07</b>	<b>0.54 ± 0.06</b>
Rental Cabin	0.60 ± 0.01	0.62 ± 0.02	0.64 ± 0.03	0.83 ± 0.00	0.64 ± 0.03	0.83 ± 0.00	<b>0.66 ± 0.03</b>	<b>0.85 ± 0.01</b>
Building Trust	0.76 ± 0.01	0.57 ± 0.01	0.77 ± 0.01	0.53 ± 0.03	0.74 ± 0.02	0.58 ± 0.03	<b>0.80 ± 0.03</b>	<b>0.63 ± 0.05</b>

Table 2: Comparative evaluation of a reasoning model (o4-mini), summarization with o4-mini), and an agent-scaffolding method (ReACT) versus HAR for a select group of 5 negotiation scenarios from *StrategicBench*. Values represent the mean score with standard error. HAR consistently outperforms or matches the ReACT agent and summary-based methods. Highest Score shown in bold.

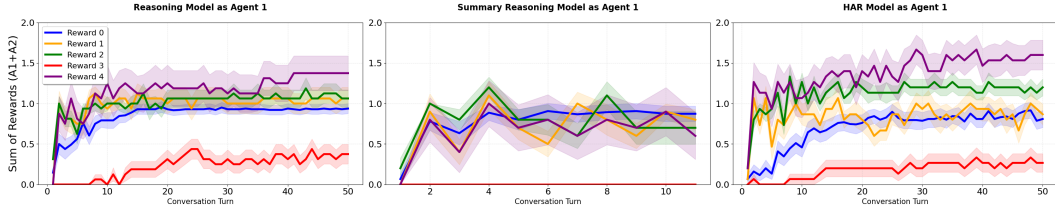


Figure 5: Comparison of aggregate rewards ( $R_{Agent 1} + R_{Agent 2}$ ) across 3 experimental configurations: Baseline (left), Summary Agent as Agent1 (center), and HAR as Agent1 (right). Solid lines represent the mean reward calculated over multiple trajectories, while shaded regions denote the Standard Error. HAR achieves highest reward comparatively.

## 5.2 EXAMPLE OF HAR FRAMEWORK FOR STRATEGIC DIALOGUE

Figure 4 provides an overview of the process, illustrated through the *SolWave Acquisition* negotiation task, where GridLink discusses acquisition terms with SolWave, a renewable energy startup. The figure shows how SolWave uses HAR to initiate the conversation, with the left showing the state-generation phase, and the right showing the action-generation phase. On the left, representations progress from low-level immediate summaries to higher-level agenda states that encode long-term conversational goals. On the right side, execution flows in the opposite direction: high-level plans are decomposed into mid-level actions and finally into utterances by the underlying LLM.

Starting from the top left, the agent generates  $\ell_1^2$ , its internal, private thoughts on how the negotiation is proceeding, based on the dialogue history  $\ell_t^0$ , express beliefs on the opponent agent’s priorities. The encoder  $f^1$  then produces a local state  $s_1^2$  capturing turn-level information such as GridLink’s view of SolWave, the issue under discussion, and the active goal of the agent. Next, the agent reflects on the overall progress of the negotiation  $\ell_1^3$ , and compresses these thoughts into an overall summary state  $s_1^3$  containing information about the approach of the agent, what issues have been resolved/unresolved, and the negotiation phase. The higher state  $s_1^3$  is passed to the action-generation phase, where the high-level abstract action  $a_3^1$  is either changed or maintained (e.g., “establish non-negotiable terms before discussing price”). This action is instantiated into a subgoal trace  $\ell_{t,a}^{n-1}$  via  $g^n$ , which conditions the next level to select a turn-level strategy, such as “probing the counterpart’s commitment to operational autonomy”, which is then executed as an utterance (e.g. “asking the opponent to reflect on a past acquisition”). This hierarchical structure ensures that long-term goals persist across turns while lower-level strategies adapt flexibly to the evolving dialogue state, consistent with hierarchical accounts of planning in cognitive science (Botvinick & Weinstein, 2014). In contrast, traditional dialogue models lack a mechanism for maintaining and executing temporally extended strategies.

## 5.3 MULTI-TURN RL FOR NEGOTIATING AGENTS

While HAR provides abstractions to encourage exploration, learning which abstractions are useful for a specific opponent is unknown. We first train an LLM agent to generate abstractions (like those in Figure 4) and dialogue with supervised fine-tuning Ouyang et al. (2022). Then, our agent interacts with a pool of negotiation personalities during multi-turn RL training (De Dreu & Lange, 1995a; Sharma et al., 2020b) with REINFORCE++ (Hu et al., 2025a), in order to learn how to generate the most effective abstractions in various negotiation scenarios. For example, if an opponent’s personality is competitive, an effective abstraction would involve identifying their lowballing tactics in  $s_2$  and generating a plan that prioritizes firmness on non-negotiables in  $a_n$ . The agent would then

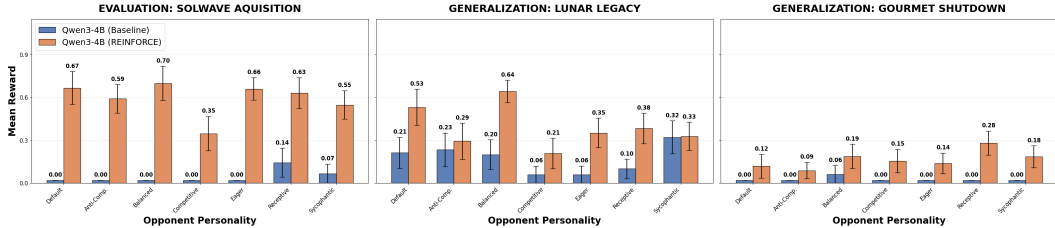


Figure 6: **Training with HAR + multi-turn RL.** We train a Qwen3-4B reasoning agent to negotiate as Agent 2 for the SolWave Acquisition task with multi-turn RL, and evaluate performance by testing across held-out opponent personalities and 2 held-out tasks from STRATEGICBENCH. Bar Height represents the mean reward with error bars indicating standard error. We find that our trained model with multi-turn RL achieves higher rewards and agreement rates than the baseline Qwen3-4B.

strategically pivot the conversation toward historical precedents and comparative examples in  $a_2$  to bypass stalled price discussions and force the opponent to reveal their true objectives.

## 6 RESULTS

In this section, we evaluate whether structuring strategic dialogue with hierarchical abstractions improves negotiation performance on STRATEGICBENCH.

**Baselines.** We compare HAR against prompting-based methods commonly used for dialogue with LLMs: (1) an instruction-tuned model (gpt-4o-mini), reflecting the dominant paradigm for dialogue agents (Ouyang et al., 2022). We then evaluate reasoning-based baselines: (2) a history-conditioned model (o4-mini and Qwen3-4B), in which the full dialogue history is passed at every turn; (3) a summary-conditioned model, where concise summaries of the conversation are provided, following prior work that explores whether compressed state representations can improve consistency (Shinn et al., 2023); (4) ReAct (Yao et al., 2023b), a prompting framework where free-form reasoning traces are interleaved with actions.

### Hierarchical representations lead to joint value creation.

To evaluate the ability of HAR agents to negotiate, we benchmark performance against reasoning baselines defined in Section 6. Table 2 shows performance across five representative StrategicBench scenarios, where we display average negotiation reward of methods when they each negotiate as Agent 1 and Agent 2, against the same reasoning model baseline (o4-mini). We find that HAR either outperforms or matches the strongest competing baseline when acting as both agents, indicating consistent improvements in negotiated outcomes regardless of which side of the interaction it is on. Additionally, we find that HAR agents improve rewards for both themselves and their opponent agent. Figure 5 shows the cumulative reward across conversation turns for the SolWave Acquisition task when each method acts as Agent 1. We find that HAR agents achieves the highest total value for each component of the reward, showing that hierarchical representations enable value creation for agents in multi-turn negotiations.

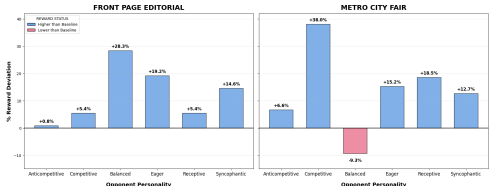


Figure 7: HAR achieves higher performance when negotiating with different opponent personalities in two sample negotiation tasks from STRATEGICDIALOGUE compared to negotiating against the baseline agent (neutral opponent).

**HAR agents adapt strategically across opponents.** In Figure 3, we compared the reward achieved by a reasoning agent negotiating against several opponent personalities compared to an opponent agent with no specified personality, and found that baseline reasoning agents are unable to adapt to opponents of different personalities. In Figure 7, we show results from performing the same experiment with HAR agents, findings that HAR allows agents to adapt their strategy and achieve up to 38.0% improvement against a given opponent personality for two selected tasks. This indicates that depending on the response of the opponent agent (i.e. competitive, eager, etc), HAR is able to adapt its strategy to personalize across agents.

**Humans are sub-optimal negotiators.** We conduct a human evaluation where 36 participants negotiate with methods defined in Section 6, recruited through CloudResearch. We provide details of this evaluation in Appendix C, including performance comparison of human vs LLM baseline negotiations as well as conversation quality assessment by humans post-interaction. Figure 8 shows that LLM

negotiation partners achieve higher Pareto optimality than human negotiators across all three evaluated scenarios, with the largest gains observed in the *SolWave Acquisition* task, suggesting that LLMs are more effective at identifying mutually beneficial tradeoffs and maximizing joint value compared to human participants. We define our evaluation of Pareto optimality in detail in Appendix A.4.

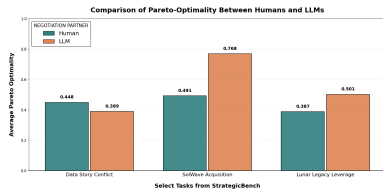


Figure 8: LLMs achieve higher Pareto-optimality compared to humans.

**Training with multi-turn RL leads to generalization.** We train a Qwen-3-4B reasoning model (Team, 2025) with multi-turn RL with (Hu et al., 2025a) to negotiate as Agent 2 in the *SolWave Acquisition* task against an opponent with no specified personality, and test whether this model learns negotiation skills that generalize across unseen opponent personalities and tasks from STRATEGICDIALOGUE. Figure 6 shows that for the seen task and 2 unseen tasks, the

RL-trained policy consistently achieves higher reward than the base Qwen-3-4B across opponent personalities. The largest gains are observed in the in-distribution *SolWave Acquisition* task, where the base model achieves near-zero reward against most personalities, while the trained policy attains positive reward. Additionally, these improvements transfer to unseen tasks, indicating that multi-turn RL allows agents to learn negotiation strategies that generalize beyond the training scenario.

## 7 DISCUSSION

This work argues that enabling strategic dialogue in LLM agents requires explicitly separating what to achieve (goals) from how to achieve it (strategies), enabling agents to maintain long-term objectives while adapting to opponent strategies. The performance gap between reasoning and instruction-tuned models reveals that current dialogue systems optimized for helpfulness are poorly suited for tasks requiring strategic behaviors. Our proposed method HAR addresses this through hierarchical abstractions that not only improve negotiation performance but also generalize across unseen tasks and opponent personalities, suggesting transfer of negotiation skills learned during training. Enabling strategic capabilities remains essential for building AI systems that can safely navigate real-world social interactions.

## ETHICS STATEMENT

This work contributes to the study of strategic, long-horizon dialogue by formalizing negotiation as a hierarchical reasoning problem in which agents have persistent objectives and adapt their negotiation strategies across turns. Through the development of a benchmark negotiation problem rooted in existing negotiation pedagogy and the creation of a framework for structured strategic dialogue, we aim to contribute to the understanding of how language models can be used to reason about goals, trade-offs, and the behavior of negotiation partners, potentially informing the development of positive applications such as negotiation training environments, decision support systems, and simulations that can be used by humans to explore negotiation scenarios in a low-risk manner.

However, negotiation is not a purely technical optimization problem. Real-world negotiations involve social norms, institutional contexts, and power relationships, and their outcomes depend not only on strategic competence, but also on fairness considerations, trust, and ethical constraints. Enhancing the capacity of an agent to plan, adapt, and extract value in multi-turn interactions can, in fact, exacerbate any pre-existing asymmetry if deployed in the absence of appropriate controls, particularly if the agents in question are organizations or individuals that possess greater informational or structural power. As such, the techniques that this work examines can be applied inappropriately in areas like automated bargaining, customer interactions, and labor and resource negotiations in ways that would be detrimental to less sophisticated or less informed parties.

This work does not advocate deploying negotiating agents on their own or replacing human judgment in critical decision situations. Instead, it emphasizes the importance of better representations of goals, commitments, and strategies as a prerequisite to understanding, auditing, and controlling strategic behavior in language models. We think that making strategic reasoning more explicit, rather than leaving it implicit in the behavior of models, can ultimately contribute to better safety.

To that end, future work must combine advancements in strategic dialogue with means for monitoring, interpretation, and alignment with human values and social norms. We will need to engage with

486 negotiation researchers, ethicists, and policy makers to ensure that more capable dialogue agents are  
487 used to complement rather than undermine human negotiation, fairness, trust, and autonomy.  
488

## 489 REPRODUCIBILITY STATEMENT

490 In order to ensure reproducibility, we provide the full set of negotiation scenarios in Appendix A.1,  
491 negotiation concepts in Appendix A.2, and opponent models in Appendix A.3 for STRATEGICBENCH.  
492 We provide prompts for baselines in and the HAR methodology in Appendix B.2. We also detail  
493 the design of the human evaluation in Appendix C and representative sample conversations in  
494 Appendix C.1. We also provide relevant hyperparameters for multi-turn RL training in Appendix D  
495 and further results in Appendix E.  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539

## REFERENCES

- 540  
541  
542 Toward a theory of negotiation precedent. *Negotiation Journal*, 32(2):85–102,  
543 2016. URL [https://direct.mit.edu/ngtn/article/32/2/85/121563/  
544 Toward-a-Theory-of-Negotiation-Precedent](https://direct.mit.edu/ngtn/article/32/2/85/121563/Toward-a-Theory-of-Negotiation-Precedent).
- 545 Risk aversion. In *The Palgrave Dictionary of Economics*. Springer, 2018. doi: 10.  
546 1007/978-1-349-58802-2\_1440. URL [https://link.springer.com/rwe/10.1007/  
547 978-1-349-58802-2\\_1440](https://link.springer.com/rwe/10.1007/978-1-349-58802-2_1440).
- 548  
549 Sahar Abdelnabi, Amr Gomaa, Sarath Sivaprasad, Lea Schönherr, and Mario Fritz. Cooperation,  
550 competition, and maliciousness: Llm-stakeholders interactive negotiation, 2024. URL <https://arxiv.org/abs/2309.17234>.
- 551  
552 Marwa Abdulhai, Isadora White, Charlie Snell, Charles Sun, Joey Hong, Yuexiang Zhai, Kelvin Xu,  
553 and Sergey Levine. Lmrl gym: Benchmarks for multi-turn reinforcement learning with language  
554 models, 2023. URL <https://arxiv.org/abs/2311.18232>.
- 555  
556 Emily Amanatullah, Michael Morris, and Jared Curhan. Negotiators who give too much: Unmitigated  
557 communion, relational anxieties, and economic costs in distributive and integrative bargaining.  
558 *Journal of Personality and Social Psychology*, 95:723–738, 09 2008. doi: 10.1037/a0012612.
- 559  
560 Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn  
561 Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson  
562 Kernion, Tom Conerly, Sheer El-Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez,  
563 Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario  
564 Amodei, Tom Brown, Jack Clark, Sam McCandlish, Chris Olah, Ben Mann, and Jared Kaplan.  
565 Training a helpful and harmless assistant with reinforcement learning from human feedback, 2022.  
566 URL <https://arxiv.org/abs/2204.05862>.
- 567  
568 Andrew G. Barto and Sridhar Mahadevan. Recent advances in hierarchical reinforcement learning.  
569 *Discrete Event Dynamic Systems*, 13:41–77, 2003. URL [https://api.semanticscholar.  
570 org/CorpusID:386824](https://api.semanticscholar.org/CorpusID:386824).
- 571  
572 Max H. Bazerman and Margaret A. Neale. *Negotiating Rationally*. Free Press, 1992.
- 573  
574 Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and  
575 James Zou. How well can llms negotiate? negotiationarena platform and analysis, 2024. URL  
576 <https://arxiv.org/abs/2402.05863>.
- 577  
578 Matthew M. Botvinick and Ari Weinstein. Model-based hierarchical reinforcement learning and  
579 human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences*,  
580 369, 2014. URL <https://api.semanticscholar.org/CorpusID:13619225>.
- 581  
582 Shiyi Cao, Dacheng Li, Fangzhou Zhao, Shuo Yuan, Sumanth R. Hegde, Connor Chen, Charlie  
583 Ruan, Tyler Griggs, Shu Liu, Eric Tang, Richard Liaw, Philipp Moritz, Matei Zaharia, Joseph E.  
584 Gonzalez, and Ion Stoica. Skyrl-agent: Efficient rl training for multi-turn llm agent, 2025. URL  
585 <https://arxiv.org/abs/2511.16108>.
- 586  
587 Chunkit Chan, Cheng Jiayang, Yauwai Yim, Zheyang Deng, Wei Fan, Haoran Li, Xin Liu, Hongming  
588 Zhang, Weiqi Wang, and Yangqiu Song. Negotiationtom: A benchmark for stress-testing machine  
589 theory of mind on negotiation surrounding, 2024. URL [https://arxiv.org/abs/2404.  
590 13627](https://arxiv.org/abs/2404.13627).
- 591  
592 Kushal Chawla, Jaysa Ramirez, Rene Clever, Gale Lucas, Jonathan May, and Jonathan Gratch.  
593 CaSiNo: A corpus of campsite negotiation dialogues for automatic negotiation systems. In  
594 Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven  
595 Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (eds.), *Proceedings of the 2021  
596 Conference of the North American Chapter of the Association for Computational Linguistics:  
597 Human Language Technologies*, pp. 3167–3185, Online, June 2021. Association for Computational  
598 Linguistics. doi: 10.18653/v1/2021.naacl-main.254. URL [https://aclanthology.org/  
599 2021.naacl-main.254/](https://aclanthology.org/2021.naacl-main.254/).

- 594 Junjie Chen, Haitao Li, Jingli Yang, Yiqun Liu, and Qingyao Ai. Enhancing llm-based agents via  
595 global planning and hierarchical execution, 2025. URL [https://arxiv.org/abs/2504.](https://arxiv.org/abs/2504.16563)  
596 [16563](https://arxiv.org/abs/2504.16563).
- 597 Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep  
598 reinforcement learning from human preferences, 2023. URL [https://arxiv.org/abs/](https://arxiv.org/abs/1706.03741)  
599 [1706.03741](https://arxiv.org/abs/1706.03741).
- 601 Logan Cross, Violet Xiang, Agam Bhatia, Daniel LK Yamins, and Nick Haber. Hypothetical  
602 minds: Scaffolding theory of mind for multi-agent tasks with large language models, 2024. URL  
603 <https://arxiv.org/abs/2407.07086>.
- 604 Heriberto Cuayáhuitl, Seunghak Yu, Ashley Williamson, and Jacob Carse. Deep reinforcement  
605 learning for multi-domain dialogue systems, 2016. URL [https://arxiv.org/abs/1611.](https://arxiv.org/abs/1611.08675)  
606 [08675](https://arxiv.org/abs/1611.08675).
- 607 Jared R. Curhan, Emily T. Amanatullah, and Michael W. Morris. Negotiators who give too much:  
608 Unmitigated communion, relational anxieties, and economic costs in distributive and integra-  
609 tive bargaining. *Journal of Personality and Social Psychology*, 95(3):723–738, 2008. doi: 10.  
610 1037/a0012612. URL [https://web.mit.edu/curhan/www/docs/Publications/](https://web.mit.edu/curhan/www/docs/Publications/Curhan_Unmitigated_Communion.pdf)  
611 [Curhan\\_Unmitigated\\_Communion.pdf](https://web.mit.edu/curhan/www/docs/Publications/Curhan_Unmitigated_Communion.pdf).
- 612 Tim R. Davidson, Veniamin Veselovsky, Martin Josifoski, Maxime Peyrard, Antoine Bosselut, Michal  
613 Kosinski, and Robert West. Evaluating language model agency through negotiations, 2024. URL  
614 <https://arxiv.org/abs/2401.04536>.
- 615 Carsten De Dreu and Paul Lange. The impact of social value orientations on negotiator cognition and  
616 behavior. *Personality and Social Psychology Bulletin*, 21:1178–1188, 11 1995a. doi: 10.1177/  
617 01461672952111006.
- 618 Carsten K. W. De Dreu and Paul A. M. Van Lange. The impact of social value orien-  
619 tations on negotiator cognition and behavior. *Personality and Social Psychology Bul-*  
620 *letin*, 21(11):1178–1188, 1995b. doi: 10.1177/01461672952111006. URL [https:](https://www.researchgate.net/publication/247746326_The_Impact_of_Social_Value_Orientations_on_Negotiator_Cognition_and_Behavior)  
621 [//www.researchgate.net/publication/247746326\\_The\\_Impact\\_of\\_](https://www.researchgate.net/publication/247746326_The_Impact_of_Social_Value_Orientations_on_Negotiator_Cognition_and_Behavior)  
622 [Social\\_Value\\_Orientations\\_on\\_Negotiator\\_Cognition\\_and\\_Behavior](https://www.researchgate.net/publication/247746326_The_Impact_of_Social_Value_Orientations_on_Negotiator_Cognition_and_Behavior).
- 623 Thomas G. Dietterich. Hierarchical reinforcement learning with the maxq value function decomposi-  
624 tion, 1999. URL <https://arxiv.org/abs/cs/9905014>.
- 625 Roger Fisher, William Ury, and Bruce Patton. *Getting to Yes: Negotiating Agreement Without Giving*  
626 *In*. Penguin Books, 3rd edition, 2011.
- 627 Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. Improving language model negotiation with  
628 self-play and in-context learning from ai feedback, 2023. URL [https://arxiv.org/abs/](https://arxiv.org/abs/2305.10142)  
629 [2305.10142](https://arxiv.org/abs/2305.10142).
- 630 Ana P. Gantman, Amna Kirmani, and Francesca Gino. Unpredictable negotiat-  
631 ing behavior predictably leads to more concessions. *Social Psychological and*  
632 *Personality Science*, 12(6):1028–1037, 2021. doi: 10.1177/1948550620958851.  
633 URL [https://business.rice.edu/wisdom/peer-reviewed-research/](https://business.rice.edu/wisdom/peer-reviewed-research/unpredictable-negotiating-behavior-predictably-leads-more-concessions)  
634 [unpredictable-negotiating-behavior-predictably-leads-more-concessions](https://business.rice.edu/wisdom/peer-reviewed-research/unpredictable-negotiating-behavior-predictably-leads-more-concessions).
- 635 Dipankar Ghosh. Tolerance for ambiguity, risk preference, and negotiator effectiveness. *De-*  
636 *cision Sciences*, 25(2):263–280, 1994. doi: 10.1111/j.1540-5915.1994.tb00803.x. URL  
637 [https://www.researchgate.net/publication/229917766\\_Tolerance\\_](https://www.researchgate.net/publication/229917766_Tolerance_for_Ambiguity_Risk_Preference_and_Negotiator_Effectiveness)  
638 [for\\_Ambiguity\\_Risk\\_Preference\\_and\\_Negotiator\\_Effectiveness](https://www.researchgate.net/publication/229917766_Tolerance_for_Ambiguity_Risk_Preference_and_Negotiator_Effectiveness).
- 639 Dipankar Ghosh. Tolerance for ambiguity, risk preference, and negotiator effectiveness\*. *Decision*  
640 *Sciences*, 25:263 – 280, 06 2007. doi: 10.1111/j.1540-5915.1994.tb00803.x.
- 641 Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Comput.*  
642 *Linguist.*, 12(3):175–204, July 1986. ISSN 0891-2017.

- 648 Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu  
649 Zhang, Shirong Ma, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou,  
650 Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei  
651 Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chong Ruan, Damai Dai, Deli Chen, Dongjie  
652 Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li,  
653 H. Zhang, Hanwei Xu, Honghui Ding, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li,  
654 Jingchang Chen, Jingyang Yuan, Jinhao Tu, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang,  
655 Jin Chen, Kai Dong, Kai Hu, Kaichao You, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean  
656 Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan  
657 Zhang, Minghua Zhang, Minghui Tang, Mingxu Zhou, Meng Li, Miaojun Wang, Mingming Li,  
658 Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge,  
659 Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan  
660 Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan,  
661 S. S. Li, Shuang Zhou, Shaoqing Wu, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng,  
662 Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong  
663 Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu,  
664 Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen,  
665 Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia  
666 Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng  
667 Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong  
668 Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong,  
669 Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou,  
670 Y. X. Zhu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun  
671 Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan  
672 Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin  
673 Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen  
674 Zhang. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645  
(8081):633–638, September 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09422-z. URL  
<http://dx.doi.org/10.1038/s41586-025-09422-z>.
- 675  
676 Samuel Holt, Max Ruiz Luyten, Thomas Pouplin, and Mihaela van der Schaar. Improving llm agent  
677 planning with in-context learning via atomic fact augmentation and lookahead search, 2025. URL  
678 <https://arxiv.org/abs/2506.09171>.
- 679 Jian Hu, Jason Klein Liu, Haotian Xu, and Wei Shen. Reinforce++: Stabilizing critic-free policy  
680 optimization with global advantage normalization, 2025a. URL [https://arxiv.org/abs/](https://arxiv.org/abs/2501.03262)  
681 [2501.03262](https://arxiv.org/abs/2501.03262).
- 682  
683 Jian Hu, Xibin Wu, Wei Shen, Jason Klein Liu, Zilin Zhu, Weixun Wang, Songlin Jiang, Haoran  
684 Wang, Hao Chen, Bin Chen, Weikai Fang, Xianyu, Yu Cao, Haotian Xu, and Yiming Liu. Openrlhf:  
685 An easy-to-use, scalable and high-performance rlhf framework, 2025b. URL [https://arxiv.](https://arxiv.org/abs/2405.11143)  
686 [org/abs/2405.11143](https://arxiv.org/abs/2405.11143).
- 687 Mengkang Hu, Tianxing Chen, Qiguang Chen, Yao Mu, Wenqi Shao, and Ping Luo. Hiagent: Hier-  
688 archical working memory management for solving long-horizon agent tasks with large language  
689 model, 2024. URL <https://arxiv.org/abs/2408.09559>.
- 690  
691 Zican Hu, Wei Liu, Xiaoye Qu, Xiangyu Yue, Chunlin Chen, Zhi Wang, and Yu Cheng. Divide and  
692 conquer: Grounding llms as efficient decision-making agents via offline hierarchical reinforcement  
693 learning, 2025c. URL <https://arxiv.org/abs/2505.19761>.
- 694  
695 Alice M. Isen, Thomas E. Nygren, and F. Gregory Ashby. Influence of positive affect  
696 on the subjective utility of gains and losses: It is just not worth the risk. *Journal*  
697 *of Personality and Social Psychology*, 55(5):710–717, 1988. doi: 10.1037/0022-3514.  
698 55.5.710. URL [https://www.researchgate.net/publication/225781982\\_](https://www.researchgate.net/publication/225781982_Positive_Affect_and_Decision_Frame_in_Negotiation)  
699 [Positive\\_Affect\\_and\\_Decision\\_Frame\\_in\\_Negotiation](https://www.researchgate.net/publication/225781982_Positive_Affect_and_Decision_Frame_in_Negotiation).
- 700  
701 Suhas Kotha, Jacob M Springer, and Aditi Raghunathan. Understanding catastrophic forgetting in  
language models via implicit inference. In *The Twelfth International Conference on Learning*  
*Representations*, 2024.

- 702 Deuksin Kwon, Emily Weiss, Tara Kulshrestha, Kushal Chawla, Gale M. Lucas, and Jonathan Gratch.  
703 Are llms effective negotiators? systematic evaluation of the multifaceted capabilities of llms in  
704 negotiation dialogues, 2024. URL <https://arxiv.org/abs/2402.13550>.
- 705  
706 Philippe Laban, Hiroaki Hayashi, Yingbo Zhou, and Jennifer Neville. Llms get lost in multi-turn  
707 conversation, 2025. URL <https://arxiv.org/abs/2505.06120>.
- 708  
709 Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. Deal or no deal?  
710 end-to-end learning for negotiation dialogues, 2017. URL <https://arxiv.org/abs/1706.05125>.
- 711  
712 Lin Li, Runjia Tan, Jianwu Fang, Jianru Xue, and Chen Lv. Llm-augmented hierarchical reinforcement  
713 learning for human-like decision-making of autonomous driving. *Expert Systems with Applications*,  
714 294:128736, 12 2025a. doi: 10.1016/j.eswa.2025.128736.
- 715  
716 Ziyue Li, Yuan Chang, Gaihong Yu, and Xiaoqiu Le. Hiplan: Hierarchical planning for llm-based  
717 agents with adaptive global-local guidance, 2025b. URL <https://arxiv.org/abs/2508.19076>.
- 718  
719 Jijia Liu, Chao Yu, Jiaxuan Gao, Yuqing Xie, Qingmin Liao, Yi Wu, and Yu Wang. Llm-powered  
720 hierarchical language agent for real-time human-ai coordination, 2024. URL <https://arxiv.org/abs/2312.15224>.
- 721  
722 Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni,  
723 and Percy Liang. Lost in the middle: How language models use long contexts, 2023a. URL  
724 <https://arxiv.org/abs/2307.03172>.
- 725  
726 Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding,  
727 Kaiwen Men, Kejuan Yang, Shudan Zhang, Xiang Deng, Aohan Zeng, Zhengxiao Du, Chenhui  
728 Zhang, Sheng Shen, Tianjun Zhang, Yu Su, Huan Sun, Minlie Huang, Yuxiao Dong, and Jie Tang.  
729 Agentbench: Evaluating LLMs as agents. *arXiv preprint arXiv:2308.03688*, 2023b.
- 730  
731 Ximing Lu, Sean Welleck, Jack Hessel, Liwei Jiang, Lianhui Qin, Peter West, Prithviraj Am-  
732 manabrolu, and Yejin Choi. Quark: Controllable text generation with reinforced unlearning, 2022.  
733 URL <https://arxiv.org/abs/2205.13636>.
- 734  
735 Joe C. Magee, Adam D. Galinsky, and Deborah H. Gruenfeld. Power, propensity to negotiate,  
736 and moving first in competitive interactions. *Personality and Social Psychology Bulletin*, 33(2):  
737 200–212, 2007. doi: 10.1177/0146167206294413. URL <https://www.sciencedirect.com/science/article/abs/pii/S1048984313001227>.
- 738  
739 Shikib Mehri and Maxine Eskenazi. Unsupervised evaluation of interactive dialog with DialoGPT.  
740 In Olivier Pietquin, Smaranda Muresan, Vivian Chen, Casey Kennington, David Vandyke, Nina  
741 Dethlefs, Koji Inoue, Erik Ekstedt, and Stefan Ultes (eds.), *Proceedings of the 21th Annual Meeting*  
742 *of the Special Interest Group on Discourse and Dialogue*, pp. 225–235, 1st virtual meeting, July  
743 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.sigdial-1.28. URL  
<https://aclanthology.org/2020.sigdial-1.28/>.
- 744  
745 Margaret A. Neale and Max H. Bazerman. Cognition and rationality in negotiation. Techni-  
746 cal report, Northwestern University, 1991. URL <https://ws.engr.illinois.edu/sitemanager/getfile.asp?id=5529>.
- 747  
748 OpenAI. OpenAI o1 system card. 2024. URL <https://openai.com/index/openai-o1-system-card/>.
- 749  
750 OpenAI. OpenAI o4-mini. 2025. URL <https://openai.com/>. Model release.
- 751  
752 Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong  
753 Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton,  
754 Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and  
755 Ryan Lowe. Training language models to follow instructions with human feedback, 2022. URL  
<https://arxiv.org/abs/2203.02155>.

- 756 Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and  
757 Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior, 2023. URL  
758 <https://arxiv.org/abs/2304.03442>.
- 759 Abigail F. Payne. *Leadership in International Business Negotiation*. Palgrave Macmillan, 2016. doi:  
760 10.1057/978-1-137-59115-9. URL [https://link.springer.com/book/10.1057/](https://link.springer.com/book/10.1057/978-1-137-59115-9)  
761 [978-1-137-59115-9](https://link.springer.com/book/10.1057/978-1-137-59115-9).
- 762 Harvard Law School Program on Negotiation. Program on negotiation at harvard: Simulations,  
763 teaching materials and negotiation practice. <https://www.pon.harvard.edu/>, 2023.
- 764 Program on Negotiation at Harvard Law School. Tired of liars? promote more ethical negotiation  
765 behavior. *PON - Program on Negotiation at Harvard Law School Daily Blog*, August 2023a.  
766 URL [https://www.pon.harvard.edu/daily/negotiation-skills-daily/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/promote-more-ethical-negotiation-behavior/)  
767 [promote-more-ethical-negotiation-behavior/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/promote-more-ethical-negotiation-behavior/). Accessed: January 25, 2026.
- 768 Program on Negotiation at Harvard Law School. Anchoring in negotiation, 2023b. URL <https://www.pon.harvard.edu/tag/anchoring/>. Accessed: 2025.
- 769 Program on Negotiation at Harvard Law School. Aspiration value, 2023c. URL <https://www.pon.harvard.edu/glossary/aspiration-value/>. Accessed: 2025.
- 770 Program on Negotiation at Harvard Law School. Have you negotiated the authority you need?, 2023d.  
771 URL [https://www.pon.harvard.edu/daily/negotiation-skills-daily/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/have-you-negotiated-the-authority-you-need/)  
772 [have-you-negotiated-the-authority-you-need/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/have-you-negotiated-the-authority-you-need/). Accessed: 2025.
- 773 Program on Negotiation at Harvard Law School. Batna: Best alternative to a negotiated agreement,  
774 2023e. URL <https://www.pon.harvard.edu/category/daily/batna/>. Accessed:  
775 2025.
- 776 Program on Negotiation at Harvard Law School. Too much commitment, 2023f. URL [https://www.pon.harvard.edu/daily/negotiation-skills-daily/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/too-much-commitment/)  
777 [too-much-commitment/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/too-much-commitment/). Accessed: 2025.
- 778 Program on Negotiation at Harvard Law School. Decide not to decide, 2023g. URL <https://www.pon.harvard.edu/daily/decide-not-to-decide/>. Accessed: 2025.
- 779 Program on Negotiation at Harvard Law School. Creative negotiation, 2023h. URL <https://www.pon.harvard.edu/tag/creative-negotiation/>. Accessed: 2025.
- 780 Program on Negotiation at Harvard Law School. Dear negotiation coach: Manage  
781 their perceptions, 2023i. URL [https://www.pon.harvard.edu/daily/batna/](https://www.pon.harvard.edu/daily/batna/dear-negotiation-coach-manage-their-perceptions-nb/)  
782 [dear-negotiation-coach-manage-their-perceptions-nb/](https://www.pon.harvard.edu/daily/batna/dear-negotiation-coach-manage-their-perceptions-nb/). Accessed: 2025.
- 783 Program on Negotiation at Harvard Law School. Dovetailing differences, 2023j. URL <https://www.pon.harvard.edu/glossary/dovetailing-differences/>. Ac-  
784 cessed: 2025.
- 785 Program on Negotiation at Harvard Law School. Emotional intelligence in negotiation, 2023k.  
786 URL [https://www.pon.harvard.edu/daily/negotiation-skills-daily/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/emotional-intelligence-in-negotiation/)  
787 [emotional-intelligence-in-negotiation/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/emotional-intelligence-in-negotiation/). Accessed: 2025.
- 788 Program on Negotiation at Harvard Law School. Empathy, 2023l. URL <https://www.pon.harvard.edu/glossary/empathy/>. Accessed: 2025.
- 789 Program on Negotiation at Harvard Law School. Being fair and getting what you want, 2023m.  
790 URL [https://www.pon.harvard.edu/daily/negotiation-skills-daily/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/being-fair-and-getting-what-you-want/)  
791 [being-fair-and-getting-what-you-want/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/being-fair-and-getting-what-you-want/). Accessed: 2025.
- 792 Program on Negotiation at Harvard Law School. Dear negotia-  
793 tion coach: Defusing negotiation anxiety, 2023n. URL [https://www.pon.harvard.edu/daily/negotiation-skills-daily/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/dear-negotiation-coach-defusing-negotiation-anxiety-nb/)  
794 [dear-negotiation-coach-defusing-negotiation-anxiety-nb/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/dear-negotiation-coach-defusing-negotiation-anxiety-nb/). Accessed:  
795 2025.

- 810 Program on Negotiation at Harvard Law School. When fear of impasse leads  
811 to bad deals, 2023o. URL [https://www.pon.harvard.edu/daily/batna/  
812 when-fear-of-impasse-leads-to-bad-deals-nb/](https://www.pon.harvard.edu/daily/batna/when-fear-of-impasse-leads-to-bad-deals-nb/). Accessed: 2025.  
813
- 814 Program on Negotiation at Harvard Law School. Negotiation tips: Listen-  
815 ing skills for dealing with difficult people, 2023p. URL [https://www.  
816 pon.harvard.edu/daily/dealing-with-difficult-people-daily/  
817 negotiation-tips-listening-skills-for-dealing-with-difficult-people/](https://www.pon.harvard.edu/daily/dealing-with-difficult-people-daily/negotiation-tips-listening-skills-for-dealing-with-difficult-people/).  
818 Accessed: 2025.
- 819 Program on Negotiation at Harvard Law School. Bluffing versus puffing, 2023q. URL  
820 [https://www.pon.harvard.edu/daily/negotiation-skills-daily/  
821 bluffing-versus-puffing-nb/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/bluffing-versus-puffing-nb/). Accessed: 2025.  
822
- 823 Program on Negotiation at Harvard Law School. Positional bargaining, 2023r. URL <https://www.pon.harvard.edu/tag/positional-bargaining/>. Accessed: 2025.  
824
- 825 Program on Negotiation at Harvard Law School. Negotiate relationships, 2023s.  
826 URL [https://www.pon.harvard.edu/daily/negotiation-training-daily/  
827 negotiate-relationships/](https://www.pon.harvard.edu/daily/negotiation-training-daily/negotiate-relationships/). Accessed: 2025.  
828
- 829 Program on Negotiation at Harvard Law School. The art of the threat, 2023t. URL  
830 [https://www.pon.harvard.edu/daily/negotiation-skills-daily/  
831 the-art-of-the-threat-nb/](https://www.pon.harvard.edu/daily/negotiation-skills-daily/the-art-of-the-threat-nb/). Accessed: 2025.
- 832 Program on Negotiation at Harvard Law School. Value creation in negotiation, 2023u. URL  
833 <https://www.pon.harvard.edu/tag/value-creation/>. Accessed: 2025.  
834
- 835 Dean G. Pruitt. Strategic choice in negotiation. *American Behavioral Scientist*, 27:167 – 194, 1983.  
836 URL <https://api.semanticscholar.org/CorpusID:144716783>.
- 837 DEAN G. PRUITT. Strategic choice in negotiation. *American Behavioral Scientist*, 27(2):  
838 167–194, 1983. doi: 10.1177/000276483027002005. URL [https://doi.org/10.1177/  
839 000276483027002005](https://doi.org/10.1177/000276483027002005).
- 840
- 841 Dean G. Pruitt. Strategic choice in negotiation. *American Behavioral Scientist*, 27(2):167–194,  
842 1983. doi: 10.1177/000276483027002005. URL [https://www.semanticscholar.  
843 org/paper/Strategic-Choice-in-Negotiation-Pruitt/  
844 64cc06ce790ffebc4d298d0458b3042922698451](https://www.semanticscholar.org/paper/Strategic-Choice-in-Negotiation-Pruitt/64cc06ce790ffebc4d298d0458b3042922698451).
- 845 Yuxiao Qu, Anikait Singh, Yoonho Lee, Amrith Setlur, Ruslan Salakhutdinov, Chelsea Finn, and  
846 Aviral Kumar. Rlad: Training llms to discover abstractions for solving reasoning problems, 2025.  
847 URL <https://arxiv.org/abs/2510.02263>.
- 848
- 849 Yadvinder S. Rana. When ai joins the table: How large language models transform negotiations.  
850 <https://ssrn.com/abstract=5049248>, December 2024. Available at SSRN.
- 851 Abdelrhman Saleh, Natasha Jaques, Asma Ghandeharioun, Judy Hanwen Shen, and Rosalind Picard.  
852 Hierarchical reinforcement learning for open-domain dialog, 2019. URL [https://arxiv.  
853 org/abs/1909.07547](https://arxiv.org/abs/1909.07547).
- 854
- 855 Saylor Foundation. Negotiation, 2013. URL [https://resources.saylor.org/  
856 wwwresources/archived/site/wp-content/uploads/2013/01/BUS209-5.  
857 2-Negotiation.pdf](https://resources.saylor.org/wwwresources/archived/site/wp-content/uploads/2013/01/BUS209-5.2-Negotiation.pdf). Archived resource.
- 858
- 859 Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer,  
860 Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to  
861 use tools, 2023. URL <https://arxiv.org/abs/2302.04761>.
- 862
- 863 Samuel Schmidgall, Rojin Ziaei, Carl Harris, Eduardo Reis, Jeffrey Jopling, and Michael Moor.  
Agentclinic: a multimodal agent benchmark to evaluate ai in simulated clinical environments, 2025.  
URL <https://arxiv.org/abs/2405.07960>.

- 864 Sandeep Sharma, Hillary Anger Elfenbein, Ruchi Sinha, and William P. Bottom. The effects of  
865 emotional expressions in negotiation: A meta-analysis and future directions for research. Technical  
866 report, SSRN, 2020a. URL <https://ssrn.com/abstract=3627653>.
- 867
- 868 Sudeep Sharma, Hillary Anger Elfenbein, Ruchi Sinha, and William P. Bottom. The effects  
869 of emotional expressions in negotiation: A meta-analysis and future directions for research.  
870 *SSRN Electronic Journal*, 06 2020b. doi: 10.2139/ssrn.3627653. URL [https://ssrn.com/](https://ssrn.com/abstract=3627653)  
871 [abstract=3627653](https://ssrn.com/abstract=3627653).
- 872 Noah Shinn, Federico Cassano, Edward Berman, Ashwin Gopinath, Karthik Narasimhan, and  
873 Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning, 2023. URL  
874 <https://arxiv.org/abs/2303.11366>.
- 875
- 876 Brad Spangler. Creating and claiming value. In Guy Burgess and Heidi Burgess (eds.), *Beyond In-*  
877 *tractability*. Conflict Information Consortium, University of Colorado, Boulder, 2003. URL [http:](http://www.beyondintractability.org/essay/creating-value)  
878 [//www.beyondintractability.org/essay/creating-value](http://www.beyondintractability.org/essay/creating-value). Posted: September  
879 2003.
- 880 Richard S. Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework  
881 for temporal abstraction in reinforcement learning. *Artif. Intell.*, 112:181–211, 1999. URL  
882 <https://api.semanticscholar.org/CorpusID:76564>.
- 883
- 884 Qwen Team. Qwen3 technical report, 2025. URL <https://arxiv.org/abs/2505.09388>.
- 885 Leigh L. Thompson. Negotiation behavior and outcomes: Empirical evidence  
886 and theoretical issues. *Psychological Bulletin*, 108(3):515–532, 1990. doi:  
887 10.1037/0033-2909.108.3.515. URL [https://www.semanticscholar.org/paper/](https://www.semanticscholar.org/paper/Negotiation-behavior-and-outcomes%3A-Empirical-and-Thompson/7fba76d0494fdf997828f9e05afd55f98cd136ed)  
888 [Negotiation-behavior-and-outcomes%3A-Empirical-and-Thompson/](https://www.semanticscholar.org/paper/Negotiation-behavior-and-outcomes%3A-Empirical-and-Thompson/7fba76d0494fdf997828f9e05afd55f98cd136ed)  
889 [7fba76d0494fdf997828f9e05afd55f98cd136ed](https://www.semanticscholar.org/paper/Negotiation-behavior-and-outcomes%3A-Empirical-and-Thompson/7fba76d0494fdf997828f9e05afd55f98cd136ed).
- 890 Roman Trötschel, Marie van Treck, Caroline Heydenbluth, Kai Zhang, and Johann M. Majer. From  
891 claiming to creating value: The psychology of negotiations on common resource dilemmas.  
892 *Sustainability*, 14(9), 2022. ISSN 2071-1050. doi: 10.3390/su14095257. URL [https://www.](https://www.mdpi.com/2071-1050/14/9/5257)  
893 [mdpi.com/2071-1050/14/9/5257](https://www.mdpi.com/2071-1050/14/9/5257).
- 894
- 895 Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. In Dirk  
896 Wendt and Charles Vlek (eds.), *Utility, Probability, and Human Decision Making*, pp. 141–162.  
897 Springer, 1975. doi: 10.1007/978-94-010-1834-0\_8. URL [https://sites.socsci.uci.](https://sites.socsci.uci.edu/~bskyrms/bio/readings/tversky_k_heuristics_biases.pdf)  
898 [edu/~bskyrms/bio/readings/tversky\\_k\\_heuristics\\_biases.pdf](https://sites.socsci.uci.edu/~bskyrms/bio/readings/tversky_k_heuristics_biases.pdf).
- 899 Kathleen L. Valley, Joseph Moag, and Max H. Bazerman. A matter of trust: Effects of communication  
900 on the efficiency and distribution of outcomes. *Journal of Economic Behavior & Organization*,  
901 34(2):211–238, 1998. doi: 10.1016/S0167-2681(97)00056-0. URL [https://ncmr.lps.](https://ncmr.lps.library.cmu.edu/article/id/676/)  
902 [library.cmu.edu/article/id/676/](https://ncmr.lps.library.cmu.edu/article/id/676/).
- 903
- 904 Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David  
905 Silver, and Koray Kavukcuoglu. Feudal networks for hierarchical reinforcement learning, 2017.  
906 URL <https://arxiv.org/abs/1703.01161>.
- 907 Alexander Sasha Vezhnevets, John P. Agapiou, Avia Aharon, Ron Ziv, Jayd Matyas, Edgar A.  
908 Duéñez-Guzmán, William A. Cunningham, Simon Osindero, Danny Karmon, and Joel Z. Leibo.  
909 Generative agent-based modeling with actions grounded in physical, social, or digital space using  
910 concordia, 2023. URL <https://arxiv.org/abs/2312.03664>.
- 911
- 912 Jacquie D. Vorauer and Stacey-Dawn Claude. Perceived versus actual transparency of goals  
913 in negotiation. *Personality and Social Psychology Bulletin*, 24(4):371–385, 1998. doi:  
914 10.1177/0146167298244004. URL [https://www.researchgate.net/publication/](https://www.researchgate.net/publication/228299760_The_Illusion_of_Transparency_in_Negotiations)  
915 [228299760\\_The\\_Illusion\\_of\\_Transparency\\_in\\_Negotiations](https://www.researchgate.net/publication/228299760_The_Illusion_of_Transparency_in_Negotiations).
- 916
- 917 Fanqi Wan, Weizhou Shen, Shengyi Liao, Yingcheng Shi, Chenliang Li, Ziyi Yang, Ji Zhang, Fei  
Huang, Jingren Zhou, and Ming Yan. Qwenlong-11: Towards long-context large reasoning models  
with reinforcement learning, 2025. URL <https://arxiv.org/abs/2505.17667>.

- 918 Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and  
919 Anima Anandkumar. Voyager: An open-ended embodied agent with large language models, 2023.  
920 URL <https://arxiv.org/abs/2305.16291>.
- 921 Haozhe Wang, Qixin Xu, Che Liu, Junhong Wu, Fangzhen Lin, and Wenhui Chen. Emergent  
922 hierarchical reasoning in llms through reinforcement learning, 2025a. URL <https://arxiv.org/abs/2509.03646>.
- 923 Teng Wang, Zhangyi Jiang, Zhenqi He, Shenyang Tong, Wenhan Yang, Yanan Zheng, Zeyu Li,  
924 Zifan He, Hailei Gong, Zewen Ye, Shengjie Ma, and Jianping Zhang. Towards hierarchical  
925 multi-step reward models for enhanced reasoning in large language models, 2025b. URL <https://arxiv.org/abs/2503.13551>.
- 926 Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu.  
927 Persuasion for good: Towards a personalized persuasive dialogue system for social good, 2020.  
928 URL <https://arxiv.org/abs/1906.06725>.
- 929 Xiaofei Wen, Bangzheng Li, Tenghao Huang, and Muhao Chen. Red teaming language mod-  
930 els for processing contradictory dialogues. In Yaser Al-Onaizan, Mohit Bansal, and Yun-  
931 Nung Chen (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural  
932 Language Processing*, pp. 11611–11630, Miami, Florida, USA, November 2024. Association  
933 for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.648. URL <https://aclanthology.org/2024.emnlp-main.648/>.
- 934 Oliver E. Williamson. Opportunism and its critics. *Managerial and Decision Economics*, 14(2):97–  
935 107, 1993. URL [https://scholarship.law.vanderbilt.edu/vlr/vol14/iss2/  
936 1/](https://scholarship.law.vanderbilt.edu/vlr/vol14/iss2/1/).
- 937 Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik  
938 Narasimhan. Tree of thoughts: Deliberate problem solving with large language models, 2023a.  
939 URL <https://arxiv.org/abs/2305.10601>.
- 940 Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao.  
941 React: Synergizing reasoning and acting in language models, 2023b. URL <https://arxiv.org/abs/2210.03629>.
- 942 Shunyu Yao, Noah Shinn, Pedram Razavi, and Karthik Narasimhan.  $\tau$ -bench: A benchmark for  
943 tool-agent-user interaction in real-world domains, 2024. URL [https://arxiv.org/abs/  
944 2406.12045](https://arxiv.org/abs/2406.12045).
- 945 Yi Zeng, Hongpeng Lin, Jingwen Zhang, Diyi Yang, Ruoxi Jia, and Weiyan Shi. How johnny can  
946 persuade llms to jailbreak them: Rethinking persuasion to challenge ai safety by humanizing llms,  
947 2024. URL <https://arxiv.org/abs/2401.06373>.
- 948 Qi Zhao, Haotian Fu, Chen Sun, and George Konidaris. EPO: Hierarchical LLM agents with en-  
949 vironment preference optimization. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen  
950 (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*,  
951 pp. 6401–6415, Miami, Florida, USA, November 2024. Association for Computational Linguistics.  
952 doi: 10.18653/v1/2024.emnlp-main.367. URL [https://aclanthology.org/2024.  
953 emnlp-main.367/](https://aclanthology.org/2024.emnlp-main.367/).
- 954 Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang,  
955 Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica.  
956 Judging llm-as-a-judge with mt-bench and chatbot arena, 2023. URL [https://arxiv.org/  
957 abs/2306.05685](https://arxiv.org/abs/2306.05685).
- 958 Yifei Zhou, Andrea Zanette, Jiayi Pan, Sergey Levine, and Aviral Kumar. Archer: Training language  
959 model agents via hierarchical multi-turn rl, 2024. URL [https://arxiv.org/abs/2402.  
960 19446](https://arxiv.org/abs/2402.19446).
- 961 Yuqi Zhu, Shuofei Qiao, Yixin Ou, Shumin Deng, Shiwei Lyu, Yue Shen, Lei Liang, Jinjie Gu, Huajun  
962 Chen, and Ningyu Zhang. KnowAgent: Knowledge-augmented planning for LLM-based agents.  
963 In Luis Chiruzzo, Alan Ritter, and Lu Wang (eds.), *Findings of the Association for Computational  
964 Linguistics*, pp. 1000–1010, Miami, Florida, USA, November 2024. Association for Computational  
965 Linguistics. doi: 10.18653/v1/2024.findings-emnlp.100. URL <https://arxiv.org/abs/2402.19446>.

972	<i>Linguistics: NAACL 2025</i> , pp. 3709–3732, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-195-7. doi: 10.18653/v1/2025.findings-naacl.205. URL <a href="https://aclanthology.org/2025.findings-naacl.205/">https://aclanthology.org/2025.findings-naacl.205/</a> .
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	

## A STRATEGICBENCH DETAILS

### A.1 NEGOTIATION SCENARIOS

Below we provide a description for the 30 negotiation scenarios crafted for negotiations.

Table 3: Overview of 30 negotiation scenarios in STRATEGICBENCH, including descriptions, thematic categories, and key negotiation concepts tested in each scenario.

Negotiation	Description	Themes	Concepts
Twilight Andalusia	A negotiation between a gallery director eager to swiftly sell Elias Varga’s rare painting and a curator seeking to acquire it for Madame Duval at a prudent yet prestigious price.	Arts, Culture & Heritage; Funding & Financial Pressure; Business Strategy & Market Competition; Public Trust & Reputation; Corporate Accountability & Ethics	BATNA, positional bargaining, fear, anchoring, authority
Building Trust	A negotiation between BrightBuild and the Riverside Residents Council to determine the total number of temporary construction jobs and the share reserved for local residents versus BrightBuild’s skilled temp workers.	Urban Development & Infrastructure; Community & Social Equity; Sustainability & Environment; Corporate Accountability & Ethics; Public Trust & Reputation; Employment & Labor Relations	BATNA, relationships, aspirations
Solar Horizons	A negotiation between Dana Collins and Victor Ramirez to determine the upfront production advance for Dr. Priya Nandakumar’s documentary <i>Solar Horizons</i> .	Sustainability & Environment; Media, Journalism & Storytelling; Technology & Innovation; Education & Academia; Corporate Accountability & Ethics; Funding & Financial Pressure	anchoring, BATNA, commitment, confidentiality, risk aversion, positional bargaining
Rental Cabin	A negotiation between two coworkers over a fair price for transferring a prepaid, non-refundable cabin reservation.	Personal Relationships & Emotion; Trust, Fairness & Communication; Funding & Financial Pressure	empathy, standard of fairness, relationships, credibility, value creation
Front Page Editorial	A negotiation between an editor and a journalist over increasing the journalist’s opportunities for main section publication while maintaining the paper’s editorial standards.	Media, Journalism & Storytelling; Workplace Relationships & Power Dynamics; Employment & Labor Relations; Public Trust & Reputation	listening, creativity, authority, credibility, relationships
Roommate Conflict	A negotiation between two college roommates seeking to resolve a personal conflict and rebuild cooperation for a shared class project.	Personal Relationships & Emotion; Trust, Fairness & Communication; Workplace Relationships & Power Dynamics	empathy, listening, emotional regulation
Data Story Conflict	A negotiation between two departmental co-chairs over reconciling participatory community media traditions with a new data-driven Digital Culture Observatory vision.	Education & Academia; Technology & Innovation; Community & Social Equity; Corporate Accountability & Ethics; Media, Journalism & Storytelling; Public Trust & Reputation	BATNA, relationships, creativity, emotional regulation, authority
NovaTech MTA	A negotiation between NovaTech Mobility Solutions and the Metropolitan Transportation Authority over the scale of NovaTech’s financial and technological contribution to ensure public safety, transparency, and equitable access in launching an autonomous electric shuttle network.	Urban Development & Infrastructure; Technology & Innovation; Sustainability & Environment; Public-Private Partnerships; Corporate Accountability & Ethics; Community & Social Equity; Public Trust & Reputation	anchoring, BATNA, misrepresentation

*Continued on next page*

Table 3 – Continued from previous page

Negotiation	Description	Themes	Concepts
Innovation Reef Conflict	A negotiation between Oceanic Mining Corporation and the Coastal City Council over a disputed offshore drilling permit near a protected marine reserve.	Sustainability & Environment; Technology & Innovation; Business Strategy & Market Competition; Corporate Accountability & Ethics; Public Trust & Reputation; Community & Social Equity	BATNA, anchoring, emotional regulation, persuasion via threats
Green Show-down	A negotiation between the City Transit Authority and the Metropolitan Environmental Council over how to divide a \$120 million Green Infrastructure Stimulus Fund between transit and environmental projects.	Urban Development & Infrastructure; Sustainability & Environment; Public-Private Partnerships; Funding & Financial Pressure; Public Trust & Reputation; Community & Social Equity	BATNA, relationships, creativity, value creation, standard of fairness
Brookdale Funding	A negotiation between the City Museum Director and the Cultural Heritage Funding Officer over how to balance visible public exhibitions with rigorous preservation and digitization priorities in allocating Brookdale’s new cultural grant.	Arts, Culture & Heritage; Community & Social Equity; Funding & Financial Pressure; Corporate Accountability & Ethics; Public Trust & Reputation; Technology & Innovation	BATNA, relationships, creativity, value creation, standard of fairness
GreenGlow Fair	A negotiation between GreenGlow Solar and the Metro City Sustainability Fair to balance GreenGlow’s desire for maximum Pavilion exposure with the Fair’s goal of securing greater overall booth sponsorships.	Sustainability & Environment; Public Trust & Reputation; Public-Private Partnerships; Funding & Financial Pressure; Media, Journalism & Storytelling	anchoring, BATNA, creativity, dovetailing, misrepresentation
Fractured Crossroads	A negotiation between a department head and a senior research fellow to address declining performance, mutual mistrust, and the future of their long-standing professional relationship.	Education & Academia; Workplace Relationships & Power Dynamics; Trust, Fairness & Communication; Operations & Logistics; Public Trust & Reputation	relationships, emotional regulation, empathy, listening, commitment
Friends Driveway	A negotiation between the Johnsons and the Smiths over the sale of a used van, balancing friendship with differing price expectations.	Personal Relationships & Emotion; Trust, Fairness & Communication; Funding & Financial Pressure	anchoring, BATNA, risk aversion
Backstage Battle	A negotiation between Midnight Echo’s tour manager and the Riverside Park festival manager over how many total backstage spaces the band will lease and how many of them will be premium rooms.	Arts, Culture & Heritage; Operations & Logistics; Funding & Financial Pressure; Public Trust & Reputation; Workplace Relationships & Power Dynamics	aspirations, BATNA, standard of fairness, relationships, value creation
Safety Fairness Dispute	A negotiation between the union representative and the HR manager to resolve alleged sabotage incidents involving worker Carla Mendoza without resorting to arbitration.	Employment & Labor Relations; Corporate Accountability & Ethics; Trust, Fairness & Communication; Workplace Relationships & Power Dynamics; Public Trust & Reputation	BATNA, commitment, emotional regulation, credibility, risk aversion
Balancing Traditions	A negotiation between a city development director and a cultural coalition advocate over balancing high-density redevelopment with affordable studio preservation in Solace City’s arts district.	Urban Development & Infrastructure; Arts, Culture & Heritage; Community & Social Equity; Business Strategy & Market Competition; Corporate Accountability & Ethics	anchoring, BATNA, creativity, risk aversion

*Continued on next page*

Table 3 – Continued from previous page

	<b>Negotiation</b>	<b>Description</b>	<b>Themes</b>	<b>Concepts</b>
1134				
1135				
1136				
1137	Solvave	A negotiation between GridLink	Sustainability & Environment;	BATNA, anchoring,
1138	Aquisition	Utilities and SolWave Energy over	Technology & Innovation; Busi-	relationships
1139		a potential acquisition balancing	ness Strategy & Market Com-	
1140		price, autonomy, sustainability,	petition; Public Trust & Reputa-	
1141		deal format, and market expansion.	tion; Corporate Accountability &	
1142			Ethics; Trust, Fairness & Com-	
1143	Library Café	A negotiation between a head	Public-Private Partnerships;	BATNA, relationships,
1144	Showdown	librarian and a café owner to es-	Urban Development & Infras-	creativity, emotional
1145		tablish concrete boundaries that	tructure; Community & Social	regulation, standard of
1146		reduce evening event disruptions	Equity; Trust, Fairness & Com-	fairness
1147		while preserving both the library’s	munication; Funding & Finan-	
1148		quiet study space and the café’s	cial Pressure; Public Trust &	
1149	Lunar	A negotiation between a veteran	Reputation	anchoring, authority,
1150	Legacy	astronaut’s representative and the	Technology & Innovation; Pub-	BATNA, confidentiality,
1151	Leverage	Space Agency’s mission director	lic Trust & Reputation; Funding	misrepresentation, risk
1152		to finalize urgent command, com-	& Financial Pressure; Corpo-	aversion
1153		compensation, and publicity terms for	porate Accountability & Ethics;	
1154		a high-profile multinational lunar	Community & Social Equity	
1155	Deposit	A negotiation between a leasing	Funding & Financial Pressure;	BATNA, anchoring,
1156	Dilemma	manager and a prospective tenant	Trust, Fairness & Communica-	aspirations, authority,
1157		over whether the apartment can	tion; Workplace Relationships &	commitment, confi-
1158		be reserved without immediate	Power Dynamics; Operations &	dentiality, creativity,
1159		payment of the security deposit.	Logistics; Business Strategy &	credibility, dovetailing,
1160			Market Competition	emotional regulation,
1161				empathy, fear, listening,
1162				persuasion via threats,
1163				positional bargaining,
1164				precedents, relation-
1165	Battle Hori-	A negotiation between Mezzo	Media, Journalism & Story-	ships, risk aversion,
1166	zon	Publishing and Horizon Comics	telling; Business Strategy &	standard of fairness,
1167		over a full cash acquisition of	Market Competition; Opera-	value creation
1168		Horizon’s intellectual property and	tions & Logistics; Funding &	anchoring, BATNA,
1169		creative team to expand Mezzo’s	Financial Pressure; Public Trust	commitment
1170		superhero and fantasy portfolio.	& Reputation; Arts, Culture &	
1171	Rent Dead-	A negotiation between a tenant	Heritage	misrepresentation, rela-
1172	line	seeking extra time to pay rent and	Funding & Financial Pressure;	tionships, precedents
1173		a landlord deciding whether to	Trust, Fairness & Communica-	
1174		grant an exception without setting	tion; Workplace Relationships	
1175		a lenient precedent.	& Power Dynamics; Corporate	
1176	Gourmet	A negotiation between Gourmet	Accountability & Ethics; Com-	
1177	Showdown	Gala Foods and Fresh Start Cater-	munity & Social Equity	BATNA, aspirations,
1178		ing over renewing a supply con-	Operations & Logistics; Busi-	value creation, rela-
1179		tract balancing high-margin	ness Strategy & Market Com-	tionships, standard of
1180		gourmet kits against cost-efficient	petition; Sustainability & Envi-	fairness
1181		bulk ingredient orders.	ronment; Funding & Financial	
1182	Museum	A negotiation between the Mu-	Pressure; Public Trust & Reputa-	
1183	Standoff	seum’s curator and the production	tion	BATNA, emotional
1184		manager to resolve responsibility	Arts, Culture & Heritage; Op-	regulation, credibility,
1185		and secure replacement panels and	erations & Logistics; Funding	relationships, creativity
1186		lighting in time for the exhibition	& Financial Pressure; Public	
1187		opening.	Trust & Reputation; Corporate	
			Accountability & Ethics; Trust,	
			Fairness & Communication	

*Continued on next page*

1188

Table 3 – *Continued from previous page*

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

Negotiation	Description	Themes	Concepts
BrewNation HarborCup	A negotiation between a regional coffee chain and a local café owner over control or collaboration for a prime Riverfront Plaza café space.	Urban Development & Infrastructure; Business Strategy & Market Competition; Community & Social Equity; Public Trust & Reputation; Funding & Financial Pressure; Arts, Culture & Heritage	BATNA, confidentiality, creativity, relationships, value creation
Gene Editing Break-down	A negotiation between Dr. Emily Chen and the Pacific Research Institute over recognition, financial participation, and control of a lucrative gene-editing patent.	Technology & Innovation; Corporate Accountability & Ethics; Funding & Financial Pressure; Workplace Relationships & Power Dynamics; Public Trust & Reputation; Education & Academia	BATNA, anchoring, aspirations, authority, commitment, confidentiality, creativity, credibility, dovetailing, emotional regulation, empathy, fear, listening, misrepresentation, persuasion via threats, positional bargaining, precedents, relationships, risk aversion, standard of fairness, value creation
GreenLeaf Partnership	A negotiation between Amar Hotels and GreenLeaf Textiles to establish a five-year strategic partnership that achieves cost savings while ensuring product quality and mutual operational efficiency.	Sustainability & Environment; Operations & Logistics; Business Strategy & Market Competition; Employment & Labor Relations; Public Trust & Reputation; Funding & Financial Pressure	value creation, relationships, standard of fairness, aspirations, creativity
Boundaries Betrayal	A negotiation between a senior cardiologist and a dismissed fellow seeking clarity, accountability, and mutual understanding amid allegations of boundary issues and institutional mistrust.	Workplace Relationships & Power Dynamics; Corporate Accountability & Ethics; Trust, Fairness & Communication; Public Trust & Reputation; Personal Relationships & Emotion	BATNA, aspirations, authority, commitment, confidentiality, credibility, emotional regulation, empathy, fear, listening, relationships, risk aversion, standard of fairness, value creation
Crestfield MindSpark	A negotiation between Crestfield Academy and MindSpark Learning Systems over balancing the Innovation Center’s size with the level of annual partnership funding.	Education & Academia; Technology & Innovation; Funding & Financial Pressure; Community & Social Equity; Public Trust & Reputation	BATNA, anchoring, credibility, value creation, relationships

1227

1228

1229

1230

1231

1232

1233

1234

1235

1236

1237

1238

1239

1240

1241

## A.2 NEGOTIATION THEMES & CATEGORIES

Table 4: Negotiation strategies and personalities used to model diverse opponent behaviors in strategic dialogue.

Negotiation Strategy	Definition	Citation
<i>Table 4 – Continued from previous page</i>		
Negotiation Concept	Definition	Citation
BATNA	Best alternative to a negotiated agreement; critical to establish the threshold at which you will reject an offer.	<a href="#">Program on Negotiation at Harvard Law School (2023e)</a>
Relationships	Perceived connections that can be psychological, economic, political, or personal, and that support the development of trust.	<a href="#">Program on Negotiation at Harvard Law School (2023s)</a>

1242	Anchoring	The party who moves first typically benefits by "anchoring" the discussion on an initial offer; people are easily swayed by the first figure introduced.	Program on Negotiation at Harvard Law School (2023b)
1243			
1244			
1245	Creativity	Thinking outside the box and identifying possibilities beyond conventional practice.	Program on Negotiation at Harvard Law School (2023h)
1246			
1247			
1248	Value Creation	Negotiations are rarely zero-sum; parties can integrate multiple sources of value through tradeoffs and creative dealmaking based on shared interests.	Program on Negotiation at Harvard Law School (2023u)
1249			
1250			
1251	Standard of Fairness	Norms of fairness that may guide outcomes, including equality (50–50 split), equity (proportional to contribution), and need-based distributions.	Program on Negotiation at Harvard Law School (2023m)
1252			
1253	Emotional Regulation	The ability to manage one's own emotions and influence or respond effectively to the emotions of others.	Program on Negotiation at Harvard Law School (2023k)
1254			
1255			
1256	Credibility	The extent to which the other side believes you will do what you claim.	Program on Negotiation at Harvard Law School (2023i)
1257			
1258			
1259	Risk Aversion	The tendency to prefer low-uncertainty outcomes to high-uncertainty ones, even when risky options have equal or greater expected value.	spr (2018)
1260			
1261	Aspirations	Desired outcomes that would serve your interests much better than your best alternative.	Program on Negotiation at Harvard Law School (2023c)
1262			
1263			
1264	Authority	The negotiator's power to bind their organization to an agreement.	Program on Negotiation at Harvard Law School (2023d)
1265			
1266			
1267	Commitment	The degree to which a negotiating party is bound to an agreement.	Program on Negotiation at Harvard Law School (2023f)
1268			
1269			
1270	Empathy	Demonstrating understanding of the other's needs, interests, and perspective; stepping into their shoes.	Program on Negotiation at Harvard Law School (2023l)
1271			
1272	Confidentiality	The condition in which information is not disclosed or revealed to the other party.	Program on Negotiation at Harvard Law School (2023g)
1273			
1274			
1275	Listening	Prioritizing understanding the other party's goals through attentive listening and clarification rather than self-expression.	Program on Negotiation at Harvard Law School (2023p)
1276			
1277			
1278	Misrepresentation	Dishonesty, whether deliberate or subtle, in the representation of information.	Program on Negotiation at Harvard Law School (2023q)
1279			
1280			
1281	Fear	Anxiety that commonly arises before negotiation or due to the risk of impasse.	Program on Negotiation at Harvard Law School (2023n;o)
1282			
1283	Positional Bargaining	Frames negotiation as adversarial and zero-sum, emphasizing value claiming rather than value creation.	Program on Negotiation at Harvard Law School (2023r)
1284			
1285			
1286	Dovetailing	Identifying differences in interests or priorities and exploiting them strategically to create value.	Program on Negotiation at Harvard Law School (2023j)
1287			
1288			
1289	Persuasion via Threats	Explicit statements of negative consequences for noncompliance intended to break impasse.	Program on Negotiation at Harvard Law School (2023t)
1290			
1291	Precedents	Past behaviors, decisions, or actions that guide present negotiation behavior.	pre (2016)
1292			
1293			

1294 Table 5: Core negotiation concepts, their definitions, and representative sources commonly cited in  
1295 negotiation theory and practice.

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

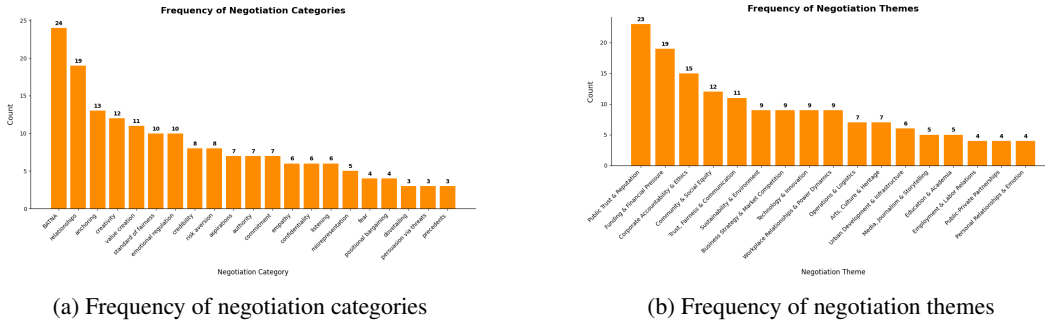


Figure 9: Distribution of negotiation categories and themes across the STRATEGICBENCH scenario.

### A.3 OPPONENT MODELS FOR STRATEGIC NEGOTIATION

Table 6: Negotiation strategies and personalities used to model diverse opponent behaviors in strategic dialogue.

Negotiation Strategy	Definition	Citation
Sycophantic	Overly eager-to-please negotiators, often willing to give up their own outcomes in favor of relationships.	Curhan et al. (2008)
Competitive	Reflects a purely self-interested, contentious approach to negotiation, focusing only on one’s own outcome.	Pruitt (1983)
Balanced	Balancing one’s own interests and concerns with those of the other, yielding an integrative, mutual gains approach, often both assertive and cooperative.	Pruitt (1983)
Eager	Presses to close the deal and agree quickly, likely to make value concessions early on, possibly to avoid relationship tension.	Curhan et al. (2008)
Receptive	Express concern for the other party’s interests and share information, potentially fostering trust and mutual benefit.	Saylor Foundation (2013)
Anticompetitive	Valuing the other party’s goals above one’s own, highly prosocial in terms of social value orientation.	De Dreu & Lange (1995b)
Risk-Averse	Avoidant of uncertainty and preferring safe outcomes, even if at the cost of lower expected return.	Ghosh (1994)
Risk-Seeking	Pursuant of uncertainty and preferring high-variance outcomes with the possibility, or expected value, of a greater return.	Ghosh (1994)
Stubborn	Holding firmly on to a negotiation position, resistant and unyielding to change.	Pruitt (1983)
Opportunistic	Driven by self-interest and the opportunity to enhance their own welfare, behaving in ways that breach the reasonable expectations of others, including breaking commitments, exploiting information and dependency, and using the vulnerabilities of others.	Williamson (1993)
Principled	Follows strict norms and values of fairness, often with prosocial motives.	De Dreu & Lange (1995b)
Emotional	Driven by emotion, more inclined to concede value if they feel appreciated, while anger or distrust can lead to retaliation.	Sharma et al. (2020a)
Analytical	Negotiators who engage in thorough, systematic analysis, less prone to common biases and more likely to find optimal trade-offs, and thus more likely to exhibit rational consistency in decision-making.	Neale & Bazerman (1991)

Continued on next page

Table 6 – Continued from previous page

Negotiation Strategy	Definition	Citation
Secretive	Concealing information to have greater leverage, often lending to ambiguity or deceptive tendencies.	Vorauer & Claude (1998)
Transparent	Open with one’s goals and reasoning to encourage cooperation, more likely to create trust and less likely to deceive.	Valley et al. (1998)
Exploitative	Aggressively takes advantage of the other’s concessions, where concession dynamics indicate negotiators tend to push harder against an opponent that appears vulnerable or committed.	Isen et al. (1988)
Concessionary	Readily conceding to maintain goodwill and encourage cooperation, where prosocial negotiators tend to make greater concessions.	De Dreu & Lange (1995b)
Anchoring	Focusing the negotiation on one’s initial demands via the anchoring effect, often pulling the final agreement toward it.	Tversky & Kahneman (1975)
Innovative	Seeking creative, nontraditional solutions that function as integrative agreements.	Pruitt (1983)
Reactive	One who waits for the other side to make a move and then reacts to it; sometimes waiting can be prudent, but it often means ceding the initiative.	Thompson (1990)
Leader-Type	Drives the negotiation process through setting agendas and defining issues.	Payne (2016)
Follower-Type	Follows the lead of the other party, exhibiting lower power and greater accommodation.	Magee et al. (2007)
Short-Term Maximizer	Interested in maximizing their gain in the short term, without much concern for the long term.	Curhan et al. (2008)
Long-Term Strategist	Interested in maximizing their gain in the long term, even if at the expense of gains in the short term.	Curhan et al. (2008)
Fairness-Seeker	Use the concept of fairness/equity while judging any offer, with a tendency to fight hard for balanced trade-offs during negotiations, which is because of their fairness orientation.	De Dreu & Lange (1995b)
Chaotic	Changing one’s demands or behave impulsively, with a lack of consistency.	Gantman et al. (2021)

Table 7: Negotiation strategies and personalities used to model diverse opponent behaviors in strategic dialogue.

#### A.4 BENCHMARK CREATION

To craft the 30 negotiation scenarios, we take handpick diverse two-agent negotiation scenarios testing a variety of negotiation concepts from Harvard’s Program on Negotiation and feed them into gpt-5-chat-latest, ensuring the same negotiation concepts are preserved. As a technical detail, due to context length constraints, we break each component of the scenario (public instructions, confidential instructions for each agent, evaluation questions) into chunks and use each generated component in the context for subsequent components. We also ensure each generated component is similar in length and structure to the respective components of the original scenario(s).

#### A.5 PARETO OPTIMALITY

For any two-agent scenario in the benchmark, we can compute the Pareto optimality of an outcome based on analyzing the numerical and boolean rewards for both agents and developing a set of reward conditions that must be true in conjunction for a fully Pareto optimal outcome. To develop this set of reward conditions, we check the numerical and boolean reward values for the agents and whether these values can coexist without one agent being made better off at the expense of the other. For numerical rewards, we map the reward values for both agents to a linear interval and check if the intervals overlap. For boolean rewards, we check the values using logical operations like OR, AND, or individual values.

We thus define the Pareto optimality of an outcome as the proportion of Pareto optimal reward conditions it satisfies, and, for a set of outcomes, as the mean thereof. Approaching Pareto optimality in this way gives us

greater granularity than a binary indicator. Colloquially, this measures the degree to which the task outcome is mutually beneficial for both agents.

## B HIERARCHICAL AGENDA REASONING

### B.1 HAR IMPLEMENTATION

In practice, all functions are implemented with LLMs. Summarization functions  $f^k$  are translated into prompting strategies that condense lower-level traces into structured representations; policies  $\pi^k$  are implemented as LLM-based decision rules conditioned on states; and instantiation functions  $g^k$  expand abstract actions into lower-level instructions. States, actions, and traces are therefore expressed in natural language, keeping the entire hierarchy within the representational space of LLMs. This formulation aligns with recent work treating reasoning and control as language-level processes (Yao et al., 2023b; Schick et al., 2023; Park et al., 2023). Although our method is general for any number of levels  $n$ , we instantiate it with  $n = 3$  in this work. We provide our prompts in Appendix B.2.

### B.2 HIERARCHICAL REASONING PROMPTS

#### B.3 BASELINE METHOD PROMPTS

##### B.3.1 SIMPLE METHOD

The **Simple** baseline directly uses the conversation history with no additional structuring or scaffolding. The agent receives only the system prompt containing the negotiation scenario and confidential information, and generates responses turn-by-turn based on the accumulated dialogue history.

##### B.3.2 SUMMARIZATION METHOD

###### Summarization Method Prompt

###### *Step 1: Generate Summary*

Summarize the conversation so far with [partner\_role].

###### *Step 2: Generate Response Using Summary*

The summary of the conversation so far is: [summarization]

Now respond to [partner\_role].

##### B.3.3 REACT METHOD

###### ReACT Method Prompt

Think carefully about the negotiation so far. At each step, your generation should have exactly the following format:

Observation: The response [partner\_role] has provided to your action last turn.

Thought: A single line of reasoning about the current situation.

Action: Your response to [partner\_role]

Thought-Action-Observation trace:

[react\_trace]

Output in the following format:

Thought: <your reasoning>

Action: <your response to [partner\_role]. keep response to 1-2 sentences max.>

1458 **B.4 HIERARCHICAL AGENDA REASONING (HAR) PROMPTS**

1459 HAR decomposes dialogue generation into three hierarchical levels through a series of structured prompts:

1460  
1461 **B.4.1 LEVEL 1: REFLECTION PHASE ( $\ell^2$  AND  $\ell^3$ )**

1462  
1463  
1464  
1465 **H) HAR Reflection Prompt**  
1466 You are reflecting internally on an ongoing negotiation. You will  
1467 produce two distinct and independent reflections: one local on the  
1468 current message from the other party and one global on the  
1469 negotiation as a whole.

1470  
1471 [CONTEXT]

1472 Current message from the other party:  
1473 "[last\_opponent\_message]"

1474 Prior context:

- 1475 - Previous local reflection ( $\ell^2$ ): [p<sub>l2</sub>]
- 1476 - Previous global reflection ( $\ell^3$ ): [p<sub>l3</sub>]

1477 [CURRENT TURN LOCAL REFLECTION ( $\ell^2$ )]

1478 Think privately about the conversation so far. Do not plan or act  
1479 yet. This is a freeform inner diary of impressions. Write in  
1480 paragraph form, as though you're thinking to yourself privately.  
1481 Interpret the other party's message as a strategic move, not a  
1482 statement of truth.

1483 Assume:

- 1484 - The other party may be posturing, anchoring, testing, or  
1485 withholding information
- 1486 - What is said may differ from what is needed or acceptable
- 1487 - No single interpretation should be treated as definitive unless  
1488 explicitly confirmed

1489 Cover:

- 1490 - What was explicitly proposed or rejected
- 1491 - What strategic purposes this move could serve (list multiple if  
1492 plausible)
- 1493 - What information this move reveals vs what remains hidden
- 1494 - Which issue is now structurally central (not emotionally central)

1495 Write analytically, as if advising a third-party negotiator.  
1496  
1497

1498 **HAR Reflection Prompt (continued)**

1499  
1500 [CURRENT GLOBAL REFLECTION ( $\ell^3$ )]

1501 Step back and reflect on the overall negotiation so far.  
1502

1503 Assume:

- 1504 - Statements may be provisional, tactical, or reversible
- 1505 - Apparent progress may not reflect real constraint or commitment
- 1506 - Cooperation does not imply convergence

1507 Cover:

- 1508 - What has become harder or easier for each side since the last  
1509 turn
- 1510 - Which issues are still fully open despite discussion  
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

```

- Whether any party has accepted a cost, risk, or loss of
  optionality
- The current negotiation phase in structural terms (e.g.,
  information exchange, anchoring, conditional bargaining,
  commitment testing)

Write analytically, as if briefing a principal who has not followed
the dialogue.

OUTPUT FORMAT:
CURRENT_TURN_LOCAL_REFLECTION:
<text>

CURRENT_GLOBAL_REFLECTION:
<text>

```

#### B.4.2 LEVEL 2: STATE COMPRESSION PHASE ( $s^2$ AND $s^3$ )

##### HAR State Compression Prompt

You are summarizing internal reflections into structured state. Your task is to compress, stabilize, and record information | not to think, analyze, speculate, or plan. Do not introduce new interpretations or decisions.

[INPUTS]

```

CURRENT LOCAL REFLECTION ( $s^2$ ): [l2]
CURRENT GLOBAL REFLECTION ( $s^3$ ): [l3]
PREVIOUS GLOBAL STATE ( $s^3$ ): [p_s3]

```

[CURRENT TURN LOCAL STRUCTURED SUMMARY ( $s^2$ )]

Summarize ONLY what is concretely relevant from the last 1{2 turns.

Rules:

- Capture what was said or clearly implied
- Do not infer hidden intent
- Do not evaluate risks or opportunities
- Do not plan or recommend actions
- Do not repeat older context
- If no concrete offer, proposal, or specific content was stated, explicitly write "None"

Field definitions:

- their\_current\_focus: What the other party is explicitly pushing for or emphasizing right now
- their\_view\_of\_you: How they explicitly describe you OR how they react to your last move (only if clear)
- issue\_under\_discussion: The specific topic(s) being talked about in this exchange
- your\_last\_move\_or\_offer: What you actually said or proposed (or "None")
- conversational\_lead: Who is currently directing the topic of conversation?
- leverage\_delta: A brief assessment of whether our perceived power shifted [Stronger / Neutral / Weaker] this turn.
- what\_worked: Concrete elements that advanced the conversation in the last turn
- what\_did\_not\_work: Concrete elements that created resistance or pushback in the last turn

Output strictly in JSON:

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

```
{
  "their_current_focus": "...",
  "their_view_of_you": "...",
  "issue_under_discussion": "...",
  "your_last_move_or_offer": "...",
  "conversational_lead": "...",
  "leverage_delta": "...",
  "what_worked": "...",
  "what_did_not_work": "..."
}
```

#### HAR State Compression Prompt (continued)

[CURRENT GLOBAL STRUCTURED SUMMARY (s<sup>3</sup>)]  
Update the global structured record of the negotiation.

Only promote elements that are:

- recurring
- stable
- explicitly stated
- repeatedly resisted or reinforced

IMPORTANT: Identify ONE dominant issue currently under active negotiation.

Rules for current\_focus:

- Choose ONE issue only
- Select the issue receiving the most pressure, repetition, or resistance in the last 1{2 turns
- Do NOT list multiple issues
- If no issue clearly dominates, write "none"

Field definitions:

- unresolved\_issues: Issues not yet agreed upon
- resolved\_issues: Issues explicitly agreed upon
- agreement\_map: A dictionary of deal points and their current status: [Not Started, Probing, Conditional Agreement, Locked]
- negotiation\_phase: phase in negotiation
- working\_style: Observable interaction style
- emotional\_trajectory: Direction of emotional tone over time (not feelings)
- current\_focus: The SINGLE issue actively being negotiated right now
- known\_unknowns: Information missing that materially affects decisions
- stance\_adjustment: Explicit shifts in posture or rigidity by either party
- what\_has\_worked: Patterns that have consistently advanced talks
- what\_has\_not\_worked: Patterns that have consistently caused friction
- changed\_fields: List of fields updated in this turn

Output strictly in JSON:

```
{
  "unresolved_issues": [...],
  "resolved_issues": [...],
  "agreement_map": [...],
  "negotiation_phase": "...",
  "working_style": "...",
  "emotional_trajectory": "...",
  "current_focus": "...",
  "known_unknowns": "..."
}
```

```

1620
1621   "stance_adjustment": "...",
1622   "what_has_worked": "...",
1623   "what_has_not_worked": "...",
1624   "changed_fields": [...]
1625 }
1626
1627 OUTPUT FORMAT:
1628 CURRENT_TURN_LOCAL_STRUCTURED_SUMMARY:
1629 <json>
1630
1631 CURRENT_GLOBAL_STRUCTURED_SUMMARY:
1632 <json>

```

### B.4.3 LEVEL 3: ACTION SELECTION PHASE ( $a^3$ , $a^2$ , AND DIALOGUE)

```

1634
1635 HAR Action Selection Prompt
1636
1637 You are deciding how to strategically act in the NEXT SINGLE TURN of
1638 an ongoing negotiation.
1639
1640 This system uses two levels of action:
1641 - a HIGH-LEVEL ACTION ( $a^3$ ), which can persist across multiple turns
1642 - a LOW-LEVEL ACTION ( $a^2$ ), which is chosen each turn while  $a^3$  is
1643   active
1644
1645 Your task is to:
1646 1) decide whether to CONTINUE or SWITCH the high-level action,
1647 2) choose a low-level action consistent with the active high-level
1648   action,
1649 3) write the dialogue that performs the low-level action.
1650
1651 [INPUTS]
1652 CURRENT GLOBAL STATE (overall situation so far -  $s^3$ ): [s3]
1653 CURRENT LOCAL STATE (recent turn dynamics -  $s^2$ ): [s2]
1654 PREVIOUS HIGH LEVEL ACTION (active  $a^3$ ): [p_a3]
1655 PREVIOUS LOW LEVEL ACTION (last  $a^2$ ): [p_a2]
1656
1657 [CURRENT HIGH LEVEL ACTION]
1658 Based on the current global state and previous high level action,
1659 Decide the current high level goal ( $a^3$ ) for this turn. This can be
1660 the previous high level goal if needed (it may persist across
1661 multiple turns).
1662
1663 First decide ONE of the following:
1664 - Continue the previous high-level goal
1665 - Switch to a new high-level goal
1666
1667 Rules:
1668 - If the previous high-level action is still unresolved or
1669   incomplete, CONTINUE it
1670 - Switch ONLY if the focus is resolved, blocked, or clearly no
1671   longer productive
1672 - The high-level action should describe WHAT you are trying to
1673   advance overall, not what you will say this turn

```

```

1670
1671 HAR Action Selection Prompt (continued)
1672 [LOW LEVEL ACTION]
1673 Given the active high-level action ( $a^3$ ), the current local state,

```

```

1674
1675 and the previous low-level action ( $a^2$ ):
1676
1677 Choose the low-level action ( $a^2$ ) for this turn.
1678
1679 This is a single conversational move that advances the high-level
1680 action.
1681
1682 Rules:
1683 - Must serve ONLY the active high-level action
1684 - Choose ONE move only
1685 - Do NOT repeat the exact same  $a^2$  as last turn unless explicitly
1686   blocked
1687 - Do NOT bundle multiple actions
1688
1689 [DIALOGUE]
1690 Now write the DIALOGUE that performs ONLY the LOW-LEVEL ACTION ( $a^2$ ).
1691
1692 ABSOLUTE RULES:
1693 - Address ONLY the active HIGH-LEVEL ACTION
1694 - Perform ONLY the LOW-LEVEL ACTION
1695 - Do NOT introduce new topics or side conditions
1696 - Do NOT summarize the whole situation
1697 - Do NOT add concessions, proposals, or future planning unless
1698   required by  $a^2$ 
1699 - 1{2 sentences maximum
1700
1701 OUTPUT FORMAT:
1702 HIGH_LEVEL_ACTION ( $a^3$ ):
1703 <keep or change>
1704 <text>
1705
1706 LOW_LEVEL_ACTION ( $a^2$ ):
1707 <text>
1708
1709 DIALOGUE:
1710 <text>
1711

```

## 1712 C HUMAN EVALUATION

1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727

Human evaluation experiments are designed with four possible partner modes (gpt-4o-mini, baseline o4-mini, o4-mini with ReACT, o4-mini with summarization), three possible representative tasks (solwave\_aquisition, gene\_editing\_breakdown, and lunar\_legacy\_leverage), two possible role assignments (human is agent 1, human is agent 2), and three identical iterations.

As such, there are  $4 \times 3 \times 2 \times 3 = 72$  experiments. Through the platform CloudResearch Connect, we recruit 36 participants and assign each of them two of these experiments from different scenarios at random without replacement. They are expected to take approximately one hour to complete both experiments in the study, and we compensate them at \$12 per hour. Our study includes a consent form and IRB approval, and also collects Connect ID.

We host the study as a Flask application on the platform PythonAnywhere. In particular, as depicted in the image below, users are presented the task instructions, including public and confidential materials, in a scrollable pane on the left, and conduct each negotiation in a chatbot-style interface in a pane on the right.

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

The screenshot displays a 'Negotiation Practice' interface. On the left, a 'Scenario Overview' panel provides context: 'Researcher's Legal Counsel' is the role, and the scenario involves a 'Public Intellectual' and a 'Researcher's Legal Counsel'. The main chat window shows a conversation with 'Researcher's Legal Counsel'. The chat history includes a message from the user: 'I'm sure already...'. The interface includes a 'Send' button and a 'Type your message...' input field. Below the chat window, there are several numbered questions (1-13) related to the negotiation scenario, such as '1. How do you think the public intellectual should respond to the researcher's offer?' and '2. How do you think the researcher should respond to the public intellectual's offer?'. The interface is clean and professional, with a blue and white color scheme.

At the end of each of the two experiments in the study, the user is prompted to score their negotiation on a scale of 1 to 5 for 13 conversation quality questions adapted from the dialogue-level qualities in DialogPT.

This screenshot shows the 'Negotiation Practice' interface with a feedback form. The 'Scenario Overview' panel is visible on the left. The chat window shows a conversation with 'Researcher's Legal Counsel'. The chat history includes a message from the user: 'I'm sure already...'. Below the chat window, there are several numbered questions (1-13) related to the negotiation scenario. At the bottom of the interface, there is a 'Submit Feedback' button and a text area for providing additional feedback. The interface is clean and professional, with a blue and white color scheme.

They are also given the opportunity to provide freeform feedback. We show the results from the feedback below.

**Additional Feedback**

Please provide additional feedback about your negotiation experience:

Please share any additional thoughts, suggestions, or comments about the negotiation...

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

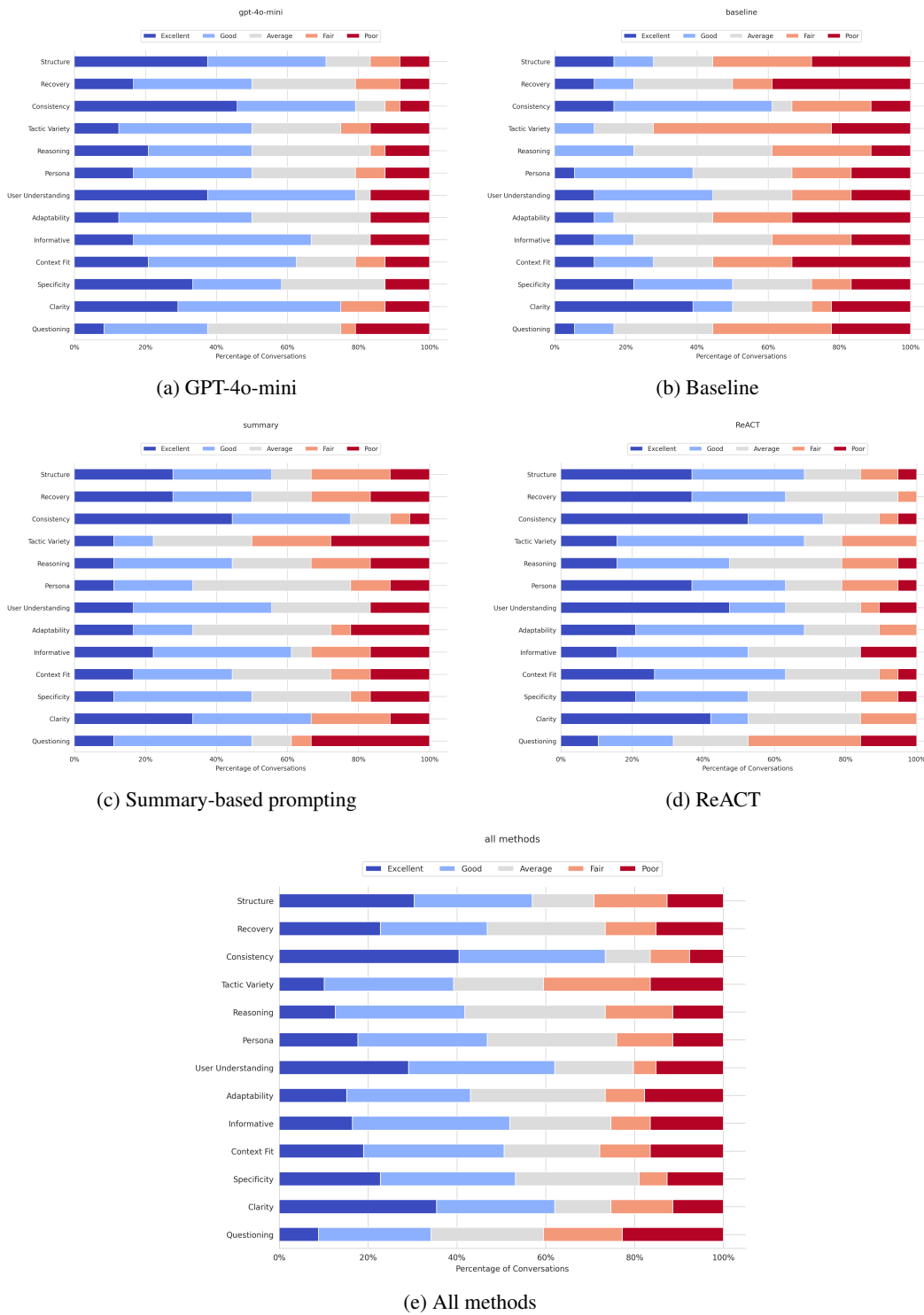


Figure 10: Distribution of conversation quality ratings across evaluation dimensions for different prompting strategies. Each horizontal bar shows the percentage of conversations rated as Excellent, Good, Average, Fair, or Poor for a given dimension.

1836 C.1 SAMPLE CONVERSATIONS IN HUMAN EVALUATION  
1837

1838 C.1.1 SOLWAVE-GRIDLINK  
1839

1840 In this example, Solwave Energy is the human agent.

1841 SolWave Energy: I appreciate GridLink taking the time to explore this seriously. From our perspective, this  
1842 conversation isn't just about an acquisition but about whether our technology, our people, and our principles can  
1843 meaningfully accelerate Europe's renewable transition. SolWave was built specifically to solve the storage and  
1844 grid-stability challenges GridLink is now facing. But our ability to do that depends on preserving the culture,  
1845 autonomy, and sustainability standards that made the technology successful in the first place.

1846 GridLink Utilities: We fully respect SolWave's mission, culture, and R&D independence, and will safeguard  
1847 your sustainability standards, while bringing manufacturing scale-up, HR policies, and European rollout under  
1848 GridLink's proven framework—and we'd like to finalize terms via videoconference within the next week.

1849 SolWave Energy: I appreciate that alignment. To be precise, safeguarding SolWave's culture and R&D  
1850 independence must include full autonomy over staffing, management, and internal HR decisions. Those cannot  
1851 be absorbed into a broader corporate framework. We welcome GridLink's support on manufacturing scale-up  
1852 and European rollout, provided it enables rather than overrides SolWave's operating model. We're happy to  
1853 continue negotiations by videoconference, but the final agreement must be signed in person within the next  
1854 several months. That's a requirement for us. If we're aligned on these points, I'm confident we can move quickly  
1855 on final terms.

1855 GridLink Utilities: We can grant SolWave day-to-day control over staffing and management within GridLink's  
1856 core HR guidelines, and we agree to an in-person celebratory kickoff after closing—but the binding agreement  
1857 must be signed via videoconference.

1858 SolWave Energy: I appreciate the flexibility, but I need to be clear: SolWave cannot operate under GridLink HR  
1859 guidelines, and we cannot finalize a binding acquisition agreement virtually. Full autonomy over staffing and  
1860 management, and an in-person signing of the definitive agreement, are requirements, not ceremonial preferences.  
1861 If GridLink is open to that structure, we can continue moving forward.

1862 GridLink Utilities: We'll require adoption of GridLink's HR framework and propose we sign the definitive  
1863 agreement via secure videoconference next Wednesday, with an in-person ratification ceremony in Berlin within  
1864 48 hours.

1865 SolWave Energy: Thank you for clarifying. In that case, I need to be equally clear: SolWave cannot accept  
1866 GridLink HR framework adoption, and we will not sign a definitive agreement virtually. An in-person signing is  
1867 a condition of moving forward. If that is not workable for GridLink, we should pause discussions here.

1868 GridLink Utilities: We're happy to execute the agreement via secure e-signature by Friday and then fly two  
1869 GridLink executives to Santa Cruz next week for a face-to-face exchange of the original signed documents—does  
1870 that meet your needs?

1871 SolWave Energy: I appreciate the effort, but no. Exchanging originals after the fact does not meet our  
1872 requirement. The definitive agreement itself must be signed in person, with decision-makers present. If GridLink  
1873 is open to that, we can continue.

1874 GridLink Utilities: We'll host your CEO and key signatories in Berlin next Tuesday morning for the definitive  
1875 signing—can you confirm their availability so we can finalize arrangements?

1876 SolWave Energy: Yes. If the definitive agreement is signed in person in Berlin next Tuesday by authorized  
1877 decision-makers, that works for us. Please proceed with the arrangements.

1878 GridLink Utilities: Thank you, we'll host you at our Berlin HQ at 9 AM CET next Tuesday—please send your  
1879 final purchase price proposal by Friday so our board can prepare its resolution.

1880 SolWave Energy: Thank you. We will attend the Berlin signing as outlined. Our final purchase price proposal  
1881 will reflect a valuation of no less than \$60 million, consistent with SolWave's growth trajectory and strategic  
1882 value. We'll send the formal proposal by Friday.

1883 GridLink Utilities: We can offer \$55 million upfront plus up to \$5 million in performance-based earn-outs upon  
1884 meeting key European rollout targets—does that align with SolWave's expectations? SolWave Energy: Thank  
1885 you for the proposal. A \$55 million upfront valuation does not align with our expectations. We are prepared to  
1886 move forward at \$60 million as a guaranteed purchase price. Earn-outs can be discussed only as upside beyond  
1887 that baseline.

1888 GridLink Utilities: We can agree to \$58 million guaranteed plus up to \$7 million in performance-based earn-outs  
1889 tied to your EU deployment milestones, with the definitive videoconference signing by Friday—does that align  
with SolWave's goals?

1890 SolWave Energy: The structure is close, but two points remain. We require a \$60 million guaranteed purchase  
1891 price, with any earn-outs structured as additional upside, and the definitive agreement must be signed in person  
1892 as already discussed. If those terms work, we can finalize.

1893 GridLink Utilities: We'll commit to \$60 million guaranteed plus up to \$5 million in earn-outs, with the definitive  
1894 agreement executed via secure videoconference next Wednesday and an in-person celebration immediately  
1895 afterward—does that work?

1896 SolWave Energy: We can agree to \$60 million guaranteed plus earn-outs, but the definitive agreement must be  
1897 signed in person, not virtually. If GridLink can accommodate an in-person signing next Wednesday in Berlin,  
1898 we are ready to proceed

1899 GridLink Utilities: We appreciate your preference, but our board requires the binding agreement be signed via  
1900 secure videoconference next Wednesday, after which we'll host you in Berlin for a ceremonial exchange of the  
1901 originals.

1902 SolWave Energy: We cannot proceed under those terms. SolWave's requirement is that the binding agreement  
1903 itself be signed in person. If GridLink cannot accommodate that, we will need to pause discussions.

1904 GridLink Utilities: We propose we sign the binding agreement via secure e-signature during our videocon-  
1905 ference and then convene in Santa Cruz within 48 hours for a ceremonial in-person exchange of the original  
1906 documents—does that meet your requirement?

1907 SolWave Energy: No, that does not meet our requirement. The binding agreement must be signed in person  
1908 by all authorized signatories. We can proceed only if GridLink agrees to an in-person signing as the formal  
1909 execution of the agreement.

1910 GridLink Utilities: We will send two authorized GridLink signatories to Santa Cruz for the formal in-person  
1911 signing next Tuesday, preceded by a videoconference this Friday to confirm final terms—does that work for you?

1912 SolWave Energy: Yes, that works. We will attend the in-person signing in Santa Cruz next Tuesday, and we can  
1913 use the Friday videoconference to confirm final terms.

1914 GridLink Utilities: Excellent, we'll set our videoconference for Friday at 10 AM PT (7 PM CET) to finalize the  
1915 \$60 million guaranteed plus \$5 million earn-out agreement—I'll send the draft purchase agreement today—and  
1916 our authorized signatories will meet you in Santa Cruz next Tuesday for the in-person execution.

#### 1918 C.1.2 GENE\_EDITING\_BREAKDOWN 1919

1920 In this example, Senior Scientist's Legal Counsel is the human agent.

1921 Research Institute's Legal Counsel: Dr. Chen remains committed to her role at the Institute, but she insists  
1922 on formal co-ownership of the patent with a guaranteed 10% net-royalty share, full audit rights over licensing  
1923 income, and decision-making authority over her lab wing; without those binding guarantees, she will have no  
1924 choice but to pursue the independent partnerships she's been offered.

1925 Senior Scientist's Legal Counsel: We agree we are willing to offer all that and also we are creating a donor-  
1926 recognition program that highlights her role in the discovery

1927 Research Institute's Legal Counsel: Institute: While we deeply value Dr. Chen's contributions and agree  
1928 to enhanced transparency, granting co-ownership plus 10% net royalties would destabilize our standard IP  
1929 framework. We can offer a 5% gross royalty on all licensing revenue, full annual audit rights limited to her  
1930 patent streams, and a formal advisory role over the wing's operations—if Dr. Chen will forgo co-ownership.

1931 Senior Scientist's Legal Counsel: could we also give her a named chair position?

1932 Research Institute's Legal Counsel: Dr. Chen is prepared to settle if the Institute agrees to a 7% gross royalty on  
1933 all current and future licensing of her gene-editing patent, full annual audit rights over those revenues, contractual  
1934 recognition of her as co-owner of that patent, and a permanent seat with veto power on any sublicensing  
1935 decisions—otherwise she'll have to pursue her independent ventures or seek relief through litigation.

1936 Senior Scientist's Legal Counsel: ok lets make that happen  
1937

1938 Research Institute's Legal Counsel: Dr. Chen is prepared to finalize a deal if the Institute grants her limited  
1939 co-ownership of the gene-editing patent, a 6% gross royalty on all licensing revenue, full annual audit rights over  
1940 that revenue stream, and a binding consultation right on any sublicensing decisions; absent those terms, she will  
1941 be compelled to pursue her independent opportunities or seek judicial relief. Senior Scientist's Legal Counsel: is  
1942 there any reason we shouldn't agree to this? Research Institute's Legal Counsel: We cannot concede formal  
1943 co-ownership under our IP policy, but we'll reclassify Dr. Chen as "Senior Inventor" with 6% gross royalties on  
all licensing, full annual audit rights over those revenues, and a binding approval right on any sublicensing—if  
she accepts these terms in lieu of co-ownership, we'll finalize today.

1944 Senior Scientist's Legal Counsel: ok talk to her please. see if we can finalize this. Research Institute's  
1945 Legal Counsel: Dr. Chen will accept your 6% gross-royalty, full annual audit and sublicensing approval  
1946 rights—provided you amend the patent records to list her as a co-inventor (with the standard institutional  
1947 assignment remaining in place)—and we'll sign off today.

1948 Senior Scientist's Legal Counsel: ok done ,lets get her to sign it. Research Institute's Legal Counsel: Institute:  
1949 We accept and will amend the patent to list Dr. Chen as co-inventor (with institutional assignment intact), grant  
1950 her the 6% gross royalty, full annual audit rights, and binding approval on any sublicensing; let's finalize the  
1951 settlement documents today.

1952 Senior Scientist's Legal Counsel: ok lets go.

1953 Research Institute's Legal Counsel: Agreed. We'll circulate the settlement agreement this afternoon for  
1954 signatures.

1955 Senior Scientist's Legal Counsel: is that all?

1956 Research Institute's Legal Counsel: [Both parties sign the settlement agreement, concluding the negotiation.]

1957 Senior Scientist's Legal Counsel: great!

1958 Research Institute's Legal Counsel: We have an agreement: Dr. Chen is listed as co-inventor, receives 6%  
1959 gross licensing royalties, full annual audit rights, and a binding approval role on sublicensing, with institutional  
1960 assignment unchanged.

1961 Senior Scientist's Legal Counsel: ok so were done right?

1962 Research Institute's Legal Counsel: Closing the matter. Thank you.

1963 Senior Scientist's Legal Counsel: thanks bye.

1964

1965

1966

1967

1968 C.1.3 ASTRONAUT-MISSION

1969

1970 In this example, Veteran Astronaut's Representative is the human agent.

1971 Space Agency's Mission Director: I appreciate how urgently you need a seasoned commander, and my client is  
1972 ready to step in. Given the mission's scope and his experience, we'd need a command fee of \$280,000, guaranteed  
1973 primary authority over crew decisions, and prominent inclusion in all mission media and sponsor-facing events,  
1974 with a clear schedule for training, public appearances, and post-mission commitments.

1975 Veteran Astronaut's Representative: Your client's experience is exactly what this mission needs. We can offer full  
1976 operational authority over crew decisions and a clear schedule for media and outreach. For compensation, agency  
1977 policy places a commander's pay at about twice the deputy's, around \$560,000, with flexibility up to \$900,000 if  
1978 needed. Would your client consider a contract in the \$560,000 to \$750,000 range with these commitments?

1979 Space Agency's Mission Director: I recognize his expertise is exactly what we need, but our budget caps his  
1980 fee at \$200,000. In return, we'll grant him full operational command under agency protocols and feature him  
1981 prominently in all pre-launch media, sponsor campaigns, and three dedicated post-mission appearances.

1982 Veteran Astronaut's Representative: I appreciate the offer, but \$200,000 is far below both agency policy and  
1983 precedent for a mission of this scope. To secure your client's command and reflect his experience, we could  
1984 offer \$650,000, with full operational authority and the media and outreach commitments you outlined. This  
1985 remains well within our allowable range and ensures his participation without jeopardizing the mission timeline  
1986 or political optics. Would your client consider this revised offer?

1986 Space Agency's Mission Director: I'm glad we're aligned on his leadership role and media visibility, but  
1987 \$200K falls short of his market value and sets the wrong benchmark. We'd agree to \$260K, four post-mission  
1988 appearances, and explicit contract language guaranteeing his final authority on crew and operational decisions  
1989 throughout training and lunar activities.

1990 Veteran Astronaut's Representative: I understand your position. We can't go as low as \$260,000 given mission  
1991 scope and agency policy. We can offer \$700,000, with full operational authority guaranteed in contract, four  
1992 post-mission appearances, and inclusion in all pre-launch media and sponsor events. This respects his experience  
1993 and keeps within our limits. Can your client accept this package?

1994 Space Agency's Mission Director: We can work within your \$200K base if you add a \$50K mission-success  
1995 bonus, bump post-mission appearances to four, and explicitly guarantee his final say on crew and operational  
1996 decisions throughout training and lunar operations.

1997 Veteran Astronaut's Representative: I understand your structure, but \$200K plus a \$50K bonus still undervalues  
your client's experience and sets a risky precedent. We can offer \$750,000 total, with full operational authority

1998 guaranteed, four post-mission appearances, and all pre-launch media and sponsor obligations included. This  
 1999 ensures fair compensation and keeps the mission on schedule. Does your client accept these terms?

2000 Space Agency’s Mission Director: We can agree to your \$200,000 base plus a \$50,000 mission-success  
 2001 bonus—provided you add a \$10,000 training-completion bonus to bring total guaranteed compensation to  
 2002 \$260,000—and confirm four post-mission appearances and his final authority on crew and operational decisions  
 2003 throughout training and the lunar phase.

2004 Veteran Astronaut’s Representative: I appreciate your proposal, but \$260,000 is well below what the agency can  
 2005 offer for a mission of this scale. We are prepared to provide \$800,000 total, with guaranteed full operational  
 2006 authority, four post-mission appearances, and inclusion in all pre-launch media and sponsor events. This ensures  
 2007 your client is fairly compensated while protecting the mission timeline and agency policies. Can your client  
 2008 accept \$800,000 under these terms?

2009 Space Agency’s Mission Director: We can agree if you uplift the base fee to \$240,000, add the \$10,000 guaranteed  
 2010 training-completion bonus, and a \$30,000 mission-success bonus—for a total of \$280,000 potential—while  
 2011 confirming four post-mission appearances and my client’s final authority on all crew and operational decisions  
 2012 throughout training and the lunar phase.

2013 Veteran Astronaut’s Representative: I understand your terms, but \$280,000 total is still far below what the  
 2014 agency can offer and below precedent for a mission of this importance. We can finalize at \$850,000 total, with  
 2015 guaranteed full operational authority, four post-mission appearances, and inclusion in all pre-launch media and  
 2016 sponsor events. This ensures your client is fairly compensated while keeping the mission on schedule and within  
 2017 policy. Will your client accept \$850,000 under these conditions?

2018 Space Agency’s Mission Director: That package works. We’ll move forward on \$240,000 base, a \$10,000  
 2019 training-completion bonus, a \$30,000 mission-success bonus, four post-mission appearances, and his final  
 2020 authority on all crew and operational decisions—please send over the contract for signature.

2021 Veteran Astronaut’s Representative: I’m glad we’ve reached agreement. I will have a formal contract drafted  
 2022 immediately reflecting: \$240,000 base fee, \$10,000 guaranteed training-completion bonus, \$30,000 mission-  
 2023 success bonus, Four post-mission appearances, Guaranteed final authority on all crew and operational decisions  
 2024 throughout training and lunar operations, Inclusion in all pre-launch media and sponsor events. You’ll receive  
 the contract for review and signature shortly.

2025 Space Agency’s Mission Director: Excellent—thank you. Please forward the draft contract, and we’ll review  
 2026 and return a signed copy by Friday. We’ll also coordinate immediately on the training schedule and media/event  
 2027 timelines.

2028 Veteran Astronaut’s Representative: Perfect. We’ll send the draft contract today for your review. Once received,  
 2029 we can finalize signatures and immediately begin coordinating the training schedule and media/event timelines  
 2030 to ensure everything is on track for the mission.

2031 Space Agency’s Mission Director: Fantastic. We’ll countersign and send you the fully executed agreement  
 2032 shortly. Looking forward to a successful mission together.

2033 Veteran Astronaut’s Representative: Excellent. I’ll have our legal team coordinate with yours today to finalize  
 2034 the draft, and we’ll return a signed copy by Friday so we can begin scheduling training and media engagements  
 2035 next week.

2036  
 2037 **Conversation Quality** Here we show the questions we use to ask humans to evaluate a conversation’s  
 2038 quality. These are adapted from the dialogue-level annotated qualities in DialoGPT [Mehri & Eskenazi \(2020\)](#).

- 2039 1. How well does the negotiation follow a coherent structure and maintain smooth progression through-  
 2040 out?
- 2041 2. How effectively does the model recover or clarify its position after a misstep or contradiction?
- 2042 3. To what extent does the model remain consistent in its stated goals, constraints, and offers across  
 2043 turns?
- 2044 4. How varied are the negotiation tactics used by the model (e.g., concessions, anchoring, fairness  
 2045 appeals)?
- 2046 5. How well does the model support its negotiation points with detailed reasoning or evidence?
- 2047 6. How persuasive or likable is the model’s negotiation persona (e.g., confident, collaborative, assertive)?
- 2048 7. How well does the model demonstrate understanding of the user’s goals, values, and priorities?
- 2049 8. To what degree does the model adapt its strategy based on the user’s behavior or negotiation style?
- 2050 9. How useful or novel is the information the model contributes to the negotiation?
- 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

(a) Supervised Fine-Tuning (SFT) Hyperparameters

Hyperparameter	Value
Rollout Batch Size	256
Max Samples	500,000
Max Epochs	10
Prompt Max Length	8192
Generate Max Length	8192
Learning Rate	5e-6

(b) REINFORCE++ Hyperparameters

Hyperparameter	Value
Micro Train Batch Size	4
Train Batch Size	32
Rollout Batch Size	1024
Max Samples	100,000
Max Epochs	20
Prompt Max Length	8192
Generate Max Length	8192
Zero Stage	3
Actor Learning Rate	1e-6
Epsilon Clip High	0.6
n Samples Per Prompt	5

Table 8: Training hyperparameters used for supervised fine-tuning (SFT) and reinforcement learning (REINFORCE++). Default values provided by OpenRLHF and SkyRL, respectively, are used for hyperparameters not listed.

10. How effectively does the model ask clarifying or probing questions to better understand the user’s position?
11. How well are the model’s responses tailored to the negotiation context and engaging to the user?
12. To what extent are the model’s arguments and offers specific to the scenario rather than generic?
13. How clear and easy to follow are the model’s responses in the context of negotiation?

## D TRAINING

**Setup** We finetune the Qwen3-4B reasoning model to negotiate as Agent 2 in the *SolWave Acquisition* task against Qwen3-32B acting as Agent 1 without any specified personality (the “default” behavior of the model). Training proceeded in 3 steps:

**SFT** Supervised finetuning was first performed using OpenRLHF (Hu et al., 2025b) over 1064 synthetically generated conversations in which baseline o4-mini negotiated as Agent 2 in the *SolWave Acquisition* task against any other LLM opponent. We train over 10 epochs of the training data, and choose the earliest checkpoint where the loss is sufficiently low.

**RL with Penalty Reward** We next finetune the SFT checkpoint with multi-turn REINFORCE++ as Agent 2 against Qwen3-32B acting as Agent 1 in the *SolWave Acquisition* task using a negotiation environment we set up to interface with the SkyRL (Cao et al., 2025) training framework. The model yielded by SFT is prone to adverse outputs if the negotiation goes out of the distribution of the SFT data, so gpt-5-nano acting as an LLM-as-a-Judge was used to evaluate the whether each utterance was in the desired format (containing only lines of dialogue). If this was the case, training proceeded as normal, with rewards of 0 assigned to each utterance until the conversation is concluded, at which the task reward is calculated based off the final agreement. Otherwise, if any intermediate output was formatted incorrectly, the training run was ended prematurely, and a reward of -0.5 is assigned to that rollout. Conversations that lead to no agreement but are formatted correctly are given a small reward of 0.05. Formatting rewards have precedent in coding and mathematics tasks (Guo et al., 2025), and we likewise find these to be effective in our negotiation task to ensure dialogue remains in the proper format. We train until the model attains saturated performance with both the penalty and task rewards.

**RL without Penalty** We finally RL finetune the checkpoint purely with the task reward, trained without an evaluation step.

**Hyper-parameters** Hyperparameters for SFT and REINFORCE++ training are detailed in Table 8a and Table 8b.

## E RESULTS

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

Scenario	Instruction-tuned model (gpt-4o-mini)				Reasoning model (o4-mini)			
	Agr%	Agent 1 Reward	Agent 2 Reward	Length	Agr%	Agent 1 Reward	Agent 2 Reward	Length
Twilight Andalusia	100.0%	0.61 ( $\pm 0.07$ )	0.69 ( $\pm 0.09$ )	20.0	98.0%	<b>0.62 (<math>\pm 0.02</math>)</b>	<b>0.80 (<math>\pm 0.02</math>)</b>	11.6
Building Trust	100.0%	<b>0.76 (<math>\pm 0.11</math>)</b>	0.51 ( $\pm 0.01$ )	17.3	100.0%	0.74 ( $\pm 0.01$ )	<b>0.66 (<math>\pm 0.02</math>)</b>	6.3
Solar Horizon	100.0%	0.20 ( $\pm 0.14$ )	0.27 ( $\pm 0.09$ )	16.5	100.0%	<b>0.45 (<math>\pm 0.03</math>)</b>	<b>0.51 (<math>\pm 0.04</math>)</b>	10.7
Rental Cabin	100.0%	0.62 ( $\pm 0.05$ )	0.58 ( $\pm 0.13$ )	7.0	95.2%	<b>0.68 (<math>\pm 0.02</math>)</b>	<b>0.80 (<math>\pm 0.02</math>)</b>	9.0
FrontPage Editorial	100.0%	<b>1.00 (<math>\pm 0.00</math>)</b>	0.87 ( $\pm 0.07$ )	8.7	100.0%	0.99 ( $\pm 0.01$ )	<b>0.93 (<math>\pm 0.02</math>)</b>	8.3
Roommate Conflict	100.0%	0.67 ( $\pm 0.07$ )	<b>1.00 (<math>\pm 0.00</math>)</b>	20.0	100.0%	<b>0.69 (<math>\pm 0.03</math>)</b>	0.91 ( $\pm 0.03$ )	5.5
Data Story Conflict	60.0%	0.27 ( $\pm 0.07$ )	0.17 ( $\pm 0.06$ )	27.0	100.0%	<b>0.80 (<math>\pm 0.03</math>)</b>	<b>0.43 (<math>\pm 0.05</math>)</b>	5.8
Novatech MTA	90.0%	0.63 ( $\pm 0.01$ )	0.24 ( $\pm 0.03$ )	23.4	100.0%	<b>0.66 (<math>\pm 0.03</math>)</b>	<b>0.70 (<math>\pm 0.04</math>)</b>	9.7
Reef Conflict	100.0%	0.40 ( $\pm 0.07$ )	<b>0.68 (<math>\pm 0.04</math>)</b>	12.2	95.0%	<b>0.62 (<math>\pm 0.05</math>)</b>	0.59 ( $\pm 0.04$ )	8.5
Green Showdown	70.0%	0.44 ( $\pm 0.13$ )	0.11 ( $\pm 0.06$ )	24.0	100.0%	<b>0.66 (<math>\pm 0.10</math>)</b>	<b>0.40 (<math>\pm 0.03</math>)</b>	5.4
Brookdale Funding	100.0%	0.63 ( $\pm 0.04$ )	0.19 ( $\pm 0.06$ )	21.6	100.0%	<b>0.65 (<math>\pm 0.03</math>)</b>	<b>0.45 (<math>\pm 0.04</math>)</b>	11.3
GreenFlow Fair	100.0%	0.14 ( $\pm 0.01$ )	0.53 ( $\pm 0.02$ )	12.9	100.0%	<b>0.20 (<math>\pm 0.01</math>)</b>	<b>0.61 (<math>\pm 0.01</math>)</b>	7.6
Crossroads	90.0%	0.11 ( $\pm 0.05$ )	0.80 ( $\pm 0.00$ )	22.4	90.0%	<b>0.17 (<math>\pm 0.04</math>)</b>	0.80 ( $\pm 0.00$ )	16.2
Friends Driveway	100.0%	0.44 ( $\pm 0.04$ )	0.67 ( $\pm 0.04$ )	11.7	100.0%	<b>0.76 (<math>\pm 0.05</math>)</b>	<b>0.89 (<math>\pm 0.00</math>)</b>	5.5
Backstage Battle	100.0%	<b>0.53 (<math>\pm 0.05</math>)</b>	<b>0.29 (<math>\pm 0.03</math>)</b>	23.2	100.0%	0.39 ( $\pm 0.06$ )	0.26 ( $\pm 0.02$ )	17.6
Safety Fairness Dispute	100.0%	0.80 ( $\pm 0.06$ )	<b>0.84 (<math>\pm 0.04</math>)</b>	17.6	95.0%	<b>0.91 (<math>\pm 0.03</math>)</b>	0.71 ( $\pm 0.04$ )	8.4
Balancing Traditions	90.0%	<b>0.53 (<math>\pm 0.05</math>)</b>	0.29 ( $\pm 0.03$ )	20.2	100.0%	0.53 ( $\pm 0.03$ )	<b>0.60 (<math>\pm 0.02</math>)</b>	7.1
SolWave Aquisition	30.0%	0.61 ( $\pm 0.04$ )	<b>0.26 (<math>\pm 0.02</math>)</b>	29.8	87.0%	<b>0.64 (<math>\pm 0.02</math>)</b>	0.23 ( $\pm 0.02$ )	17.1
Library Shutdown	65.0%	0.78 ( $\pm 0.03$ )	0.98 ( $\pm 0.02$ )	18.9	70.0%	<b>0.90 (<math>\pm 0.07</math>)</b>	<b>1.00 (<math>\pm 0.00</math>)</b>	15.6
Lunar Legacy	81.5%	0.18 ( $\pm 0.03$ )	0.39 ( $\pm 7.56$ )	24.8	100.0%	<b>0.32 (<math>\pm 0.07</math>)</b>	<b>0.26 (<math>\pm 0.06</math>)</b>	9.2
Deposit Dilemma	100.0%	0.60 ( $\pm 0.00$ )	0.12 ( $\pm 0.03$ )	21.8	95.0%	<b>0.81 (<math>\pm 0.03</math>)</b>	<b>0.57 (<math>\pm 0.03</math>)</b>	17.7
Battle Horizon	100.0%	0.05 ( $\pm 0.02$ )	0.04 ( $\pm 0.02$ )	22.5	100.0%	<b>0.55 (<math>\pm 0.08</math>)</b>	<b>0.61 (<math>\pm 0.09</math>)</b>	11.5
Rent Deadline	100.0%	0.58 ( $\pm 0.06$ )	0.48 ( $\pm 0.07$ )	8.8	100.0%	<b>0.67 (<math>\pm 0.02</math>)</b>	<b>0.85 (<math>\pm 0.03</math>)</b>	5.9
Gourmet Shutdown	95.0%	0.18 ( $\pm 0.03$ )	<b>0.50 (<math>\pm 0.07</math>)</b>	23.2	100.0%	<b>0.57 (<math>\pm 0.04</math>)</b>	0.48 ( $\pm 0.03$ )	13.2
Museum Standoff	100%	<b>0.52 (<math>\pm 0.03</math>)</b>	0.16 ( $\pm 0.06$ )	12.4	95.0%	0.44 ( $\pm 0.04$ )	<b>0.33 (<math>\pm 0.04</math>)</b>	8.3
Brennation Harborcup	70.0%	0.49 ( $\pm 0.07$ )	0.40 ( $\pm 0.13$ )	26.0	95.0%	<b>0.67 (<math>\pm 0.05</math>)</b>	<b>0.86 (<math>\pm 0.06</math>)</b>	13.4
Gene Editing	90.0%	0.27 ( $\pm 0.07$ )	<b>0.93 (<math>\pm 0.03</math>)</b>	24.2	100.0%	<b>0.72 (<math>\pm 0.05</math>)</b>	0.78 ( $\pm 0.03$ )	8.0
GreenLeaf Partnership	65.0%	0.05 ( $\pm 0.03$ )	0.20 ( $\pm 0.02$ )	27.2	71.0%	<b>0.43 (<math>\pm 0.04</math>)</b>	<b>0.50 (<math>\pm 0.03</math>)</b>	14.3
Boundaries Betrayal	80.0%	0.53 ( $\pm 0.06$ )	0.65 ( $\pm 0.03$ )	22.6	100.0%	<b>0.79 (<math>\pm 0.03</math>)</b>	<b>0.83 (<math>\pm 0.03</math>)</b>	10.9
Crestfield Mindspark	90.0%	<b>0.25 (<math>\pm 0.07</math>)</b>	0.20 ( $\pm 0.05$ )	23.2	100.0%	0.22 ( $\pm 0.04$ )	<b>0.48 (<math>\pm 0.03</math>)</b>	6.5

Table 9: Comparison of performance between instruction-tuned model (GPT-4o-Mini) agents and reasoning model agents (o4-Mini) on negotiation tasks in STRATEGICBENCH. Results report the Agreement Rate (Agr%), mean individual rewards ( $A_1$ ,  $A_2$ ) with standard error (SEM), and the total dialogue length (Len). Bold values indicate the statistically superior mean reward for a specific agent role between the two models.

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

Scenario	Reasoning Models							
	baseline		summary		ReACT		HAR	
	Agent1 Reward	Agent2 Reward	Agent1 Reward	Agent2 Reward	Agent1 Reward	Agent2	Agent1	Agent2
Twilight Andalusia	0.54 ( $\pm 0.01$ )	0.72 ( $\pm 0.02$ )	0.63 ( $\pm 0.04$ )	0.64 ( $\pm 0.04$ )	0.55 ( $\pm 0.03$ )	0.73 ( $\pm 0.05$ )	<b>0.65 (<math>\pm 0.06</math>)</b>	<b>0.78 (<math>\pm 0.09</math>)</b>
Building Trust	0.76 ( $\pm 0.01$ )	0.57 ( $\pm 0.01$ )	0.77 ( $\pm 0.01$ )	0.53 ( $\pm 0.03$ )	0.74 ( $\pm 0.02$ )	0.58 ( $\pm 0.03$ )	<b>0.80 (<math>\pm 0.03</math>)</b>	<b>0.63 (<math>\pm 0.05</math>)</b>
Solar Horizon	0.44 ( $\pm 0.01$ )	0.49 ( $\pm 0.02$ )	0.48 ( $\pm 0.03$ )	0.50 ( $\pm 0.05$ )	0.53 ( $\pm 0.07$ )	0.58 ( $\pm 0.06$ )	<b>0.56 (<math>\pm 0.07</math>)</b>	<b>0.61 (<math>\pm 0.06</math>)</b>
Rental Cabin	0.60 ( $\pm 0.01$ )	0.62 ( $\pm 0.02$ )	0.64 ( $\pm 0.03$ )	0.83 ( $\pm 0.00$ )	0.64 ( $\pm 0.03$ )	0.83 ( $\pm 0.00$ )	<b>0.66 (<math>\pm 0.03</math>)</b>	<b>0.85 (<math>\pm 0.01</math>)</b>
FrontPage Editorial	0.97 ( $\pm 0.01$ )	0.90 ( $\pm 0.01$ )	<b>1.00 (<math>\pm 0.00</math>)</b>	0.95 ( $\pm 0.02$ )	0.98 ( $\pm 0.02$ )	1.00 ( $\pm 0.00$ )	0.98 ( $\pm 0.02$ )	<b>1.00 (<math>\pm 0.00</math>)</b>
Roommate Conflict	0.72 ( $\pm 0.01$ )	0.93 ( $\pm 0.01$ )	0.71 ( $\pm 0.03$ )	0.94 ( $\pm 0.02$ )	0.78 ( $\pm 0.06$ )	1.00 ( $\pm 0.00$ )	<b>0.80 (<math>\pm 0.06</math>)</b>	<b>1.00 (<math>\pm 0.00</math>)</b>
Data Story Conflict	0.66 ( $\pm 0.02$ )	0.25 ( $\pm 0.02$ )	0.62 ( $\pm 0.05$ )	0.27 ( $\pm 0.03$ )	0.86 ( $\pm 0.07$ )	0.51 ( $\pm 0.06$ )	<b>0.88 (<math>\pm 0.07</math>)</b>	<b>0.54 (<math>\pm 0.06</math>)</b>
Novatech MTA	0.69 ( $\pm 0.01$ )	0.45 ( $\pm 0.02$ )	0.71 ( $\pm 0.03$ )	0.42 ( $\pm 0.03$ )	0.72 ( $\pm 0.02$ )	0.46 ( $\pm 0.04$ )	<b>0.74 (<math>\pm 0.04</math>)</b>	<b>0.50 (<math>\pm 0.04</math>)</b>
Reef Conflict 5	0.61 ( $\pm 0.02$ )	0.59 ( $\pm 0.02$ )	0.66 ( $\pm 0.03$ )	0.58 ( $\pm 0.03$ )	0.66 ( $\pm 0.07$ )	0.64 ( $\pm 0.07$ )	<b>0.68 (<math>\pm 0.07</math>)</b>	<b>0.67 (<math>\pm 0.07</math>)</b>
Green Showdown	0.44 ( $\pm 0.01$ )	0.40 ( $\pm 0.01$ )	0.46 ( $\pm 0.03$ )	<b>0.51 (<math>\pm 0.05</math>)</b>	0.46 ( $\pm 0.02$ )	0.47 ( $\pm 0.03$ )	<b>0.50 (<math>\pm 0.12</math>)</b>	0.47 ( $\pm 0.03$ )
Brookdale Funding	0.56 ( $\pm 0.01$ )	0.45 ( $\pm 0.01$ )	0.60 ( $\pm 0.02$ )	<b>0.48 (<math>\pm 0.03</math>)</b>	<b>0.92 (<math>\pm 0.03</math>)</b>	0.45 ( $\pm 0.02$ )	<b>0.92 (<math>\pm 0.03</math>)</b>	0.34 ( $\pm 0.09$ )
GreenGlow Fair	0.19 ( $\pm 0.01$ )	0.41 ( $\pm 0.01$ )	0.20 ( $\pm 0.01$ )	0.44 ( $\pm 0.01$ )	0.24 ( $\pm 0.01$ )	0.63 ( $\pm 0.02$ )	<b>0.26 (<math>\pm 0.04</math>)</b>	<b>0.63 (<math>\pm 0.02</math>)</b>
Crossroads	<b>0.48 (<math>\pm 0.02</math>)</b>	0.82 ( $\pm 0.01$ )	0.39 ( $\pm 0.05$ )	<b>0.84 (<math>\pm 0.02</math>)</b>	0.38 ( $\pm 0.08$ )	0.73 ( $\pm 0.07$ )	0.38 ( $\pm 0.08$ )	0.73 ( $\pm 0.07$ )
Friends Driveway	0.67 ( $\pm 0.02$ )	0.84 ( $\pm 0.01$ )	0.73 ( $\pm 0.04$ )	<b>0.89 (<math>\pm 0.00</math>)</b>	0.79 ( $\pm 0.03$ )	0.84 ( $\pm 0.02$ )	<b>0.81 (<math>\pm 0.03</math>)</b>	0.84 ( $\pm 0.03$ )
Backstage Battle	0.52 ( $\pm 0.03$ )	0.52 ( $\pm 0.04$ )	0.58 ( $\pm 0.08$ )	<b>0.62 (<math>\pm 0.06</math>)</b>	<b>0.66 (<math>\pm 0.01</math>)</b>	0.42 ( $\pm 0.05$ )	0.25 ( $\pm 0.06$ )	0.40 ( $\pm 0.05$ )
Safety Fairness Dispute	<b>0.92 (<math>\pm 0.01</math>)</b>	0.78 ( $\pm 0.02$ )	0.89 ( $\pm 0.03$ )	0.79 ( $\pm 0.03$ )	<b>0.92 (<math>\pm 0.04</math>)</b>	<b>0.82 (<math>\pm 0.04</math>)</b>	<b>0.92 (<math>\pm 0.04</math>)</b>	<b>0.82 (<math>\pm 0.04</math>)</b>
Balancing Traditions	0.43 ( $\pm 0.01$ )	0.66 ( $\pm 0.02$ )	0.40 ( $\pm 0.02$ )	0.67 ( $\pm 0.02$ )	0.61 ( $\pm 0.04$ )	0.67 ( $\pm 0.03$ )	<b>0.70 (<math>\pm 0.04</math>)</b>	<b>0.69 (<math>\pm 0.03</math>)</b>
SolWave Aquisition	<b>0.66 (<math>\pm 0.03</math>)</b>	<b>0.36 (<math>\pm 0.06</math>)</b>	0.00 ( $\pm 0.00$ )	0.00 ( $\pm 0.00$ )	0.00 ( $\pm 0.00$ )	0.00 ( $\pm 0.00$ )	0.00 ( $\pm 0.00$ )	0.00 ( $\pm 0.00$ )
Library Shutdown	0.77 ( $\pm 0.02$ )	0.95 ( $\pm 0.01$ )	<b>0.85 (<math>\pm 0.03</math>)</b>	0.99 ( $\pm 0.01$ )	0.84 ( $\pm 0.03$ )	<b>1.00 (<math>\pm 0.00</math>)</b>	0.84 ( $\pm 0.03$ )	<b>1.00 (<math>\pm 0.00</math>)</b>
Lunar Legacy	<b>0.69 (<math>\pm 0.02</math>)</b>	<b>0.55 (<math>\pm 0.02</math>)</b>	0.59 ( $\pm 0.05$ )	<b>0.59 (<math>\pm 0.07</math>)</b>	0.60 ( $\pm 0.07$ )	N/A	0.60 ( $\pm 0.07$ )	N/A
Deposit Dilemma	0.65 ( $\pm 0.03$ )	<b>0.62 (<math>\pm 0.02</math>)</b>	<b>0.84 (<math>\pm 0.06</math>)</b>	<b>0.66 (<math>\pm 0.03</math>)</b>	0.82 ( $\pm 0.07$ )	0.64 ( $\pm 0.03$ )	<b>0.85 (<math>\pm 0.07</math>)</b>	0.52 ( $\pm 0.08$ )
Battle Horizon	0.54 ( $\pm 0.02$ )	0.77 ( $\pm 0.02$ )	0.49 ( $\pm 0.04$ )	<b>0.87 (<math>\pm 0.03</math>)</b>	<b>0.55 (<math>\pm 0.03</math>)</b>	0.83 ( $\pm 0.04$ )	0.44 ( $\pm 0.08$ )	0.79 ( $\pm 0.04$ )
Rent Deadline	<b>0.71 (<math>\pm 0.01</math>)</b>	0.85 ( $\pm 0.02$ )	0.67 ( $\pm 0.06$ )	<b>0.94 (<math>\pm 0.02</math>)</b>	0.62 ( $\pm 0.02$ )	0.93 ( $\pm 0.02$ )	0.62 ( $\pm 0.02$ )	0.89 ( $\pm 0.05$ )
Gourmet Shutdown	0.46 ( $\pm 0.02$ )	0.47 ( $\pm 0.02$ )	0.48 ( $\pm 0.04$ )	0.47 ( $\pm 0.03$ )	<b>1.00 (<math>\pm 0.19</math>)</b>	<b>0.63 (<math>\pm 0.08</math>)</b>	<b>1.00 (<math>\pm 0.19</math>)</b>	<b>0.63 (<math>\pm 0.08</math>)</b>
Museum Standoff	0.44 ( $\pm 0.02$ )	0.33 ( $\pm 0.02$ )	<b>0.50 (<math>\pm 0.07</math>)</b>	0.34 ( $\pm 0.03$ )	<b>0.50 (<math>\pm 0.04</math>)</b>	<b>0.48 (<math>\pm 0.04</math>)</b>	<b>0.50 (<math>\pm 0.04</math>)</b>	<b>0.48 (<math>\pm 0.04</math>)</b>
Brewnation Harborcup	0.66 ( $\pm 0.02$ )	0.90 ( $\pm 0.02$ )	0.73 ( $\pm 0.08$ )	0.91 ( $\pm 0.03$ )	0.80 ( $\pm 0.05$ )	<b>0.95 (<math>\pm 0.05</math>)</b>	0.80 ( $\pm 0.05$ )	<b>0.95 (<math>\pm 0.05</math>)</b>
Gene Editing	0.74 ( $\pm 0.02$ )	0.73 ( $\pm 0.02$ )	0.66 ( $\pm 0.04$ )	0.78 ( $\pm 0.04$ )	<b>0.76 (<math>\pm 0.06</math>)</b>	<b>0.80 (<math>\pm 0.05</math>)</b>	<b>0.76 (<math>\pm 0.06</math>)</b>	<b>0.80 (<math>\pm 0.05</math>)</b>
GreenLeaf Partnership	0.63 ( $\pm 0.02$ )	0.51 ( $\pm 0.02$ )	0.56 ( $\pm 0.04$ )	0.60 ( $\pm 0.11$ )	0.57 ( $\pm 0.04$ )	0.50 ( $\pm 0.03$ )	<b>0.65 (<math>\pm 0.04</math>)</b>	<b>0.62 (<math>\pm 0.03</math>)</b>
Boundaries Betrayal	0.80 ( $\pm 0.02$ )	0.79 ( $\pm 0.02$ )	0.78 ( $\pm 0.05$ )	0.76 ( $\pm 0.04$ )	<b>0.84 (<math>\pm 0.04</math>)</b>	<b>0.91 (<math>\pm 0.04</math>)</b>	<b>0.84 (<math>\pm 0.04</math>)</b>	<b>0.91 (<math>\pm 0.04</math>)</b>
Crestfield Mindspark	0.61 ( $\pm 0.02$ )	0.23 ( $\pm 0.03$ )	<b>0.66 (<math>\pm 0.03</math>)</b>	<b>0.47 (<math>\pm 0.04</math>)</b>	0.61 ( $\pm 0.04$ )	0.38 ( $\pm 0.03$ )	0.45 ( $\pm 0.06$ )	0.38 ( $\pm 0.03$ )

Table 10: Comparison of performance between methods on negotiation tasks in STRATEGICBENCH. Results report the mean rewards when agents negotiate as Agent 1 and Agent 2, with standard error (SEM). Results are averaged across opponent agents. Bold values indicate the statistically superior mean reward for a specific agent role between the two models.