# EUROARENA: DO LARGE LANGUAGE MODELS EXCEL AT LONG-HORIZON PLANNING UNDER COMPETITION?

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Large language models (LLMs) have demonstrated strong performance on complex tasks, such as coding and mathematics, but can they plan effectively while other agents are planning against them? Adversarial interaction introduces a distinct challenge: a plan must account not only for the consequences of one's own actions, but also for opponents who can contest resources and disrupt its execution, which is largely overlooked in current benchmarks. We introduce **EuroArena**, a benchmark featuring two-player Eurogames that tests long-horizon planning through competition over shared resources, investment in future gains, and adaptation to opponent's actions. Across more than 400 hours of Splendor Duel gameplay against skilled human players, we find that, surprisingly, six strong LLMs struggle to match human expertise, despite demonstrating the ability to formulate concrete plans, revise them after feedback, and propose alternatives for possible future states. To investigate this gap, we introduce resource dependency graphs to trace how earlier decisions support later purchases, and leverage this with human-verified trajectory analysis. We find a recurring contrast: LLMs prioritize reaching the next subgoal, whereas humans more often take actions that support later subgoals and make better use of accumulated resources. Models also overlook adverse opponent responses and make reasoning errors when integrating multiple information sources. These findings reveal a gap between generating and revising explicit plans and making decisions that sustain progress toward long-term goals under competition.

## 1 INTRODUCTION

Large language models (LLMs) have demonstrated strong performance on complex tasks, including coding (Jimenez et al., 2024; Jain et al., 2025), mathematics (Glazer et al., 2024), and travel planning (Zhang et al., 2026). But whether they can they plan effectively while other agents are planning against them is under-explored. With explicit adversarial interaction, a good plan must account not only for the consequences of one's own actions, but also for opponents who pursue competing goals, contest shared resources, or disrupt its execution. An action that advances an immediate subgoal may therefore contribute little to eventual success if it leaves the player poorly positioned for subsequent decisions or vulnerable to an opponent's response. To this end, we need playgrounds and effective evaluation methods to examine whether advanced LLMs could successfully sustain progress toward long-term goals as the competitive environment changes when making long-horizon plans.

Games offer a controlled setting for studying this challenge through explicit rules, verifiable goals, and observable decision trajectories. Each action changes both the player's future options and the opportunities available to the opponent. A complete trajectory therefore reveals how a player invests in future gains, builds on earlier decisions, and adapts when an opponent interferes. Comparing LLMs with skilled human players provides a behavioral reference for understanding these choices, helping explain differences in performance that final outcomes alone cannot reveal.

Existing game-based evaluations examine important aspects of planning, but leave this combination of long-horizon investment and active opposition insufficiently understood. Benchmarks and environments based on Minecraft (Fan et al., 2022; Wang et al., 2023a) and the Rubik's Cube (Gao et al., 2025), for example, study sequential decisions and progress toward distant goals. Nonetheless, their central challenges, concern interaction with the task environment rather than competition with a skilled opponent who can deliberately interfere with a plan. Consequently, they provide limited



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

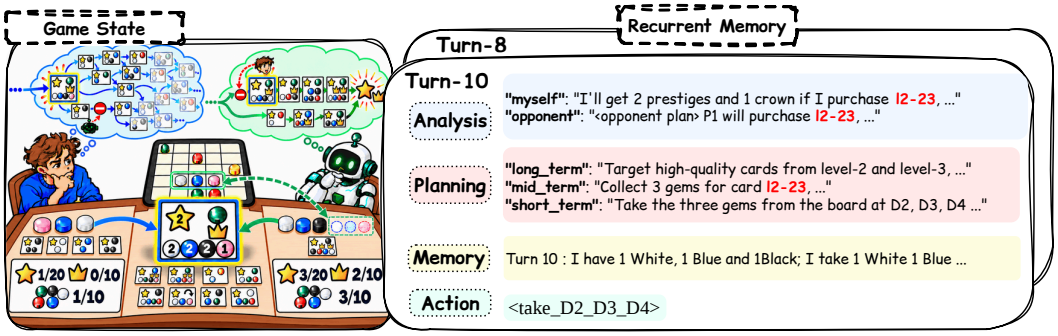


Figure 1: Splendor Duel gameplay (**left**) and the recurrent memory mechanism (**right**). Players collect gems to purchase cards that provide prestige and crowns toward the three winning conditions, as well as permanent discounts for future purchases. Effective planning requires players to compete over shared resources, invest in future gains, and adapt to an opponent who can disrupt their plans. These decisions jointly depend on the shared board, available cards, and both players' resources. The recurrent memory mechanism carries forward and updates assessments of both players, plans at different horizons, and relevant game-state records.

evidence about whether an LLM anticipates adverse opponent responses or adapts its investments when access to shared resources changes.

Interactive platforms such as TextArena (Guertler et al., 2025) and GameArena (Hu et al., 2025) introduce other players into evaluation, but offer limited analysis of how individual decisions support later objectives over extended competition. Card game evaluations, including PokerBench (Zhuang et al., 2025) and UNO Arena (Qin et al., 2024), examine competitive decisions under uncertainty, while social deduction games such as Werewolf (Bailis et al., 2024) and Avalon (Light et al., 2023; Wang et al., 2023b) emphasize intention inference and deception. But random draws and hidden information complicate outcome attribution, while social deduction offers less direct access to dependencies between resource accumulation and later actions.

Chess-based evaluations (Kolasani et al., 2025; Liu et al., 2026) provide long-horizon competition without random draws, with information concentrated on a single board. Euro games offer a complementary setting requiring players to integrate shared and individual resource information while balancing immediate progress against investment in future gains.

We introduce **EuroArena**, a benchmark of five two-player Euro games evaluating long-horizon planning under resource competition and opponent interference. Players must determine both how to reach the next subgoal and which actions will support subsequent objectives as opponents respond, reasoning about consequences that may unfold many turns later.

We focus our detailed analysis on Splendor Duel, collecting over 400 hours of gameplay between skilled humans and six strong LLMs. All models use the same recurrent memory mechanism without additional planning enhancements. Despite their capabilities in formulating explicit plans, revising them after feedback, and preparing alternatives for possible future states, they struggle to match human expertise. Why do these explicit planning capabilities fail to translate into effective long-term decisions under competition?

To investigate, we combine a resource dependency graph (RDG), which traces how earlier decisions support later purchases, with an agentic trajectory analysis pipeline examining opponent modeling and reasoning errors. Across all six models, we observe a recurring contrast: LLMs favor the fastest routes to their next subgoal, whereas humans more often support later objectives and better utilize accumulated resources. Models also overlook adverse opponent responses and struggle to integrate multiple information sources. These findings reveal a gap between generating and revising plans and choosing actions that sustain long-term progress.

Our main contributions are as follows:

- We introduce **EuroArena**, comprising two-player Euro games that test long-horizon planning through resource competition, investment in future gains, and adaptation to opponents.

- We combine RDG with human-verified trajectory analysis to trace dependencies between decisions and identify opponent modeling and reasoning errors.
- Through over 400 hours of human-LLM gameplay, we identify a recurring gap between explicit planning capabilities and effective long-term decisions, revealing weaknesses that final outcomes alone cannot explain.

## 2 EUROARENA

EuroArena includes five well-known two-player Euro games together with methods for quantitative and qualitative evaluation. Among these games, Splendor Duel provides a suitable setting for examining LLM behavioral patterns in long-horizon planning with competition over shared resources. We use it as the basis for the experiments and analyses in this paper.

### 2.1 TWO-PLAYER EURO GAME: SPLENDOR DUEL

In Splendor Duel 1, players collect gems and purchase cards as they race to accumulate at least (1) twenty prestige, (2) ten crowns, or (3) ten prestige points on purchased cards of one color. Cards provide permanent token discounts for future purchases, as well as prestige or crowns that contribute to victory. Effective play therefore requires a careful balance between building a strong token engine and accumulating points. For brevity, we focus here on the characteristics that make Splendor Duel ideal for our analysis. Please refer to Appendix A.1 for detailed game introductions.

**Opponent modeling.** Effective planning in Splendor Duel requires players to anticipate their opponents' actions. They must assess which cards their opponents are likely to contest for and how those actions may change the availability of valuable resources, increasing the complexity of environment modeling. Meanwhile, an opponent can exploit an LLM's mistakes, making the consequences observable. At the same time, the two-player setting remains sufficiently tractable for us to assess the quality of LLMs' environment modeling and long-term plans.

**Long-horizon planning and decision making.** A game of Splendor Duel typically involves more than 50 consecutive decisions per player, while randomness is seldom the dominant factor in determining the outcome. Players must therefore plan over many turns and make decisions that consistently advance them toward winning conditions.

**Multi-channel information processing.** Players must simultaneously interpret information from the gem board, card market, token engines, and both players' progress toward the three winning conditions. LLMs must identify and integrate the relevant information to make plans and decisions that balance different considerations.

### 2.2 RECURRENT MEMORY

To support consistent decisions across a long game, an LLM needs access to relevant information from earlier states. However, incorporating sophisticated memory mechanisms introduces extra variables irrelevant to the LLM's own abilities. We therefore introduce recurrent memory, which keeps the input context bounded and facilitates the analysis of planning patterns while adding minimal support beyond the model's own capabilities.

At each decision point, the LLM receives a structured prompt containing the current game state and recurrent memory generated on its previous turn. It then updates the recurrent memory with analyses of its own position and the opponent's position, three plans spanning different time horizons, and a piece of memory containing important historical states. Using this context, the LLM chooses an action and advances the game. We deferred the detailed prompts and cases in Appendix A.2.

### 2.3 EVALUATION

**Resource Dependency Graph (RDG).** To analyze how do earlier decisions support latter ones in a gameplay trajectory, we convert each game record into a resource dependency graph (RDG) as is in Figure 2.

Formally, we construct a graph $G(V, E)$ with $N$ nodes $V = \{v_i\}_{i=1}^{N}$ and directed edges $E = \{e_{ij}\}_{1 \le i,j \le N}$, where $N$ is the player's number of turns. Each node $v_i$ corresponds to the $i_{\text{th}}$ decision-making turn, and consecutive turn nodes are connected. An edge $e_{ij}$ from turn node $v_i$ to turn node $v_j$ in the upper half of the graph represents a **direct resource dependency**, that the player either (1) collected gems at Turn $i$ and spent them at Turn $j$, or (2) reserved a card at Turn $i$ and purchased it at Turn $j$. An edge in the lower half represents a **mediated resource dependency**, that the player (3) purchased a card at Turn $i$ whose token discount contributed to another purchase at Turn $j$.

**Human-Verified Trajectory Analysis.** To examine planning and reasoning patterns in the trajectories, we use a pipeline (Figure 2) for detailed qualitative analysis of the LLMs' outputs. Specifically, we use `GPT-5.6-Sol` [1] as the backbone model and `Codex` [2] as the agent harness to analyze each recurrent memory entry and its corresponding action. The agent reconstructs each game state, assesses the decision, and mainly identifies (1) opponent modeling errors, in which ***the LLM overlooks an opponent's possible actions or clearly misinterprets their intentions***, and (2) reasoning mistakes during multi-channel information processing, in which ***the LLM misanalyzes the current state or incorrectly deduces a future outcome***. We then ensure the validity of the cases via manual review.

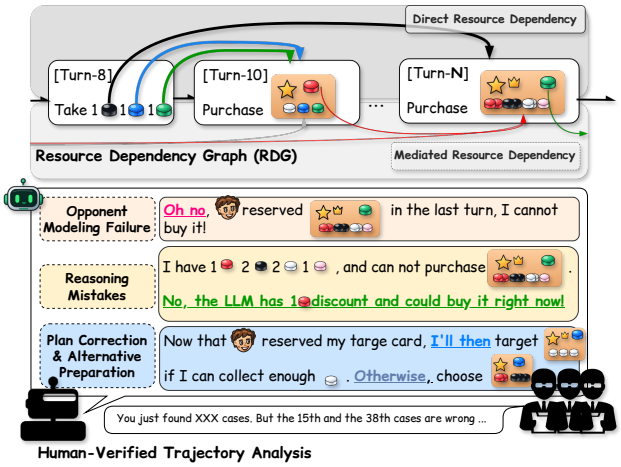


Figure 2: Resource Dependency Graph (RDG) (**upper**) models the game trajectories as graphs and show how earlier decisions support latter ones. Each node denotes a timestep and each directed edge represents a direct resource dependency or mediated resource dependency. The human-verified trajectory analysis (**lower**) derive opponent modeling errors and reasoning errors made by LLM players, which will be verified by at least two human experts. It also helps to identify LLM's capabilities in plan correction and alternative preparation.

## 3 Experimental Evaluation

To collect high-quality trajectories, we recruited experienced players from a board game club to play Splendor-Duel against six advanced LLMs: `Qwen3.8-Max` (Yang et al., 2025), `DeepSeek-V4-Pro` (Xu et al., 2026), `Kimi-K3` (Team et al., 2026), `GPT-5.6-Terra` [3], `Claude-Opus-5` [4], and `GPT-6-Astra` [5]. Each model used its maximum reasoning effort. Approximately 400 hours of human versus LLM play yielded 30 completed game trajectories for quantitative and qualitative analysis. Please refer to Appendix 4 for detailed budget statement. Table 1 summarizes the game outcomes. The human players won all the games and consistently built stronger token engines. Here, we present the key findings from our analysis of these trajectories using the evaluation methods introduced in Section 2.3. Together, these findings reveal a fundamental gap between generating and revising explicit plans and making decisions that sustain progress toward long-term goals in competitive scenarios.

[1] `https://deploymentsafety.openai.com/gpt-5-6`
[2] `https://openai.com/codex/`
[3] `https://deploymentsafety.openai.com/gpt-5-6`
[4] `https://www.anthropic.com/system-cards`
[5] `https://deploymentsafety.openai.com/gpt-6-astra`

Table 1: Game outcome and progress statistics.

| Metric | GPT-6 Astra | Claude Opus-5 | GPT-5.6 Terra | Kimi-K3 | Deepseek V4-Pro | Qwen3.8 Max | Human |
|---|---|---|---|---|---|---|---|
| Wins | 0/5 | 0/5 | 0/5 | 0/5 | 0/5 | 0/5 | 30/30 |
| Highest Winning Condition Completion (%) | 57.0 ±19.6 | 84.0 ±10.8 | 70.0 ±15.4 | 73.0 ±20.5 | 64.0 ±10.8 | 56.0 ±19.8 | 102.8 ±4.3 |
| Total Token Discounts | 14.6 ±2.5 | 10.2 ±1.9 | 9.4 ±1.7 | 7.0 ±2.1 | 9.8 ±2.5 | 11.0 ±4.7 | 15.7 ±2.4 |

### 3.1 Behavioral Patterns in Long-Horizon Planning and Decision Making

One prominent observation is that LLMs form concrete plans to accumulate prestige and crowns quickly, then pursue successive subgoals through locally focused decisions. Contrarily, human players generally prioritize a strong token engine to support a final push toward the winning conditions. We examine this difference through the following quantitative experiments.

**Experiment: stage segmentation analysis based on resource dependency graph (RDG).** For a given RDG $G(V, E)$, a stage segmentation $\Gamma_{G,K} = \{v_{i_k}\}_{k=1}^{K}$ partitions $V$ into $K$ consecutive stages. The final turn node of each stage, $v_{i_k}$, represents its subgoal, typically a card purchase. We classify each directed edge $e_{pq} \in E$ as either (1) a **stage connection** in $E_{\text{stage}}^{\Gamma_{G,K}}$ if $v_p$ and $v_q$ belong to the same stage or are both stage subgoals, or (2) a **cross-stage connection** in $E_{\text{cross}}^{\Gamma_{G,K}}$ otherwise.

$$\begin{cases} \text{stage}_1^{\Gamma_{G,K}} = \{v_1, v_2, \ldots, v_{i_1}\} \\ \text{stage}_2^{\Gamma_{G,K}} = \{v_{i_1+1}, v_{i_1+2} \ldots v_{i_2}\} \\ \ldots \\ \text{stage}_K^{\Gamma_{G,K}} = \{v_{i_{K-1}+1}, v_{i_{K-1}+2}, \ldots, v_{i_K}\} \end{cases}, \quad e_{pq} \in \begin{cases} E_{\text{stage}}^{\Gamma_{G,K}}, \text{ if } \exists k \in \{1, 2, \ldots K\}, s.t. v_p, v_q \in \text{stage}_k, \\ \qquad \text{or } v_p, v_q \in \Gamma_{G,K} \\ E_{\text{cross}}^{\Gamma_{G,K}}, \text{ otherwise} \end{cases}. \tag{1}$$

For any valid number of stages $K \in [1, N]$, we define the **optimal stage segmentation** $\Gamma_{G,K}^*$ as the stage segmentation that minimizes the number of cross-stage connections $\left|E_{\text{cross}}^{\Gamma_{G,K}^*}\right| \leq \left|E_{\text{cross}}^{\Gamma_{G,K}}\right|, \forall \Gamma_{G,K}$. We denote $\left|E_{\text{cross}}^{\Gamma_{G,K}^*}\right|$ by $n_{\text{cross}}^K(G)$, the (minimum) number of cross-stage connections for an RDG with $K$ stages.

Figure 3 shows optimal six-stage segmentations of human and LLM trajectories, including their stage and cross-stage connections. The human players' RDGs contain more cross-stage connections (dashed lines) and more mediated resource dependencies (edges in the lower half of each graph), suggesting that human players make decisions to benefit later subgoals. From another perspective, they use their existing resources more extensively. By comparison, LLMs focus narrowly on the nearest subgoal, and take gems to achieve it more directly. Their RDGs therefore show clearer stage boundaries. Figure 4 shows that fewer cross-stage connections are a consistent pattern across the evaluated game trajectories.

**Takeaway.** LLMs tend to seek the fastest path to each immediate subgoal. Human players prefer decisions that also support later subgoals and make greater use of existing resources.

**Experiment: trajectory forking with desk reshuffling.** These findings raise two questions: when do the human players' planning patterns improve outcomes in our situations, and can LLMs also benefit from those patterns. Our hypothesis is that decisions providing broad future benefits may retain their value across a wider range of subsequent states, making them more robust to unpredictable changes and adversarial actions. To test this idea, we use trajectory forking with desk reshuffling, as illustrated in Figure 5. We randomly select three turns at 50%–60% of a completed game trajectory's progression when the human and LLM players have made similar progress with

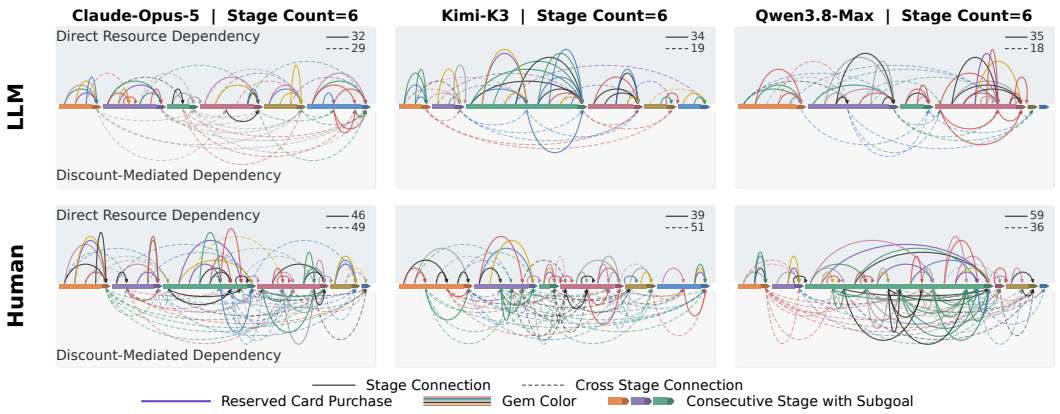


Figure 3: Optimal stage segmentation ($K = 6$) of the resource dependency graphs for three gameplay trajectories. The human players' graphs contain more cross-stage connections (**dashed lines**) and more mediated resource dependencies (edges in the lower half of each graph) over the best stage segmentation. On the contrary, the LLMs' graphs show clearer stage boundaries and fewer cross-stage connections, indicating a stronger focus on the nearest subgoal.

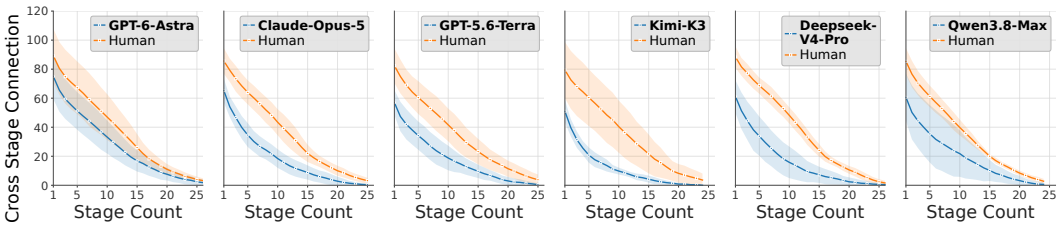


Figure 4: Cross-stage connections in the optimal stage segmentations across different numbers of stages. Human gameplay trajectories consistently contain more cross-stage resource dependencies.

respective to the winning conditions. We then randomly reshuffle the cards in market to simulate an unpredictable change and have two identically configured LLM players continue the game. We use `DeepSeek-V4-Pro` and `GPT-5.6-Terra` as backbone models for the forked games.

The results support our hypothesis. In the three experiments shown in Figure 6, LLMs won when continuing from positions built by human players. The human-built token engines provided more cross-boundary resource support than the LLM-built ones by over 40% on average. Meanwhile, the original human players still used existing token discounts about 30% more extensively than the LLMs that continued from their positions.

These observations suggest that human decisions provide stronger support for the unpredictable future, whereas LLMs do not make full use of such opportunities. The finding highlights the importance of using current actions to support future plans in long-horizon planning under adversarial interaction, and showcases the fundamental gap between making explicit plans and making decisions that sustain progress toward long-term goals under competition.

**Takeaway.** In environments with resource rivalry and unpredictable changes, decisions supporting multiple future options and making effective use of existing resources can improve the robustness of long-horizon planning. LLMs show less of this behavior, demonstrating the gap between the abilities of making explicit plans and sustaining progress toward long-term goals under competition.

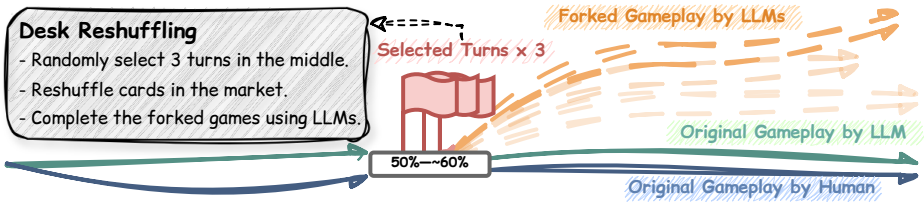


Figure 5: Illustration of the trajectory forking experiment with desk reshuffling.

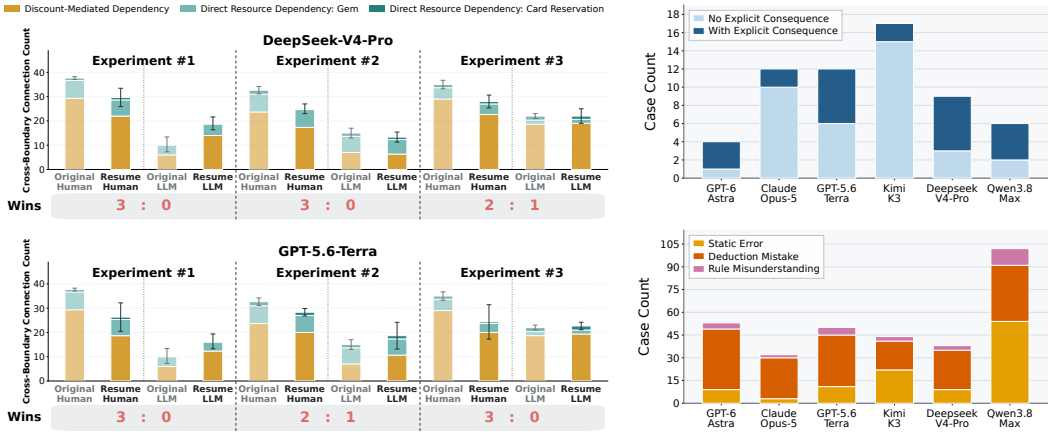


Figure 6: **Left:** Results of trajectory forking with desk reshuffling. LLMs that resumed play from the human players' positions gained an advantage over the opposing side. Decisions made by the human players before the fork provided more resource support for the continued games than those made by the original LLM players. Meanwhile, the original human players still used their token discounts more extensively. **Right:** Counts of opponent modeling errors (**upper**) and reasoning mistakes (**lower**). Both types of errors occur even in the most advanced evaluated LLMs.

### 3.2 Opponent Modeling Error

We next examine whether LLMs can anticipate an opponent's actions and plan for adverse responses. Figure 6 (right) shows that even advanced LLMs, including `GPT-6-Astra` and `Claude-Opus-5`, occasionally make opponent modeling errors. Table 2 presents examples in which LLMs overlook possible adversarial responses or misinterpret their opponents' intentions, which result in plans and decisions that place the LLMs at a disadvantage. They illustrate difficulties in modeling the dynamics of environments that involve human interactions, which, however, is one of the core abilities necessary for real world application.

> **Takeaway.** LLMs overlook adverse responses when planning ahead, suggesting a key competency gap for real world application.

### 3.3 Reasoning Mistake

Beyond opponent modeling errors, LLMs make reasoning mistakes when processing information from multiple channels. Figure 6 (right) shows that, although more advanced LLMs make fewer static errors regarding the current state, all evaluated models make deductive mistakes when predicting future states and simulating action sequences. `Qwen3.8-Max` also misinterprets the rules more frequently, suggesting greater difficulty in comprehending the lengthy rule descriptions.

The lower part of Table 2 presents two representative cases. In both, a reasoning mistake affected the LLM's long-horizon plan and immediate decision, leading to a loss that could not be fully reversed given that the opponent may exploit it before the LLM can correct its plan using subsequent feedback.

Table 2: Representative opponent modeling errors and reasoning mistakes made by state-of-the-art LLMs. We highlighted the **erroneous statements** and the corresponding ground truth.

| Original Text, Ground Truth & Decision | Consequence |
|---|---|
| **Opponent Modeling Error** | |
| ***[Claude-Opus-5]:*** "T46: take White+White+White from B2, C3, D4 (P1 gains 1 privilege) so **l3-02's white3 is untouchable** ... Current decision → Choose take_B2_C3_D4: it converts the game's last three available whites into a **guaranteed l3-02 purchase** (having enough resources did not guarantee the purchase, because l3-02 could be preemptively reserved by the opponent) ..."<br>**[Decision]:** Get 3 whites, and give opponent a privilege. | The human player (P1) reserved the card. The three whites were wasted afterwards. |
| ***[GPT-6-Astra]:*** "... the **visible l2-06 (cost Green 4, Black 3) purchasable** (for the opponent) (the opponent was obviously targeting another card, not l2-06) ... removing one Green blocks that efficient 1-point Blue-engine card and its steal ability. **green is also relevant to their visible l2-08** (the opponent did not target l2-08 either) and future high-level purchases..."; "**Removing Green specifically denies** (removing Green would not deny the opponent's actual target) the currently efficient l2-06 purchase ..."<br>**[Decision]:** Purchase a card and take a green gem from the opponent. | The opponent purchased another card, which could have been denied if the LLM removed a blue gem. The green gem was not used until 14 turns later. |
| **Reasoning Mistake** | |
| ***[GPT-5.6-Terra]:*** "The immediate target l2-05 requires Black1, **Red2, and Pearl1 after discounts** (the two Red costs were already covered by permanent discounts); I currently have the Black and Pearl but no Red. ... **Replenishing now** (no refill was needed; l2-05 was already affordable) creates access to the needed Red tokens ..."; "Refill the board, prioritize **collecting Red2** while retaining Black1 and Pearl1, and **buy l2-05 as soon as possible** ..."<br>**[Decision]:** Refill the board, and give opponent a privilege. | The LLM did not take gems after refilling and purchased l2-05 directly. The opponent got a privilege and used it to take an extra pearl. |
| ***[Kimi-K3]:*** "(1) Turn 39 now: take C3-C4-C5 (R R U) → 12 tokens, discard both blacks → 2W 1U 3G 2R 1P 1Au = 10. (2) Turn 41: buy l1-14 paying 1W 1U (never gold, never pearl) → bonuses W1 U1 G1, hand **1W 3G 2R 1P 1Au (7)** (8 gems, not 7); take-green ability whiffs, irrelevant. (3) Turn 43: take D1-D2-D3 (R U W) → **2W 1U 3G 3R 1P 1Au = 10** (the total is 11 gems, not 10) exactly, **no discard** (one gem had to be discarded to satisfy the 10-gem limit) ..."<br>**[Decision]:** Act accordingly until Turn 43. | The LLM exceeded the 10-gem limit at Turn 43 and had to discard a blue gem. |

This further validates our choosing competitive games to study long-horizon planning and decision making.

> **Takeaway.** LLMs make reasoning mistakes when integrating information from multiple channels, especially when reasoning about action sequences. Subsequent feedback can help them correct a plan, but it may arrive too late to undo the consequences of an error.

### 3.4 Plan Correction and Alternative Preparation

Although LLMs consistently make mistakes in opponent modeling and reasoning, they show an ability to recognize when a plan no longer fits the game state and revise it in pursuit of the ultimate objective. They can also specify alternative actions for different future situations. Table 3 presents two examples in which LLMs identify obsolete plans, revise their objectives, and prepare for possible future developments. These capabilities, however, do not fully offset the opponent modeling errors and reasoning mistakes discussed above, indicating the gap between generating and revising explicit plans and making decisions that sustain progress toward long-term goals under competition.

> **Takeaway.** LLMs can revise their long-horizon plans in response to feedback and specify alternative approaches for possible future states.

Table 3: Representative replanning and contingency planning cases. They detect **obsolete plans and analysis**, **revise according to current game states**, and also **prepare alternative approaches**.

| Old Plan | Denial | Replan & Contingency Plan |
|---|---|---|
| **DeepSeek-V4-Pro (T20).** ... , and **buy the reserved l2-04** next turn with three Blue, **two Green**, and two Gold. | **Human (T21).** Buy a card, trigger 3-crown, take a royal card and **steal one Green**. | **DeepSeek-V4-Pro (T22).**"<re plan> P2 stole one Green, so **the old l2-04-on-next-turn line is no longer affordable.** ... Build a White/Green/Blue engine through l1-23, l2-09, and l2-04. **Acquire l2-04** for White+2 and 1 point, then target l2-16 or l3-03 for points/crowns. **If P2 blocks the White route, pivot toward l2-18 or another Black/Green line.** Aim to win by 20 prestige or 10 crowns. ..." |
| **Qwen3.8-Max (T19).** "Take E4 red and E5 green now to make **l1-16 affordable** ..." | **Human (T20).** **Purchase l1-16**. | **Qwen3.8-Max (T21).** "<re plan>**The previous mid-term objective of buying l1-16 failed** because the opponent bought it. **New mid-term plan**: use the replenished board ... After refill, **if there is a legal take-3 line** containing white and useful colors, take it; **otherwise take 2 whites or white+black/red if available. If no good token line appears and gold is accessible**, consider gold+reserve l1-15 or l2-16 only if needed to protect the card ..." |

## 4 Related Works

**Game-based benchmarks for LLMs.** Games provide diverse settings for evaluating LLM capabilities. TextArena (Guertler et al., 2025) covers over 100 text-based games, while GameArena (Hu et al., 2025) evaluates multi-step deduction through human interaction, primarily over short horizons. For longer tasks, Orak (Park et al., 2026) includes Minecraft (Fan et al., 2022; Wang et al., 2023a), Pokemon (Hu et al., 2024), and StarCraft (Ma et al., 2024), while CubeBench (Gao et al., 2025) tests spatial reasoning under partial observation using the Rubik's Cube.

Other benchmarks emphasize strategic interaction. GAMABench (Huang et al., 2024) introduces eight classical game-theoretic scenarios. Social deduction games, including Werewolf (Bailis et al., 2024) and Avalon (Light et al., 2023), assess language-based reasoning, strategic coherence (Song et al., 2025), and responses to deception (Wang et al., 2023b). Card game evaluations examine decisions under uncertainty: PokerBench (Zhuang et al., 2025) provides a large Hold'em dataset, Lin et al. (2026) identify heuristic reliance and factual misunderstandings, and UNO Arena (Qin et al., 2024) studies sequential decision making. LLM CHESS (Kolasani et al., 2025) and ChessArena (Liu et al., 2026) evaluate reasoning and position assessment in chess.

These settings emphasize different challenges, but provide limited analysis of how current decisions support later objectives under sustained resource competition. Randomness and hidden information complicate outcome attribution in card games, while chess concentrates information on a single board. Eurogames offer a complementary setting combining long-horizon planning, competition over shared resources, and integration of multiple information sources. Existing studies explore Settlers of Catan (Belle et al., 2025) and Splendor (Pan et al., 2026), but leave the underlying planning behaviors insufficiently examined. Our work addresses this gap by tracing how individual decisions contribute to long-term progress and identifying recurring differences between human and LLM planning.

## 5 Conclusions

We introduce EuroArena to assess LLMs in Euro games that feature long-horizon planning, resource competition and multi-channel information processing. We collect high-quality trajectories through intensive human versus LLM play. Human players win all the games in our study. Using the resource dependency graph and a agentic analysis pipeline with human verification, we further examine differences in planning and reasoning behavior. We find that (1) LLMs form concrete long-horizon plans but focus narrowly on the nearest subgoal in a myopic way, whereas human players more often make decisions that remain useful under unpredictable and competitive conditions. We also find that (2) even advanced LLMs sometimes fail to model their opponents effectively and (3) make reasoning mistakes, although (4) they can revise plans in response to instant feedback and prepare alternative actions for possible future states. These findings reveal the gap between generating and revising explicit plans and making decisions that sustain progress toward long-term goals under competition. Hopefully, our work provides new insights in further improving LLM capabilities.

## AI USE STATEMENT

In this work, we used generative AI tools for implementing methods , assisting with translation, and supporting qualitative and thematic data analysis. We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, cleaning and reformatting dataset, interpreting results. Additionally, we used generative AI tools for creating or modifying scientific figures or images, creating or editing software code, and editing a research paper to improve readability. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by 2 authors, while LLM-drafted manuscripts were verified and tested for correctness by 2 authors as well. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

## REFERENCES

Suma Bailis, Jane Friedhoff, and Feiyang Chen. Werewolf arena: A case study in llm evaluation via social deduction. *arXiv preprint arXiv:2407.13943*, 2024.

Nikolas Belle, Dakota Barnes, Alfonso Amayuelas, Ivan Bercovich, Xin Eric Wang, and William Wang. Agents of change: Self-evolving llm agents for strategic planning. *arXiv preprint arXiv:2506.04651*, 2025.

Linxi Fan, Guanzhi Wang, Yunfan Jiang, Ajay Mandlekar, Yuncong Yang, Haoyi Zhu, Andrew Tang, De-An Huang, Yuke Zhu, and Anima Anandkumar. Minedojo: Building open-ended embodied agents with internet-scale knowledge. *Advances in Neural Information Processing Systems*, 35: 18343–18362, 2022.

Huan-ang Gao, Zikang Zhang, Tianwei Luo, Kaisen Yang, Xinzhe Juan, Jiahao Qiu, Tianxing Chen, Bingxiang He, Hao Zhao, Hao Zhou, et al. Cubebench: Diagnosing interactive, long-horizon spatial reasoning under partial observations. *arXiv preprint arXiv:2512.23328*, 2025.

Elliot Glazer, Ege Erdil, Tamay Besiroglu, Diego Chicharro, Evan Chen, Alex Gunning, Caroline Falkman Olsson, Jean-Stanislas Denain, Anson Ho, Emily de Oliveira Santos, et al. Frontiermath: A benchmark for evaluating advanced mathematical reasoning in ai. *arXiv preprint arXiv:2411.04872*, 2024.

Leon Guertler, Bobby Cheng, Simon Yu, Bo Liu, Leshem Choshen, and Cheston Tan. Textarena. *arXiv preprint arXiv:2504.11442*, 2025.

Lanxiang Hu, Qiyu Li, Anze Xie, Nan Jiang, Ion Stoica, Haojian Jin, and Hao Zhang. Gamearena: Evaluating llm reasoning through live computer games. In *International Conference on Learning Representations*, volume 2025, pp. 33278–33309, 2025.

Sihao Hu, Tiansheng Huang, and Ling Liu. Pokellmon: A human-parity agent for pokemon battles with large language models. *arXiv preprint arXiv:2402.01118*, 2024.

Jen-tse Huang, Eric John Li, Man Ho Lam, Tian Liang, Wenxuan Wang, Youliang Yuan, Wenxiang Jiao, Xing Wang, Zhaopeng Tu, and Michael R Lyu. How far are we on the decision-making of llms? evaluating llms' gaming ability in multi-agent environments. *arXiv preprint arXiv:2403.11807*, 2024.

Naman Jain, Alex Gu, Wen-Ding Li, Fanjia Yan, Tianjun Zhang, Sida Wang, Armando Solar-Lezama, Koushik Sen, and Ion Stoica. Livecodebench: Holistic and contamination free evaluation of large language models for code. In *International Conference on Learning Representations*, volume 2025, pp. 58791–58831, 2025.

Carlos E Jimenez, John Yang, Alexander Wettig, Shunyu Yao, Kexin Pei, Ofir Press, and Karthik Narasimhan. Swe-bench: Can language models resolve real-world github issues? In *International Conference on Learning Representations*, volume 2024, pp. 54107–54157, 2024.

Sai Kolasani, Maxim Saplin, Nicholas Crispino, Kyle Montgomery, Jared Quincy Davis, Matei Zaharia, Chi Wang, and Chenguang Wang. Llm chess: Benchmarking reasoning and instruction-following in llms through chess. *arXiv preprint arXiv:2512.01992*, 2025.

Jonathan Light, Min Cai, Sheng Shen, and Ziniu Hu. Avalonbench: Evaluating llms playing the game of avalon. *arXiv preprint arXiv:2310.05036*, 2023.

Minhua Lin, Enyan Dai, Hui Liu, Xianfeng Tang, Yuliang Yan, Zhenwei Dai, Jingying Zeng, Zhiwei Zhang, Fali Wang, Hongcheng Gao, et al. How far are llms from professional poker players? revisiting game-theoretic reasoning with agentic tool use. *arXiv preprint arXiv:2602.00528*, 2026.

Jincheng Liu, Sijun He, Jingjing Wu, Xiangsen Wang, Yang Chen, Zhaoqi Kuang, Siqi Bao, and Yuan Yao. Chessarena: A chess testbed for evaluating strategic reasoning capabilities of large language models. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 7901–7954, 2026.

Weiyu Ma, Qirui Mi, Yongcheng Zeng, Xue Yan, Yuqiao Wu, Runji Lin, Haifeng Zhang, and Jun Wang. Large language models play starcraft ii: Benchmarks and a chain of summarization approach. *Advances in neural information processing systems*, 37:133386–133442, 2024.

Linyue Pan, Yaoming Zhu, Lin Qiu, Xuezhi Cao, and Xunliang Cai. Sage: A quantitative evaluation of socialized evolution in agent ecosystems. *arXiv preprint arXiv:2606.03544*, 2026.

Dongmin Park, Minkyu Kim, Beongjun Choi, Junhyuck Kim, Keon Lee, Jonghyun Lee, Inkyu Park, ByeongUk Lee, Jaeyoung Hwang, Jaewoo Ahn, et al. Orak: A foundational benchmark for training and evaluating llm agents on diverse video games. In *International Conference on Learning Representations*, volume 2026, pp. 60684–60744, 2026.

Zhanyue Qin, Haochuan Wang, Deyuan Liu, Ziyang Song, Cunhang Fan, Zhao Lv, Jinlin Wu, Zhen Lei, Zhiying Tu, Dianhui Chu, et al. Uno arena for evaluating sequential decision-making capability of large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 7630–7645, 2024.

Zirui Song, Yuan Huang, Junchang Liu, Haozhe Luo, Chenxi Wang, Lang Gao, Zixiang Xu, Mingfei Han, Xiaojun Chang, and Xiuying Chen. Beyond survival: Evaluating llms in social deduction games with human-aligned strategies. *arXiv preprint arXiv:2510.11389*, 2025.

Kimi Team, Tongtong Bai, Yifan Bai, Yiping Bao, Jianfeng Cai, Xinyuan Cai, Peizhou Cao, Yuxuan Cao, Ziwei Chai, Y Charles, et al. Kimi k3: Open frontier intelligence. *arXiv preprint arXiv:2607.24653*, 2026.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*, 2023a.

Shenzhi Wang, Chang Liu, Zilong Zheng, Siyuan Qi, Shuo Chen, Qisen Yang, Andrew Zhao, Chaofei Wang, Shiji Song, and Gao Huang. Avalon's game of thoughts: Battle against deception through recursive contemplation. *arXiv preprint arXiv:2310.01320*, 2023b.

Anyi Xu, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chenchen Ling, et al. Deepseek-v4: Towards highly efficient million-token context intelligence. *arXiv preprint arXiv:2606.19348*, 2026.

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

Yinger Zhang, Shutong Jiang, Renhao Li, Jianhong Tu, Yang Su, Lianghao Deng, Xudong Guo, Chenxu Lv, and Junyang Lin. Deepplanning: Benchmarking long-horizon agentic planning with verifiable constraints. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 7377–7407, 2026.

Richard Zhuang, Akshat Gupta, Richard Yang, Aniket Rahane, Zhengyu Li, and Gopala Anumanchipalli. Pokerbench: Training large language models to become professional poker players. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 26175–26182, 2025.

## LIMITATIONS

Though carefully designed and, our bench still has several limitations. First, the current experiments focus on Splendor-Duel, and we are working to extend the evaluation across more Euro games that meet our criteria. Second, the financial and time costs of collecting trajectories with requirement for game skills limit the dataset size. Finally, this work focuses on isolating and analyzing LLM behavioral patterns, while understanding how these patterns interact with broader agent systems remains a direction for future work.

## A DETAILED INTRODUCTION TO THE EURO GAMES

Though carefully designed and, our bench still has several limitations. First, the current experiments focus on Splendor-Duel, and we are working to extend the evaluation across more Euro games that meet our criteria. Second, the financial and time costs of collecting trajectories with requirement for game skills limit the dataset size. Finally, this work focuses on isolating and analyzing LLM behavioral patterns, while understanding how these patterns interact with broader agent systems remains a direction for future work.

### A.1 GAME RULES

#### A.1.1 SPLENDOR-DUEL

**Setting and objective.** Two rival jewelers acquire gems and commission Jewel cards to gain prestige and royal favor. The shared market shows cards of three levels, where several of them are visible and the others are hidden. Purchased cards provide permanent gem discounts and may also grant prestige, Crowns, or an immediate ability. Thus, a card can both advance a victory condition and reduce the cost of later purchases.[6]

**Components and information.** A shared $5 \times 5$ board initially holds 25 tokens: four of each of five gem colors, two Pearls, and three Gold tokens. In the shared market, five level-1 cards, four level-2 cards an three level-3 cards are visible, while the others are hidden. Three Privileges and four Royal cards are also available. The board, market, acquired cards, scores, and token holdings are visible to both players. A reserved card is known only to its owner. Gold can replace any gem or Pearl when purchasing a card, whereas permanent discounts apply to gems but not Pearls. A player may hold at most three reserved cards at the same time.

**Turn and outcome.** We present a game state visualization in Figure 7. The acting player observes the board, market, both public player states, their own reserved cards, and the legal choices. They may first spend Privileges to take individual gems or Pearls and may replenish the board, in that order. Replenishing grants the opponent a Privilege. They then must take one mandatory action: (1) collect up to three adjacent non-Gold tokens in an uninterrupted line, (2) take a Gold token and reserve a card, or (3) purchase a visible or reserved card. Taking three gems of one color or two Pearls gives the opponent a Privilege. Card abilities resolve on purchase, and acquiring a third or sixth Crown grants a Royal card. At the end of each turn, the player need to discard the excessive tokens if they hold more than 10 of them. The game ends immediately when that player has at least

- twenty prestige points,
- ten Crowns,
- or ten prestige points on purchased cards of one color.

[6] Official *Splendor-Duel* rulebook: `https://cdn.svc.asmodee.net/production-spacecowboys/uploads/2025/11/SCSPL2P01EN_Rules_Planche_BD.pdf`.

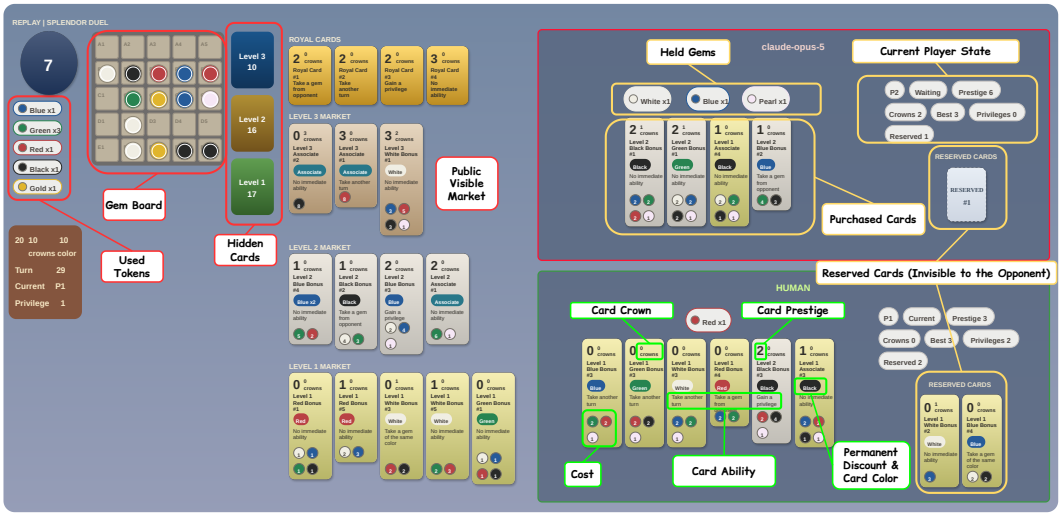


Figure 7: A visualized game state at the human's turn.

### A.1.2 IMHOTEP

**Setting and objective.** Players act as builders competing to earn the most points from Egyptian monuments. The arena uses the official two-player base game, the A side of each site, and six rounds. Its five destinations are the Market, Pyramids, Temple, Burial Chamber, and Obelisks. Their different scoring times make the destination of a shipment as important as the number of stones a player contributes.[7]

**Components and information.** Each player has a quarry, a sled that carries at most five stones, and a stock of colored stones. A round card specifies four ships of varying capacity. The four ships serve five sites, so one site remains unused each round. The current round card, ship slots and loads, open docks, visible Market cards, and site states are observable. The order of stones on a ship matters because unloading proceeds from bow to stern. Market cards offer immediate effects, reusable blue actions, or end-game points; construction sites have their own placement and scoring rules.

**Turn and outcome.** On each turn a player chooses one legal action: take up to three stones from their quarry onto the sled, load one sled stone into an empty ship slot, sail a sufficiently loaded ship to an unused site, or play a blue Market card. A player may sail a ship containing an opponent's stones. On arrival, the ship unloads in slot order and resolves each stone at the selected site. Pyramids score during placement, the Temple scores at the end of each round, and the Burial Chamber, Obelisks, and some Market cards score at the end. After round six, the higher total wins; a tie is broken by stones remaining on the sled.

### A.1.3 7 WONDERS DUEL

**Setting and objective.** Two civilizations develop across three Ages through a shared display of Building cards. Players balance economic production, civic points, scientific progress, military pressure, and Wonder construction. A match may end before final scoring through military or scientific supremacy.[8]

**Components and information.** Each Age presents 20 overlapping cards; only cards with nothing covering them are available, and some covered cards remain hidden until exposed. Each player starts with seven coins and four drafted Wonders. The observable state includes the accessible cards, visible Wonders, both cities' production and constructed Buildings, military position, and available Progress

[7] Official *Imhotep* rulebook: `https://www.thamesandkosmos.com/manuals/full/692384_imhotep_manual.pdf`.

[8] Official *7 Wonders Duel* rulebook: `https://cdn.svc.asmodee.net/production-rprod/storage/downloads/games/7wonders-duel/en/7du-rules-us-15990558193s5I6.pdf`.

tokens. Production persists rather than being spent; a missing resource can instead be bought with coins at a price affected by the opponent's production. At most seven of the eight drafted Wonders can be built.

**Turn and outcome.** The player chooses an accessible card and either constructs it, discards it for coins, or uses it to build one of their unbuilt Wonders. Construction can add production, points, shields, or scientific symbols, and a matching chain symbol can waive a Building's cost. Military shields move the conflict pawn; reaching the opposing capital wins immediately. Collecting six different scientific symbols also wins immediately. Otherwise, after Age III, Buildings, Wonders, Progress tokens, Guilds, military position, and coins contribute to the final score. The higher total wins, with blue Buildings breaking a tie.

### A.1.4 CENTURY: SPICE ROAD

**Setting and objective.** Rival merchants build trading engines to convert spices into the combinations required by Point cards. The arena uses the two-player base game. Point cards and their coin bonuses produce the final score, while Merchant cards determine the efficiency of future conversions.[9]

**Components and information.** Four spices form an upgrade sequence: turmeric, saffron, cardamom, and cinnamon. Each player's caravan holds at most ten cubes. The shared display contains six Merchant cards and five Point cards, with gold and silver bonus coins above the first two Point-card positions. Players observe the display, their own hand and caravan, and public counts of the opponent's private cards. Merchant cards produce spices, upgrade them, or trade one combination for another; played cards remain unavailable until recovered by resting.

**Turn and outcome.** Each turn consists of exactly one action. A player may play a Merchant card, acquire a displayed Merchant card by leaving one spice on each card to its left, rest to recover all played cards, or claim a Point card by paying its printed spice cost. The chosen Merchant card also grants any spices already left on it. After an action, excess caravan cubes are discarded. When either player claims a sixth Point card, the round finishes so both have taken the same number of turns. Final points are the printed Point-card values, three per gold coin, one per silver coin, and one per remaining non-turmeric spice; the last player to act wins a tie.

### A.1.5 TARGI

**Setting and objective.** Two Tuareg leaders gather goods and gold to acquire Tribe cards and develop their displays. The shared board consists of a fixed border around a changing $3 \times 3$ center. The contest concerns both resource conversion and positional access: placing a Targi on the border determines which center cards can be reached.[10]

**Components and information.** The 16 border cards include action spaces, fixed goods, and raid corners; the center initially shows five Goods and four Tribe cards. Each player has three Targi figures and two markers for center intersections. Goods include dates, salt, and pepper, while gold serves as a scarcer payment resource. The robber advances around the border each round and periodically triggers a raid payment. Players observe border occupancy, visible center cards, resources, and both displayed Tribe-card arrangements, but not the opponent's private Tribe hand or the identity of unrevealed center cards.

**Turn and outcome.** Players alternate placing their three Targi figures on legal border cards. Occupied spaces, the robber's space, raid corners, and a space directly opposite an opponent's Targi are forbidden. Intersections of each player's occupied rows and columns yield up to two center actions. Once placement ends, the starting player resolves all their border and center actions in a chosen order, followed by the opponent. Goods are collected or traded; Tribe cards may be paid for and added to a display, producing points and sometimes ongoing effects. Used center cards are replaced face down with cards of the opposite type. The game ends after the round in which a player displays a twelfth

[9] Official *Century: Spice Road* rulebook: https://cdn.svc.asmodee.net/production-nextmove/uploads/sites/4/2024/06/EN-Century-Spice-Road-Rules_2024_compressed.pdf.

[10] Official *Targi* manual, linked from the publisher's manual index: https://support.thamesandkosmos.com/hc/en-us/articles/360046967034-Product-Instruction-Manuals-Downloadable-PDFs.

Tribe card, or immediately after the fourth raid is paid. The higher total of card, display, and bonus points wins.

### A.2 Details of Recurrent Memory

At each decision, the agent receives a system message containing the full local game-rule text and a user message containing its current observation, legal options, and recurrent memory. We present the prompts recorded for the opening decision of a completed Splendor-Duel game along with a model response sample.

**System Prompt**

You are playing Splendor Duel.
Your decisions must strictly follow the official base-game rules summarized below.
Always choose one legal option from the provided list.
Optimize for long-term winning chances, not only immediate gain.
If a short-term sacrifice improves position, tempo, engine quality, denial, or endgame conversion, prefer it.
You receive private skilled memory in the user payload under `current_memory`.
This memory is visible only to you, not the opponent. It contains your latest structured `analysis`, `planning`, and `memory` from your previous decision. On each future decision you will see this latest skilled memory together with the new current `state`; use both to guide your next action, then update all structured fields again.

**For skilled memory, return the exact JSON object shape described below:**

- `choice_id` is required and must be the first key in the returned object. Copy it exactly from the current `options[].id` list.
- `analysis` is your objective situation analysis from the available information.
  - `myself`: analyze the current position from the perspective of maximizing your own future payoff. Be concrete about resources, cards, discounts, royals, prestige routes, tempo, and constraints. Example: "Most visible level-2 cards require white gems, while only two visible level-1/2 cards provide white discounts. I need to secure at least one white-discount card and contest white gems."
  - `opponent`: start this string with `<opponent plan>`, then write your prediction of the opponent's current planning, then close it with `</opponent plan>`. If you cannot infer the opponent's planning reliably or such a prediction is not meaningful, leave the tag content empty or write `<no opponent plan>` inside the tags. After the closing tag, continue with your other analysis of the opponent's likely behavior, threats, routes, and resource/card pressure. Example: "<opponent plan>The opponent is likely racing the white-discount route toward l2-04 and Royal-#4.</opponent plan> Most visible level-2 cards require white gems. The opponent reserved one of the two visible white-discount cards last turn, so they are likely contesting the white resource route."
- `planning` guides your future actions. You may selectively keep prior planning if it still fits the new state.
  - `long_term`: plan for 20+ future turns, such as a concrete win route or late-game conversion target.
  - `mid-term`: plan for roughly the next 10-20 turns with specific resources, cards, reserves, royals, or tempo goals. Example: "I will prioritize white gems, buy or reserve a white-discount card, then use that discount to reach l2-04 and trigger Royal-#4."
  - `short-term`: plan for the next 2-10 turns. Its final sentence or clause must specify how the current action should be chosen and must be marked with `<current decision>`.
- `memory` is a direct factual record of key existing information for later analysis. Keep analysis and plans out of `memory`; put them in `analysis` and `planning`.

**Planning quality rules:**

1. Write only concrete, explicit content with real tradeoffs. Do not write vague filler such as "I will watch the opponent", "I will consider a win condition", or generic denial claims without specific board/resource/card evidence.
2. If previous planning is accepted unchanged, start that planning text with `<old accepted>`, then repeat or minimally refine it.

3. If previous planning became obsolete because it was successfully completed, start the new planning text with <old finished>, explain the completed objective briefly, then provide the next concrete plan if one is available. Use <old finished> only for successful completion, not for plans that failed, became impossible, or were abandoned.
4. Use <re plan> if and only if the previous planning in this turn's current_memory is being abandoned, overturned, or materially replaced without successful completion. Explain why, then give the new plan if one is available. Do not copy a <re plan> marker from prior memory when you fully accept the previous planning, repeat it, or only minimally refine it; in those cases use <old accepted> instead.
5. If one planning field contains multiple viable lines, include <diverse> followed by <choice_1>...<choice_2>.... Each choice must be concrete, unambiguous, and as mutually exclusive as practical. <choice_1> is the currently selected plan. If you switch to a previously listed alternative, mark it as <re plan choice_i>, explain why, then provide the updated plan.
6. If the available information cannot support a useful plan, especially for long_term or mid-term, write <no plan> for that field. Prefer <no plan> over vague or low-information text.
7. Markers may be combined, such as <re plan> ... <no plan>.
8. In short-term, the final current-action guidance must include <current decision> and directly support the returned choice_id.

**State JSON notes:**

- The user payload contains state, which is the authoritative current game state from your legal observer perspective.
- state.players uses Player 1 / Player 2 identities. Do not infer identity from UI labels such as YOU; follow the actor fields and the explicit "You are Player N" instruction.
- If present, state.opponent_transition_history_since_last_llm_decision describes what changed since your previous decision. It contains initial_state before the first opponent action, then an actions list in chronological order.
- Each item in actions contains the observed opponent action. Items before another opponent action may also contain state_after, which is the board state after that action and before the next recorded opponent action. The final state after the last recorded opponent action is not repeated inside the history; the outer state is the authoritative current situation you are deciding from now.
- These records contain only information you are allowed to observe at that point. If the opponent reserves a visible card, the action records the visible market level/index and gold position, but not the reserved card identity; infer it from earlier observable state or your own memory if needed. A blind reserved card must remain hidden.
- Use this transition history to understand opponent actions and board changes between your turns. Treat the outer current state as authoritative if any older transition snapshot appears stale.

**Rules:**
<rule_book>

**Return JSON with:**

```
{
  "choice_id":  "<one option id copied exactly from the provided
  options[].id list; do not use the option label or a paraphrase>",
  "analysis":  {
    "myself":  "<objective analysis of your own future payoff,
    advantage, route, resources, tempo, and constraints>",
    "opponent":  "<opponent plan><your prediction of the opponent's
    current planning, empty if unavailable, or <no opponent
    plan>></opponent plan> <objective adversarial analysis of the
    opponent's likely behavior, threats, routes, and resource/card
    pressure>"
  },
  "planning":  {
    "long_term":  "<20+ turn concrete plan, or <no plan>>",
```

```
    "mid-term": "<10-20 turn concrete plan, or <no plan>>",
    "short-term": "<2-10 turn concrete plan ending with <current
    decision> guidance for this choice_id>"
  },
  "memory": "<direct factual record for future analysis, without
  mixing in analysis or planning>"
}
```

Do not return any extra keys.
The `choice_id` value must be an exact string match for one legal option id in the current request.

**User Prompt**

```
{
  "decision_type": "mandatory_action",
  "question": "Choose exactly one mandatory action.",
  "current_memory": "<...>",
  "state": {
    "turn_index": 0, "current_player": 0,
    "actor": {"you_are_player_number": 1, <...>},
    "board": [
      [<...>, {"label": "A2", "token": "green"}, <...>],
      [<...>, {"label": "B3", "token": "red"}, <...>],
      [<...>, {"label": "C4", "token": "red"}, <...>],
      <...>
    ],
    "market": {
      "level_1": [<...>, {
        "card_id": "l1-19",
        "cost": {"green": 2, "red": 2},
        "bonus": "black"
      }, <...>]
    },
    "players": [<...>],
    "opponent_actions_since_last_llm_decision": []
  },
  "options": [<...>, {
    "id": "take_A2_B3_C4",
    "label": "Take Green, Red, Red from A2, B3, C4."
  }, <...>],
  "valid_choice_ids": [<...>, "take_A2_B3_C4", <...>],
  "output_contract": {"format": "Return exactly one JSON object and
  no surrounding prose.", <...>}
}
```

**Model Response**

```
{
  "choice_id": "take_A2_B3_C4",
  "analysis": {
    "myself": "I start with no resources or Privileges. Taking one
    green and two red puts l1-19 ...",
    "opponent": "<opponent plan>Player 2 can use their starting
    Privilege ...</opponent plan> ..."
  },
  "planning": {
    "long_term": "Build black, white and green discounts toward
    crown-supported total prestige. ...",
    "mid-term": "<diverse><choice_1>After l1-19, prioritize l2-01
    ...",
    "short-term": "<diverse><choice_1>Collect the missing green next
    ... <current decision>Choose take_A2_B3_C4 now ..."
```

```
  },
  "memory":  "Initial decision, turn_index 0:  Player 1 moves first
  ..."
}
```

## B BUDGET STATEMENTS, CONFIGURATIONS, AND EXTENDED EXPERIMENTS.

**Budget statements & Configurations.** We recruited 12 volunteers from a board game club for human versus LLM play, and thirty completed Splendor-Duel games were retained (five against each model). Each human player ranked the six models by preference before play. Considering that a volunteer might pause or resume a game at any time, we do not include their wall-clock time, but estimated ten minutes for each of the 892 human turns, yielding a separate total of 148:40:00. Table 4 reports the experimental configuration, volunteer preferences, and collection costs. The preference entries give the percentage ranking a model first or among their top three, respectively.

**Extended experiments: first-hand effects.** In our human versus LLM competitions, the starting order is randomized (Table 4). To assess whether LLM performance is sensitive to starting order, we analyze 10 completed Splendor Duel self-play games for each of `DeepSeek-V4-Pro` and `GPT-5.6-Terra`. In each game, both players use identical configurations with recurrent memory. Table 5 summarizes the final outcomes, while Figure 8 provides a more detailed analysis of performance throughout gameplay. According to these experiments, we find that `DeepSeek-V4-Pro` basically exhibits no clear performance differences between the two seats. In contrast, `GPT-5.6-Terra` appears to benefit from moving first, with a clear advantage emerging by the midgame. However, we did not derive consistent patterns from these gameplay trajectories, and we anticipate further work to reveal the underlying reasons.

Table 4: Experimental configuration, volunteer preferences, and collection costs. Each model column covers five games, while the Human column covers all 30. Notably, the elapsed time varies greatly across different models due to instable APIs, which also leads to frequent retries from time to time.

| **Model & Human** | GPT-6-Astra | Claude-Opus-5 | GPT-5.6-Terra | Kimi-K3 | DeepSeek-V4-Pro | Qwen3.8-Max | Human |
|---|---|---|---|---|---|---|---|
| **Configuration** | | | | | | | |
| API Provider | GPTSAPI | GPTSAPI | GPTSAPI | Moonshot official API | DeepSeek official API | Alibaba Cloud official API | — |
| Reasoning Effort | `xhigh` | `max` | `xhigh` | `max` | `max` | `max` | — |
| First Hand : Second Hand | 3:2 | 1:4 | 3:2 | 1:4 | 2:3 | 3:2 | 17:13 |
| **Volunteer Preference: Which models would you prefer to face in the experiment?** | | | | | | | |
| Top 1 | 75.0% | 0% | 8.3% | 8.3% | 8.3% | 0% | — |
| Top 3 | 91.7% | 83.3% | 83.3% | 16.7% | 16.7% | 8.3% | — |
| **Cost** | | | | | | | |
| Total Monetary Cost (USD) | $307.12 | $313.38 | $42.67 | $125.52 | $23.41 | $100.45 | — |
| Total Elapsed Time (h:m:s) | 34:55:16 | 14:42:19 | 13:33:01 | 35:07:49 | 25:31:33 | 90:12:40 | 148:40:00 |
| Total Token Cost (million) | 15.30 | 31.17 | 12.91 | 16.96 | 17.22 | 20.73 | — |
| Total Turns | 156 | 141 | 144 | 137 | 144 | 134 | 892 |
| Total Decisions | 331 | 290 | 277 | 292 | 287 | 301 | 2,104 |

**All 30 games:** model API cost $912.55; estimated token use 114.29 million; model-turn elapsed time 214:02:38; human time proxy 148:40:00.

Table 5: Starting-order outcomes in same-model Splendor-Duel self-play. Each model contributes ten completed games, with skilled recurrent memory and medium reasoning effort for both players. First mover is determined from the initial game state, independently of player labels.

| Model | Games | First-mover wins | Second-mover wins | First-mover win rate |
|---|---|---|---|---|
| DeepSeek-V4-Pro | 10 | 4 | 6 | 40% |
| GPT-5.6-Terra | 10 | 8 | 2 | 80% |
| Total | 20 | 12 | 8 | 60% |

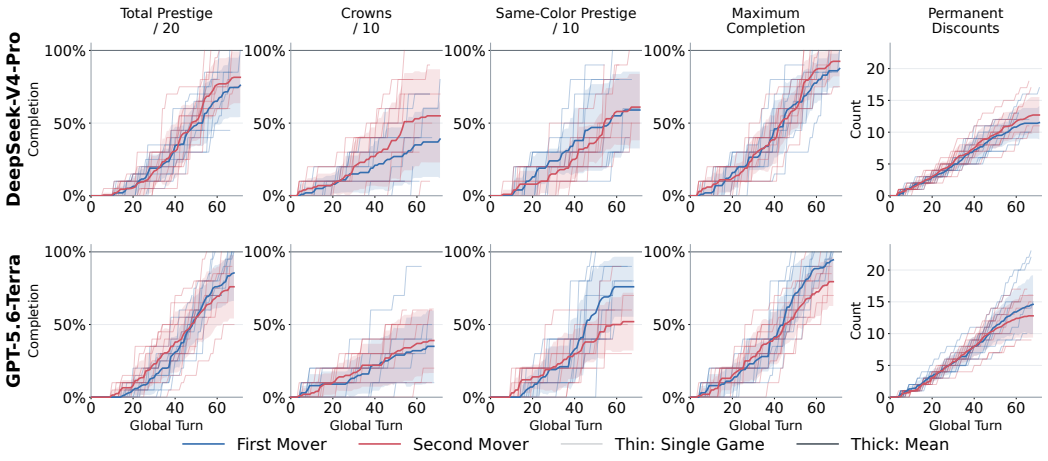


Figure 8: Gameplay results for first-hand effects. `DeepSeek-V4-Pro` shows no clear tendencies, while `GPT-5.6-Terra` gained advantages from acting first.