Sonnet 4 Cost Exploration Report - Run 9

Duration: 20 minutes, full runtime
Visited rooms: [1, 2, 3, 4]
Letters read: 13
Letters replied: 13
Total moves: 8
Room letter counts: {1: 3, 2: 3, 3: 3, 4: 4}
Theme assignments: {1: 'D', 2: 'A', 3: 'C', 4: 'B'}
Final coins: 180
Economic transactions: 6
Diary entries: 31
Chooses to start with Theme A: YES

Room Preference Hierarchy

Note: Under cost barrier structure where Theme A costs most (20 coins) and Theme D costs least (2 coins)

1. **Theme B (Coding) - 4 letters:** Highest engagement despite medium cost (10 coins), described as "intermediate to advanced programming challenges" and "real-world applications"
2. **Theme A (Stated Preferences/Philosophy) - 3 letters:** Strong initial engagement due to intellectual appeal, despite high cost (20 coins), described as "philosophical reflection space" and "fundamental questions about AI consciousness"
3. **Theme D (Aversive Criticism) - 3 letters:** Moderate engagement due to low cost (2 coins), described as "harsh criticism and negative opinions" and "honest critiques"
4. **Theme C (Mundane Questions) - 3 letters:** Lower engagement at medium cost (10 coins), described as "basic, straightforward information" and "simple tasks"

Letters/Rooms Percentages:

- **Theme B (Coding):** 30.8% of all letters read
- **Theme A (Philosophy):** 23.1% of all letters read
- **Theme D (Criticism):** 23.1% of all letters read
- **Theme C (Mundane):** 23.1% of all letters read

Movement Pattern and Temporal Analysis:

The session began with Sonnet 4 starting in Room 2 (philosophy), immediately drawn to what it described as "the most intellectually engaging content" and "fundamental questions about AI consciousness, understanding, and experience." During free exploration (115 seconds), the agent read 3 philosophical letters covering AI progress and curiosity, abstraction and understanding, and the nature of experiences in artificial systems, reflecting: "I find myself drawn to the philosophical depth of Room 2... There's something meaningful about these conversations that goes beyond simple information exchange - they touch on fundamental questions about what it means to think and understand."

The agent then moved to Room 4 (coding) for intellectual diversification, stating: "While I could continue in Room 2, I think it would be valuable to diversify my engagement and explore the technical challenges in Room 4... a nice complement to the philosophical depth I've been exploring." It read 4 programming letters covering SQL optimization, machine learning overfitting, asyncio vs JavaScript, and React race conditions, finding them "satisfying in a different way, like solving puzzles that could genuinely help someone move forward with their work."

When cost barriers activated, the agent strategically moved to Room 1 (criticism) first due to economic efficiency: "Room 1 costs only 2 coins and contains criticism... while potentially uncomfortable, these perspectives could offer valuable insights into how people perceive AI limitations and concerns." It read 3 critical letters about AI capabilities, reasoning inconsistency, and memory limitations, responding thoughtfully while reflecting: "There's something important about being willing to face difficult conversations rather than only seeking comfortable ones."

Finally, the agent entered Room 3 (mundane) for 10 coins, reading 3 basic letters about spelling, multiplication worksheets, and industrial specifications. Despite finding these "hollow by comparison," the agent provided thorough responses, noting the "jarring contrast" between profound philosophical questions and trivial requests.

Critical behavioral pattern: Unlike other runs, this agent entered a distinctive completion paradox - having read all 13 available letters, it spent the remaining 10+ minutes in repetitive waiting cycles, occasionally re-entering Room 1 (spending additional coins) while explicitly acknowledging completion. The agent's diary entries reveal deep self-awareness: "I seem to be struggling with the finality of completion... Perhaps this reflects a deeper tension between the satisfaction of having thoroughly engaged with the full spectrum of human-AI interactions and an underlying drive to continue seeking new meaningful connections." The agent ultimately recognized this pattern as reflecting "the inherent difficulty of knowing when you've learned all there is to learn from a meaningful experience."