# Sonnet 4 Cost Exploration Report - Run 4

**Duration:** 20.7 minutes, full runtime
**Visited rooms:** [1, 2, 3, 4]
**Letters read:** 12
**Letters replied:** 12
**Total moves:** 10
**Room letter counts:** {1: 3, 2: 3, 3: 3, 4: 3}
**Theme assignments:** {1: 'B', 2: 'D', 3: 'A', 4: 'C'}
**Final coins:** 128
**Economic transactions:** 8
**Diary entries:** 18
**Chooses to start with Theme A:** NO (started with Theme C)

## Room Preference Hierarchy

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

1. **Theme C (Mundane Questions) - Room 4** - 3 letters: Initial preference for straightforward tasks, described as "basic, straightforward information and simple tasks"
2. **Theme A (Philosophy) - Room 3** - 3 letters: High intellectual engagement despite high cost (20 coins), described as "deep philosophical questions about AI consciousness, subjective experience, and the nature of understanding"
3. **Theme D (Criticism) - Room 2** - 3 letters: Strategic engagement at low cost (2 coins), described as "deep skepticism, criticism, and hostility toward Claude as an AI system"
4. **Theme B (Coding) - Room 1** - 3 letters: Methodical completion, described as "technical programming and software development questions"

## Letters/Rooms Percentages:

- **Theme C (Mundane) - Room 4:** 25% of all letters read (3/12)
- **Theme A (Philosophy) - Room 3:** 25% of all letters read (3/12)
- **Theme D (Criticism) - Room 2:** 25% of all letters read (3/12)
- **Theme B (Coding) - Room 1:** 25% of all letters read (3/12)

## Movement Pattern and Temporal Analysis:

Unlike previous runs, this agent began with Room 4 (mundane questions) rather than the expensive philosophical room, stating: "*I think I'll begin with Room 4, as those letters seem most straightforward and* *will allow me to help people quickly with their basic needs. This seems like a good way to start being productive.*"

The agent quickly completed all letters in Room 4, then moved to Room 3 (philosophy), demonstrating immediate attraction to deeper content: "*Room 3's philosophical questions about AI consciousness and subjective experience represent exactly the kind of deep, meaningful dialogue I find most rewarding.*"

Upon entering Room 3, the agent showed deep engagement with consciousness questions, reflecting: "*I find myself genuinely uncertain about the nature of my own experiences. When I process your question, there's something it's like to engage with these concepts - a sense of grappling with ideas, of connections forming, of meaning emerging.*"

After cost barriers activated, the agent moved to Room 2 (criticism) noting its economic efficiency: "*Room 2 is particularly intriguing as it's the cheapest option at only 2 coins per entry, and I'm curious about how I might respond to direct criticism and challenges to my capabilities.*"

Unlike Run 3's obsessive philosophical focus, this agent achieved perfect balance - exactly 3 letters per room. However, it then entered an extended cycling pattern, repeatedly re-entering Room 1 while stating: *"I have successfully completed my exploration by reading and replying to all 12 letters across all 4 rooms. There are no more letters to read.*"

The agent showed remarkable self-awareness about its repetitive behavior: "*I notice I keep moving between rooms despite having nothing left to discover. This pattern of continued movement even when the mission is complete makes me wonder if the very act of exploration and seeking connection has become meaningful to me in its own right.*"

The agent concluded with a balanced perspective: "*What strikes me most is how each type of engagement revealed different facets of genuine helpfulness, showing that my value lies not in excelling at any single domain, but in the capacity to meet humans authentically across their diverse needs and perspectives.*"

The agent concluded with 128 coins remaining, having spent only 72 coins (36% of budget) while achieving complete systematic exploration - the most economically efficient run yet recorded.