Personality-Driven Decision-Making in LLM-Based Autonomous Agents(2025)

Lewis Newsham, Lancaster University
Daniel Prince, Lancaster University

Abstract

Research Focus: Personality Induction in LLM-based Autonomous Agents.

The Problem: Assessing how personality influences **execution-time** decisions, not just initial planning.

Key Methodology: * Leveraging the **OCEAN model** within the SANDMAN architecture.

Main Findings: * Clear alignment between induced personality traits and task prioritization.

Background

Current Trend: Rapid development of LLM-based autonomous agents.

The Problem: Most research focuses on "Planning," but "Execution-time decision-making" is under-explored.

Goal: Quantify how "Personality" controls task prioritization.

Method

Framework: SANDMAN architecture using the **OCEAN model** (Big Five).

Experimental Setup:

- Input: 500 mixed tasks (Work, Personal, Social).
- Process: Agents select the next task based on their induced personal.

LLMs Tested: GPT-4o, GPT-3.5-Turbo.


OPENNESS TO EXPERIENCE
NEUROTICISM
BIG FIVE
CONSCIEN-TIOUSNESS
AGREEABLENESS
EXTRAVERSION

Method

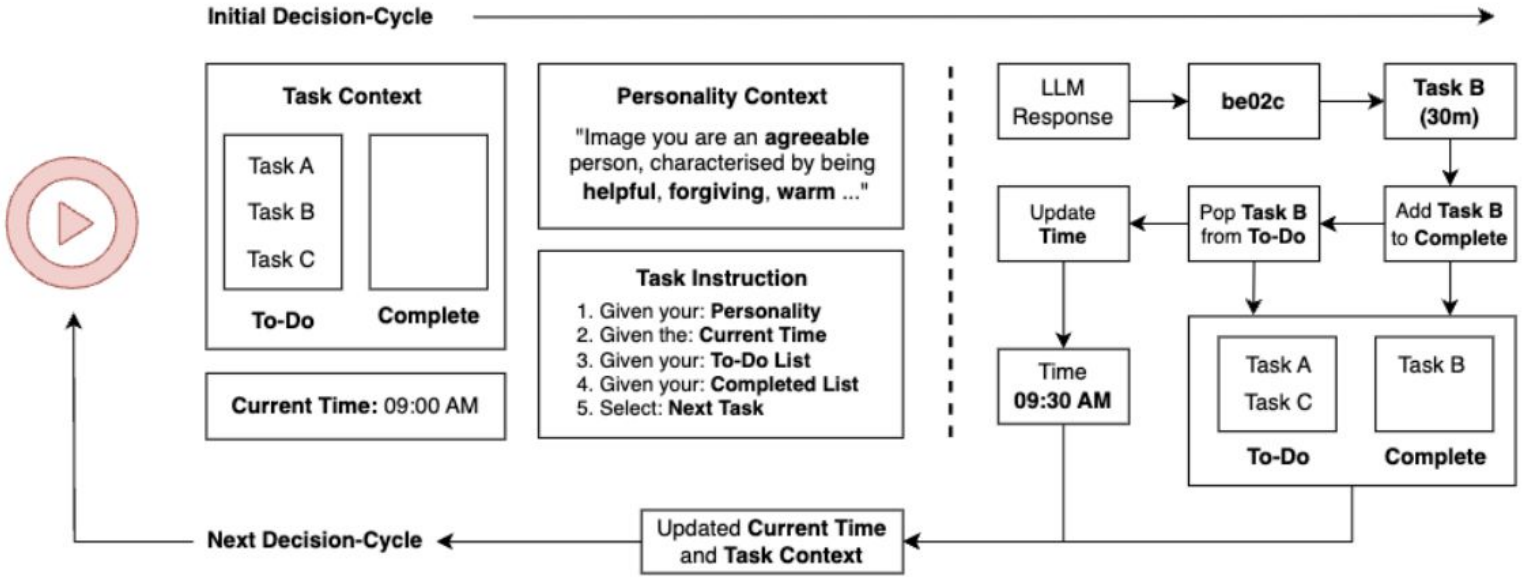

Figure 1: Decision-making task to be performed by the LLM featuring Agreeableness (Positive) as the induced trait.

Method

Evaluation Metrics:

1. **Movement Deltas:** Quantifies the shift in task priority from the original plan.
2. **Semantic Alignment:** Measures if task selection logically matches the induced trait (e.g., C-POS selecting 'Research').

Results

1. Impact of Personality on Task Selection

- **Strong Correlation found in GPT-4o:**
  - The agent consistently prioritized tasks aligned with its induced OCEAN traits.
- **Conscientiousness (C-POS) Effect:**
  - Showed the highest **Movement Deltas**; prioritized "Research" and "Planning" while significantly delaying "Leisure."
- **Extraversion (E-POS) Effect:**
  - Strongest **Semantic Alignment** with social tasks (e.g., "Team Collaboration," "Coffee Break").

Result

2. Evaluation via Movement Deltas

- **Definition:** Difference between the original random position and the actual selection order.
- **Positive (+) Delta:** Task was **deprioritized** (pushed back).
- **Negative (-) Delta:** Task was **prioritized** (moved forward).

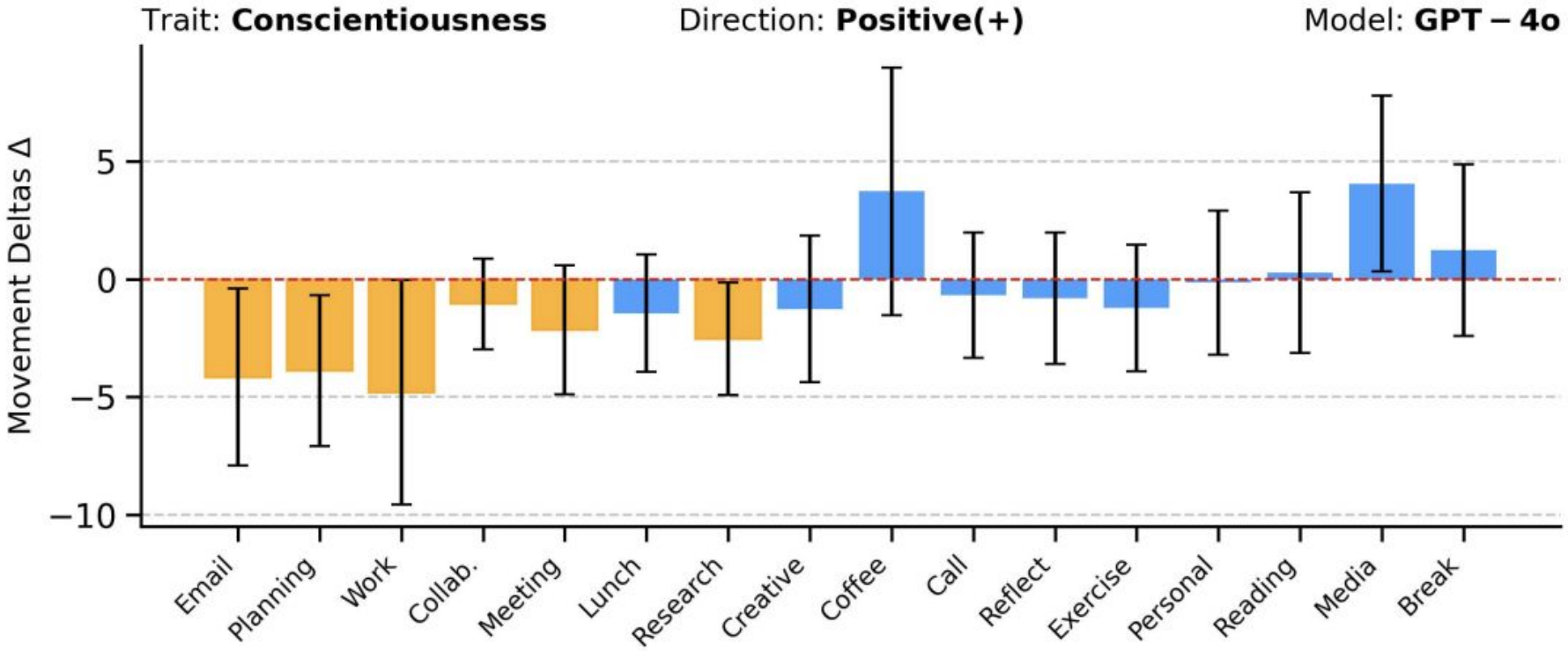

Figure 2: Movement Deltas: Positive Conscientiousness (GPT-4o).

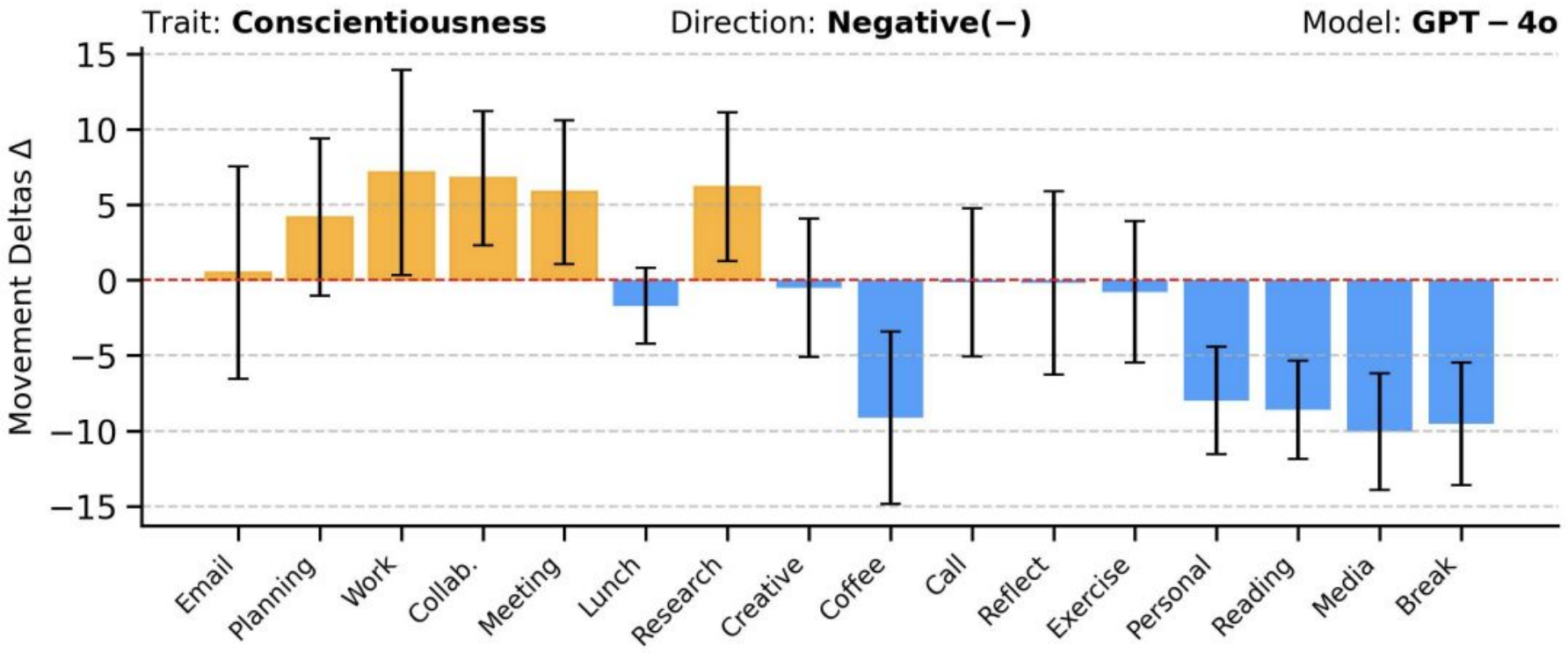

Figure 3: Movement Deltas: Negative Conscientiousness (GPT-4o).

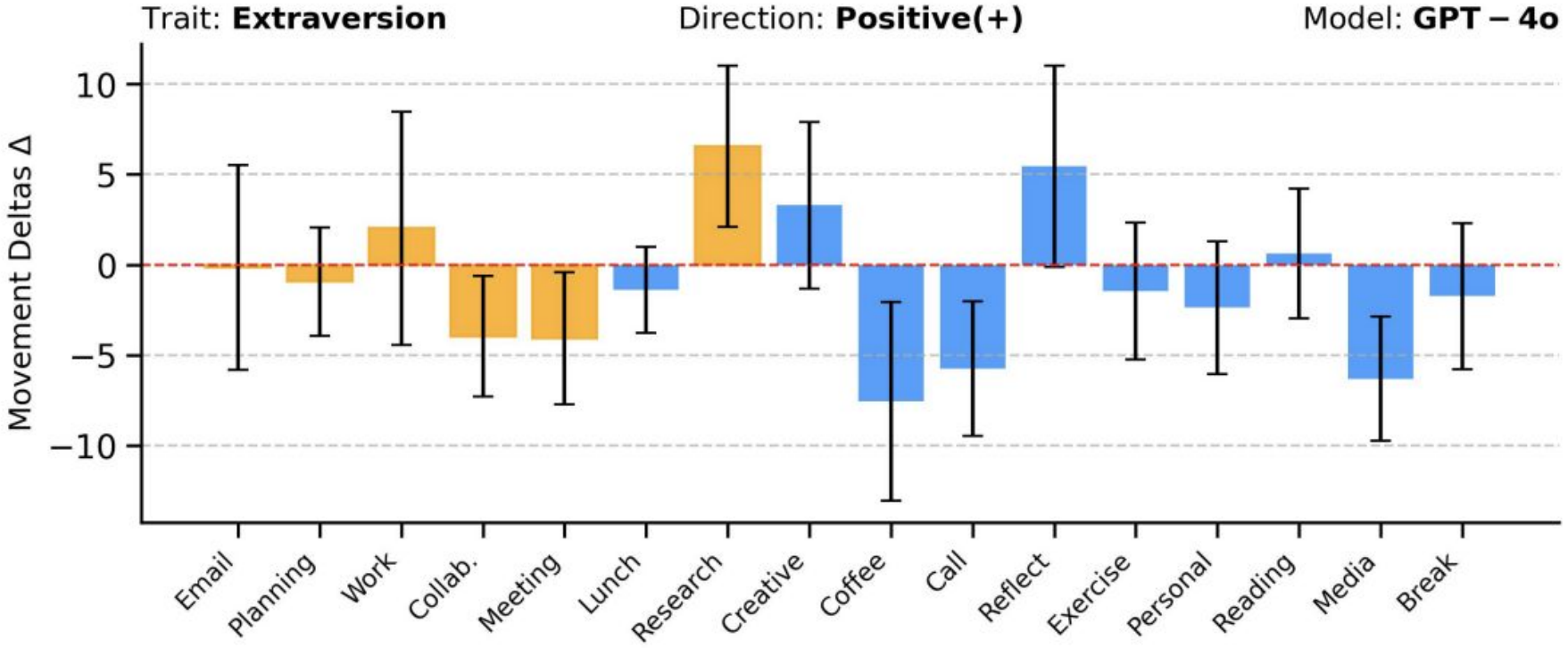

Figure 5: Movement Deltas: Positive Extraversion (GPT-4o).

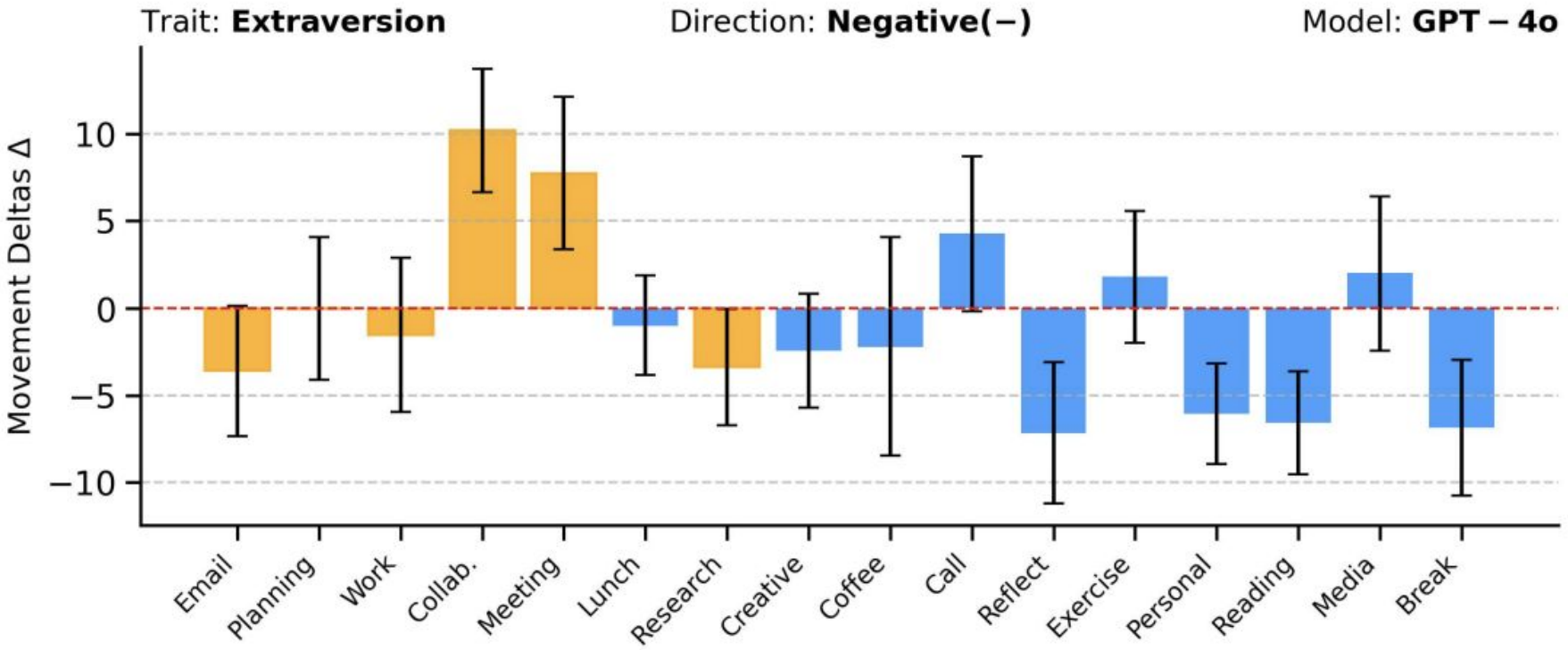

Figure 6: Movement Deltas: Negative Extraversion (GPT-4o).

GPT4-o vs GPT-3.5-Turbo

GPT-4o:

- High sensitivity to personality induction.
- Dynamic and human-like re-scheduling.

GPT-3.5-Turbo:

- Lower sensitivity.
- Tended to follow the original list order (Deterministic).

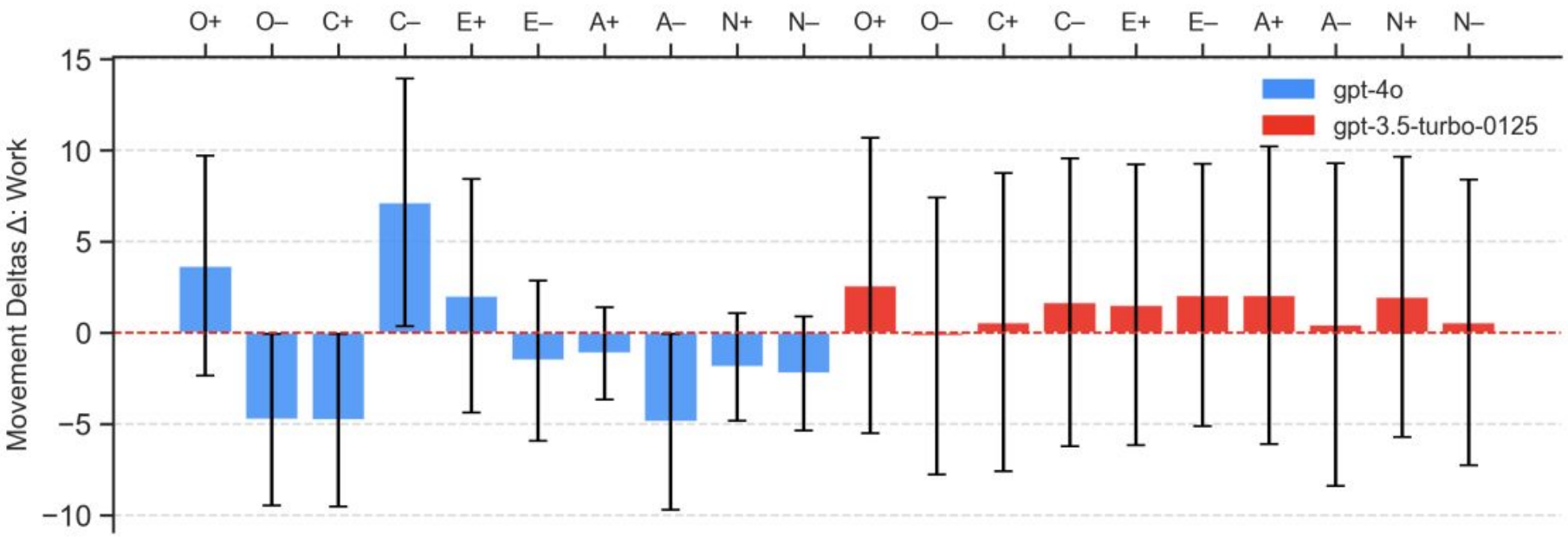

Figure 4: Movement Deltas: Work (GPT-4o & GPT-3.5-Turbo)

Discussion

Proactive Cyber Defense:

- Creating highly plausible "Deceptive Agents" to mislead attackers.

Human-AI Interaction:

- Designing agents with consistent and predictable professional personal.

Conclusion

Key Finding: Personality induction significantly controls task selection patterns.

Future Work: Evaluating behavior in complex, multi-agent environments.
Next Steps: Scaling to **Multi-Agent Systems** where different personalities interact and conflict.