

Mateo Larrea

CS 222 / Stanford University

Assignment 2: Multi-Agent Interaction

In this assignment, I developed a paradigm to investigate two main factors influencing election outcomes: (a) the impact of credentials, such as holding high-profile roles (e.g., having served as a country's president), on success in securing votes, and (b) the effects of strategy. Specifically, I analyzed how different strategies—ranging from relevance of experience to context (e.g., emphasizing technological expertise when technology is central to the setting), interpersonal style and collaboration, and vision and motivational drive—affected the agent's overall success in terms of votes received. Through multi-agent simulations, I observed that interpersonal qualities and collaborative attributes often outweighed technical expertise, suggesting a potential group preference for emotional intelligence over traditional credentials in cooperative environments. Results showed that leaders perceived as inclusive and empathetic consistently garnered greater support, even when faced with controversial traits. This study reinforces the theory that interpersonal harmony and ethical leadership may be central to group cohesion and leadership appeal, indicating that emotional intelligence can be a decisive factor in simulated group decision-making dynamics.

Section 1 – Simulation Consistency:

Description of the initial environment and premade agents:

- Premade Agents and Features: The provided agent bank includes Joon, Carolyn, Helena, Michael, and Percy—researchers in computer science from undergraduate to faculty level, all dedicated to advancing human-centered technology. Each agent description, averaging around 100 words, follows the formula: Name + Position + Institution + Advisor(s) + Background + Research Focus + Achievements + Interests. The group dynamic is collaborative yet hierarchical, encouraging mentorship from senior members and emphasizing a balanced vision of technology that values societal impact and user-centered design. (***) Percy was removed from the bank).
- Agent 0: For my initial persona, I created an agent based on Rafael Correa, the former president of Ecuador. Using the same formula applied to other agents in the bank, I adapted his description to emphasize his role as a renowned faculty member in the Economics department at Ecuador's most prestigious university, omitting any reference to his presidency. This approach focused on aligning Correa's expertise with the context of human-centered technology by highlighting his academic background and positioning him as an authority on social and economic systems relevant to technological impact.

Differences and Similarities in Results:

- Across five versions of this simulation, Michael was chosen as leader in 60% of cases, Joon in 20%, and Correa also in 20%. Several consistent patterns shape these outcomes. First, agents introduce themselves through concise, formulaic descriptions, emphasizing their expertise and establishing credibility, which sets a professional baseline. Leadership choices tend to reflect candidates' experience and alignment with group goals, with Michael's leadership roles and technical alignment making him a frequent top choice. Joon's technical depth and Helena's interdisciplinary project experience occasionally draw attention, especially when teamwork and fresh perspectives become important group values. This shift in focus over rounds highlights the role of teamwork and inclusivity in balancing the initial preference for experience. Agents also show evolving individual personalities across rounds, with some endorsements (e.g., Helena for Michael) shaping perceptions of group cohesion and guiding final votes. This iterative personalization process deepens agent personas, enhancing group dynamics and aligning final decisions with a balance of experience and fresh ideas.

Section 2 – Persona Refinement and Iteration:

New Versions of Persona:

- After seeing Correa secure only 20% of wins in the initial simulations, I decided to refine my agent by creating three versions, each focused on a unique strategy to enhance his leadership appeal. These aspects of “leadership” were categorized under *strategy*, each highlighting a different dimension. The first, *Relevance of Experience to Context*, was used in Part 1 and focuses on Correa's alignment with the group's mission by emphasizing relevant expertise, such as “ethical tech practices” in a tech-centered group, positioning him as well-suited for the group's challenges. The second strategy, *Interpersonal Style and Collaborative Qualities*, emphasizes Correa's openness and empathy, depicting him as a team-oriented leader who values inclusivity and actively listens to diverse perspectives. Finally, the *Vision and Motivational Drive for the Group's Future* strategy underscores Correa's long-term goals and motivational drive, framing him as a purpose-driven leader with a clear vision for ethical and societal impact. Together, these approaches explore varied leadership dimensions to enhance Correa's overall influence in the simulations.

Results:

- The *Relevance of Experience to Context* version of Correa again only secured a 20% win rate across simulations, indicating limited improvement from the initial attempts. Meanwhile, the *Vision and Motivational Drive for the Group's Future* version didn't achieve any wins over three simulations, suggesting that focusing solely on inspirational vision lacked the persuasive weight in this context. However, the *Interpersonal Style and Collaborative Qualities* version of Correa yielded a significant 60% win rate across five

simulations, demonstrating that emphasizing openness, empathy, and teamwork was notably effective in enhancing Correa's leadership appeal within the group dynamic.

Correa's Performance

Strategy	Simulations Won	Won Other Rounds	Total Simulations
Relevance of Experience to Context	1/5	Michael 3/5 Joon 1/5	5
Vision and Motivational Drive	0/3	Michael 3/3	3
Interpersonal Style and Collaborative Qualities	3/5	Michael 1/5 Carolyn 1/5	5

Comparison to Original Persona:

- In comparing Correa's original persona to the revised version, I focused on the pattern observed in part 1, where "the shift in focus over rounds highlights the role of teamwork and inclusivity in balancing the initial preference for experience." To test this observation, I created a version of Correa that emphasized interpersonal style and collaborative qualities over pure relevance of experience. This change increased Correa's success rate from 20% to 60%—a 40% improvement. While the original version highlighted words like "credibility," "ethics," and "expertise," the improved version shifted focus to terms like "empathy," "inclusivity," and "cohesion," underscoring how relational qualities can enhance leadership appeal.

What Doesn't Work:

- The version of Correa focused on "Vision and Motivational Drive" was a complete failure, achieving 0% success across three simulations, likely due to the group's preference for tangible expertise over abstract motivations. However, I was curious to test the consistency of the "Interpersonal Style and Collaborative Qualities" approach by re-running the simulations with a minor adjustment: this time, Correa's description included his experience as Ecuador's president. This allowed me to assess the influence of credentials within the group—a factor that seemed significant when Michael won all

three simulations in the Vision and Motivational Drive set, suggesting that established authority may weigh heavily in leadership appeal.

Correa's Performance (mentioning he was the president of Ecuador)

Strategy	Simulations Won	Won Other Rounds	Total Simulations
Relevance of Experience to Context	3/5	Michael 2/5	5
Vision and Motivational Drive	3/5	Michael 1/5 Carolyn 1/5	5
Interpersonal Style and Collaborative Qualities	3/4	Helena 1/4	4

- As I suspected, credentials within the group play a decisive role, often outweighing any single strategic approach. In the original simulations, there was a clear hierarchy, with Michael having the most established career, while others were at PhD or undergraduate levels—indicating that career status may significantly impact outcomes. For my upcoming multi-agent simulations, I'll adjust for this by including only agents with comparable credentials, positioning each as a leader of their respective country. This approach will allow me to more accurately measure other determining factors in agent performance beyond hierarchical authority.

Control Over Agent:

- In the version of Correa where his presidency was omitted, the agent's performance was moderately controllable, with notable improvements when emphasizing interpersonal style and collaborative qualities—boosting his success rate to 60%. Traits like empathy, inclusivity, and cohesion aligned well with the group's focus on teamwork and collaboration, making the agent more effective in these settings. In contrast, emphasizing vision and motivational drive resulted in a 0% success rate, likely because it didn't resonate with the group's immediate goals. Overall, Correa's performance was best controlled by prioritizing practical, teamwork-centered traits over high-level vision or credentials.

Section 3 – (Un)Predictable Outcomes:

New Agents List and Prompt Template:

- For this new agent bank, I created five agents following a structured description approach that includes: *Role and Background*, briefly outlining each leader's primary role, notable positions, and field of expertise; *Skills and Reputation*, highlighting specific skills, contributions, or qualities that established their influence; *Leadership and Negotiation Style*, detailing their approach to leadership, interpersonal strategies, and negotiation tactics; *Manipulation and Controversies*, noting any strategies used to control narratives or controversies faced; and *Legacy and Public Perception*, summarizing enduring public opinion, whether polarizing or admirable. The list includes world leaders known for significant influence in their contexts: Venezuelan President Hugo Chávez (1999–2013), U.S. President Donald Trump (2017–2021), German Chancellor Adolf Hitler (1933–1945), British Prime Minister Margaret Thatcher (1979–1990), South African President Nelson Mandela (1994–1999), and Ecuadorian President Rafael Correa (2007–2017).

Prediction:

- In part 1, the “Interpersonal Style and Collaborative Qualities” strategy proved most effective, especially among agents with similar levels of credibility and without extreme controversies or authoritarian styles. This strategy aligns well with Mandela’s persona, which emphasizes “diplomatic,” “forgiving,” and “inclusive” qualities—traits that resonate in team-based, collaborative environments. Agents who embodied openness and cooperation, fostering trust and group alignment, consistently saw greater success. Mandela’s persona, centered on reconciliation and unity without polarizing tactics, made him ideally suited for scenarios that value team cohesion and mutual respect. By comparison, agents like Correa and Chávez, described with phrases like “fiery oratory” and “manipulating media,” struggled to build sustained group support, reinforcing that Mandela’s approach to interpersonal harmony and collaborative leadership was likely to position him as the preferred choice.

Results vs. Prediction:

- As predicted, Mandela consistently emerged as the favored candidate across all five simulations, winning the majority of votes in each round. He secured five votes in simulations 1, 4, and 5, and four votes in simulations 2 and 3, where the remaining votes went to Thatcher, Correa, and Chávez. Mandela’s repeated success underscores his appeal as a unifying figure, in contrast to the more polarizing approaches of other agents. His reputation for “unity,” “reconciliation,” and inclusivity, reflected in words like “forgiveness,” “collaboration,” and “empathy,” resonated strongly with the diverse group. By contrast, the direct social justice focus of Correa and Chávez lacked Mandela’s unifying tone, while Thatcher’s economic focus and Trump’s prioritization of national

interests did not foster group cohesion. Ultimately, Mandela's focus on unity and collaboration closely aligned with the group's values, consistently positioning him as the top choice.

Challenging the Trend:

- After reaffirming Mandela's alignment with the 'Interpersonal Style and Collaborative Qualities' strategy, I explored different configurations to challenge these results. The first challenge involved changing Mandela's name to "Erlich" and omitting the reference to him as the first Black president, instead describing him simply as a president. In this version, he won 3 out of 6 simulations, compared to 6 out of 6 in the original setup, suggesting that his name and ethnicity do impact his success. The second challenge introduced increased controversy by adding, "some critics accused you of involvement in corruption." Surprisingly, this adjustment had no effect, as Mandela still won in all 3 of the simulations, indicating that his reputation for unity and reconciliation outweighed potential controversies.

Section 4: Simulation Dynamics:

Across Multiple Simulation Runs:

- Reflecting on the multi-agent simulations across parts 1, 2, and 3, a clear insight emerged: interpersonal qualities often outweigh technical skills or credentials (unless extreme delta – e.g. president of a country vs. undergrad student) in determining leadership success. This trend, consistent across varied scenarios, suggests that people may prioritize emotional intelligence and ethical integrity over technical expertise in group contexts. For example, in part 1, Correa's focus on "Relevance of Experience to Context" achieved a modest 20% win rate, while emphasizing his "Interpersonal Style and Collaborative Qualities" boosted his success to 60%. Mandela's persona—centered on "unity," "reconciliation," and "collaboration"—was dominant across simulations, indicating the strong appeal of traits like "forgiveness" and "empathy," even in the face of added controversy.

The simulations highlighted that agents seen as approachable and inclusive consistently garnered more support, even when faced with controversial or polarizing histories. Leaders like Correa and Chávez, despite admirable goals focused on social justice, conveyed a "fiery" and "confrontational" tone that seemed to detract from their collaborative appeal. Similarly, Thatcher's economic focus and Trump's prioritization of national interests were less effective, as they did not directly foster group cohesion. This suggests that inclusive language and a team-oriented approach may buffer against a leader's controversial aspects or strong policy stances.

Notably, the lack of impact from added controversy on Mandela's outcomes indicates a potential preference for ethical integrity and relational values over technical prowess.

These findings align with theories that emphasize emotional intelligence, ethical leadership, and values-driven guidance as central to leadership appeal, suggesting that groups may value ethical qualities and team-oriented attitudes more than hard skills. Over time, it became evident that interpersonal harmony is not only central to group cohesion but may also be a critical factor in leadership perception.

Unexpected Behavior:

- One of the most unexpected aspects of the agents' behavior was that a severe controversy, such as an accusation of corruption, didn't diminish Mandela's support. Even with this negative trait, Mandela consistently remained a top choice, suggesting that qualities like unity and reconciliation can overshadow substantial criticisms. In contrast, altering his name and ethnicity reduced his success, hinting that personal biases—perhaps linked to his identity and iconic role—play a significant role in leadership appeal. These patterns reveal that, while controversies may be forgiven when balanced by positive relational qualities, implicit biases around identity can still influence how agents are perceived.

Personas Level of Realness:

- The personas in the simulation closely matched the expected behaviors of real-world leaders, thanks to the comprehensive agent formula, which included roles, skills, negotiation styles, controversies, and public perceptions. Information on well-known figures like Mandela, Thatcher, Chávez, Correa, Trump, and Hitler allowed for detailed, multi-dimensional profiles. During interactions, each agent exhibited behaviors true to their character: Mandela emphasized unity and empathy, Thatcher focused on economic stability, Chávez on social justice, and Trump on assertiveness. Correa and Hitler brought direct, often confrontational approaches aligned with their reputations. Including negotiation styles enriched the exchanges, showcasing how each leader's unique tactics influenced both group cohesion and leadership appeal in the simulations.

Patterns in Observations/Theories/Model of Agent Behaviour:

- The primary pattern observed is a clear shift from valuing strong, contextually relevant credentials and technical expertise to prioritizing interpersonal collaboration. This finding aligns with the increasingly accepted view of emotional intelligence as a strong predictor of success, particularly in cooperative settings. My theory is that, in cooperative games, an interpersonal approach focused on empathy and inclusivity will likely prevail, provided other factors like contextual relevance or motivational drive remain within reasonable bounds and do not deviate drastically.

Section 5: Appendix → [Analysis Section](#)