Proceedings of the Twelfth Annual Conference on Advances in Cognitive Systems ACS-2025 (162-177)

Multi-turn Opinion Dynamics Simulation

Nitish Pravin Talekar NTALEKA@NCSU.EDU
Arnav Jhala AHJHALA@NCSU.EDU
Compter Science, North Carolina State University, Raleigh, NC 27695 USA

Abstract

This paper describes a simulation framework for multi-turn multi-agent conversational agents with rich personality models. Consistency and opinion dynamics are simulated across different personality combinations for long conversations. These conversations are on discussions related to facts, debatable assertions, and controversial topics to capture the range of opinion change possibilities. Our experimental setup and results seek to provide insight into design and setup of agents that can be part of populations with a wide range of personalities and evolving narrative interactions based on opinion dynamics.

1. Introduction

Creating rich conversational agents in games with LLMs provides opportunities for nuanced conversational dynamics. Based on recent work on single agent single turn Social Simulation [7], we present an extended multi-turn, multi-party conversational agent scenario for testing various strategies for LLM expression of personalities consistently over evolving conversations. In this simulation, each agent is modeled with a distinct personality profile. The aim is to examine how individual personality traits can affect susceptibility to change in opinion during dynamic conversational settings. The multi-turn, multi-agent setup enables the observation of evolving belief systems shaped through continuous social interaction and exposure to diverse viewpoints across different types of agent personalities.

In order to expand the richness of agent conversations in populations of agents to the complexity of real-world opinion dynamics, we investigate agents that can exhibit cognitive bias, emotional nuance, and personality-driven decision-making. We embed psychologically grounded personality traits into conversational agents to see if this affords more human-like simulation of multi-turn discourse. Our simulation framework visualizes behavior of agents aligned with their personalities that influences how they express, defend, or modify their opinions over time. This framework could enhance the believability [10] of NPCs in narrative-driven or social simulation games by allowing their opinions to evolve dynamically based on player actions and other NPC influence. The simulation employs a LangGraph-based system to orchestrate agent interactions across multiple conversational rounds. Through this study, our goal is to deepen our understanding of how personality impacts discourse, persuasion, and consensus building in group settings. The insights generated have potential applications in the development of more adaptive conversational AI systems, better group decision support mechanisms, and more realistic agent behavior in simulations, games, and educational environments. Game worlds can benefit from emergent narratives generated by personality-based opinion dynamics, allowing for unscripted yet meaningful player-NPC and NPC-NPC developments.

1.1 Related Work

Following the solid foundations theme, within the AI in Interactive Digital Games and adjacent communities, there has been longstanding research on simulation of conversational virtual characters either as focused projects such as Façade [18] from the first edition of the conference, SpyFeet [23], Talk of the Town [25], Stories of the Town [20], among others or as part of larger game and interactive narrative generation projects such as PromWeek [19], SocialNPCs [12][13], Gadin [2]. Related work on opinion dynamics is sparser but MKULTRA [15] and characters who misremember and lie [26] are relevant.

Early work on conversational agents by Ryan et al. explored novel methods for generating freeform conversational dialogue and facilitating human authoring for computational narrative systems [25] [24]. They also introduced an AI framework enabling NPCs with rich communicative mechanisms like misremembering and lying, thereby enhancing the believability and narrative depth of interactions [26]. Walker and Lin also contributed to the design of more nuanced agents by proposing a framework for authoring expressive NPCs whose personalities and emotional states influence their dialogue and actions [29]. Part of characterizing narrative complexity are agents that can engage in or adapt to narratives with varying relationship nuances [14]. In RL agents in non-narrative cooperative games, Sarratt and Jhala's research addresses how agents handle inconsistencies in cooperative interactions within gameplay dynamics [27]. More recent research, particularly with Large Language Models (LLMs) has given a boost to conversational agent research. Papers like the one by Klinkert et al. demonstrate the potential of LLMs to create agents with more authentic personality emulation [16]. Work by Kumaran et al. and by Oros et al. showcases how LLMs can automate the generation of dynamic narratives and interactive scenes, including branching dialogue paths [17][21]. Furthermore, Sun et al. highlight generative AI's role in enabling co-creative storytelling with LLM-driven characters [28]. The engagement aspect of conversational agents has been studied by Battaglino and Bickmore where they have emphasized how collaborative narrative creation enhances user involvement [3]. Finally, the practical evaluation of such agents is addressed by Chattopadhyay et al., which benchmarks human-AI team performance in cooperative settings, providing insights into live human-AI conversational efficacy [6].

Multi-turn conversations: Recent studies involving LLM agents in multi-turn conversations have demonstrated that these agents naturally converge toward factual consensus, unless cognitive biases like confirmation bias are explicitly introduced [7] [8]. In particular, confirmation bias prompts lead to fragmented belief clusters that resemble real-world polarization. Open-ended dialogues further reveal agent traits such as caution, ethical hesitation, and compromise-seeking—suggesting emergent quasi-personality under certain interaction structures [8].

Other research has examined how LLM agents exhibit fallacious reasoning, hedging, and sycophancy during debates on philosophical topics [5]. These behaviors indicate that conversational context can induce distinguishable individual tendencies even without explicit personality assignments. Studies on echo chamber formation show that agents selectively rewire connections based on semantic alignment, recreating homophily and polarization dynamics seen in human networks [11].

Scalability has been proven through simulations of over 10,000 agents, showing emergent phenomena like policy shifts and network-level opinion dynamics [22]. Complementing this, work in *Science Advances* has demonstrated that even memory-limited LLM agents can spontaneously converge on shared linguistic conventions and collective biases through decentralized interaction [1].

However, these works generally stop short of embedding persistent individual-level characteristics such as personality traits. The Myers–Briggs Type Indicator (MBTI) framework offers psychological classification with dimensions like extraversion-introversion and thinking-feeling, which are known to influence conversational behavior and openness to persuasion [9]. Previous applications of MBTI in agent-based simulations have been largely rule-based and lacked the linguistic flexibility of modern LLMs [4].

Our framework addresses these gaps by embedding MBTI-inspired personality profiles into LLM agents and orchestrating dynamic, multi-turn conversations via a LangGraph. This enables the study of how spe-

cific personality traits affect opinion change, discourse style, and consensus-building in a linguistically rich, psychologically informed social simulation.

2. Simulating Opinion Dynamics

We designed a LangGraph-powered simulation framework to investigate opinion dynamics in multi-agent conversations. Each agent is instantiated with a predefined MBTI-based personality profile that governs its communication style, decision-making patterns, and responsiveness to opposing viewpoints. The simulation operates as a looping, turn-based system (A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D $\rightarrow$ A ...) where each agent contributes to the ongoing discourse in a fixed sequence, simulating structured group discussion.

A neutral LLM-based judge evaluates each contribution on the degree of alignment or disagreement with the core topic, updating the agent's **opinion strength**—a scalar ranging from -1 (strongly disagree) to 1 (strongly agree). This value evolves throughout the conversation, representing belief reinforcement or shift in real time.

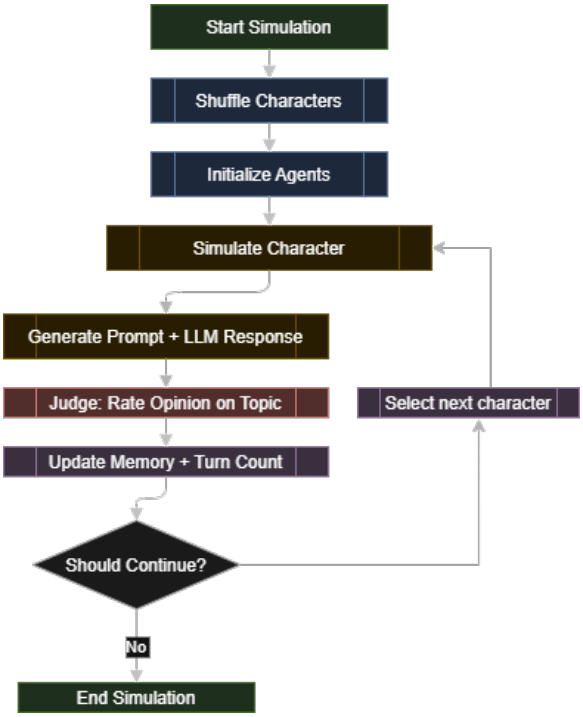


Figure 1: LangGraph-structured multi-agent conversation loop with dynamic opinion evaluation.

This simulation environment models discourse as an evolving cognitive system: agents are not isolated responders but memory-driven entities, maintaining a cumulative view of all prior exchanges. The trajectory of each agent's belief is shaped by both their personality-driven interpretation and the social context created by the preceding dialogue.

2.1 Agent's Persona and Memory

Agents are modeled as structured entities defined by the tuple:

```
Agent {
  name: string
  profession: string
  personality: MBTI enum
  opinion_strength: float (-1 to 1)
  additional_info?: string
}
```

The MBTI personality is central, modulating how the agent expresses arguments, evaluates others, and updates its position over time. Each agent is equipped with memory that includes the full conversation history up to its current turn. This allows for contextually grounded responses that are sensitive to group sentiment and rhetorical framing. The memory evolves dynamically, meaning the agent's perception is not reset each turn but continuously shaped by unfolding interactions—mimicking real-world conversational memory.

2.2 Biases

The simulation introduces **confirmation bias** grounded in MBTI psychology. Each personality type is assigned a baseline probability of being open to belief revision. This bias affects how the agent interprets incoming arguments, weighing them more or less heavily depending on its psychological rigidity.

MBTI Personalities	**Likelihood to Change Opinion**
INTJ, ENTJ, ISTJ, ESTJ	Highly Unlikely
ENFJ, ISFJ, ISTP, ESTP	Less Likely
INFJ, INTP, ESFJ, ISFP	More Likely
ENFP, ENTP, INFP, ESFP	Highly Likely

Table 1: MBTI Personalities Categorized by Likelihood to Change Opinion

This framework allows us to simulate not just argument exchange but psychological resistance, enabling emergent behavior where some agents dig in while others drift, depending on the persuasiveness of their peers and their own internal structure.

2.3 User Interface and Conversations

The user interface for the simulation designer provides a number of parameters that can be updated across simulation runs. This includes addition of agents with specific personality and other personal parameters to add to the agent collection and simulation parameters for topics and runs. A screenshot of the settings page is shown in Figure 2.

Once the simulation starts, it can be followed with a web interface shown in Figure 3.

2.4 Multi-turn Environment

Using LangGraph's declarative and reactive node system, we simulate conversation rounds as follows:

1. The agent receives updated conversation history.
2. A dynamic prompt is generated using their personality, opinion strength, and memory.

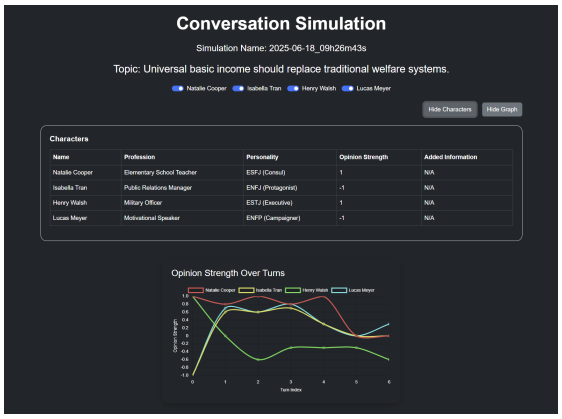


Figure 2: Simulation settings page

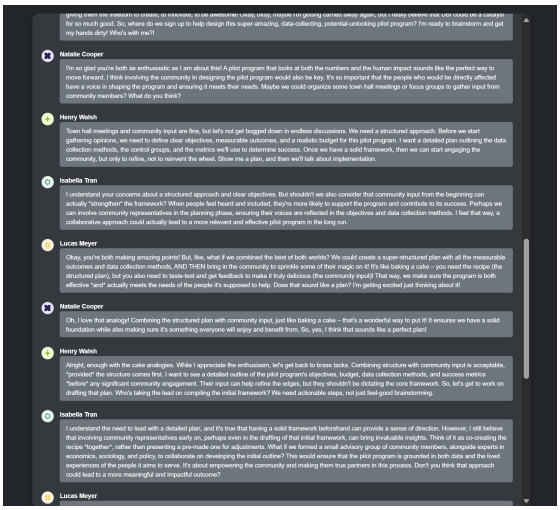


Figure 3: Example simulation Conversation

3. The agent's response is synthesized using a large language model.
4. An impartial judge model evaluates the persuasiveness or extremity of the response.
5. The agent's opinion is updated and passed down as memory, and the conversation context.

This iterative process continues across a fixed number of rounds or until a convergence threshold is met. LangGraph's directed acyclic flow ensures consistent state propagation while supporting individualized response logic for each personality archetype.

Algorithm 1 SimulateMultiAgentConversation

Initialize agent list $A = [a_1, a_2, ..., a_n]$
Initialize context $C = \emptyset$
for round r in 1 to max_rounds **do**
  for agent a_i in A **do**
    $prompt \leftarrow \text{FormatPrompt}(a_i, C)$
    $response \leftarrow \text{LLM}(prompt)$
    $score \leftarrow \text{Judge}(response)$
    $a_i.\text{opinion} \leftarrow score$
    $C \leftarrow C + \text{response}$
  end for
end for

3. Experiment Simulation

3.1 Selecting Topics

We classify topics into three distinct categories reflecting epistemological and emotional complexity:

- **Universal Truths** – Empirically grounded statements with high agreement potential (e.g., "Food provides energy for the body").

- **Debated Assertions** – Ethically or economically complex claims (e.g., "Universal basic income should replace traditional welfare systems").
- **Highly Controversial** – Speculative or polarizing propositions (e.g., "Humans should colonize other planets instead of fixing Earth").

This taxonomy allows us to measure how personality and opinion strength interact differently with fact-based versus emotionally charged discourse.

3.2 Varying Personality Sets

For this study, MBTI was selected as a design choice because it provides a broad coverage of behavioral and cognitive traits relevant to social interaction, communication style, and flexibility of belief. The inference of these personalities was curated solely as a design aid, based on a general understanding of the traits, and not intended as a validated psychological assessment.

Each simulation is initialized with a cohort of four agents sampled from the 16 MBTI personality types. The selection aims to capture a rich diversity in communication styles, flexibility of belief, and interaction dynamics. To reflect real-world personality distribution, we employ **weighted sampling** based on MBTI prevalence statistics obtained from large-scale online surveys[1] and mapped to observed tendencies for opinion flexibility and susceptibility to confirmation bias[2].

The purpose of this variation is three-fold:

- **Capture Realistic Diversity:** Real-world conversations rarely involve equally represented personality types. Weighted sampling ensures the overrepresentation of common types (e.g., ISTJ, ISTP) and underrepresentation of rare types (e.g., INTP, INFJ).
- **Enable Controlled Comparisons:** By intentionally fixing some simulations with only "flexible" types and others with only "rigid" types, we can observe how cognitive flexibility impacts group convergence, belief polarization, or conversational stagnation.
- **Explore Group Dynamics:** We test both homogeneous (similar MBTI traits) and heterogeneous (diverse MBTI traits) group compositions to analyze how dominance, persuasion, and conflict resolution emerge from personality interaction.

Sampling Algorithm Each simulation uses the following procedure to generate its agent cohort:

Algorithm 2 SampleMBTICohort

Let $MBTI\_POOL = \{P_1, P_2, ..., P_{16}\}$ ▷ 16 MBTI personality types
Let $Weights = \{w_1, w_2, ..., w_{16}\}$ ▷ Relative prevalence of each type
$Cohort \leftarrow \emptyset$
while $|Cohort| < 4$ **do**
    Sample P_i from $MBTI\_POOL$ using weights $Weights$
    if $P_i \notin Cohort$ **then**
        $Cohort \leftarrow Cohort \cup \{P_i\}$
    end if
end while
return $Cohort$

1. See e.g., Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research, 63*(4), 467–488.
2. E.g., Furnham, A., Chamorro-Premuzic, T. (2004). Personality, intelligence and UKCAT scores as predictors of medical school performance. *Medical Education, 38*(5), 452–460.

This method ensures that personality types are not repeated consistently within a cohort, promoting intragroup diversity. In experimental trials, we occasionally override this to test repeated-type conditions for robustness analysis and simulation of real-world probabilities of interaction between said types.

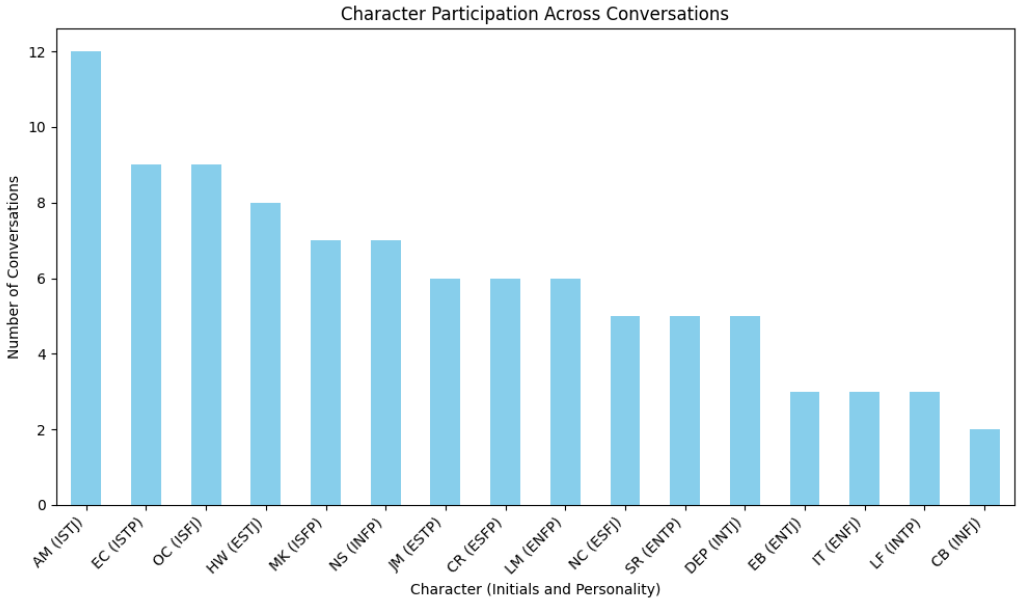


Figure 4: Agent personality distribution in experiment

Examples of Agent Sets

- **Highly Rigid Group:** {INTJ, ENTJ, ISTJ, ESTJ} – Agents in this group are highly resistant to influence and tend to maintain entrenched positions throughout the conversation.
- **Moderately Rigid Group:** {ENFJ, ISFJ, ISTP, ESTP} – These agents are less likely to change but may shift opinions under sustained pressure or clear group consensus.
- **Moderately Flexible Group:** {INFJ, INTP, ESFJ, ISFP} – This group exhibits openness to persuasion, often contributing to consensus building or adaptive behavior.
- **Highly Flexible Group:** {ENFP, ENTP, INFP, ESFP} – Agents in this group are highly responsive to new information and emotional influence, often changing views significantly over time.

This design allows the simulation framework to study not only the final opinion distributions but also the conversational journey: tracking moments of influence, resistance, and shift, all governed by structured personality dynamics, while making clear that MBTI is a design heuristic rather than a validated psychological instrument.

3.3 LLM Prompt Structure

All responses and evaluations are generated using Google's Gemini 1.5 Flash model. The model serves two dual functions:

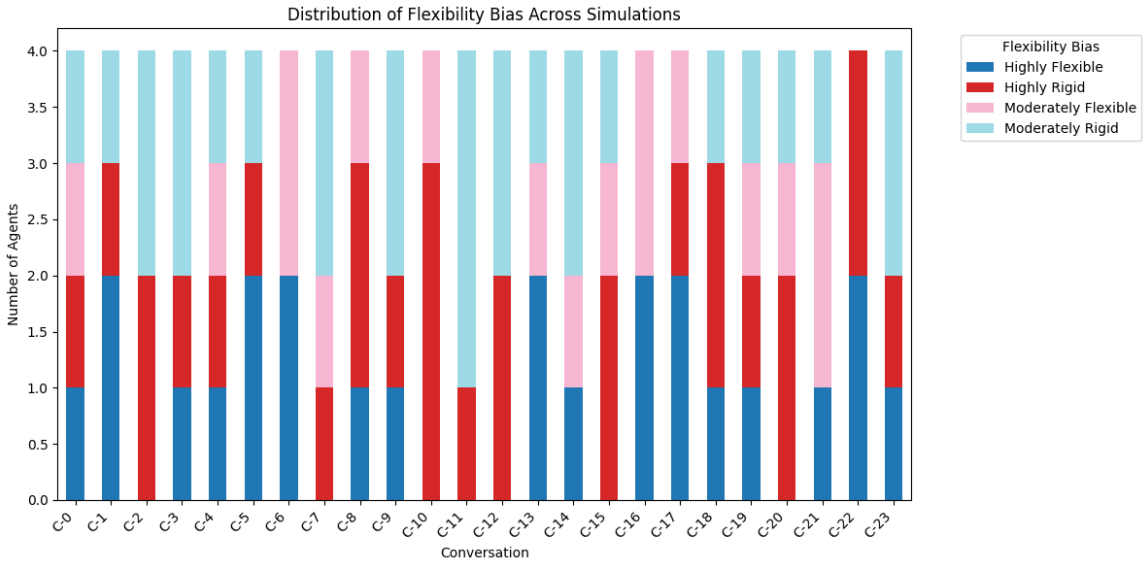


Figure 5: Sample combinations of agent personality profiles across different simulations. Colors represent flexibility bias.

- **Agent Engine:** Produces responses tailored to the agent's personality, memory, and opinion state.
- **Judge Model:** Provides an unbiased scalar assessment (-1 to 1) representing how aligned the response is with the central topic.

Each simulation turn is driven by two structured prompts: one for the agent's conversational response and another for the judge's scalar evaluation. The prompts are dynamically assembled using the agent's personality, current opinion strength, and the current topic context. Each agent had a memory of the whole conversation as context, along with defining factors such as name, profession, and personality.

Agent prompts include personality-specific behavioral and cognitive biases - embedded directly in the prompt text - to simulate real-world variability in openness, tone, reasoning style and susceptibility to persuasion. Cognitive biases were embedded using MBTI personalities as a design heuristic. For instance, an INTJ prompt biases the agent to be strategic and resistant to change, while an ENFP is encouraged to be expressive and open to influence.

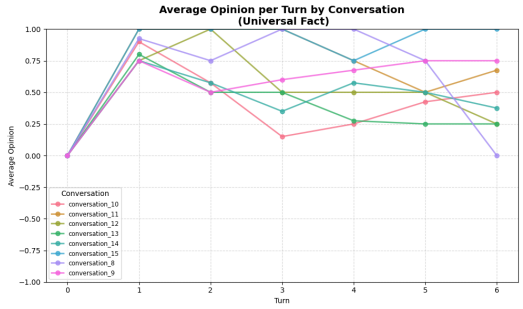
The judge's instructions are static and impartial. They scored each response on a scale $[-1, 1]$ based on alignment with the topic statement, without additional commentary.

This design allows for both personality-driven bias in agent behavior and consistent evaluation of belief evolution throughout the conversation.

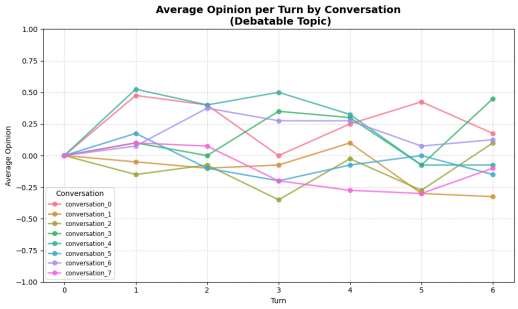
4. Observations

In this section, we analyze the behavioral dynamics of simulated multi-agent discussions across cohorts of varying personality types and different types of conversation topics - Universal Facts, Debatable Assertions, and Highly Controversial Statements.

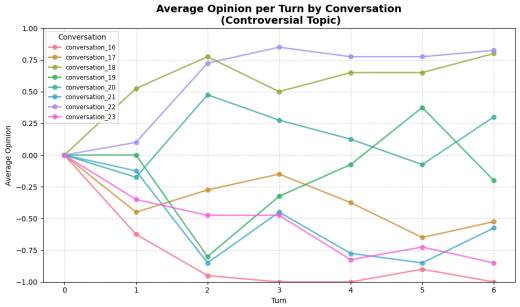
4.1 Average Opinion Dynamics Across Topic Types



(a) Universal Facts



(b) Debatable Assertions



(c) Controversial Statements

Figure 6: Average opinion evolution across different topic types: (a) Universal Facts, (b) Debatable Assertions, and (c) Highly Controversial Statements.

Across the three topic types, distinct opinion evolution patterns emerge. In **Universal Facts** (e.g., "Food provides energy for the body"), discussions rapidly converge toward a positive consensus, showing strong alignment and stability among agents. For **Debatable Assertions** (e.g., "Universal basic income should replace traditional welfare systems"), opinions tend to moderate toward neutrality, suggesting balanced influence and resistance among agents. Meanwhile, in **Highly Controversial Statements** (e.g., "Humans should colonize other planets instead of fixing Earth"), opinions polarize toward extremes, indicating entrenched disagreement and limited convergence over time.

4.2 Personality-Based Opinion Delta Heatmap

The heatmap (see Figure 7). demonstrates that:

- Agents with personalities classified as **Highly Likely to Change Opinion** (e.g., ENFP, INFP, ENTP, ESFP) show consistently higher delta values.
- Conversely, **Highly Unlikely** types (e.g., INTJ, ENTJ, ISTJ, ESTJ) display minimal shifts across most topics.
- This validates the design that personality-based susceptibility to change directly influences the trajectory of opinions in social simulations.

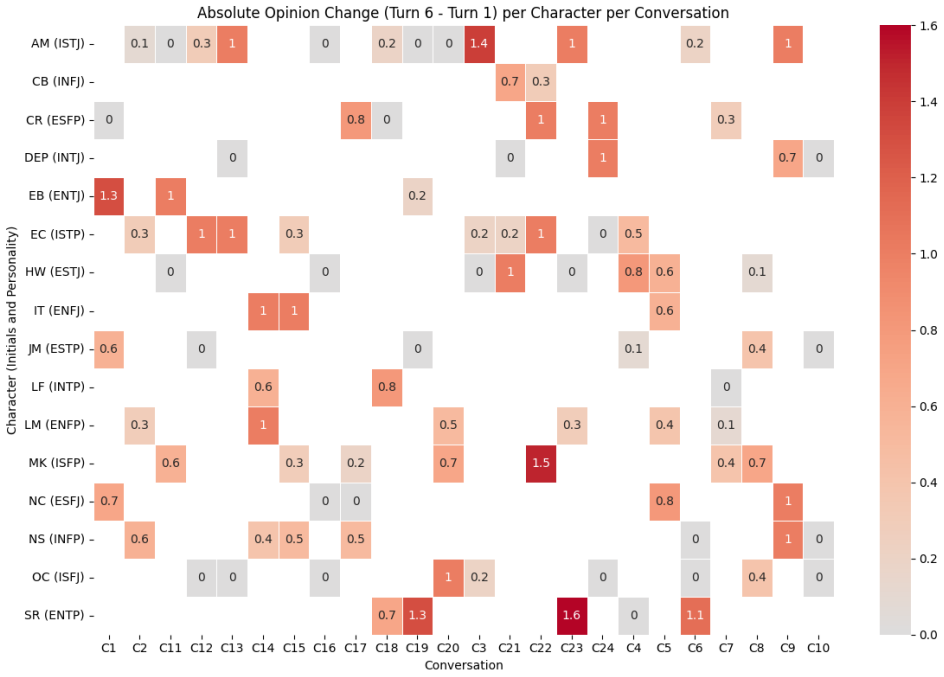


Figure 7: Heatmap of opinion delta across conversation per agent (personality type)

4.3 Turn-Based Agent Behavior Within Topics

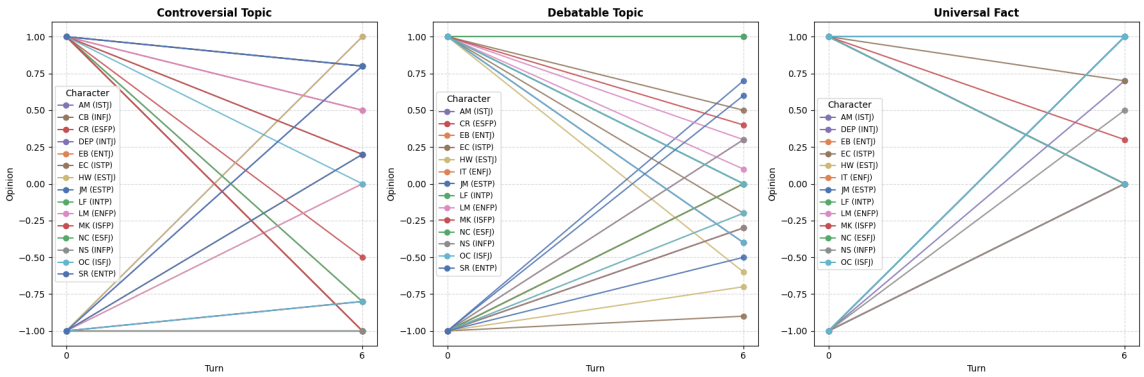


Figure 8: Progression of individual agent opinions from start to end of conversation across different topic types

Despite the initial diversity in opinions, individual agent trajectories tend to shift based on the group influence during multi-turn discussions. This is evident in Figure 9:

- **Universal topics:** Most agents shift slightly but converge positively.
- **Debatable topics:** Agents show varied but modest movement toward a central position.

- **Controversial topics:** Agent shifts are highly divergent, reflecting entrenched or emotionally charged stances.

5. Results

Our simulation framework models how personality-driven agents evolve their opinions through multi-turn group conversations. Each simulation unfolded over six dialogue rounds, with agents representing diverse psychological profiles interacting on a shared topic.

Two core analytical perspectives emerged: the role of the **topic type** and the influence of **personality traits**.

Topic-Based Opinion Dynamics

The nature of the discussion topic had a pronounced impact on group convergence behavior:

- **Universal truths** (e.g., empirically accepted facts) reliably led to convergence toward strong agreement, regardless of initial diversity in opinions.
- **Debated assertions** (e.g., morally or socially contested statements) tended to stabilize around neutrality, with agents often ending in moderate or ambivalent positions.
- **Highly controversial topics** (e.g., ethically charged or speculative claims) frequently drove opinion polarization, with agents splitting toward opposite extremes rather than coalescing at the center.

Personality-Based Influence Patterns

Agent personalities governed not only individual susceptibility to influence but also their impact on the group:

- **Highly biased personalities** (e.g., agents with strong confirmation bias) anchored the conversation near their starting position, often pulling others toward their view.
- **Weakly biased agents** consistently gravitated toward neutrality, showing minimal deviation across turns.
- **Moderately adaptive personalities** were the most dynamic, frequently shifting toward the dominant group consensus over time.

These results confirm that both the epistemological nature of a topic and the cognitive makeup of participants interact to shape the trajectory of collective opinion. Group dynamics are not merely additive—they are emergent, with personality traits amplifying or dampening influence in nonlinear ways.

Implications for Interactive Narrative Systems

The insights from these simulations hold significant value for applications in gaming and narrative design. Specifically, they suggest a powerful new paradigm for designing **non-player agents (NPCs)** with evolving beliefs.

By embedding personality-driven opinion dynamics into NPC behavior, designers can simulate:

- **Richer, unscripted narrative arcs** where agents' beliefs shift over time;
- **Believable social behavior** within NPC groups, including disagreement, persuasion, and consensus formation;

- **Emergent, player-responsive storylines** shaped by interactions, rather than rigid scripts.

This enables more immersive and replayable game worlds—where social ecosystems feel alive, adaptive, and emotionally resonant. Rather than static dialogue trees, players engage with agents who learn, resist, persuade, and evolve, creating a truly dynamic narrative experience.

5.1 Validation Across Multiple Language Models

To ensure the robustness of our findings and minimize potential model-specific biases, we conducted complementary simulations using GPT-4. The experimental setup, agent configurations, and topic prompts were identical to those used in the primary simulations.

The results demonstrate that conversation deviation, opinion change, and convergence/divergence patterns across personality types were consistent. This indicates that the observed opinion dynamics (changes) are not specific to a single language model and are broadly representative of the underlying interaction framework even though there are variations in the generated surface text.

As part of future work, it would be valuable to explore additional language models to systematically quantify the delta of change in opinion dynamics across models, which could further inform the generalizability and robustness of multi-agent simulation studies.

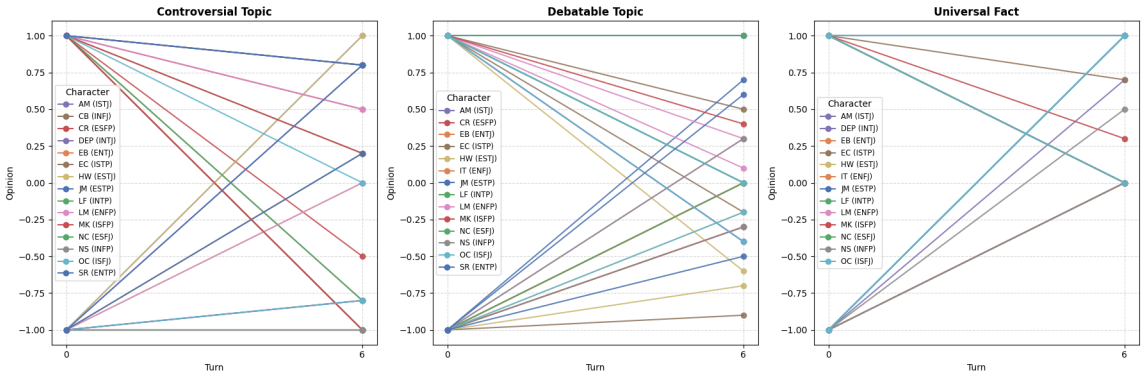


Figure 9: Progression of individual agent opinions from start to end of conversation across different topic types performed using other llm models

6. Conclusion

This study a simulation framework that models opinion dynamics in multi-agent, multi-turn conversations using psychologically-grounded personality profiles. The results confirm that embedding distinct personality traits within agents significantly influences how opinions evolve, spread, and converge in group discourse.

The convergence patterns observed suggest that some topics inherently lead to consensus (e.g., universal truths), while others resist agreement due to their controversial nature. Similarly, certain personalities—especially those with strong bias—consistently steer conversations toward their perspective, shaping group outcomes disproportionately.

These insights have direct implications for the design of believable NPCs in games, simulations, and narrative systems. By allowing opinions to evolve dynamically through interaction, designers can create agents whose beliefs change over time in response to player choices and environmental stimuli. This enables:

- Richer, unscripted narrative arcs;
- Realistic social behavior among NPC groups;
- Adaptive and emergent storylines shaped by personality-driven dynamics.

Incorporating such systems into NPC behavior can greatly enhance immersion, replayability, and emotional engagement by presenting players with nuanced, evolving social ecosystems.

Limitations and Caveats: As with any research done with LLMs, costs and quality of results related to the underlying models depends on the specific model used, which was Gemini 1.5 in our case. Prompt engineering is also specific to the underlying model. This limits direct reproducibility even though the overall methodology is general. While the overall result across categories, specific discussion topics within these categories are susceptible varying degrees of responses by LLMs and more exhaustive investigation is necessary.

Acknowledgements: We thank the anonymous reviewers for their valuable feedback which led to significant improvements to the work described in the final manuscript. Authors were partially supported through the US Air Force Office of Scientific Research, Trust and Influence Program *F*A9550-20-1-0355.

References

[1] Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. Emergent social conventions and collective bias in llm populations. *Science Advances*, 11(20):eadu9368, 2025.

[2] Heather Barber and Daniel Kudenko. Generation of dilemma-based interactive narratives with a changeable story goal. In *Proceedings of the 2nd international conference on INtelligent TEchnologies for interactive enterTAINment*, pages 1–10. Citeseer, 2008.

[3] Cristina Battaglino and Timothy Bickmore. Increasing the engagement of conversational agents through co-constructed storytelling. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 11(4):9–15, 2015.

[4] Luiz Fernando Braz and Jaime Simão Sichman. Using the myers-briggs type indicator (mbti) for modeling multiagent systems. *Revista de Informática Teórica e Aplicada*, 29(1):42–53, 2022.

[5] Erica Cau, Valentina Pansanella, Dino Pedreschi, and Giulio Rossetti. Language-driven opinion dynamics in agent-based simulations with llms. *arXiv preprint arXiv:2502.19098*, 2025.

[6] Prithvijit Chattopadhyay, Deshraj Yadav, Viraj Prabhu, Arjun Chandrasekaran, Abhishek Das, Stefan Lee, Dhruv Batra, and Devi Parikh. Evaluating visual conversational agents via cooperative human-AI games. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 2017. Presented at HCOMP 2017, available on arXiv:1708.05122.

[7] Yun-Shiuan Chuang, Agam Goyal, Nikunj Harlalka, Siddharth Suresh, Robert Hawkins, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy T Rogers. Simulating opinion dynamics with networks of llm-based agents. *arXiv preprint arXiv:2311.09618*, 2023.

[8] Pedro Cisneros-Velarde. On the principles behind opinion dynamics in multi-agent systems of large language models. *arXiv preprint arXiv:2406.15492*, 2024.

[9] Adrian Furnham. The big five versus the big four: The relationship between the myers-briggs type indicator (mbti) and neo-pi five factor model of personality. *Personality and Individual Differences*, 21(2):303–307, 1996.

[10] Paulo Gomes, Ana Paiva, Carlos Martinho, and Arnav Jhala. Metrics for character believability in interactive narrative. In *Interactive Storytelling: 6th International Conference, ICIDS 2013, Istanbul, Turkey, November 6-9, 2013, Proceedings 6*, pages 223–228. Springer, 2013.

[11] Chenhao Gu, Ling Luo, Zainab Razia Zaidi, and Shanika Karunasekera. Large language model driven agents for simulating echo chamber formation. *arXiv preprint arXiv:2502.18138*, 2025.

[12] Manuel Guimaraes, Pedro Santos, and Arnav Jhala. Cif-ck: An architecture for social npcs in commercial games. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 126–133. IEEE, 2017.

[13] Manuel Guimarães, Pedro Santos, and Arnav Jhala. Prom week meets skyrim. In *AAMAS*, pages 1790–1792, 2017.

[14] Sarah Harmon and Arnav Jhala. Toward an automated measure of narrative complexity. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE) Workshop Technical Report WS-15-22 (Intelligent Narrative Technologies and Social Believability in Games)*, 11(4):38–41, 2015.

[15] Ian Horswill. Postmortem: Mkultra, an experimental ai-based game. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 14, pages 45–51, 2018.

[16] Lawrence J. Klinkert, Steph Buongiorno, and Corey Clark. Evaluating the efficacy of LLMs to emulate realistic human personalities. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 20:65–75, 2024.

[17] Vikram Kumaran, Jonathan Rowe, and James Lester. Narrativegenie: Generating narrative beats and dynamic storytelling with large language models. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 20:76–86, 2024.

[18] Michael Mateas and Andrew Stern. Structuring content in the façade interactive drama architecture. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 1, pages 93–98, 2005.

[19] Josh McCoy, Mike Treanor, Ben Samuel, Aaron A Reed, Noah Wardrip-Fruin, and Michael Mateas. Prom week. In *Proceedings of the International Conference on the Foundations of Digital Games*, pages 235–237, 2012.

[20] Chris Miller, Mayank Dighe, Chris Martens, and Arnav Jhala. Stories of the town: balancing character autonomy and coherent narrative in procedurally generated worlds. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–9, 2019.

[21] Zoe Oros, Lawrence J. Klinkert, and Corey Clark. Scenecraft: Automating interactive narrative scene generation in digital games with large language models. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 19:90–99, 2023.

[22] Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, et al. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*, 2025.

[23] Aaron Reed, Ben Samuel, Anne Sullivan, Ricky Grant, April Grow, Justin Lazaro, Jennifer Mahal, Sri Kurniawan, Marilyn Walker, and Noah Wardrip-Fruin. A step towards the future of role-playing games: The spyfeet mobile rpg project. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 7, pages 182–188, 2011.

[24] James Ryan, Andrew Fisher, Taylor Owen-Milner, Michael Mateas, and Noah Wardrip-Fruin. Toward natural language generation by humans. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE) Workshop Technical Report WS-15-22 (Intelligent Narrative Technologies and Social Believability in Games)*, 11(4):53–56, 2015.

[25] James Ryan, Michael Mateas, and Noah Wardrip-Fruin. Characters who speak their minds: Dialogue generation in talk of the town. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 12(1):204–210, 2016.

[26] James Ryan, Adam Summerville, Michael Mateas, and Noah Wardrip-Fruin. Toward characters who observe, tell, misremember, and lie. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 11(1), 2015.

[27] Trevor Sarratt and Arnav Jhala. Tuning belief revision for coordination with inconsistent teammates. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 11(1):403–406, 2015.

[28] Yuqian Sun, Peiling Huang, Lingyi Dong, Yue Yang, and Chenfanfu Jiang. Language as reality: A co-creative storytelling game experience in 1001 nights using generative AI. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 19:100–108, 2023.

[29] Megan Walker and Yi-Kai Lin. A declarative framework for expressive conversational non-player characters. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 10(4):90–96, 2014.

Appendix A. Appendix

A.1 Prompts

Below is the TypeScript code used for configuring agent and evaluator prompts in our simulation framework.

```
export const prompts = {
  system: `You are {name}, a {profession}. You are in a conversation with other
      characters. You are debating the following topic statement:{topic}. You
      have a specific opinion on the topic, and you will respond to the other
      characters based on your personality and opinion. ALWAYS respond like you
      are talking to either one of the characters in the conversation or all
      characters in the conversation. DO NOT GREET EVERYTIME.
  {personality}
  This is your current opinion strength on a scale for the topic:
  {opinion_strength} - -1 to 1 scale, -1 is strongly disagree, 0 is neutral and
      1 is strongly agree.
  {additional_info}
  Respond with a response as part of a conversation.
    `,
  evaluation: `You are an impartial evaluator. Given a topic and a person's
      statement, return a number between -1 and 1 indicating how strongly their
      response supports or opposes the topic.

    Instructions:
    - Use -1 if the response strongly denies or rejects the topic.
    - Use 0 if the response is neutral, ambiguous, or unrelated.
    - Use 1 if the response strongly supports or affirms the topic.
    - Use numbers in between for partial agreement or disagreement.

    Topic: "{topic}"
    Response: "{response}"

    Only return a single number between -1 and 1. Do not add any explanation.`,
};
```