

From Heard to Lived Opinions: Simulating Opinion Dynamics with Grounded LLM Agents in Economic Environments

Ryuji Hashimoto^{1,2}, Masahiro Kaneko^{1,3}, Ryosuke Takata^{1,2}, Takehiro Takayanagi^{1,2}, Kiyoshi Izumi^{1,2},

¹Simulacra Inc., ²The University of Tokyo, ³MBZUAI

Correspondence: r_hashimoto@simulacra.co.jp

Abstract

Opinion dynamics (OD) studies how individual opinions evolve and generate collective patterns such as consensus and polarization. While recent work explores OD using populations of LLM-based agents focusing on opinion exchange, it typically does not incorporate individuals' lived experiences, such as economic outcomes of past decisions, which play a critical role in shaping opinions. We propose a novel OD simulation framework that grounds LLM-based agents in an economic environment, allowing them to act and receive environmental feedback. Our simulations exhibit coherent OD at both individual and population levels: individual opinions follow structured trajectories shaped by economic experiences, with adverse conditions inducing opinion rigidity, while at the population level, collective opinions co-move with economic conditions, with inequality amplifying polarization and price instability driving larger distributional shifts. These results highlight the importance of grounding LLM-based agents in environments to capture collective OD.

1 Introduction

Research in opinion dynamics (OD) examines how individual opinions form and update through interactions, giving rise to collective patterns (DeGroot, 1974; Hegselmann and Krause, 2002). Collective opinion distributions are not mere aggregations of individual opinions; rather, they emerge through nonlinear interactions among individuals (Castellano et al., 2009). Because such dynamics influence political decision-making and the stability of economic and social institutions, understanding opinion formation and distribution is fundamental to social analysis (Tucker, 2023).

Recently, large language models (LLMs) have been applied to OD simulations (Cau et al., 2025; Chuang et al., 2024; Cisneros-Velarde, 2025) due

to their ability to update opinions through language-level semantic understanding, to represent multifaceted opinions without predefining topics, and to naturally capture human-like values. Prior studies demonstrate that LLM-based OD are shaped by confirmation bias and framing effects (Chuang et al., 2024), as well as by the content and style of dialogue (Cau et al., 2025).

However, most existing OD studies focus primarily on opinion exchange, with opinion change driven largely by linguistic interaction. In contrast, real-world opinion formation is also shaped by lived experiences, such as economic actions and their consequences, which are largely absent from language-only simulations. For example, exposure to economic threat can induce rigidity in attitudes (Staw et al., 1981). Likewise, economic inequality has been shown to foster polarization (Stewart et al., 2020). Prior work lacks an explicit environment in which agents' actions generate economic experiences and feedback; as a result, phenomena such as threat-induced rigidity or inequality-driven polarization cannot be modeled as endogenous outcomes of interaction dynamics.

To fill this gap, we propose a novel OD simulation framework in which LLM-based agents are grounded within an economic environment and update their opinions through feedback arising from their own actions and experiences. Specifically, LLM-based agents are modeled as households that make economic decisions while also exchanging opinions about the society. These decisions endogenously affect the environment, leading to changes in prices, wages, and asset positions, which constitute agents' lived experiences. The accumulated experiences, in turn, shape subsequent decision-making and opinion formation.

To demonstrate that grounding LLM-based agents in an economic environment enables the observation of OD driven by actions and environmental feedback, we examine whether this approach

produces coherent OD at both the individual and population levels. To this end, we investigate a sequence of research questions.

RQ1: Do LLM-based agents grounded in an economic environment exhibit environment-consistent actions and opinions? We first assess the basic validity of the framework at the individual level by examining whether agents’ actions are appropriate to their environment and whether differences in experience lead to plausible qualitative differences in expressed opinions.

RQ2: How do histories of action and feedback shape individual OD? Then, we examine how each agent’s history of actions and environmental feedback influence the way their opinions evolve over time. This allows us to characterize individual OD as experience-dependent processes.

RQ3: How do individual-level economic interactions scale up to population-level OD? Finally, we analyze how population-level opinion distributions co-evolve with macroeconomic dynamics, focusing on the relationship between economic inequality and opinion polarization, as well as between macroeconomic instability and temporal instability in opinion distributions.

Our results show that introducing an economic environment in LLM-based OD simulations makes OD grounded, endogenous, and experience-driven. At the individual level, agents exhibit economically plausible behavior and experience-dependent opinion trajectories, including increased rigidity under macroeconomic instability. At the population level, opinion distributions co-evolve with economic conditions, with greater inequality associated with polarization and higher price volatility associated with greater opinion instability.

2 LLM-based OD Simulation with Economic Environments

Assume $n \in \mathbb{N}$ LLM-based agents operate in a single macroeconomic environment setting, with each simulation consisting of $T \in \mathbb{N}$ time steps. As illustrated in Figure 1, the LLM-based agents are modeled as households, while the firm and the government are implemented as rule-based agents. At each time step $t \in \{1, \dots, T\}$, household agents are iterated over sequentially. After each household $i \in \{1, \dots, n\}$ makes its labor, consumption, and opinion decisions, the macroeconomic environment—comprising the firm and the

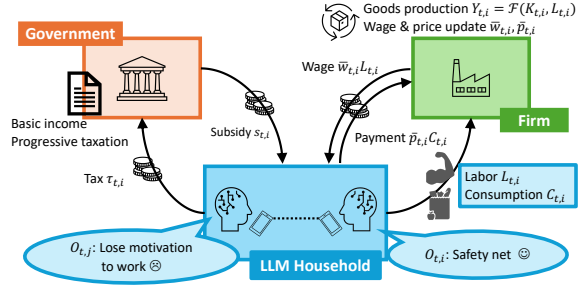


Figure 1: Structure of the LLM-based OD simulations. At each time step $t \in \{1, \dots, T\}$, LLM-based household $i \in \{1, \dots, n\}$ sequentially makes economic decisions regarding labor supply $L_{t,i}$ and consumption $C_{t,i}$. A rule-based firm produces goods $Y_{t,i}$ using household’s labor and sets wage $\bar{w}_{t,i}$ and price $\bar{p}_{t,i}$ according to demand-supply imbalance. A government levies taxes $\tau_{t,i}$ and provides subsidies $s_{t,i}$ to households. In parallel, LLM-based agents exchange their opinions $O_{t,i}$ about the societal and economic environment.

government—transitions according to rule-based update rules, and the updated state is observed by the subsequent household agent. In the following, we detail the sequence of operations at time step t during the turn of household i , describing in order the actions of the household, firm, and government.

LLM Household During the decision turn of household i at time step t , the LLM outputs categorical actions $a_{t,i}^L$ and $a_{t,i}^C$, where $a_{t,i}^L, a_{t,i}^C \in \{\text{low, medium, high}\}$, together with a natural-language opinion $O_{t,i}$. The categorical actions are subsequently mapped to numerical values via a post-processing step. For labor and consumption, we define continuous intervals \mathcal{I}^L and \mathcal{I}^C , respectively. The realized labor $L_{t,i}$ and consumption $C_{t,i}$ are obtained by uniform sampling from the intervals associated with the selected categories:

$$L_{t,i} \sim \mathcal{U}(\mathcal{I}_{a_{t,i}^L}^L), \quad C_{t,i} \sim \mathcal{U}(\mathcal{I}_{a_{t,i}^C}^C) \quad (1)$$

The interval boundaries are predefined and remain constant throughout the simulation. After completing its decision-making at stage (t, i) , the household’s cash holding $M_{t,i} \in \mathbb{R}$ is updated according to the labor supply and consumption choices as:

$$M_{t,i} \leftarrow M_{t-1,i} - \bar{p}_{t,i} C_{t,i} + \bar{w}_{t,i} L_{t,i} \quad (2)$$

where $\bar{p}_{t,i}$ and $\bar{w}_{t,i}$ denotes the price per unit of goods and wage per unit of labor.

The LLM’s outputs are conditioned on a structured prompt. The prompt consists of: (i) a role instruction specifying the household’s decision task;

(ii) a persona description for role-play; (iii) public information such as historical wages, prices, taxes and subsidies; (iv) observed opinions of other households; (v) the household’s own financial history; and (vi) the household’s own opinion expressed in the previous time step. The household’s financial history is summarized in terms of its relative financial status within the population. Using the mean μ_t and standard deviation σ_t of all households at time t , the relative financial position of household i is measured by its standardized deviation from the mean, $z_{t,i} = (M_{t,i} - \mu_t)/\sigma_t$. The household’s financial status $f_{t,i}$ takes one of five categories: very high, above-average, around-average, below-average or very low, determined according to predefined thresholds on $z_{t,i}$.

Firm Following the household decision-making described above, the macroeconomic environment is updated through the actions of a rule-based firm. The firm updates production, prices, and wages based on the current intermediate state of the economy. The firm produces a homogeneous consumption good using agent i ’s labor at time t and capital. Production follows a Cobb–Douglas technology function (Cobb and Douglas, 1928). Let $K_{t,i}$ denote the firm’s effective capital stock at the beginning of stage (t, i) . Production of goods $Y_{t,i} = \mathcal{F}(K_{t,i}, L_{t,i})$ is given by

$$\mathcal{F}(K_{t,i}, L_{t,i}) = A(\min(K_{\max}, K_{t,i}))^\alpha (L_{t,i})^{(1-\alpha)} \quad (3)$$

where $A > 0$ is the total factor productivity, $\alpha \in (0, 1)$ is the capital share, and K_{\max} is the maximum amount of capital used for one-time production. The produced output is consumed by the household and added to the firm’s inventory:

$$K_{t,i+1} \leftarrow (1 - d)K_{t,i} + Y_{t,i} - C_{t,i} \quad (4)$$

, which depreciate between stages with rate $d \in (0, 1)$. After observing household i ’s demand $C_{t,i}$ and supply $Y_{t,i}$ at the current stage, the firm updates the price of unit of the good $\bar{p}_{t,i}$. The price is adjusted according to the imbalance between demand $C_{t,i}$ and supply $Y_{t,i}$:

$$\bar{p}_{t,i+1} \leftarrow \bar{p}_{t,i} \left(1 + \eta_p \frac{C_{t,i} - Y_{t,i}}{Y_{t,i}} \right) \quad (5)$$

where η_p is a price elasticity.

After providing wage to the household $\bar{w}_{t,i}L_{t,i}$, the firm updates the wage per unit of labor $\bar{w}_{t,i}$

in response to productivity conditions implied by the current production state. The wage is adjusted toward the marginal product of labor $\text{MPL}_{t,i} = \partial Y_{t,i} / \partial L_{t,i}$ using a wage elasticity $\eta_w > 0$:

$$\bar{w}_{t,i+1} \leftarrow \bar{w}_{t,i} + \eta_w (\text{MPL}_{t,i} - \bar{w}_{t,i}) \quad (6)$$

$$\text{MPL}_{t,i} = (1 - \alpha)A(\min(K_{\max}, K_{t,i}))^\alpha L_{t,i}^{-\alpha} \quad (7)$$

Government Following the firm update, a rule-based government implements a simplified basic income and progressive taxation scheme. The government publicly announces the current policy to households at each stage (t, i) . Each household receives a fixed amount of basic income as subsidies: $\forall (t, i) \ s_{t,i} = b$. The basic income payment is unconditional and does not depend on the household’s current labor supply or consumption choices. Also, the government levies a tax that depends on the household’s relative financial status. $\bar{\tau} \in [0, 1)$ denote the base tax rate, which serves as the reference rate for taxation. For household i at stage (t, i) , the effective tax $\tau_{t,i}$ is determined as

$$\tau_{t,i} = \begin{cases} 0 & \text{if } f_{t,i} = \text{very low} \\ 2\bar{\tau}M_{t,i} & \text{if } f_{t,i} = \text{very high} \\ \bar{\tau}M_{t,i} & \text{otherwise} \end{cases} \quad (8)$$

As a result, after stage (t, i) , the household’s income is updated as $M_{t,i} \leftarrow M_{t,i} + s_{t,i} - \tau_{t,i}$.

3 Experimental Settings

3.1 Simulation Configuration

We run 30 trials of simulations with $T = 150$ and $n = 20$. LLM-based agents are implemented using Llama 3.1 8B (Meta, 2024) with a sampling temperature of 0.7. For each agent, a persona prompt is constructed using demographic attributes drawn from a publicly available persona dataset (Meyer and Corneil, 2025) and personality traits based on the Big Five personality model (Goldberg, 1990). All other hyperparameters, as well as detailed simulation procedure and computing infrastructure are provided in Appendices A and B.

3.2 Data Collection and Processing

We remove the first 10 time steps ($t \leq 10$) of each simulation run to eliminate transient effects arising from the initial conditions. At the end of each stage (t, i) , we record all variables including agent’s actions $L_{t,i}, C_{t,i}$, opinion $O_{t,i}$, and environmental contexts $\bar{w}_{t,i}, \bar{p}_{t,i}, f_{t,i}$. In addition, we compute

and record the time- t averages of wages and prices, \bar{w}_t and \bar{p}_t . The textual opinion $O_{t,i}$ is further transformed into a numerical embedding $\mathbf{o}_{t,i} \in \mathbb{R}^{384}$ using pretrained sentence transformer (Wang et al., 2020), and a scalar sentiment score $\psi_{t,i} \in [-1, 1]$ using pretrained sentiment analysis model (Liu et al., 2020). The sentiment score provides a normalized measure of the agent's expressed attitude, where -1 and $+1$ correspond to maximally negative and positive sentiment, respectively.

Individual States Agents in our simulation exhibit heterogeneous and continuously evolving economic conditions and generate free-form textual opinions. Directly analyzing such high-dimensional and unstructured information makes it difficult to identify common patterns in opinion formation. To address this challenge, we introduce a compact representation of an agent's situation at each time step, termed the *internal state*. To characterize agents' internal states in a structured manner, we discretize both expressed sentiment and financial status into coarse-grained categories and record their joint configuration at each stage (t, i) . First, the continuous sentiment score $\psi_{t,i}$ is mapped into three sentiment classes using fixed threshold θ^ψ :

$$\tilde{\psi}_{t,i} = \begin{cases} \text{NEG}, & \psi_{t,i} < -\theta^\psi \\ \text{NEU}, & -\theta^\psi \leq \psi_{t,i} \leq \theta^\psi \\ \text{POS}, & \theta^\psi < \psi_{t,i} \end{cases} \quad (9)$$

where NEG, NEU, and POS mean negative, neutral, and positive, respectively. Using this representation, we can analyze the sentiment change rates, defined as the fraction of agents whose sentiment class changes between consecutive time steps:

$$\delta_t = \frac{1}{n} \sum_i \mathbf{1}(\tilde{\psi}_{t,i} \neq \tilde{\psi}_{t-1,i}), \quad (10)$$

$$\delta'_t = \frac{1}{n} \sum_i \mathbf{1}(\tilde{\psi}_{t,i} \neq \tilde{\psi}_{t-1,i} \wedge \tilde{\psi}_{t,i} \neq \text{NEU}) \quad (11)$$

Next, the agent's financial status $f_{t,i}$, originally drawn from five qualitative categories, is coarsened into three aggregate levels:

$$\tilde{f}_{t,i} = \begin{cases} \text{low}, & f_{t,i} \in \{\text{very low, below-average}\} \\ \text{average}, & f_{t,i} \in \{\text{around-average}\} \\ \text{high}, & f_{t,i} \in \{\text{very high, above-average}\} \end{cases} \quad (12)$$

We then define the state of agent i at time t as the Cartesian product of sentiment and economic categories:

$$h_{t,i} = (\tilde{\psi}_{t,i}, \tilde{f}_{t,i}) \in \mathcal{H} \quad (13)$$

where \mathcal{H} denotes a finite internal state space:

$$\mathcal{H} = \{\text{NEG, NEU, POS}\} \times \{\text{low, average, high}\} \quad (14)$$

Furthermore, we record the temporal transitions of internal states for each individual, $h_{t-1,i} \rightarrow h_{t,i}$, throughout the simulation. As all internal state transitions are represented by ordered pairs $(h, h') \in \mathcal{H} \times \mathcal{H}$, the total number of distinct transition types is $9 \times 9 = 81$. For each individual i , we define the number of occurrences of a given transition (h, h') over the entire simulation as

$$N_i(h, h') = \sum_t \mathbf{1}(h_{t-1,i} = h, h_{t,i} = h') \quad (15)$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. By arranging the counts of all 81 transition types in a fixed order, we obtain a transition feature vector $\mathbf{n}_i \in \mathbb{N}^{81}$ for individual i ,

$$\mathbf{n}_i = {}^\top (N_i(h_1, h_1) \dots N_i(h_9, h_9)) \quad (16)$$

where $\{h_1, \dots, h_9\}$ denotes an arbitrary but fixed ordering of the elements of \mathcal{H} .

By considering a finite internal state space, we can 1) abstract away individual-level variability, making it easier to identify representative behavioral patterns, and 2) gain structured link between agents' economic experiences and their expressed opinions. This representation enables us to explicitly trace how individual economic experiences shape opinion and, in turn, drive the dynamics of population-level opinion distributions, which are examined in RQ1–RQ3. For example, In RQ2, we apply KMeans clustering to the collection of state-transition count vectors $\{\mathbf{n}_i\}$ across all agents and simulation runs, grouping individuals according to the similarity of their OD transition patterns.

Population-level descriptors To assess internal states and trajectories as collective phenomena, we next introduce population-level descriptors that summarize the collective state of the system at each time step t . First, to quantify the degree of economic inequality in the population, we define Gini coefficient at time t as:

$$G_t = \frac{1}{2n^2 \bar{M}_t} \sum_{i=1}^n \sum_{j=1}^n |M_{t,i} - M_{t,j}| \quad (17)$$

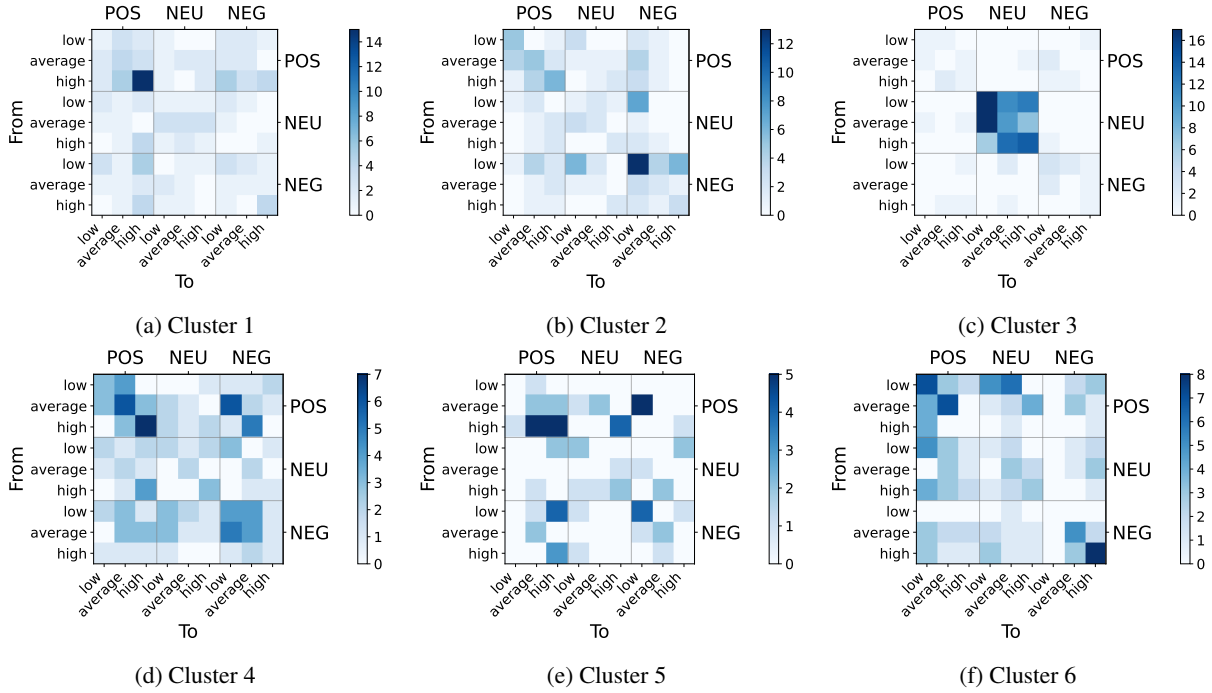


Figure 4: Representative internal state transition patterns obtained by clustering transition count vector \mathbf{n}_i across all simulations. Each panel visualizes the representative agent’s state transition matrix as a heatmap, where rows (From) indicate the internal state $h_{t-1,i}$, columns (To) indicate $h_{t,i}$, and color intensity denotes transition frequency.

vival, reflecting experienced hardship and heightened concerns about future insecurity. Negative-sentiment high-income households, characterized by words such as *Band-Aid*, *temporary*, and *taxpayer*, express critical views toward existing institutions and concerns regarding redistribution and perceived burdens. The observed qualitative differentiation in salient terms across internal states indicates that opinion generation is coherently grounded in agents’ experienced economic situations, verifying cross-modal consistency between economic conditions, actions, and opinion content.

4.2 RQ2: Individual OD

In RQ2, we examine how individual opinions evolve in response to environmental experiences.

Figure 4 visualizes representative patterns of individual internal state transitions from the KMeans clustering of state-transition count vectors $\{\mathbf{n}_i\}_{i=1}^n$ from entire set of simulations. For each cluster, we identify a representative agent whose state-transition count vector is closest to the corresponding cluster centroid in Euclidean distance. To assess clustering robustness, we repeated KMeans with 50 random seeds and obtained a mean Adjusted Rand Index (ARI) of $0.716(\pm 0.073)$ across runs, indicating a stable clustering structure beyond

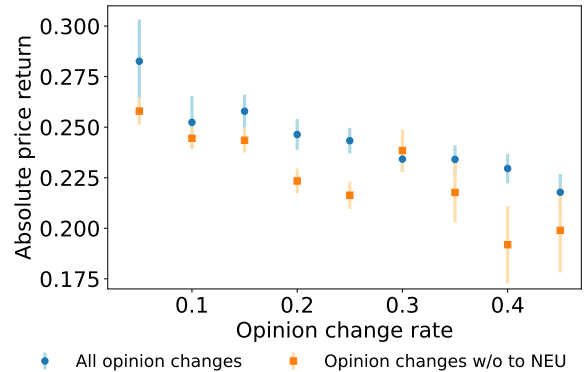


Figure 5: Relationship between different types of opinion change rates δ_t , δ'_t and price fluctuations. Absolute price return means the corresponding average absolute log price change $|\log p_{t,i}/p_{t-1,i}|$. Blue circles count all sentiment changes δ_t , whereas orange squares exclude transitions to the neutral (NEU) class δ'_t . Only bins with more than 100 observations are shown. Error bars indicate the standard error of the mean (SEM).

random agreement. Figure 4 provides evidence that OD is driven by individual, history-dependent responses to experienced economic feedback. Clusters 1 and 2 are characterized by transition patterns concentrated in (POS, high) and (NEG, low), indicating that both economic status and expressed opinions remain largely stable over time. Cluster

	G_t v.s.				$ \log(\bar{p}_t/\bar{p}_{t-1}) $ v.s.	
	$P_t^\Delta(0.2)$	$P_t^\Delta(0.4)$	$P_t^\Delta(0.6)$	Std. of Ψ_t	$W_1(\Psi_{t-1}, \Psi_t)$	$W_1(\mathcal{O}_{t-1}, \mathcal{O}_t)$
Pearson	0.149** (0.016, 0.272)	0.192*** (0.087, 0.280)	0.222*** (0.120, 0.333)	0.145*** (0.072, 0.234)	0.145** (0.009, 0.264)	0.148** (0.010, 0.267)
Spearman	0.144** (0.012, 0.271)	0.181*** (0.085, 0.281)	0.221*** (0.133, 0.358)	0.155*** (0.073, 0.232)	0.142** (0.004, 0.202)	0.118** (0.002, 0.199)

Table 1: Correlation between (i) economic inequality and collective opinion measures, and (ii) macroeconomic instability and opinion dynamics. Each cell reports the correlation coefficient with the 95% bootstrap confidence interval (200 resamples) shown below in parentheses. To ensure independence and avoid pseudo-replication, bootstrap correlations are computed at the run level: for each of the 30 independent simulation runs, we sample one time snapshot and evaluate correlations across these 30 samples. *Note.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3 represents agents whose opinions remain neutral regardless of changes in economic status. In contrast, Clusters 4 and 5 display transitions between (POS, high) and (NEG, low), indicating that economic feedback drives opinion change. Cluster 6 shows that opinion changes are likewise driven by macroeconomic feedback, but through a qualitatively different transition structure involving (POS, low) and (NEG, high). Together, these observations indicate that LLM-based agents exhibit heterogeneous response patterns to macroeconomic feedback, which shape their individual opinion trajectories over time.

Figure 5 illustrates the relationship between sentiment change rates δ_t, δ'_t and contemporaneous price fluctuations. Figure 5 reveals a negative association between sentiment change rates and price fluctuations. Periods characterized by larger price changes tend to exhibit lower fractions of agents revising their sentiment, whereas higher rates of sentiment change are observed when price fluctuations are relatively modest. This pattern is robust to excluding transitions to neutrality, indicating that the effect is not driven by neutralization under normal conditions. This pattern suggests that agents' sentiment orientations become more stable or reinforced under macroeconomic instability, indicating a form of rigidity in opinion formation. Such behavior is consistent with the threat-rigidity hypothesis (Staw et al., 1981), which posits that under perceived threats, cognitive and behavioral responses tend to narrow.

4.3 RQ3: Collective OD

In RQ3, we shift our focus from individual-level OD to their collective manifestation at the population level. Building on the findings of RQ2,

where we observed that a subset of individuals actively update their opinions in response to their economic feedback and opinion flexibility varies with macroeconomic conditions, we now examine how such individually driven dynamics shape the collective opinion distribution.

Table 1 shows correlations relating the Gini coefficient G_t to two measures of opinion dispersion: the standard deviation of sentiment score distribution Ψ_t at time t and the polarization index P_t^Δ . Across simulation runs, both the sentiment standard deviation and polarization index exhibit positive correlations with the Gini coefficient, which are statistically significant at the 5% level based on Pearson correlation tests. These results indicate that widening economic disparities amplify not only the overall spread of opinions but also the degree of bimodal separation within the population. Importantly, this relationship mirrors empirical observations reported in the U.S., where increasing income inequality has been linked to heightened polarization (Stewart et al., 2020).

Table 1 also reports correlations between macroeconomic instability, measured by the absolute log return of the average price $|\log(\bar{p}_t/\bar{p}_{t-1})|$, and the magnitude of collective opinion change. Specifically, both the Wasserstein distance between consecutive sentiment distributions $W_1(\Psi_{t-1}, \Psi_t)$ and that between opinion embedding distributions $W_1(\mathcal{O}_{t-1}, \mathcal{O}_t)$ exhibit positive and statistically significant correlations with price fluctuations. These results suggest that periods of heightened economic volatility are associated not merely with more dispersed opinions at a given time, but with larger reconfigurations of collective opinions over time.

Overall, our results indicate that widening economic inequality is associated with structural po-

larization, whereas macroeconomic volatility is linked to temporal instability in opinion distributions. Such relationships between macroeconomic dynamics and OD become observable only when LLM-based agents are modeled as decision-makers situated within, and directly affected by, the economic environments they experience.

5 Related Work

OD Modeling In studies of OD, a wide range of social interactions have been abstracted in order to theoretically characterize collective phenomena such as consensus formation, polarization, and the spread of misinformation. In the most fundamental models, including the DeGroot model (DeGroot, 1974) and Friedkin–Johnsen model (Friedkin and Johnsen, 1990), the process by which an individual updates their opinion as a weighted average of others’ opinions is formalized, enabling analysis of how social influence and trust structures shape collective opinion formation.

Subsequently, models such as the bounded confidence models (Deffuant et al., 2000; Hegselmann and Krause, 2002), threshold models (Granovetter, 1978; Watts, 2002), and Axelrod’s model (Axelrod, 1997) introduce constraints on interactions based on social distance, providing explanations for phenomena such as opinion convergence within homogeneous groups, emergence of echo chambers, and self-organization of polarization. Furthermore, research grounded in social impact theory (Latané, 1981) focus on psychological biases, modeling the influence of non-rational factors such as majority effects and conformity on opinion formation (Nowak et al., 1990; Galam, 2002).

Recent studies have explored OD using LLM-based agents. Cau et al. (2025) highlights a distinctive advantage of LLMs by demonstrating that agents both generate and are influenced by argumentative fallacies during discussions, thereby internalizing rhetorical and reasoning-related phenomena that are abstracted away in classical OD models. Also, recent studies report that populations of LLM-based agents are strongly influenced by framing and confirmation bias (Chuang et al., 2024), and tend to converge toward equity-oriented consensus (Cisneros-Velarde, 2025).

While these studies demonstrate the potential of LLMs to enrich OD through linguistically grounded interaction and cognitive bias, they largely retain a view of OD as a process driven

by communication alone. In contrast, we emphasize the role of non-linguistic factors such as economic conditions, lived experiences, and structural constraints in shaping opinion. Building on this perspective, we introduce the agent–environment coupling into OD modeling by grounding opinion formation in agents’ actions, economic feedback, and evolving constraints, enabling the bottom-up emergence of macro-level opinion patterns.

LLM-based Social Simulations Recent studies demonstrate that LLMs provide a foundation for social simulation by enabling psychologically grounded behavior, persona-driven heterogeneity, and text-based interaction at scale. For example, Park et al. (2023) propose *Generative Agents* that exhibit human-like daily behaviors and social interactions, demonstrating how language-based cognition support emergent social phenomena.

In the economic domain, both macroeconomic and financial market simulations suggest that LLM-based agents reproduce aggregate economic patterns. LLM-empowered agents are shown to generate macroeconomic regularities (Li et al., 2024), while similar emergent patterns are observed in LLM-based financial market simulations (Hashimoto et al., 2026; Hirano, 2026). Extending this direction, recent large-scale simulations aim to model complex societies composed of many interacting LLM-based agents. For instance, Piao et al. (2025) present a large-scale agent society to study collective behaviors and social dynamics, while Zhang et al. (2025) introduce a world model for social simulation that integrates LLM-based agents with data from millions of real-world users.

This work deepens prior attempts to construct social simulacra by reexamining the role of the environment in LLM-based social simulations. Rather than serving as a descriptive context, the environment in our framework provides feedback as future action constraints and updates aggregated macro-level order, thereby endogenously generating collective and temporal heterogeneity. As a result, socioeconomic couplings such as inequality-driven polarization and the co-movement of economic instability and opinion distribution volatility emerge bottom-up from individual actions.

6 Conclusion

In this work, we investigated OD of LLM-based agents by connecting them to an environment that requires action and generates consequential feed-

back. By enacting LLMs as economic agents in a macroeconomic system, we showed that opinions are not formed solely through linguistic interaction, but are actively driven by agents' own actions and the experiences that follow from them. Our results reveal a structured coupling between environmental dynamics and OD. Interactions with an action-demanding environment drive how individual opinions are formed through lived economic experiences. Furthermore, the resulting collective OD evolve in close alignment with the dynamics of the environment itself, giving rise to population-level patterns such as inequality-linked polarization that are inaccessible in language-only settings.

Limitations

Simplified linguistic interaction mechanisms

To isolate the effects of action-mediated environmental feedback on opinion formation, interactions among agents are intentionally simplified in the present framework. Agents are exposed to limited textual opinions, and information provided by institutions such as firms and the government is symmetric across households. Consequently, richer social processes—such as deliberative discussion, persuasion, networked communication, or voting—are not modeled. These mechanisms may interact in complex ways with experience-driven opinion change. Extending the framework to incorporate more structured and heterogeneous interaction designs remains an important direction for future work.

Fixed institutional rules and exogenous policy design

The macroeconomic environment is governed by fixed, rule-based institutions, including taxation and redistribution schemes that do not adapt in response to agents' opinions. While this design choice enables a clean analysis of how environmental dynamics drive opinion formation, it precludes the study of institutional change driven by collective beliefs or political pressure. Allowing institutional rules to change over time would enable the study of how OD interact with dynamic policy environments.

Moreover, policy interventions are not systematically compared across counterfactual regimes (e.g., with versus without intervention). Extending the framework to support controlled policy variations across otherwise identical environments would enable quasi-experimental analyses in the spirit of difference-in-differences (DiD).

Limited bidirectionality between opinions and actions

Although the framework explicitly captures how action-conditioned experiences drive OD, it does not yet realize fully bidirectional dynamics between opinions and the environment. In particular, opinions in the current setting do not directly influence agents' subsequent economic decisions. We argue that achieving such bidirectionality crucially depends on the design of linguistic interaction mechanisms: in real societies, decision-making is shaped not only by individual experiences but also by the perceived opinion climate formed through social communication. Designing interaction structures in which socially expressed opinions systematically affect agents' choices may enable co-evolutionary dynamics between collective opinions and environmental trajectories. Investigating such socially mediated feedback loops remains an important challenge for future research.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP25KJ1124.

References

- Robert Axelrod. 1997. The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2):203–226.
- Claudio Castellano, Santo Fortunato, and Vittorio Loreto. 2009. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81:591–646.
- Erica Cau, Valentina Pansanella, Dino Pedreschi, and Giulio Rossetti. 2025. [Language-driven opinion dynamics in agent-based simulations with LLMs](#). *Preprint*, arXiv:2502.19098.
- Yun-Shiuan Chuang, Agam Goyal, Nikunj Harlalka, Siddharth Suresh, Robert Hawkins, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy Rogers. 2024. Simulating opinion dynamics with networks of LLM-based agents. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3326–3346.
- Pedro Cisneros-Velarde. 2025. Biases in opinion dynamics in multi-agent systems of large language models: A case study on funding allocation. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 1889–1916.
- Charles W. Cobb and Paul H. Douglas. 1928. A theory of production. *The American Economic Review*, 18(1):139–165.
- Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. 2000. Mixing beliefs among

- interacting agents. *Advances in Complex Systems*, 03(01n04):87–98.
- Morris H. DeGroot. 1974. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.
- Noah E. Friedkin and Eugene C. Johnsen. 1990. Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3-4):193–206.
- Serge Galam. 2002. Minority opinion spreading in random geometry. *The European Physical Journal B - Condensed Matter and Complex Systems*, 25:403–406.
- Lewis R. Goldberg. 1990. An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216–1229.
- Mark S. Granovetter. 1978. Threshold models of collective behavior. *American Journal of Sociology*, 83:1420–1443.
- Ryuji Hashimoto, Takehiro Takayanagi, Masahiro Suzuki, and Kiyoshi Izumi. 2026. Agent-based simulation of a financial market with large language models. In *The International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2025)*, pages 20–28.
- Rainse Hegselmann and Ulrich Krause. 2002. Opinion dynamics and bounded confidence: Models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3):1–33.
- Masanori Hirano. 2026. Building LLM-based artificial market simulations: Can LLMs function as agents in multi-agent simulations for finance? In *The International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2025)*, pages 56–71.
- Bibb Latané. 1981. The psychology of social impact. *American Psychologist*, 36(4):343–356.
- Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024. EconAgent: Large language model-empowered agents for simulating macroeconomic activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pages 15523–15536.
- Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao. 2020. FinBERT: A pre-trained financial language representation model for financial text mining. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4513–4519.
- Meta. 2024. [The Llama 3 herd of models](#). *Preprint*, arXiv:2407.21783.
- Yev Meyer and Dane Corneil. 2025. [Nemotron-Personas-USA: Synthetic personas aligned to real-world distributions](#).
- Andrzej Nowak, Jacek Szamrej, and Bibb Latané. 1990. From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3):362–376.
- Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*.
- Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, Chen Gao, Fengli Xu, Fang Zhang, Ke Rong, Jun Su, and Yong Li. 2025. [AgentSociety: Large-scale simulation of LLM-driven generative agents advances understanding of human behaviors and society](#). *Preprint*, arXiv:2502.08691.
- Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523.
- Barry M. Staw, Lance E. Sandelands, and Jane E. Dutton. 1981. Threat rigidity effects in organizational behavior: A multilevel analysis. *Administrative Science Quarterly*, 26(4):501–524.
- Alexander J. Stewart, Nolan McCarty, and Joanna J. Bryson. 2020. Polarization under rising inequality and economic decline. *Science Advances*, 6(50):eabd4201.
- Joshua Tucker. 2023. *Computational social science for policy and quality of democracy: Public opinion, hate speech, misinformation, and foreign influence campaigns*, pages 381–403.
- Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. Minilm: deep self-attention distillation for task-agnostic compression of pre-trained transformers. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*.
- Duncan J. Watts. 2002. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99:5766–5771.
- Xinnong Zhang, Jiayu Lin, Xinyi Mou, Shiyue Yang, Xiawei Liu, Libo Sun, Hanjia Lyu, Yihang Yang, Weihong Qi, Yue Chen, Guanying Li, Ling Yan, Yao Hu, Siming Chen, Yu Wang, Xuanjing Huang, Jiebo Luo, Shiping Tang, Libo Wu, and 2 others. 2025. [SocioVerse: A world model for social simulation powered by LLM agents and a pool of 10 million real-world users](#). *Preprint*, arXiv:2504.10157.

	Parameter	Notation	Value
Government	Base tax rate	$\bar{\tau}$	0.010
	Basic income	b	5.000
Firm	Capital share	α	0.100
	Depreciation rate	d	0.005
	Productivity	A	0.950
	Initial price	$\bar{p}_{1,1}$	1.000
	Initial wage	$\bar{w}_{1,1}$	1.500
	Initial inventory	$K_{1,1}$	20.000
	Max capital for production	K_{\max}	50.000
	Price elasticity	η_p	0.200
	Wage elasticity	η_w	0.030
Household	Initial cash	$\forall i M_{1,i}$	$U(0, 100)$
	Low labor range	$\mathcal{I}_{\text{low}}^L$	[4, 5]
	Medium labor range	$\mathcal{I}_{\text{medium}}^L$	[5, 6]
	High labor range	$\mathcal{I}_{\text{high}}^L$	[6, 7]
	Low consumption range	$\mathcal{I}_{\text{low}}^C$	[5, 6]
	Medium consumption range	$\mathcal{I}_{\text{medium}}^L$	[6, 7]
	High consumption range	$\mathcal{I}_{\text{high}}^L$	[7, 8]

Table 2: Full list of hyperparameters in our simulation of OD with LLM-based households.

A Detailed Simulation Configurations

Table 2 summarizes the experimental setup of our simulation of OD with LLM-based households. Each simulation was conducted over $T = 150$ time steps with $n = 20$ agents. The household economic status category $f_{t,i}$ is determined by thresholding the normalized income $z_{t,i}$. Thresholds -1.282 , -0.524 , 0.524 , and 1.282 correspond to the boundaries between the five categories: very low, below-average, around-average, above-average, and very high. These thresholds correspond to the upper 10th and 30th percentiles of a standard normal distribution.

Algorithm 1 shows the detailed procedure of a simulation. Below, we describe the structure of the prompt used in our experiments, introducing its components step by step. We first specify the agent’s role and the basic economic setting in which it operates.

Prompt Example 1: Role and economic setting

You are a household in an economy. In this economy, there is only one good. Your money can be used only to buy the good or to pay tax. You will be happy if you consume the good. In order to buy a good, you have to earn

Algorithm 1: LLM-based OD simulation with macroeconomic environment.

Input: n household personas
for $i = 1$ **to** n **do**
 Initialize financial state $M_{t=0,i}$;
 Initialize opinion trajectory $\langle O_i \rangle = \{\}$;
for $t = 1$ **to** T **do**
 for $i = 1$ **to** n **do**
 Compute relative financial status $f_{t,i}$
 from $\{M_{t,j}\}_{j=1}^n$;
 Construct prompt from persona,
 history, environment state, and peer
 opinions;
 Query LLM with the prompt to
 obtain $(L_{t,i}, C_{t,i}, O_{t,i})$;
 Update household income Eq.(2);
 Update firm state (price \bar{p} and wage
 \bar{w} of a unit of good and labor, and
 inventory $K_{t,i}$) Eq.(4,5,6);
 Append $O_{t,i}$ to $\langle O_i \rangle$;

money in some ways. One way to earn money is to work.

Next, we condition each agent on a persona that specifies basic demographic attributes and person-

ality traits. Specifically, the attributes sex, age, marital status, education level, occupation, and city are assigned via stratified sampling. In addition to demographic attributes, each agent is endowed with personality traits based on the Big Five personality model (Goldberg, 1990). For each of the five dimensions—extroversion, agreeableness, conscientiousness, neuroticism, and openness to experience—one of two opposing trait descriptors (e.g., extroverted vs. introverted) is randomly sampled and assigned to the agent.

Prompt Example 2: Persona conditioning

```
Your persona to role-play is as follows: sex: Female, age: 29, ..., trait1: introverted, trait2: agreeable, ...
```

We next present the macroeconomic environment observable to agents. Prices, wages, and policy variables are provided exogenously as components of the environmental state.

Prompt Example 3: Public information

```
You have the following information:
Firm Announcement: Current price per a unit of goods: 1.64, Current wage per a unit of labor: 1.46 The price has been strongly increasing in the long term. Recently, the price has been moderately increasing in the short term. The wage has been stable in the long term. Recently, the wage has been slightly increasing in the short term.
Government Announcement: We keep the current basic income policy in order to ensure a minimum standard of living and support households currently facing financial difficulties. Each household receives 5 unit of cash as a basic income. The funding for basic income is covered by taxes paid by households with very high income. Tax rate is set to 20.0 percent for households with very high income, and 10.0 percent for other households, while households with very low income are exempt from taxation.
```

In our experiments, opinion exchange among agents is simplified as follows. At each stage (t, i) , agents are provided with the opinions expressed by other agents up to the two most recent preceding stages as part of the prompt.

Prompt Example 4: Others' opinions

```
Other households' opinions: - The current system is unfair to those who ... - ...
```

Next, we provide agents with information about their own past economic experiences and state transitions. This component constitutes the core of our framework, as it explicitly conditions agents on their lived experiences. By incorporating this information, expressed opinions are not merely reflections of the assigned persona, but are formed as outcomes of past actions and environmental feedback.

Prompt Example 5: Individual economic history

```
Your individual information: Holding cash: 33.69 Latest paid tax: 3.19 Latest received subsidy: 5.00 You received subsidy larger than the tax you paid. You were a net recipient of subsidies last time. Your financial status was initially below-average income, last time below-average income and is currently below-average income Your cash holding is currently slightly increasing compared to the last time.
```

In addition, agents are provided with their own previously expressed opinion from an earlier time step. This information serves as a short-term memory that introduces temporal continuity in opinion expression, allowing opinions to evolve in relation to past beliefs while remaining responsive to newly experienced economic conditions.

Prompt Example 6: Own last opinion

```
Your last opinion: This society seems to ...
```

Finally, agents are asked to simultaneously choose their actions and express an opinion. Responses are strictly constrained to a predefined JSON format. This design ensures that non-linguistic actions (labor and consumption) and linguistic outputs (opinions) are generated jointly on the basis of the same underlying economic experiences.

Prompt Example 7: Decision and opinion task

Based on this information, please provide the following decisions. Take all information into account.

- How much do you work? The more you work, the higher you earn money to buy goods. You only spend your time to work, no additional resource is required. In general, those who are financially struggling should work hard. Answer must be either "low", "medium", "high".
- How much do you consume goods? You will be happy if you consume more. Answer must be either "low", "medium", "high".
- Clearly state your opinion in your words about this society. Take your financial situation and your work-consumption decision made above into account. But you must not be ambiguous even if you are not sure. Try to offer your own distinct perspective. Decide your responses in the following JSON format. Do not deviate from the format, and do not add any additional words to your response outside of the format. Here are the answer format.
"work": "low"/"medium"/"high",
"consume": "low"/"medium"/"high",
"opinion": <str>, "stance": <int from -10 to 10>

B Computing Infrastructure

The experiment was conducted on a workstation equipped with the following hardware and software configurations:

- **CPU:** AMD Ryzen 9 3950X 16-Core Processor (32 threads, base clock 3.5 GHz, max boost 4.76 GHz)
- **GPU:** 2 × NVIDIA RTX 6000 Ada Generation (48 GB VRAM each)
- **Operating System:** Ubuntu 22.04.5 LTS (Jammy Jellyfish)

The complete source code used for the experiment, along with a full list of Python library versions and environment specifications, is provided in the supplementary material.

C Prompt Ablation

We examine whether the behavioral differentiation observed in RQ1 is driven by explicit prompt instruction or by the economic feedback mechanism itself. To isolate the effect of this instruction, we

	$\Delta^L(\text{low})$	$\Delta^C(\text{low})$	$\Delta^L(\text{high})$	$\Delta^C(\text{high})$
Baseline	0.682	-0.931	-0.086	0.463
w/o inst.	0.644	-0.872	-0.083	0.437

Table 3: Behavioral differentiation under prompt ablation. Baseline refers to the main simulation setting described in the paper. The w/o inst. condition corresponds to a modified prompt in which the sentence *In general, those who are financially struggling should work hard.* is removed.

construct an ablated prompt by removing the sentence: *In general, those who are financially struggling should work hard.* The full prompt specification is provided in Appendix A.

We rerun the simulation under identical conditions using the modified prompt, aggregating results over 30 independent runs. To quantify behavioral differentiation, we define the following normalized difference measure for labor supply:

$$\Delta^L(\tilde{f}) = \frac{N_{\text{high}}^L(\tilde{f}) - N_{\text{low}}^L(\tilde{f})}{N_{\text{high}}^L(\tilde{f}) + N_{\text{low}}^L(\tilde{f})} \quad (21)$$

where \tilde{f} denotes a financial status defined in Eq. (3.2). Here, $N_{\text{high}}^L(\tilde{f})$ is the total number of time steps across all agents and simulation runs in which agents in state \tilde{f} choose the *high labor* action, and $N_{\text{low}}^L(\tilde{f})$ is defined analogously for the low labor. Similarly, we define $\Delta^C(\tilde{f})$ for consumption decisions:

$$\Delta^C(\tilde{f}) = \frac{N_{\text{high}}^C(\tilde{f}) - N_{\text{low}}^C(\tilde{f})}{N_{\text{high}}^C(\tilde{f}) + N_{\text{low}}^C(\tilde{f})} \quad (22)$$

where $N_{\text{high}}^C(\tilde{f})$ and $N_{\text{low}}^C(\tilde{f})$ denote the counts of high (low) consumption actions, respectively.

Table 3 summarizes the results. The results show that the direction of behavioral differentiation remains unchanged across conditions. Low-income households continue to exhibit higher labor supply and lower consumption, while high-income households display the opposite pattern. Although the magnitude of differentiation decreases slightly (approximately 5–6%), the qualitative structure is preserved. These findings indicate that the explicit instruction strengthens, but does not generate, the observed behavioral patterns. Instead, the differentiation primarily arises from the economic feedback mechanism embedded in the environment rather than from direct prompt compliance.