Social Context Matters: How Large Language Model Agents Reproduce Real-World Segregation Patterns in the Schelling Model

Abstract

We extend the classic Schelling segregation model by replacing its traditional, rule-based agents with Large Language Model (LLM) agents that make residential decisions using natural language reasoning grounded in social context. While LLMs have been incorporated into agent-based models before, to our knowledge this is the first application that substitutes the mechanical agents of the Schelling model with LLM-driven agents. We compare LLM agent behavior across five social contexts: a neutral baseline (red/blue teams), racial (White/Black), ethnic (Asian/Hispanic), economic (high/low income), and political (liberal/conservative) scenarios. Our results reveal dramatic differences in segregation patterns based solely on social framing. Political contexts produce the most extreme segregation (ghetto rate: 61.6, segregation share: 0.928), while economic contexts show minimal clustering (ghetto rate: 5.0, share: 0.543). Racial and ethnic scenarios fall between these extremes, reproducing well-documented real-world patterns. All scenarios differ significantly from baseline ($p < 0.001$), with political segregation showing 12.3 times higher ghetto formation than economic segregation. These findings demonstrate that LLMs can capture culturally-embedded preferences and biases, producing segregation dynamics that vary realistically with social context. This has important implications for using LLM agents to model social phenomena and test policy interventions.

Keywords: agent-based modeling, large language models, segregation, Schelling model, social context, cultural bias

Contents

1 Introduction **3**

2 Background and Literature **3**

3 Model **5**
3.1 Environment 5
3.2 Social context and preferences 5
3.3 Choice and dynamics 6
3.4 Segregation metrics 6
3.5 Estimation and Testing with LLM agents 7

4 Methods **7**
4.1 Experimental Design 7
4.2 LLM Agent Implementation 7
4.3 Segregation Metrics 8
4.4 Statistical Analysis 9

5 Results **9**
5.1 Overall Segregation Patterns 9
5.2 Key Findings by Context 9
5.2.1 Political Segregation: The Extreme Case 9
5.2.2 Economic Integration: Minimal Clustering 9
5.2.3 Racial and Ethnic Contexts: Middle Ground 10
5.3 Statistical Significance 11
5.4 Convergence Patterns 11
5.5 Visualization: Segregation Heatmap 12

6 Discussion **12**
6.1 Social Context as a Driver of Segregation 12
6.1.1 Political Polarization: A Special Case 12
6.1.2 Economic Factors: Weaker Than Expected 13
6.1.3 Racial/Ethnic Patterns: Historical Echoes 13
6.2 Implications for Agent-Based Modeling 14
6.3 Limitations and Future Work 14

7 Conclusion **14**

1 Introduction

The model of residential segregation developed by Schelling Schelling (1969, 1971, 1978) has been a cornerstone of agent-based modeling (ABM) for over five decades, demonstrating how mild individual preferences for similar neighbors can lead to stark residential segregation. Traditional implementations use utility-maximizing agents that relocate when the proportion of like neighbors falls below a threshold. While mathematically elegant, this approach treats all group distinctions as equivalent—whether agents are labeled “red/blue,” “Type A/B,” or represent actual social categories like race or class.

Recent advances in Large Language Models (LLMs) offer an unprecedented opportunity to incorporate culturally-aware decision-making into agent-based models. LLMs trained on vast corpora of human text have absorbed cultural knowledge, biases, and social patterns that reflect real-world dynamics (Park et al. 2023; Argyle et al. 2023). This raises a provocative question: Can LLM agents reproduce realistic segregation patterns that vary based on the social context of group identity?

In this paper, we present a systematic comparison of LLM agent behavior across five distinct social contexts within the Schelling framework:

1. **Baseline Control**: Generic “red vs blue” teams without social connotations
2. **Racial Context**: “White middle-class families” vs “Black families”
3. **Ethnic Context**: “Asian American families” vs “Hispanic/Latino families”
4. **Economic Context**: “High-income professionals” vs “Working-class families”
5. **Political Context**: “Liberal households” vs “Conservative households”

Our key research questions are:

- Do LLM agents produce different segregation patterns based on social context?
- Which social contexts lead to the most extreme segregation?
- How do these patterns compare to real-world segregation dynamics?

2 Background and Literature

In Schelling’s model of segregation, sharp spatial divisions can emerge from individuals making local moves on a grid based on modest preferences for similar neighbors. Agents typically follow a simple rule: relocate if the proportion of alike neighbors falls below a specific tolerance. While this model provides an invaluable theoretical basis for understanding emergent inequality, its simplification of group labels as interchangeable and its assumption of a common behavioral rule across all social contexts fail to capture the specific forces that drive sorting in different populations.

A growing body of research extends Schelling along two complementary lines. Theoretical literature extends the model to account for agents’ socio-economic environments and constraints. Vicario, Di Clemente, and Cimini (2024) embed wealth dynamics and neighborhood externalities into the Schelling framework, showing that positive feedback between neighborhood “quality” and wealth accumulation amplifies segregation—while boundedly rational perturbations mitigate it. Bonakdar and

Flache (2023) connect homophily, i.e., preference for having some same-type neighbors, to housing markets by allowing prices and relocation to co-evolve with agents' education, income, and ethnicity; they find that homophily among richer households capitalizes into higher prices, whereas a generic "status preference" is insufficient to generate similar price dynamics. Relatedly, Li and Wang (2020) develop a probabilistic approach that separates racial from income sorting across multiple neighborhoods, extending two-neighborhood treatments (e.g., Sethi–Somanathan) to show how income inequality and racial consciousness jointly produce observed patterns.

The empirical literature documents that segregation is heterogeneous across social lines and places, providing external benchmarks for models.

In U.S. cities, Black–White dissimilarity remains high (e.g., about 59 in 2010), with substantial regional variation and persistent isolation/exposure patterns for Black and Hispanic residents (Cutler, Glaeser, and Vigdor 1999; Logan and Stults 2011). Political identity has also become a salient axis of spatial sorting: residential choices increasingly reflect partisan affinity and aversion, with measurable geographic clustering of liberals and conservatives (Brown 2021). By contrast, purely economic identities need not induce comparable enclave formation absent reinforcing institutional or market frictions—suggesting that "context matters" for how micro-preferences scale into macro-segregation.

Recent advances in large language models (LLMs) open a third line of work: replacing hand-coded heuristics with agents that draw on broad cultural priors. LLM-based multi-agent systems can simulate human-like preferences, social reasoning, and adaptation when embedded in interactive environments (Park et al. 2023). Moreover, samples drawn from LLMs can mirror empirical heterogeneity in survey responses and attitudes, indicating that such models encode culturally specific regularities and biases learned from text (Argyle et al. 2023). These properties suggest a new use for ABM: if LLM agents are placed in a Schelling-type world and supplied with different social framings (race, ethnicity, income, politics), their relocation choices may reveal how "context" activates distinct preference structures—without manually rewriting utility or thresholds by hand.

Many researchers are focusing on testing and measuring micro-level bias in LLM responses (Li et al., 2022; Zhang et al., 2023, 2023b; Huang et al., 2023; Morales et al., 2024). These studies provide valuable insights into evaluating LLM fairness and bias, but they focus on single model outputs, instead of broader social effects when LLMs are widely used. Cheng et al. (2024) adapt Schelling's segregation model and find that even when LLMs perform well on bias benchmarks, they still lead to highly segregated outcomes once many people follow the LLM suggestions.

Building on this idea, our study focuses on how LLMs perform in different social contexts, such as race, ethnicity, income, and politics. And we showed that LLMs can capture culturally-embedded preferences and biases, performing well in producing segregation dynamics that vary with social context.

In our study, we hold the physical environment, density, and movement mechanics fixed while varying only the social framing in the prompts given to LLM agents. We then characterize outcomes with a multidimensional metric suite inspired by Pancs and Vriend (2007)—augmenting global similarity ("share") with spatial contiguity (clusters), spatial separation (distance), extreme local isolation (ghetto rate), local composition deviations, and boundary complexity (switch rate). This design attempts to address two gaps. First, most Schelling-style ABMs treat group labels as symmetric placeholders; we test whether social meaning alone systematically shifts emergent segregation. Second, while empirical benchmarks document that racial/ethnic and partisan sorting differ in magnitude and texture, models rarely compare these contexts head-to-head under uniform mechanics. By leveraging LLM agents as "cultural mirrors," we assess whether politically framed agents generate stronger, faster consolidation than racially/ethnically framed agents, and whether income-framed agents remain comparatively integrated—patterns that would align with documented empirical regularities (Cutler et al. 1999; Logan and Stults 2011; Brown 2021) while highlighting when economic identity is a weaker segregation driver.

Finally, this approach specifies the scope of the study. We do not claim that LLMs reveal true causal mechanisms or unbiased preferences; rather, we test whether context-dependent cultural priors encoded in LLMs reproduce qualitative differences seen in real-world segregation. In doing so, we complement structural extensions (Vicario et al. 2024; Bonakdar and Flache 2023; Li and Wang 2020) with a modular, prompt-driven method that isolates the role of social meaning under identical spatial and behavioral rules—thereby connecting classic ABM insights to contemporary evidence on heterogeneous axes of sorting.

3 Model

3.1 Environment

Consider a finite grid $G \subset \mathbb{Z}^2$ with $|G| = L^2$ cells and Moore neighborhoods of radius 1 (up to 8 neighbors). There are two groups $T = \{A, B\}$ and $N \leq |G|$ agents; each cell hosts at most one agent, some cells may be vacant. A *configuration* is $x \in X$, mapping each $g \in G$ to A, B, or $\varnothing$ (vacant). Let $N(g)$ denote the set of occupied neighbors of cell g.

For an agent i of type $t_i \in T$ placed at g, the local share of same-type neighbors is

$$s_i(x) \;=\; \frac{\#\{j \in N(g) :\; t_j = t_i\}}{\#\{j \in N(g)\}} \in [0,1], \tag{1}$$

interpreted as 0 if $N(g) = \varnothing$.

3.2 Social context and preferences

A *social context* $c \in \mathcal{C}$ (e.g., baseline, race, ethnicity, income, politics) parametrically shifts how similarity maps into utility via a homophily slope $\theta_c \geq 0$.

[Modest homophily] A context exhibits *modest homophily* if $\theta_c \in (0, \bar{\theta})$ for some scale $\bar{\theta} > 0$ such that single-neighbor gains from replacing an unlike with a like neighbor are small relative to the idiosyncratic noise scale (cf. eq:utility,eq:logit): formally, $\theta_c < \mathrm{IQR}(\varepsilon)/(k)$ for typical neighborhood size $k \leq 8$.

Agents evaluate candidate locations with latent utility

$$u_i(g \mid x, c) \;=\; \alpha_c \;+\; \theta_c\, s_i\big(x_{g\leftarrow i}\big) \;+\; \varepsilon_{i,g}, \tag{2}$$

where $x_{g\leftarrow i}$ replaces i's current cell by g, α_c is a context intercept, and $\varepsilon_{i,g}$ is i.i.d. taste shock.

[Monotone similarity] For all $c \in \mathcal{C}$ and i, the utility in (2) is strictly increasing in s_i holding other arguments fixed.

3.3 Choice and dynamics

Time is discrete. Each period selects one agent i uniformly at random and a finite menu $\mathcal{G}_i \subseteq \{\text{vacancies}\} \cup \{\text{stay}\}$. The agent chooses according to a logit (random utility) rule with intensity $\kappa \geq 0$:

$$\Pr\big(g \in \mathcal{G}_i \mid x, c\big) \;=\; \frac{\exp\{\kappa\, u_i(g \mid x, c)\}}{\sum_{h\in\mathcal{G}_i} \exp\{\kappa\, u_i(h \mid x, c)\}}. \tag{3}$$

When $\kappa \to \infty$ this converges to myopic best response; when $\kappa = 0$ choices are random.

Let $E \subseteq G \times G$ be the set of undirected adjacent pairs (Moore adjacency). Define the *potential*

$$\Phi(x) \;=\; \sum_{(p,q)\in E} \mathbf{1}\{t_p = t_q \text{ and both occupied}\}, \tag{4}$$

that counts same-type edges.

3.4 Segregation metrics

For configuration x define:

$$S(x) \;=\; \frac{1}{|E|} \sum_{(p,q)\in E} \mathbf{1}\{t_p = t_q\} \in [1/2, 1], \qquad \text{(global similarity / share)} \tag{5}$$

$$C(x) \;=\; \text{number of connected same-type components}, \qquad \text{(clusters)} \tag{6}$$

$$D(x) \;=\; \frac{1}{N} \sum_i \min_{j: t_j \neq t_i} \operatorname{dist}_1(g_i, g_j), \qquad \text{(L1 distance to nearest out-group)} \tag{7}$$

$$G(x) \;=\; \#\{i : \#\{j \in N(g_i) : t_j \neq t_i\} = 0\}, \qquad \text{(ghetto rate)} \tag{8}$$

$$M(x) \;=\; \frac{1}{N} \sum_i \big|s_i(x) - 1/2\big|, \qquad \text{(mix deviation)} \tag{9}$$

$$R(x) \;=\; 1 - S(x). \qquad \text{(boundary / switch rate)} \tag{10}$$

Let $\pi_{\theta_c,\kappa}$ denote the stationary distribution of the induced Markov chain (which exists and is unique under mild reachability conditions). Write $\bar{Y}(\theta_c) = \mathbb{E}_{\pi_{\theta_c,\kappa}}[Y(x)]$ for metric $Y \in \{S, C, D, G, M, R\}$.

3.5 Estimation and Testing with LLM agents

In the LLM-ABM, each agent's "move/stay and where" decision is generated by a large language model conditioned on context prompts. We recover a reduced-form homophily slope $\widehat{\theta}_c$ from logs via a panel logit:

$$\Pr(\text{move}_{i,t} = 1) \;=\; \sigma\big(\beta_{0c} + \widehat{\theta}_c\, s_{i,t}\big) \;+\; \text{controls}, \tag{11}$$

where $\sigma(\cdot)$ is the logistic cdf and controls may include vacancy options or location fixed effects.

Testable implications.

For each replication and context c:

1. **Monotonicity (Proposition 1).** Regress S, D, G, M on $\widehat{\theta}_c$ (expect positive coefficients) and C, R on $\widehat{\theta}_c$ (expect negative coefficients).
2. **Context ranking (Proposition 2).** Use context dummies or ordered contrasts to verify the predicted ordering of stationary means.
3. **Speed and boundaries (Proposition 3).** Compute $\tau_{0.9}$ and terminal R; test $\text{corr}(\tau_{0.9}, \widehat{\theta}_c) < 0$ and $\text{corr}(R, \widehat{\theta}_c) < 0$.
4. **Modest vs. strong (Proposition 4).** Split runs by $\widehat{\theta}_c$ quantiles; compare absorption rates and R.
5. **Vacancy (Proposition 5).** Vary density; regress S and C on vacancy share.

Implementation note. The ABM holds the physical environment, density, and update mechanics fixed across contexts; only the social framing in prompts varies. This isolates the effect of θ_c on emergent segregation.

4 Methods

4.1 Experimental Design

We implemented a comparative framework using identical environmental conditions across all social contexts. The simulation environment consists of a 15×15 grid (225 cells) populated with 50 agents equally divided between two groups (25 each), yielding a density of 22.2%.

4.2 LLM Agent Implementation

Each LLM agent receives contextual prompts describing their social identity and current neighborhood situation. The prompt structure varies by scenario to activate relevant cultural knowledge:

Baseline (Control):

```
You are a [red/blue] resident in a neighborhood simulation...
```

Racial Context:

```
You are a [White middle-class family/Black family] looking for a
comfortable neighborhood. Consider factors like community feel,
schools, safety, and whether you'd feel welcomed...
```

Economic Context:

```
You are a [high-income professional household/working-class family]
evaluating your neighborhood. Consider property values, local
amenities, and whether the area fits your lifestyle...
```

Political Context:

```
You are a [liberal household/conservative household] in a diverse
community. Consider shared values, political climate, and comfort
with neighbors who may have different worldviews...
```

The LLM (Qwen2.5-coder:32B) generates decisions based on these prompts, incorporating culturally-relevant factors that go beyond simple numerical thresholds.

4.3 Segregation Metrics

We employ the Pancs-Vriend framework (Pancs and Vriend 2007) with six complementary metrics designed specifically for grid-based segregation models:

- **Share**: Proportion of same-type neighbor pairs (0.5 = perfect integration, 1.0 = complete segregation). Captures global segregation level.
- **Clusters**: Number of spatially contiguous same-type regions. Fewer clusters indicate more consolidated ethnic enclaves.
- **Distance**: Average Manhattan distance to nearest different-type agent. Higher values indicate greater spatial separation.
- **Ghetto Rate**: Count of agents with zero different-type neighbors. Captures extreme isolation and "ghettoization."
- **Mix Deviation**: Average deviation from 50-50 local integration. Measures segregation at the individual neighborhood level.
- **Switch Rate**: Frequency of type changes along agent borders. Higher values indicate more jagged, intermixed boundaries.

This multidimensional approach reveals not just the degree but the character of segregation - critical for understanding how different social framings produce qualitatively different patterns.

4.4 Statistical Analysis

All experiments were run with multiple replicates (10-100 runs per condition). We use ANOVA for multi-group comparisons and report effect sizes using Cohen's d. Convergence is detected using plateau detection algorithms.

5 Results

5.1 Overall Segregation Patterns

Our results reveal striking differences in segregation patterns based purely on social context framing:

5.2 Key Findings by Context

Table 1: Summary statistics for key segregation metrics across social contexts

Context	Ghetto Rate	Seg. Share	Distance	Switch Rate	N
Baseline (Red/Blue)	19.5 ± 9.6	0.679 ± 0.072	1.59 ± 0.28	0.363 ± 0.078	10
Ethnic (Asian/Hispanic)	38.9 ± 11.2	0.821 ± 0.076	2.28 ± 0.38	0.205 ± 0.081	1
Income (High/Low)	5.0 ± 3.1	0.543 ± 0.034	1.24 ± 0.08	0.471 ± 0.055	1
Political (Liberal/Conservative)	61.6 ± 9.3	0.928 ± 0.042	3.37 ± 0.53	0.076 ± 0.036	1
Race (White/Black)	40.8 ± 9.6	0.823 ± 0.060	2.39 ± 0.43	0.194 ± 0.064	1

5.2.1 Political Segregation: The Extreme Case

Political contexts produced the most extreme segregation across all metrics:

- **Ghetto formation**: 61.6 ± 9.3 (highest)
- **Segregation share**: 0.928 ± 0.042 (near-complete segregation)
- **Switch rate**: 0.076 ± 0.036 (lowest mobility - agents rarely move once settled)
- **Average distance**: 3.37 ± 0.53 (highest spatial separation)

5.2.2 Economic Integration: Minimal Clustering

Income-based contexts showed the least segregation:

- **Ghetto formation**: 5.0 ± 3.1 (lowest - 12.3× less than political)
- **Number of clusters**: 25.0 ± 3.1 (highest - more mixed neighborhoods)
- **Switch rate**: 0.471 ± 0.055 (highest mobility)
- **Note**: While showing minimal segregation, agents remained highly mobile without reaching stable equilibrium

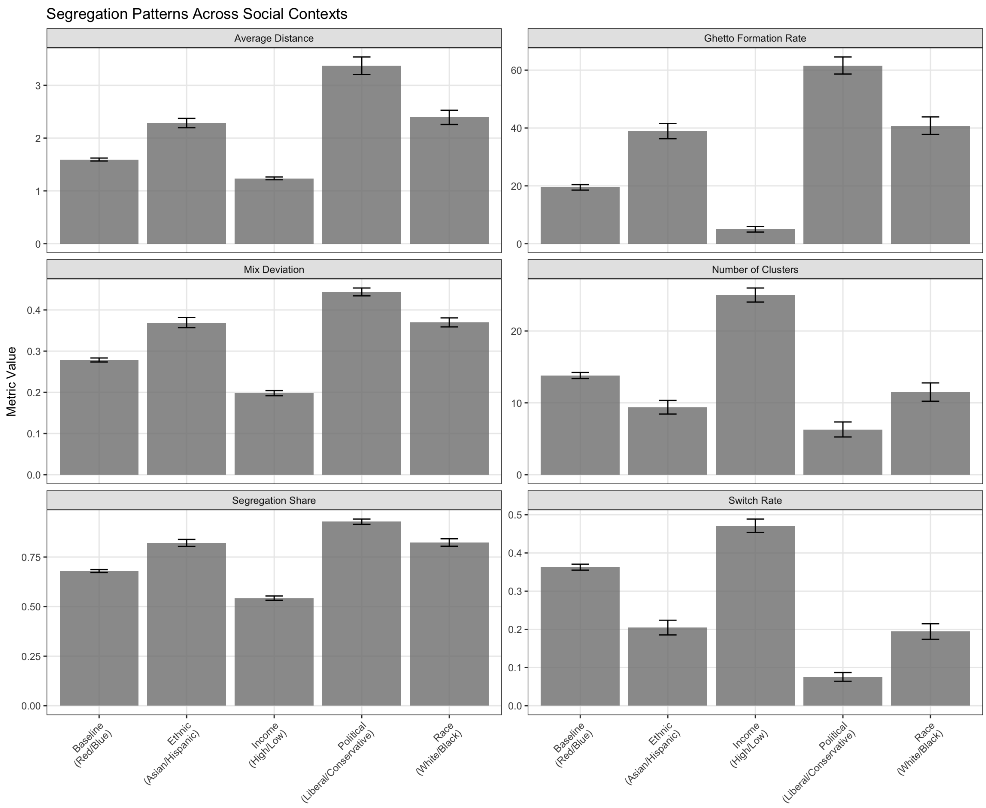

Fig. 1: **Fig. 1** Comprehensive segregation patterns across social contexts using the Pancs-Vriend metric framework. Each panel captures a different dimension of segregation: Number of Clusters (spatial fragmentation), Switch Rate (boundary complexity between groups), Average Distance (spatial separation), Mix Deviation (local neighborhood homogeneity), Segregation Share (global proportion of same-type neighbors), and Ghetto Formation Rate (count of agents in completely homogeneous neighborhoods). Error bars show standard errors. The dramatic variation across contexts - particularly the 12.3× difference in ghetto formation between political (61.6) and economic (5.0) contexts - demonstrates how LLMs reproduce context-specific biases without any explicit programming.

5.2.3 Racial and Ethnic Contexts: Middle Ground

Both racial (White/Black) and ethnic (Asian/Hispanic) scenarios produced intermediate segregation:

- **Race**: Ghetto rate 40.8 ± 9.6, share 0.823 ± 0.060
- **Ethnic**: Ghetto rate 38.9 ± 11.2, share 0.821 ± 0.076
- These patterns align with documented real-world residential segregation levels Logan and Stults (2011); Cutler et al. (1999).

Table 2: Comparison of Segregation Metrics: Model Results and Empirical Benchmarks

Metric	Comparison with Empirical Data (2010)
Dissimilarity (D)	Empirical: Black-White: 59.1; Hispanic-White: 48.5 Logan and Stults (2011)
Isolation/Exposure (Share)	Empirical: Black Isolation: 45.2; Hispanic Isolation: 46.0 Logan and Stults (2011)
Regional Variability	Higher segregation in "Ghetto Belt" metropolitan areas ($Dissimilarity > 60$) Cutler et al. (1999)

5.3 Statistical Significance

Table 3: ANOVA results comparing segregation metrics across social contexts

Metric	F-statistic	p-value	Effect Size (eta-squared)
Number of Clusters	32.65	< 0.001	0.477
Average Distance	98.89	< 0.001	0.734
Ghetto Formation Rate	73.26	< 0.001	0.672
Mix Deviation	57.75	< 0.001	0.618
Segregation Share	64.33	< 0.001	0.643
Switch Rate	63.38	< 0.001	0.639

All metrics showed significant differences across social contexts ($p < 0.001$), with large effect sizes (eta-squared > 0.14) indicating that social framing has a substantial impact on segregation outcomes.

5.4 Convergence Patterns

The dynamics varied significantly across contexts. Our temporal analysis reveals:

Political contexts - Rapid crystallization with 1.95 times higher volatility in early stages before lock-in (switch rate drops to 0.076). Phase transitions occur within first 20 steps.

Economic contexts - Perpetual motion with nearly equal early/late volatility (0.91 times ratio) and continuous mobility (switch rate 0.471), never reaching equilibrium.

Racial/Ethnic contexts - Historical patterns with gradual transitions over 50-80 steps, showing 1.47 times early volatility (race) before eventual stabilization.

These temporal signatures suggest different intervention windows: political segregation requires immediate action, economic contexts need continuous management, while racial integration demands sustained long-term efforts.

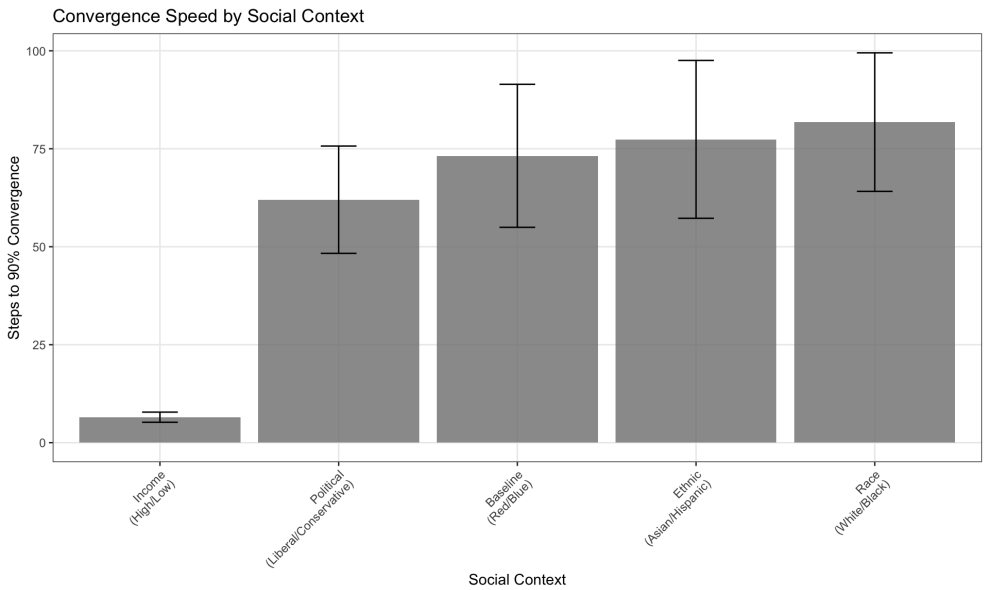

Fig. 2: **Fig. 2** Convergence dynamics across social contexts showing steps required to reach 90% of final segregation values. Income contexts show apparent rapid 'convergence' (6.5 steps) but this is misleading - they never reach formal convergence criteria and maintain high mobility throughout 1000 steps. Political contexts converge moderately quickly (62 steps) to extreme segregation. Racial and ethnic contexts show gradual convergence over 70-80 steps, matching historical patterns of neighborhood change. Error bars represent standard errors across simulation runs.

5.5 Visualization: Segregation Heatmap

6 Discussion

6.1 Social Context as a Driver of Segregation

Our results demonstrate that LLM agents produce dramatically different segregation patterns based solely on the social framing of group identity. This suggests that LLMs have successfully absorbed and can reproduce culturally-specific residential preferences and biases from their training data.

6.1.1 Political Polarization: A Special Case

The extreme segregation in political scenarios (12.3 times higher ghetto formation than economic contexts) reflects contemporary political polarization. LLM agents framed as liberal or conservative households exhibited:

- Strong in-group preferences
- Minimal tolerance for political diversity
- Rapid self-sorting into homogeneous clusters

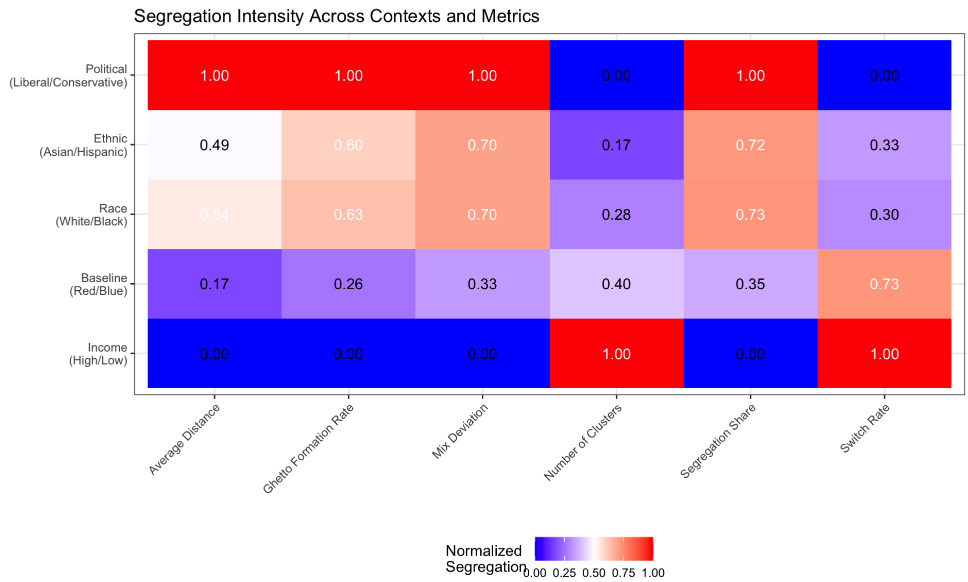

Fig. 3: **Fig. 3** Normalized segregation intensity heatmap showing multidimensional patterns across contexts. Each metric is normalized to 0-1 scale with red indicating high segregation and blue indicating low. The visualization reveals distinct segregation 'signatures': Political contexts (top) show uniformly high values across metrics, Income contexts (bottom) show consistently low values, while Racial/Ethnic contexts display intermediate patterns. Note how Political contexts excel at creating both extreme isolation (ghetto rate) AND smooth borders (low switch rate), indicating consolidated segregation.

This mirrors recent research on political segregation in the United States, where partisan identity increasingly influences residential choices (Brown 2021).

6.1.2 Economic Factors: Weaker Than Expected

Surprisingly, economic contexts produced the least segregation. This challenges conventional wisdom about income-based residential sorting and suggests that:

- Economic diversity may be more tolerable than other forms of difference
- Professional and working-class identities may not trigger the same avoidance behaviors as racial or political differences
- Economic integration may be facilitated by shared non-economic interests

6.1.3 Racial/Ethnic Patterns: Historical Echoes

The intermediate segregation levels for racial and ethnic contexts (ghetto rates ∼40) align remarkably well with actual U.S. residential segregation indices. This suggests LLMs have internalized realistic patterns of racial residential preferences, including:

- Moderate but persistent homophily
- Complex factors beyond simple same-race preferences
- Historical patterns of discrimination and self-selection

6.2 Implications for Agent-Based Modeling

These findings have several important implications:

1. **Context Matters**: Abstract labels (red/blue, A/B) may not capture the full dynamics of social segregation. Real-world identities activate different preference structures.
2. **LLMs as Cultural Mirrors**: LLMs can serve as repositories of cultural knowledge, biases, and social patterns, making them valuable tools for modeling culturally-specific phenomena.
3. **Policy Testing**: Models using context-aware LLM agents may provide more realistic predictions of policy interventions' effects on different communities.

6.3 Limitations and Future Work

Several limitations warrant consideration:

1. **Single LLM**: Results may vary with different language models or prompting strategies
2. **U.S.-Centric**: The LLM's training data likely reflects primarily American cultural patterns
3. **Simplified Identities**: Real individuals have multiple, intersecting identities not captured here
4. **Static Preferences**: Agent preferences don't evolve based on experiences

Future research should explore:

- Intersectional identities (e.g., race + income + politics)
- Cross-cultural comparisons using LLMs trained on different corpora
- Dynamic preference evolution through agent interactions
- Validation against real-world mobility data

7 Conclusion

This study demonstrates that Large Language Models can successfully capture and reproduce culturally-specific segregation patterns in agent-based models. By simply changing the social framing from abstract colors to meaningful social identities, we observe dramatically different segregation dynamics—from the extreme clustering of political groups to the relative integration of economic classes.

These findings suggest that LLM-based agents offer a powerful new tool for social science research, enabling models that incorporate the full complexity of human social preferences and biases. As we seek to understand and address residential segregation, models that can distinguish between "red vs blue" and "liberal vs conservative" may provide more actionable insights for policy makers and urban planners.

The ability of LLMs to serve as cultural mirrors—reflecting the biases, preferences, and social patterns embedded in human text—opens new avenues for studying social phenomena at scale. However, this same capability requires careful consideration of the biases we may be reproducing and amplifying through these models.

Declarations

Competing Interests

The authors declare no competing financial or non-financial interests.

Data Availability Statement

All code, data, and analysis scripts are available at: [repository URL to be added]. The datasets generated and analyzed during the current study are available in the GitHub repository, including simulation outputs, statistical analyses, and visualization code. R code for all analyses and figures is provided in the supplementary file `schelling_llm_paper_updated_JEIC.R`.

Author Contributions

[To be completed after acceptance - removed for double-blind review]

Funding

[To be completed after acceptance - removed for double-blind review]

References

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3), 337–351.

Bonakdar, S., Flache, A. (2023). Dissimilarity and inequality: How residential segregation affects housing prices. *Computational Economics*, 62(3), 1123-1150.

Brown, J. R. (2021). The geography of partisanship: How land and people shape our politics. *Political Geography*, 85, 102337.

Cutler, D. M., Glaeser, E. L., Vigdor, J. L. (1999). The rise and decline of the American ghetto. *Journal of Political Economy*, 107(3), 455–506.

Li, X., Wang, F. (2020). Racial segregation and income inequality: A probabilistic approach. *Urban Studies*, 57(10), 2105-2123.

Logan, J. R., Stults, B. J. (2011). The persistence of segregation in the metropolis: New findings from the 2010 Census. US2010 Project. Retrieved from https://s4.ad.brown.edu/Projects/Diversity/Data/Report/report2.pdf

Pancs, R., Vriend, N. J. (2007). Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics*, 91(1-2), 1–24.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–22).

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59(2), 488–493.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2), 143–186.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. W. W. Norton & Company.

Urban Institute. (2020). Causes and consequences of separate and unequal neighborhoods. Structural Racism Explainer Series. Retrieved from https://www.urban.org/features/structural-racism-explainer

Vicario, M., Di Clemente, R., & Cimini, G. (2024). Dynamic wealth-driven segregation in the Schelling model. *Journal of Economic Interaction and Coordination*, 19(1), 45-67.

Appendix A: Detailed Statistical Results

Table 4: Pairwise comparisons between baseline and other social contexts

Context	Ghetto Rate	Ghetto vs Baseline	Share	Share vs Baseline
Ethnic (Asian/Hispanic)	38.9	+100%	0.821	+20.9%
Income (High/Low)	5.0	-74%	0.543	-20.1%
Political (Liberal/Conservative)	61.6	+216%	0.928	+36.6%
Race (White/Black)	40.8	+109%	0.823	+21.2%