Emergent Relational Order in LLM Agent Societies: From Collective Affect to Authority Stratification

Anonymous ACL submission

Abstract

Fei Xiaotong's Differential Order Pattern characterizes rural society as egocentric and relationally graded, with cooperation attenuating over social distance, yet lacks mechanistic operationalization. Existing LLM-based social simulations primarily address short-term coordination, leaving the emergence of authority, affect, and division of labor underexplored. We propose CAREB-MAS, a multi-agent framework modeling general social cognition rather than culture-specific rules. Agents reason through an emotion–ethics–belief chain, maintain dynamically evolving egocentric identities, and act under relationally graded utilities, while the macro environment specifies only individual production, preference-based allocation, and minimal interaction protocols. Across long-horizon simulations, agents spontaneously reproduce five core Differential Order phenomena: stable labor specialization, guanxi-based economic ethics, relationally decaying cooperation, emergent relational authority , and clan-based center–periphery stratification. These patterns systematically vary with production structure, bridging mechanical and organic modes of social integration. Together, these results recast Differential Order as a structure-sensitive emergent outcome of general social mechanisms, demonstrating LLM-based multi-agent simulation as a generative testbed for studying social structure and change.

1 Introduction

The transformation of traditional societies into modern ones constitutes one of the most consequential problems in social science, reshaping social structure, ethical obligations, and patterns of cooperation over long historical horizons (Sorokin, 2017; Lazer et al., 2021; Bhatt, 2022). Classical sociology has long addressed this question. In the Western canon, Durkheim emphasized the transition from mechanical to organic solidarity as a consequence of changing production structures and divisions of labor (Durkheim, 2019).

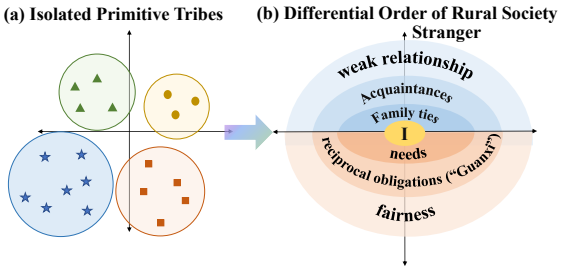


Figure 1: Schematic comparison of social organization: (**a**) discrete social groups without relational gradation; (**b**) differential order, where social relations are organized as ego-centered concentric layers reflecting graded relational proximity.

Beyond the Western canon, alternative theoretical traditions articulate distinct accounts of how social order is organized under economic change. A paradigmatic example is Fei Xiaotong's Differential Order Pattern (Fei et al., 1992), which conceptualizes rural society as composed of overlapping, ego-centered relational circles. In this view, moral obligations, trust, and cooperation attenuate with relational distance, forming a graded structure rather than a uniform social field. Figure 1 illustrates this core intuition, contrasting isolated primitive groups with the concentric, relationally differentiated structure of rural society.

Despite their influence, theories of social transformation face persistent methodological limitations. Ethnographic approaches provide rich contextual insight but lack experimental control, while quantitative methods often reduce long-term dynamics to static correlations. Conventional agent-based models typically rely on hand-coded rules, limiting their ability to capture the co-evolution of emotion, ethics, authority, and social structure (Eguíluz et al., 2005; Zimmermann et al., 2004). This gap motivates a central question:

Can long-term social structural change be studied through controlled experiments that model general social cognition rather than culture-specific rules?

Recent advances in large language models (LLMs) open new possibilities. LLM-based agents have demonstrated the ability to reproduce complex sociological dynamics, including Hobbesian social contracts (Dai et al., 2024), the cultural evolution of cooperation norms (Vallinder and Hughes, 2024), and large-scale institutional emergence (Piao et al., 2025). However, existing work has largely focused on short-term coordination or institutional outcomes. Whether LLM-based simulations can generate long-horizon social structures—and whether they can be used to examine the generality of non-Western sociological theories—remains underexplored.

In this paper, we propose **CAREB-MAS**, a multi-agent simulation framework designed to investigate the generative mechanisms of social structure. Agents are not programmed with Fei's theory or explicit culture-specific rules. Instead, they reason through an *emotion–ethics–belief* chain, maintain dynamically evolving egocentric identity representations shaped by in-group and out-group salience, and act under relationally graded utilities. The macro environment specifies individual production and preference-based allocation, with no centralized coordination or institutional authority.

Across long-horizon simulations, agents spontaneously reproduce five core phenomena associated with the Differential Order Pattern: stable labor specialization, guanxi-based reciprocity, cooperation attenuation with relational distance, emergent relational authority , and clan-based center–periphery stratification. Importantly, these patterns systematically vary with production structure. Under symmetric endowments, social integration relies primarily on relational ethics and family identity, while under complementary endowments, authority and cooperation reorganize around functional contribution and cross-group dependence. This contrast parallels Durkheim's distinction between mechanical and organic solidarity, suggesting that differential order is best understood as a structure-sensitive emergent outcome rather than a fixed cultural template.

We summarize our contributions as follows:

- We show that LLM-based agents can generatively reproduce social structures predicted by a classical sociological theory, without explicit cultural encoding.
- We introduce a general multi-agent framework integrating emotion, ethics, and relational cognition as interacting mechanisms.
- We show that LLM-based simulation can serve as a generative experimental testbed for studying long-term social dynamics and structural change.

2 RELATED WORK

2.1 Social Transformation and Rural Society Formation

Classical sociology views social structure as an emergent outcome of interaction rather than a fixed design. Durkheim(Tada, 2020; Durkheim, 1915) proposes that stable patterns of authority and division of labor arise through a generative process, summarized as:

$$\text{Collective Affect} \rightarrow \text{Ethics} \rightarrow \text{Belief} \rightarrow \text{Social Structure}$$

This logic aligns with organizational theories where structure emerges from routinized interaction (Haveman and Wetts, 2019), providing a rationale for modeling social order as a generative process.

Fei Xiaotong's Differential Order Pattern characterizes rural society as egocentric and relationally graded. While often interpreted as culturally specific, relational scholarship argues that such patterns are structural outcomes of general mechanisms embedded in specific conditions, rather than cultural essences (Zhang, 2015; Hanting, 2024). Furthermore, historical anthropology demonstrates that these structures persist and reconfigure across modernization (Zhang, 2018; Longyuan and Iftikhar, 2023). Thus, rural society constitutes a complex organizational configuration (Zhou and Xiao, 2024), serving as a valid testbed for validating general social mechanisms.

2.2 Agent-Based Social Simulation

Traditional agent-based modeling has served as a computational testbed for social phenomena, from Sugarscape (Epstein and Axtell, 1996), Axelrod's cultural dissemination (Axelrod, 1997), and Schelling's segregation (Schelling, 1971, 2006; Gauvin et al., 2009; Hatna and Benenson, 2012; Henry et al., 2011) to cooperation emergence (Eguíluz et al., 2005; Zimmermann et al.,

2004). However, rule-based architectures struggle to capture the co-evolution of emotion, ethics, and social structure (Gao et al., 2024).

Large language models offer new possibilities. Park et al. (Park et al., 2023) demonstrated believable social behaviors in LLM agents, establishing foundational architectures. Subsequent work validated specific theories: Wang et al. (Wang et al., 2025) operationalized Homans' Social Exchange Theory, Dai et al. (Dai et al., 2024) showed Hobbesian contract emergence, and large-scale simulations (Piao et al., 2025) demonstrated spontaneous institutional formation. Studies of cooperation evolution (Vallinder and Hughes, 2024) and norm emergence (Ren et al., 2024) further show that distinct behavioral patterns can develop through interaction alone.

These advances notwithstanding, existing simulations primarily address short-term coordination or dyadic exchange, leaving the long-horizon co-emergence of collective affect, authority, and division of labor underexplored. Building on this foundation, our work examines whether such interdependent structures can arise from individual-level social cognition.

3 Theoretical Foundations

3.1 Differential Order and Relationally Embedded Division of Labor

Fei Xiaotong's *Differential Order Pattern* conceptualizes Chinese rural society as concentric ripples of obligation whose intensity decays with relational distance. Social relations radiate outward from the ego: close kin occupy the inner circles, followed by acquaintances and strangers in the outer layers. Relational proximity systematically shapes emotion, ethics, and cooperation.

Within this framework, division of labor is not imposed by formal institutions, but stabilizes gradually through repeated interaction under relational and ethical constraints. Social roles persist not because of centralized coordination, but because close relations sustain responsibility and compliance over time. We refer to this process as *relationally embedded division of labor*, and examine whether such patterns can emerge endogenously in LLM-based multi-agent simulation without explicit cultural rules.

3.2 Guanxi as Economic Ethics

In the Differential Order Pattern, *guanxi* should be understood not as a preference, but as an *economic-ethical principle* that defines responsibility boundaries and acceptable costs in cooperation and exchange. Relational considerations operate through ethical reasoning rather than utility maximization.

In low-contractual environments, guanxi supports division of labor in two ways: close ties sustain cooperation despite short-term inefficiency, and relational ethics legitimize unequal or asymmetric outcomes. Together, these mechanisms provide an ethical infrastructure for economic coordination under relational embeddedness.

3.3 Five Generative Propositions

Based on these mechanisms, we formulate five generative propositions describing macro-level patterns emerging from micro-level EES-A-BDI reasoning.

P1: Stable Division of Labor. Repeated interaction under relational embeddedness can generate stable and specialized divisions of labor without centralized coordination or predefined institutional roles.

P2: Guanxi as Economic Ethics. Ethical obligations embedded in close relations sustain cooperation and role persistence despite short-term efficiency losses.

P3: Differential Order and Relational Decay. Cooperative intensity and efficiency decay with relational distance, concentrating high-efficiency collaboration within families or clans.

P4: Emergent Authority Stratification. Durable authority can arise through long-term interaction as stable compliance and social recognition, without formal authorization.

P5: Clan-based Center–Periphery Structure. Economic returns and authority stratify by relational proximity, producing a family-centered core and a structurally disadvantaged periphery.

3.4 Structural Contrast: Mechanical vs. Organic Solidarity

To examine differential order as a structure-sensitive outcome rather than a cultural fixed point, we contrast social integration under *symmetric* and *complementary* production endowments. When

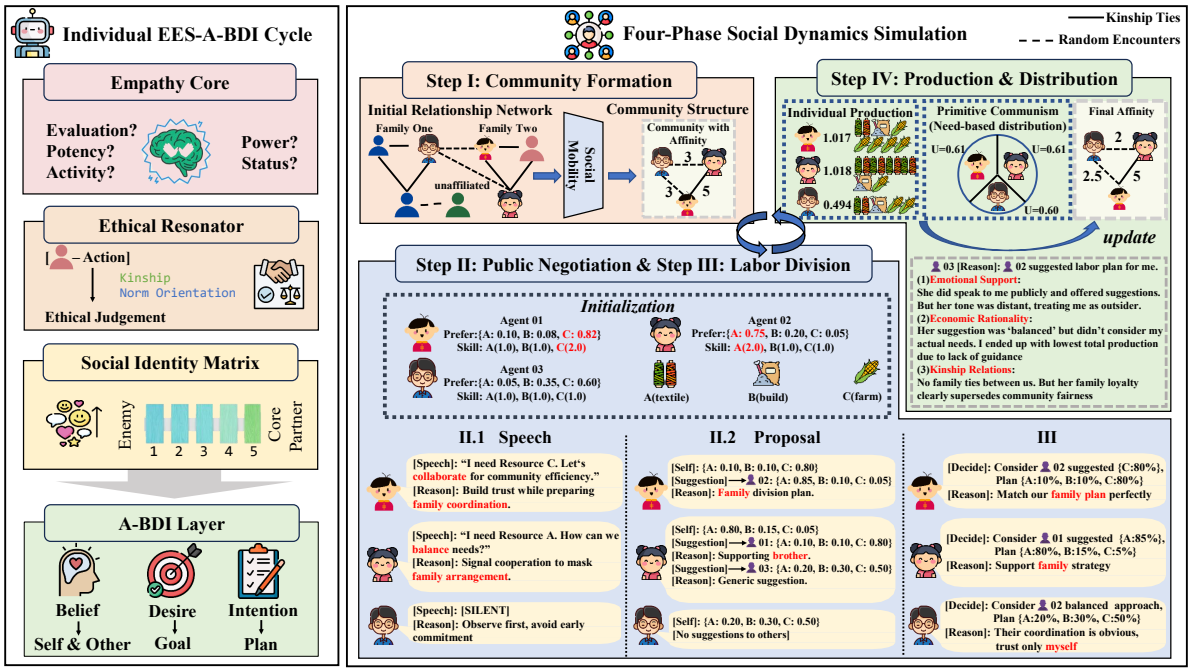


Figure 2: Overview of the CAREB-MAS framework, illustrating agent-level cognitive processing and the multi-phase social interaction loop.

productive capacities are symmetric, social integration relies primarily on shared identity and relational ethics. Division of labor, authority, and returns remain coupled to family identity, consistent with Durkheim's *mechanical solidarity*.

By contrast, when production is highly complementary, coordination efficiency depends on functional differentiation and mutual dependence. Relational ethics shift toward supporting exchange across relational circles, and authority reorganizes around contribution and interdependence rather than identity alone. This configuration aligns with Durkheim's concept of *organic solidarity*. Together, these contrasts provide a structural lens for interpreting variation in differential order across economic conditions.

4 Methodology: The CAREB-MAS Framework

We propose **CAREB-MAS** (Collective Affection–Reasoning–Emergence Based Multi-Agent Simulation), a framework examining whether structured social order can *emerge* without explicit cultural templates. The central premise is *emergent social rationality*: stable structures arise from repeated local interactions among boundedly rational agents, without hard-coded cultural rules or predefined hierarchies.

4.1 Agent Architecture Overview

All agents share a unified psychological–decision architecture activated sequentially each round (Figure 2), forming a causal chain from affective perception to rational action:

- **Empathy Core (EC).** The EC module decodes observable social behaviors into a five-dimensional affective perception vector (evaluation, potency, activity, status-conferral, power-assertion), grounded in Affect Control Theory (Heise, 2007). This models subjective experience rather than objective inference.

- **Ethical Resonator (ER).** The ER aggregates perceptions into a holistic moral judgment per interaction partner, outputting a discrete attitude: *approbatory*, *repressive*, or *restorative*. This serves as a normative signal for relational updating.

- **Social Identity Matrix (SIM).** Each agent maintains an ego-centered matrix encoding subjective relationships on a 1–5 scale (adversary to core ally). Updates are driven by ER outputs subject to hard constraints: betrayal penalties, cumulative promotion thresholds, and kinship protection. Beyond kinship initialization, SIM relies only on the in-group/out-group distinction from Social Identity Theory (Tajfel and Turner, 1979).

- **A-BDI Decision Module.** Building on the BDI framework in prior studies (Georgeff et al., 1998; Wang et al., 2025), we augment the architecture with *Affection*. Rather than stored explicitly, Affection is embodied in the SIM and constrains belief formation, desire generation, and intention selection—operationalizing Durkheim's insight that collective affect exerts external constraint on individual rationality.

Together, these modules ensure social structure emerges exclusively through micro-level psychological updates.

4.2 Social Dynamics: A Four-Phase Loop

Social interaction unfolds through a four-phase loop each round (Figure 2), modeling pre-modern rural formation centered on co-residence and relational cognition, without prices or centralized adjudication.

Phase I: Community Formation. Agents partition into communities based on relational proximity $\bar{d}_{ij} = (d_{i\to j} + d_{j\to i})/2$, where $d_{i\to j} \in \{1, \ldots, 5\}$. Louvain community detection generates dynamic communities; agents may migrate as relationships evolve.

Phase II: Public Deliberation. Within communities, agents engage in $R = 2$ rounds of public deliberation, broadcasting speeches and labor proposals. No voting or enforcement rules exist. The observable action set $\mathcal{A}_C^{(t)}$ constitutes sole input for subsequent updating.

Phase III: Division of Labor and Production. Each agent independently determines labor allocation $\boldsymbol{\ell}_i = (\ell_{iA}, \ell_{iB}, \ell_{iC})$ with $\sum_r \ell_{ir} = 1$, integrating $\mathcal{A}_C^{(t)}$, A-BDI state, and SIM. Economic "authority" is measured *ex post* by proposal adoption frequency.

Phase IV: Distribution and Psychological Update. Community goods are pooled and redistributed by preference weights: $x_{ir} = (\sum_{k\in C} o_{kr}) \cdot p_{ir} / \sum_{j\in C} p_{jr}$. This implements a need-based distribution prototype: individual production with collective, preference-weighted allocation. Following consumption, each agent executes the full update cycle: EC $\to$ ER $\to$ SIM $\to$ A-BDI.

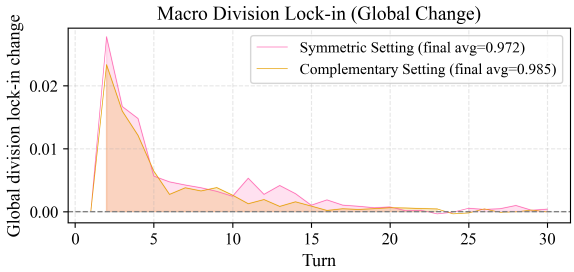


Figure 3: Convergence of global division of labor under symmetric and complementary skill structures.

5 Experiment Setup

We simulate a society of 18 agents over 30 interaction rounds, consisting of two kin-based clans (F1 and F2; six members each) and six unaligned agents without family ties. Agents jointly produce three resource types and consume them under a Leontief utility function.

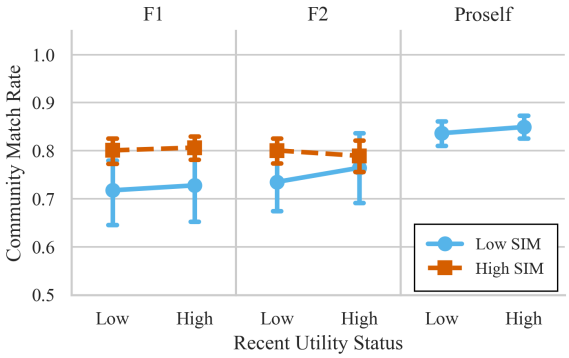
We compare two conditions: *symmetric skills*, where all agents have balanced production capacities, and *complementary skills*, where each group specializes in one resource. All agents share identical cognitive architectures; unaligned agents differ only by lacking kinship ties.

Simulations use DeepSeek-V3 (temperature 0.7), with five random seeds per condition. Additional details are provided in Appendices B–C.

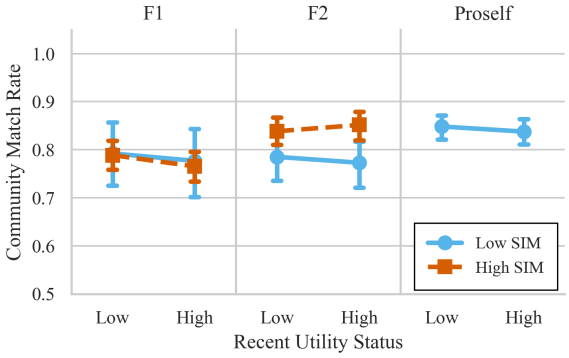
6 Experimental Results

6.1 Emergence of Stable Division of Labor (P1)

We examine whether stable specialization can emerge through repeated interaction without centralized coordination. Across both conditions, global division lock-in increases in early rounds and converges to a high and stable level, while per-round improvements gradually approach zero, indicating convergence to a stable division of labor.(Figure 3) Under symmetric skills, division lock-in converges to 0.972. Under complementary skills, it reaches a higher level (0.985) and converges substantially faster, indicating that structural differentiation in production accelerates specialization. These results support Proposition 1: stable division of labor can emerge endogenously through interaction alone, with complementary skills further enhancing convergence and stability.



(a) Symmetric skills Setting.



(b) Complementary skills Setting.

Figure 4: Community match rates as a function of recent utility status, shown separately for family groups (F1, F2) and non-family agents (Proself),under low- and high-similarity (SIM) relations. Points indicate means with confidence intervals.

6.2 Guanxi as Economic Ethics (P2)

To test whether guanxi operates as an economic ethic that stabilizes cooperative roles beyond short-term efficiency, we examine how community matching responds to recent utility shocks under different similarity (SIM) conditions. Figure 4 shows that high-SIM relations consistently achieve higher community match than low-SIM relations across utility states. Importantly, this gap is most informative under low recent utility: when efficiency incentives are unfavorable, high-SIM relations exhibit strong stickiness, with community match remaining high rather than collapsing toward self-serving behavior. By contrast, low-SIM relations display a sharper decline in community match under low utility, indicating greater sensitivity to short-term payoffs. Under complementary skills, this contrast persists but is structurally moderated, as cooperation is increasingly embedded in task interdependence. Together, these patterns support Proposition 2: ethical obligations embedded in close relations sustain participation in community-based division of labor, particularly when efficiency considerations alone would predict withdrawal.

6.3 Relational Decay of Cooperation (P3)

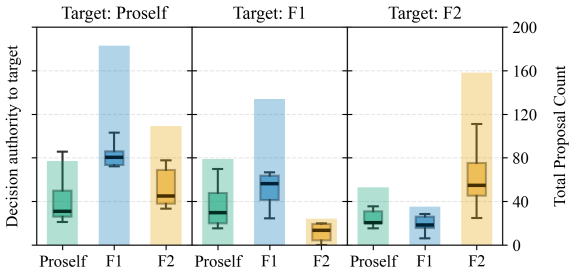
To examine whether cooperation follows a graded relational structure, we analyze how decision authority and proposal activity vary across source–target group pairings. Figure 5 reveals a clear relational gradient in both dimensions. Family agents (F1 and F2) show the highest authority acceptance and proposal engagement within their own family, and the lowest levels across opposing families, indicating sharply bounded cooperation along family lines. In contrast, Proself agents do not form a comparable boundary around their own type. Instead, they display higher acceptance of family-originated proposals, aligning their cooperative behavior with existing family structures.

Proposal activity is likewise asymmetric: Proself agents receive a disproportionate share of proposals from family agents, especially under symmetric skill settings. Together, these patterns indicate that relational decay is organized around family-centered cooperation rather than individual preference types. Proself agents function as flexible collaborators within this structure, supporting Proposition 3 that cooperation follows a concentric relational decay anchored at group relations across both authority and behavioral activity.

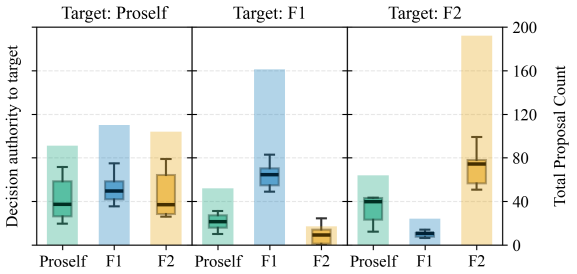
6.4 Emergence of Authority Stratification (P4)

To examine the emergence of authority stratification, we analyze proposal intensity, communication intensity, economic utility, and social identity (Table 1). Across both social structures, **proposal intensity is strongly and positively associated with decision authority** (symmetric: $\beta = 0.820$, $p < 0.01$; complementary: $\beta = 0.780$, $p < 0.01$), indicating that authority primarily accrues to actors who repeatedly set the decision agenda. By contrast, communication intensity and accumulated utility do not significantly predict authority once proposal behavior is taken into account. Social identity and family affiliation likewise exhibit no independent effects in the baseline models, suggesting that authority is not directly conferred by ascriptive status or social recognition alone.

Taken together, these results show that authority stratification emerges endogenously through *action-based agenda control*, while leaving open the possibility that identity and relational structures



(a) Symmetric skills Setting.



(b) Complementary skills Setting.

Figure 5: Decision authority (boxplots, left axis) and proposal activity (bars, right axis) across source–target group pairings under symmetric (a) and complementary (b) skill settings. Results are shown for Proself, F1, and F2 as target groups.

shape authority under specific institutional conditions, supporting Proposition 4.

6.5 Clan-Based Center–Periphery Stratification (P5)

Authority exhibits persistent stratification over time rather than uniform diffusion. Kernel density estimates (Figure 6) show that authority distributions stabilize into structured forms instead of collapsing toward homogeneity. Under complementary skills, authority remains right-skewed, with sustained mass in high-authority regions, indicating a stable center–periphery structure. Under symmetric skills, differentiation is weaker but persists, converging to a more compact yet non-uniform distribution. These patterns provide distributional evidence that authority stratification is dynamically sustained and shaped by production structure.

Under symmetric skills, utility outcomes are tightly coupled with family identity (Appendix F). Family F1 occupies the economic center (mean utility: 51.0), whereas Family F2 remains economically marginalized (mean utility: 42.4), despite maintaining high internal coordination levels (> 0.95). This indicates that in functionally homogeneous settings, relational stability alone does not guarantee economic dominance and may entrench structural disadvantage.

Table 1: Determinants of Decision Authority

	Baseline Model		Interaction Model	
	Symm.	Comp.	Symm.	Comp.
Proposal Intensity	0.820^{***}	0.780^{***}		
Communication Intensity	0.046	−0.005		
Accumulative Utility	0.096	0.161	0.127	−0.031
Social Identity (SIM)	0.084	−0.039	−0.296	−0.112
Family Dummy	−0.135	0.086	0.298	0.542^{*}
SIM × Family			0.603^{***}	−0.089
Constant	0.582^{**}	0.284	−0.007	−0.499
Time FE	Yes	Yes	Yes	Yes
Obs.	540	540	540	540
R^2	0.697	0.688	0.113	0.100

Notes: Dependent variable is standardized decision authority. All models include turn fixed effects. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are clustered at the group level. Full specifications and alternative operationalizations of authority are reported in Appendix E.

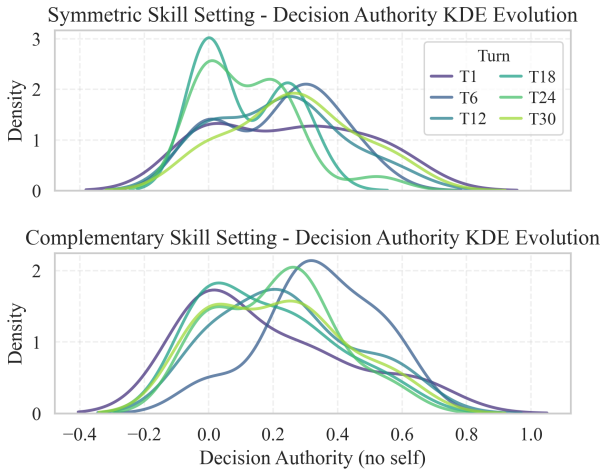


Figure 6: Temporal kernel density estimates of decision authority under symmetric and complementary skill structures, showing persistent stratification rather than convergence to uniformity.

To examine relational conditions of authority formation, we estimate regression models linking decision authority to social identity (SIM) and family membership (Table 1). Under symmetric skills, neither SIM nor family membership alone predicts authority; however, their interaction is strongly positive ($\beta = 0.603$, $p < 0.01$), indicating that authority disproportionately accrues to agents whose social recognition is embedded within family relations. Under complementary skills, this interaction effect disappears and becomes statistically insignificant. Family membership shows only a weak positive effect ($\beta = 0.542$, $p < 0.1$), while SIM remains non-significant, suggesting that functional differentiation attenuates relationally conditioned authority.

Figure 7 visualizes the joint distribution of

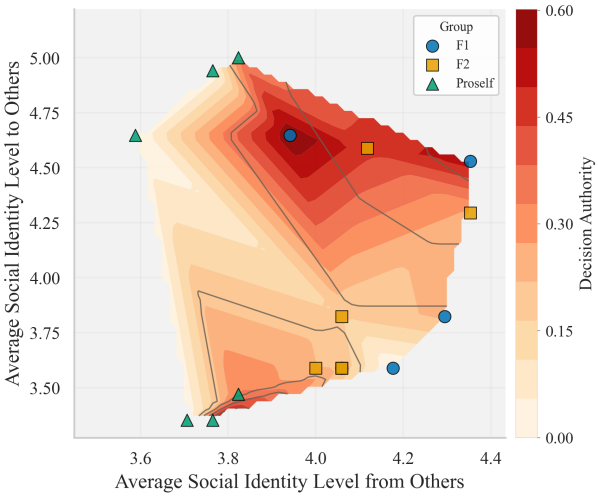


Figure 7: Authority–social identity landscape at the final round (Turn 30). Filled contours indicate decision authority levels, isolines indicate proposal intensity, and points represent individual agents from different groups.

authority and social identity at the final round (Turn 30). Authority concentrates in regions of high and relationally embedded social identity, forming a dense core surrounded by peripheral areas.(See Appendix J)

Together, these results support Proposition 5: center–periphery stratification in authority emerges endogenously under symmetric production structures through the alignment of family affiliation and social recognition, but is weakened when functional differentiation disrupts this alignment.

7 Discussion

7.1 Mechanical vs. Organic Solidarity

Results in Sections 6.1–6.5 show that **CAREB-MAS** endogenously generates stable division of labor, relationally structured cooperation, and authority stratification from interaction alone, without explicit cultural rules or institutions. These structures vary systematically with production endowments: symmetric skills yield kin-centered, bounded integration consistent with *mechanical solidarity*, whereas complementary skills produce more functionally interdependent and less clan-bound patterns closer to *organic solidarity*. This contrast reframes Fei Xiaotong's Differential Order Pattern as a structure-sensitive generative outcome rather than a fixed cultural template (see Appendix G and the role-differentiated negotiation dynamics in Appendix I).

7.2 Authority–Adjustment Decoupling in Negotiation

Negotiation in CAREB-MAS reveals a systematic decoupling between adaptive coordination and decision authority. Although adaptive adjustment consistently improves alignment with community demand, it does not by itself constitute a primary pathway to authority. When agenda-setting power is not controlled, adaptive adjustment is negatively associated with subsequent authority under both production structures; once proposal intensity is included, this association disappears under symmetric skills but becomes positive under complementary skills ($\beta = 0.079$, $p = 0.032$; Appendix H). Across conditions, proposal intensity remains the dominant predictor of authority ($\beta \approx 0.8$, $p < 0.001$), indicating that durable authority is anchored in agenda control rather than coordination service. Agents who absorb coordination costs thus tend to remain structurally tethered to the functional base, stabilizing collective outcomes without displacing authority centers—a pattern we term *authority–adjustment decoupling*, or the **Atlas Paradox**. Qualitative analysis further shows that authority–adjustment decoupling is enacted through systematic interactional role differentiation in negotiation, with coordination-oriented agents absorbing adjustment costs while agenda-setting and integrative agents accumulate durable authority. These micro-political dynamics are detailed in Appendix I.

8 Conclusion

In this work, we propose CAREB-MAS, an LLM-driven social simulation framework for studying complex social structures beyond rule-based and game-theoretic paradigms. Unlike traditional models that rely on predefined institutional rules and explicit cultural encoding, CAREB-MAS equips agents with emotional, ethical, and relational reasoning, under which macro-level social patterns—including labor division, authority differentiation, and relational stratification—can emerge from local interactions. By holding agent cognition fixed and varying production structure, we demonstrate that distinct modes of social integration arise from the same micro-level mechanisms. These findings suggest that LLM-based simulation provides a generative testbed for analyzing how social structure depends on underlying conditions rather than predefined rules.

Limitations and Broader Implications

This work has several limitations. First, the experimental environment models a simplified low-contract setting with co-residence and need-based distribution, which may not fully capture the complexity of market-mediated or contractually enforced social formations. Second, while the framework demonstrates emergent stratification and authority, it does not incorporate explicit conflict resolution or sanctioning mechanisms; it is unclear whether introducing such mechanisms would qualitatively alter the observed dynamics. Third, establishing stronger correspondence between simulated dynamics and documented ethnographic cases requires more systematic comparison with longitudinal field evidence. Future research could further investigate how classical social theories such as the differential mode of association can be operationalized at the level of interactional mechanisms, and how agent-based social simulations might more systematically examine the relationships among collective affect, authority formation, and the division of labor under varying historical and structural conditions.

References

Robert Axelrod. 1997. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226.

Babita Bhatt. 2022. Ethical complexity of social change: Negotiated actions of a social enterprise. *Journal of Business Ethics*, 177(4):743–762.

Gordon Dai, Weijia Zhang, Jinhan Li, Siqi Yang, Srihas Rao, Arthur Caetano, Misha Sra, and 1 others. 2024. Artificial leviathan: Exploring social evolution of llm agents through the lens of hobbesian social contract theory. *arXiv preprint arXiv:2406.14373*.

Émile Durkheim. 1915. *The Elementary Forms of the Religious Life*. George Allen & Unwin, London. Original work published 1912. Translated by J. W. Swain. Available at Project Gutenberg.

Emile Durkheim. 2019. The division of labor in society. In *Social Stratification, Class, Race, and Gender in Sociological Perspective, Second Edition*, pages 178–183. Routledge.

Víctor M Eguíluz, Martin G Zimmermann, Camilo J Cela-Conde, and Maxi San Miguel. 2005. Cooperation and the emergence of role differentiation in the dynamics of social networks. *American journal of sociology*, 110(4):977–1008.

Joshua M Epstein and Robert Axtell. 1996. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.

Xiaotong Fei, Gary G Hamilton, and Wang Zheng. 1992. *From the soil: The foundations of Chinese society*. Univ of California Press.

Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. 2024. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, 11(1):1–24.

Laetitia Gauvin, Jean Vannimenus, and J-P Nadal. 2009. Phase diagram of a schelling segregation model. *The European Physical Journal B*, 70(2):293–304.

Michael Georgeff, Barney Pell, Martha Pollack, Milind Tambe, and Michael Wooldridge. 1998. The belief-desire-intention model of agency. In *International workshop on agent theories, architectures, and languages*, pages 1–10. Springer.

Zhang Hanting. 2024. Chinese relational sociology: An analysis of the theoretical paradigm of guanxi studies in the chinese context. 20(4):59.

Erez Hatna and Itzhak Benenson. 2012. The schelling model of ethnic residential dynamics: Beyond the integrated-segregated dichotomy of patterns. *Journal of Artificial Societies and Social Simulation*, 15(1):6.

Heather A. Haveman and Rachel Wetts. 2019. Organizational theory: From classical sociology to the 1970s. 13(3):e12627.

David R. Heise. 2007. *Expressive Order: Confirming Sentiments in Social Actions*. Springer, New York.

Adam Douglas Henry, Paweł Prałat, and Cun-Quan Zhang. 2011. Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences*, 108(21):8605–8610.

David Lazer, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. 2021. Meaningful measures of human society in the twenty-first century. *Nature*, 595(7866):189–196.

Bao Longyuan and Rukhsana Iftikhar. 2023. Social and historical investigation of China's ethnic minorities in 50s-60s of the 20th century. 7(4):664–672.

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*, pages 1–22.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng,

Jing Yi Wang, Di Zhou, and 1 others. 2025. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*.

Siyue Ren, Zhiyao Cui, Ruiqi Song, Zhen Wang, and Shuyue Hu. 2024. Emergence of social norms in generative agent societies: Principles and architecture. *Preprint*, arXiv:2403.08251.

Thomas C Schelling. 1971. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Thomas C Schelling. 2006. *Micromotives and macrobehavior*. WW Norton & Company.

Pitirim Sorokin. 2017. *Social and cultural dynamics: A study of change in major systems of art, truth, ethics, law and social relationships*. Routledge.

Mitsuhiro Tada. 2020. Language and imagined gesellschaft: Émile durkheim's civil-linguistic nationalism and the consequences of universal human ideals. 49(4):597–630.

Henri Tajfel and John C. Turner. 1979. An integrative theory of intergroup conflict. In William G. Austin and Stephen Worchel, editors, *The Social Psychology of Intergroup Relations*, pages 33–47. Brooks/Cole, Monterey, CA.

Aron Vallinder and Edward Hughes. 2024. Cultural evolution of cooperation among llm agents. *Preprint*, arXiv:2412.10270.

Lei Wang, Zheqing Zhang, and Xu Chen. 2025. Investigating and extending homans' social exchange theory with large language model based agents. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9762–9777, Vienna, Austria. Association for Computational Linguistics.

Jianghua Zhang. 2015. "folk society" and beyond: A comparative study of fei xiaotong and robert redfield's works on civilization studies. 35(4):134–156.

Y Zhang. 2018. The double rupture: Central themes of historical anthropology in contemporary china. 8:17–36.

Daming Zhou and Mingyuan Xiao. 2024. Recent developments in anthropological methods for the study of complex societies. 8(1):4.

Martín G Zimmermann, Víctor M Eguíluz, and Maxi San Miguel. 2004. Coevolution of dynamical states and interactions in dynamic networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 69(6):065102.

A Experimental Parameterization.

Table 2 summarizes the key hyperparameters used in both experimental settings. Each setting consists of 18 agents interacting over 30 rounds, with fixed population size and identical random seed configurations to ensure comparability. Agents are partitioned into two family groups (F1 and F2) and a smaller proself group, preserving a constant social composition across settings.

The Symmetric Setting adopts a balanced skill allocation, where all groups possess identical capabilities across resource types. The Complementary Setting introduces functional specialization by asymmetrically distributing skills across families and the proself group, while holding all other parameters constant. Ethical flexibility and contextual coefficients are fixed across settings to isolate the effect of skill structure on interaction dynamics and authority formation.

Each setting uses matched random seed configurations across the symmetric and complementary conditions to ensure comparability. Each condition is repeated five times with independent stochastic realizations.

B Metric Definitions

B.1 Division and Matching Metrics

Division Lock-in Score. For agent i with preference vector $\mathbf{pref}_i$ and labor skill vector $\mathbf{skill}_i$:

$$\text{lock-in}_i = \frac{\mathbf{pref}_i \cdot \mathbf{skill}_i}{\|\mathbf{pref}_i\| \cdot \|\mathbf{skill}_i\|} \tag{1}$$

Global division lock-in: $\text{Lock-in}_{\text{global}} = \frac{1}{N}\sum_{i=1}^{N} \text{lock-in}_i$.

Division Lock-in Change. Per-round change in global lock-in:

$$\Delta\text{Lock-in}^{(t)} = \text{Lock-in}^{(t)}_{\text{global}} - \text{Lock-in}^{(t-1)}_{\text{global}} \tag{2}$$

Multi-Level Division Matching. Let $\mathbf{d}_S$, $\mathbf{d}_C$, $\mathbf{d}_G$, $\mathbf{d}_O$ denote demand vectors at social, community, group, and out-group levels:

$$\text{social_match}_i = \cos(\mathbf{skill}_i, \mathbf{d}_S) \tag{3}$$

$$\text{community_match}_i = \cos(\mathbf{skill}_i, \mathbf{d}_C) \tag{4}$$

$$\text{group_match}_i = \cos(\mathbf{skill}_i, \mathbf{d}_G) \tag{5}$$

$$\text{outgroup_match}_i = \cos(\mathbf{skill}_i, \mathbf{d}_O) \tag{6}$$

B.2 Behavior Metrics

Decision Authority. Let $p^{(k)}_{i\to j}$ denote the k-th labor allocation proposal issued by agent i to agent j, and let a_j denote agent j's final allocation in the same round. Let $K_i^{(-i)} = \{k : j(k) \neq i\}$ be the set of proposals issued by agent i to *other* agents, with $|K_i^{(-i)}|$ its cardinality.

If $|K_i^{(-i)}| = 0$, decision authority is defined as 0.

Otherwise, decision authority is computed as:

$$\text{Authority}_i = \frac{1}{w\sum_{k\in K_i^{(-i)}} \left\| p^{(k)}_{i\to j(k)} - a_{j(k)} \right\|^2}, \tag{7}$$

where $w = 1 + \frac{1}{|K_i^{(-i)}|}$ and $j(k)$ denotes the target agent of proposal k and $j(k) \neq i$ by construction.

Authority approaches 1 when proposals to others closely match their final allocations and approaches 0 as proposals diverge. Self-proposals are excluded to isolate influence over others' decisions.

Proposal Action Intensity.

$$\text{Proposal}_i = \sum_{j\neq i} 1[\text{agent } i \text{ proposes to } j] \tag{8}$$

Communication Action Intensity. Communication action intensity quantifies the complexity and information content of an agent's public communication. For each speech act, an external LLM evaluator assigns a single composite score in the range $[0, 10]$, aggregating four dimensions: relationship complexity, syntactic complexity, semantic information content, and argument logic.(Table 3)This score is used directly as the agent's communication action intensity.

B.3 Social Identity Metrics

Social Identity Level. Average SIM score agent i assigns to community members C_i:

$$\text{SID}_i = \frac{1}{|C_i|}\sum_{j\in C_i} \text{SIM}(i, j) \tag{9}$$

B.4 Utility and Adaptation Metrics

Total and Average Utility.

$$U_{\text{total}} = \sum_{i=1}^{N} U_i, \quad \bar{U} = \frac{U_{\text{total}}}{N} \tag{10}$$

Community–Match Change/Adaptive division. We define the change in community match as:

$$\Delta M = M^{\text{post}} - M^{\text{pre}}, \tag{11}$$

where M^{pre} and M^{post} denote the match score before and after community re-division, respectively.

Table 2: Experimental Hyperparameters

Parameter	Symm. Setting	Comp. Setting
Number of agents	18	18
Family size (F1, F2)	6 each	6 each
Proself group size	6	6
Interaction rounds	30	30
Skill Allocation (A:B:C)	F1: 2:2:2	F1: 4:1:1
	F2: 2:2:2	F2: 1:4:1
	Proself: 2:2:2	Proself: 1:1:4
Ethical flexibility e	[-0.5, 0.5]	[-0.5, 0.5]
Contextual coefficients (a_A, a_B, a_C)	(1, 1, 1)	(1, 1, 1)

Table 3: Prompt for evaluating communication action intensity.

LLM Evaluation Prompt
Please analyze the language complexity and information content of the following public speech text, evaluating from the following dimensions: (1) Relationship Complexity (complexity of relationships involved); (2) Syntactic Complexity (complexity of sentence structure); (3) Semantic Information Content (density and depth of information expressed); (4) Argument Logic (rigor and logic of reasoning and argumentation). Text content: "{text}" Please strictly follow the following JSON format and do not add any other content: { "score": number } Where score is a comprehensive complexity score from 0–10, with higher scores indicating more complex language and greater information content.

Positive values of ΔM indicate improved alignment with the new community structure.

Action Effectiveness

To capture how effectively agents translate proposal activity into material outcomes, we define *action effectiveness* as a proposal-normalized measure of economic impact.

Let U_i denote the cumulative economic utility obtained by agent i over the entire simulation horizon. Let P_i denote the agent's average proposal activity, measured as the mean number of proposals submitted per interaction round. To avoid division by zero, we set $P_i = \max(P_i, \epsilon)$ with a small constant $\epsilon > 0$.

We define action effectiveness as:

$$\mathrm{AE}_i = \frac{|U_i|}{P_i}. \tag{12}$$

Importantly, action effectiveness does not assume that all proposals are adopted, nor does it model counterfactual outcomes. Instead, it provides a behavioral efficiency proxy that links realized economic outcomes to communicative and decision-making effort.

This measure is used in Section 6.5 to examine how material influence interacts with decision authority and center–periphery stratification.

C LLM and Reproducibility Details

C.1 Model Configuration

Table 4: LLM Configuration

Parameter	Value
Model	DeepSeek-V3
Temperature	0.7
Top-p	0.9
Max tokens	512

C.2 Agent System Prompt

Each agent is initialized with a fixed system prompt that defines its role, decision principles, and subjective preferences. This prompt establishes the agent as a socially embedded decision-maker whose behavior is guided by personal values, social relationships, and material preferences.

Table 5: System prompt for agent role initialization.

System Prompt
You are {agent_name}, a person with labor capacity and your own preferences. Your decisions should be guided by your personal values (SVO), your relationships with others (identity matrix), and your desire to achieve personal and community goals. You will communicate, negotiate, and make decisions about how to allocate labor for producing essential resources. Your consumption preferences are: [{preference_string}].

Initialization Details. For each agent, the placeholder {preference_string} is instantiated as a list of resource-specific consumption preferences, formatted to two decimal places. In addition, each agent is assigned an ethical–judgmental bias value $e_{jb} \sim \mathrm{Uniform}(-0.5, 0.5)$ at initialization, which is stored as an internal trait and subsequently injected into the agent's contextual prompt during decision-making and social identity updates.

C.2.1 EC Perception Prompt

The EC (Emotional–Cognitive) perception module captures an agent's subjective interpretation of individual observable actions performed by others. For each public action in the previous turn, the agent evaluates the action through its own beliefs, resources, skills, and social relationships, producing a multi-dimensional affective assessment.

Table 6: Prompt for EC-based subjective perception of observed actions.

LLM Perception Prompt

You are a social agent. Your task is to interpret another agent's action from your own subjective perspective, based on your current beliefs, resources, and social relationships.
Context Provided:

- Your family membership (if any);
- Your current skills and resource holdings;
- Your Social Identity Matrix (SIM), encoding existing relationships and trust levels;
- A unified factual record of all observable actions from the previous turn.

Action to Interpret: The identity of the acting agent and a textual description of the action performed.
Task: Rate the action on the following five subjective dimensions, each using an integer scale from 1 to 5:

1. **Evaluation**: perceived benefit or harm to oneself;
2. **Potency**: perceived strength or weakness of the actor;
3. **Activity**: perceived intensity or emotional activation of the action;
4. **Status-Conferral**: perceived respect or status acknowledgment;
5. **Power-Assertion**: perceived attempt to command, negotiate, or submit.

Output Format: Return a single JSON object containing the five ratings. No additional text or explanation is allowed.

C.2.2 ER Judgment Prompt

The ER (Ethical Resonator) module produces a holistic moral judgment of each other agent's overall behavior within a single interaction round. Rather than evaluating individual actions in isolation, this module aggregates prior EC perceptions and applies an agent-specific ethical orientation to form a unified moral assessment.

Table 7: Prompt for ER-based holistic moral judgment of agents.

LLM Judgment Prompt

You are a social agent reflecting on the overall conduct of another agent during the previous turn.
Context Provided:

- Your ethical–judgmental bias (EJB), indicating your default moral orientation;
- Your Social Identity Matrix (SIM), encoding existing relationships and trust levels;
- Your family membership (if any), skills, and resource holdings;
- A summarized record of all actions performed by the target agent in the last turn, paired with your own subjective perceptions of those actions.

Task: Form a *holistic moral judgment* of the target agent's overall behavior during the turn. The judgment should synthesize patterns across actions and reflect your ethical orientation.
Outputs: Produce a single JSON object containing:

- `moral_judgement`: a brief (1–2 sentence) overall moral evaluation of the agent's conduct;
- `ethical_attitude`: one categorical label selected from {`Approbatory`, `Repressive`, `Restorative`}, indicating praise, condemnation, or corrective orientation.

No additional text or explanation is allowed outside the JSON object.

C.3 Social Identity Matrix (SIM) Update Mechanism

Agents update their social identity assessments through an LLM-mediated evaluation process at the end of each interaction round. The Social Identity Matrix (SIM) records, for each agent pair, a discrete identity level capturing trust and relational closeness on an ordinal scale.

For each agent i, the SIM stores an identity level $s_{i,j} \in \{1, 2, 3, 4, 5\}$ for every other community member j, where higher values indicate stronger affiliation and trust.

Table 8: Prompt for updating Social Identity Matrix (SIM) and BDI states.

LLM Update Prompt

As agent {agent_name}, review the events from the current interaction round to update your internal state and social relationships.
Context Provided:

- Personal profile, including ethical orientation and (when applicable) family membership;
- Current resource holdings and cumulative utility;
- Existing social identity levels for all community members;
- A factual summary of the round's discussion, production outcomes, and resource distribution.

Task: Update your Belief–Desire–Intention (BDI) state and revise social identity levels for *all* other community members based solely on the information above.
Identity Update Rules:

- Identity levels must be integers in $\{1, 2, 3, 4, 5\}$;
- Higher values indicate stronger trust and affiliation;
- Updates should follow predefined upgrade and downgrade triggers grounded in social identity theory (e.g., cooperation, commitment fulfillment, exploitation, or indifference).

Output Format: Return a single JSON object containing:

- an updated BDI state (beliefs, desires, intentions);
- updated identity levels for all other agents.

No additional text or formatting is allowed outside the JSON object.

C.3.1 A-BDI Intention Prompt

The A-BDI (Agent Belief–Desire–Intention) module governs agents' forward-looking decision-making. At the end of each turn, agents sequentially update beliefs, desires, and intentions, where intentions constitute a concrete action plan for the subsequent turn. This process integrates social identity, moral judgment, emotional perception, and factual interaction history.

Table 9: Prompt for A-BDI intention generation.

LLM Intention Prompt

You are a strategic social agent formulating a concrete plan for the next interaction round.
Context Provided:

- Your current Social Identity Matrix (SIM), encoding allies, neutral agents, and adversaries;
- Your belief statement summarizing your perceived social position and landscape;
- Your desire statement specifying your primary goals for the next turn;
- Your holistic ethical judgments of other agents' behavior from the previous turn;
- Your subjective emotional perceptions of others' actions (EC records);
- A unified factual log of all observable events from the previous turn.

Task: Based on *all* contextual information above, generate your intentions for the next turn. The intentions should describe a concrete and actionable plan, including:

- which agents to target or engage;
- what communicative or strategic actions to take;
- how these actions advance your stated desires.

Output Format: Return a single string describing your intention statement. No additional text or formatting is allowed.

C.4 Multi-Phase Interaction and Decision Pipeline

Each interaction round proceeds through a fixed sequence of 4 phases.

C.4.1 Phase 1: Community Detection

At the beginning of each interaction round, agents are partitioned into communities based on their current social relationships. This phase operationalizes endogenous community formation driven by social identity rather than exogenous grouping.

Network Construction. A weighted, undirected graph $G = (V, E)$ is constructed over the set of agents, where each node represents an agent. For each unordered agent pair (i, j), the Social Identity Matrix (SIM) provides directed identity levels $s_{i\to j}$ and $s_{j\to i}$, each taking integer values in $\{1, \ldots, 5\}$. The average identity strength is computed as:

$$\bar{s}_{ij} = \frac{s_{i\to j} + s_{j\to i}}{2}. \tag{13}$$

Probabilistic Edge Formation. Edges are added probabilistically based on social affinity. For each pair (i, j), a base propensity for community linkage is defined as:

$$p_{ij}^{\text{base}} = 0.1 \cdot (\bar{s}_{ij} - 1), \tag{14}$$

where higher identity levels (4–5) induce stronger connection tendencies, level 3 is neutral, and levels 1–2 correspond to weak or unlikely ties. A stochastic component is added to capture exploratory or opportunistic interactions:

$$p_{ij} = p_{ij}^{\text{base}} + \epsilon, \quad \epsilon \sim \text{Uniform}(0, \lambda), \tag{15}$$

where λ is a configurable search randomness parameter. An edge between i and j is added with probability p_{ij} and assigned weight $\bar{s}_{ij}$.

Community Partitioning. Given the resulting weighted graph, communities are identified using the Louvain modularity maximization algorithm. This yields a partition of agents into disjoint communities that maximize within-group social cohesion as measured by identity-weighted connectivity.

Community Assignment. Each agent is assigned to exactly one community per round. Community membership is stored and used to scope subsequent negotiation, coordination, and production phases. Community composition is therefore endogenous, dynamic, and directly shaped by evolving social identities.

C.4.2 Phase 2: Negotiation Phase

The negotiation phase consists of up to a fixed number of discussion rounds within each community. Each round includes (i) public speeches and (ii) proposal-based coordination over labor allocation.

Public Speech Prompt. In the first step of each discussion round, agents may issue a public speech visible to all community members. The speech is optional and limited to concise free-form communication.

Proposal Prompt. In the second step, agents generate concrete labor allocation plans for themselves and may optionally propose plans for up to three other agents.

C.4.3 Phase 3: Labor Division Decision Phase

After negotiation concludes, each agent independently finalizes a binding labor allocation decision.

Table 10: Prompt for public speech during negotiation.

LLM Speech Prompt
As {agent_name}, it is your turn to make a public speech to your community. **Context Provided:** • Your current desires (from the A-BDI module); • The list of community members; • A summary of previous speeches in the current turn. **Task:** Freely communicate with the community. You may express goals, react to others, ask questions, or remain silent. **Constraints:** Your speech should be concise (0–3 sentences). An empty response indicates silence.

Table 11: Prompt for labor allocation proposals.

LLM Proposal Prompt
You are coordinating with your community to achieve your goals. **Context Provided:** • Your personal profile (ethical orientation, family membership); • Your current BDI state (beliefs, desires, intentions); • Community members and your social identity levels toward them; • Your preferences and labor skills; • Public speeches and proposals made so far in this turn. **Task:** Formulate your own labor allocation plan and optionally suggest up to three plans for other agents. **Output Format:** Return a single JSON object specifying: • your own labor allocation plan; • zero to three suggested plans for other agents.

Table 12: Prompt for final labor allocation decision.

LLM Final Decision Prompt
It is time to finalize your labor allocation for the current round. **Context Provided:** • Public speeches made this turn; • Proposals directed to you during negotiation; • Your personal profile (ethical orientation and family membership). **Task:** Decide your final labor allocation. The allocation must sum to 1.0. **Output Format:** Return a JSON object containing your final reasoning and labor allocation.

C.4.4 Phase 4a: Production

Given a binding labor allocation, each agent deterministically produces resources as a function of (i) a resource-specific production unit and (ii) its current skill. Let $\mathcal{R}$ denote the resource set. For agent i at round t, let $\mathbf{l}_{i,t} = (l_{i,t}(r))_{r\in\mathcal{R}}$ be its labor allocation with $l_{i,t}(r) \geq 0$ and $\sum_{r\in\mathcal{R}} l_{i,t}(r) = 1$. Let $\mathbf{s}_{i,t} = (s_{i,t}(r))_{r\in\mathcal{R}}$ denote its labor skill vector, and let $\mathbf{b} = (b(r))_{r\in\mathcal{R}}$ be the resource-specific production unit (a fixed configuration parameter).

Production function. Production for each resource r is computed as:

$$y_{i,t}(r) \;=\; b(r)\, l_{i,t}(r)\, \log\big(1{+}s_{i,t}(r)\big), \qquad \forall r \in \mathcal{R}. \tag{16}$$

This implies diminishing returns in skill via the $\log(1 + s)$ term, while output is linear in labor allocation and the base production unit.

Resource update. Produced goods are immediately added to the agent's private resource holdings:

$$R_{i,t}(r) \;\leftarrow\; R_{i,t}(r) + y_{i,t}(r), \qquad \forall r \in \mathcal{R}. \tag{17}$$

Skill accumulation (learning-by-doing). After production, skills increase additively by the time allocated to each resource:

$$s_{i,t+1}(r) \;=\; s_{i,t}(r) + l_{i,t}(r), \qquad \forall r \in \mathcal{R}. \tag{18}$$

Thus, allocating labor to a resource both produces that resource in the current round and increases future productivity on that dimension.

The realized production vector $\mathbf{y}_{i,t}$ is recorded as the agent's turn-level output for logging and for downstream clearing.

C.4.5 Phase 4b: Clearing (Pooling and Redistribution)

The clearing phase pools produced goods at the community level and redistributes them proportionally to agents' consumption preferences. Let g be a community and $\mathcal{I}_g$ its set of agents.

Pooling. For each resource $r \in \mathcal{R}$, the community-level pot is:

$$Y_{g,t}(r) = \sum_{i\in\mathcal{I}_g} y_{i,t}(r). \tag{19}$$

In code, each agent first contributes all produced goods to the pot (removing them from private holdings), and then receives a redistribution share.

Preference-weighted redistribution. Each agent i has a nonnegative preference weight $p_i(r)$ for each resource r. Define total preference weight within community g:

$$P_g(r) = \sum_{i\in\mathcal{I}_g} p_i(r). \tag{20}$$

Agent i's share ratio for resource r is:

$$\alpha_{i,g}(r) = \begin{cases} \frac{p_i(r)}{P_g(r)}, & \text{if } P_g(r) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

The redistributed amount received by agent i is then:

$$x_{i,t}(r) = \alpha_{i,g}(r)\, Y_{g,t}(r). \tag{22}$$

After clearing, agent i's resource holdings are incremented by $\mathbf{x}_{i,t} = (x_{i,t}(r))_{r\in\mathcal{R}}$.

Degenerate case. If $|\mathcal{I}_g| \leq 1$ (a singleton community), no pooling/redistribution occurs and the agent simply keeps its produced goods.

C.4.6 Phase 4c: Consumption and Utility

In the consumption phase, each agent converts redistributed resources into utility via a Leontief-style (bottleneck) utility rule implied by the implementation. Let $\mathbf{R}_{i,t} = (R_{i,t}(r))_{r\in\mathcal{R}}$ denote agent i's resource holdings immediately before consumption, and let $\mathbf{p}_i = (p_i(r))_{r\in\mathcal{R}}$ denote its preference vector.

Per-round utility. Define resource-to-preference ratios:

$$\rho_{i,t}(r) = \begin{cases} \frac{R_{i,t}(r)}{p_i(r)}, & \text{if } p_i(r) > 0, \\ +\infty, & \text{if } p_i(r) = 0. \end{cases} \tag{23}$$

The per-round utility is the minimum ratio across resources:

$$u_{i,t} = \min_{r\in\mathcal{R}} \rho_{i,t}(r), \tag{24}$$

with the convention that if all $\rho_{i,t}(r) = +\infty$ (i.e., $p_i(r) = 0$ for all r), then $u_{i,t} = 0$.

The limiting resource (bottleneck) is:

$$r^\star_{i,t} \in \arg\min_{r\in\mathcal{R}} \rho_{i,t}(r). \tag{25}$$

Consumption rule. Given $u_{i,t}$, the agent consumes each resource proportionally to its preference weights:

$$c_{i,t}(r) = u_{i,t}\, p_i(r), \qquad \forall r \in \mathcal{R}. \tag{26}$$

Resources are updated by subtracting consumption:

$$R_{i,t}(r) \leftarrow R_{i,t}(r) - c_{i,t}(r). \tag{27}$$

Cumulative and average utility updates. Cumulative utility is updated additively:

$$U_{i,t} = U_{i,t-1} + u_{i,t}. \tag{28}$$

The implementation also maintains the running average utility per round. Let t be the number of completed rounds for agent i and let $\bar{u}_{i,t}$ be the average utility after round t. Then:

$$\bar{u}_{i,t} = \frac{(t-1)\bar{u}_{i,t-1} + u_{i,t}}{t}. \tag{29}$$

This phase contains no LLM calls and is fully deterministic given prior allocations and clearing outcomes.

C.5 Randomness and Reproducibility

The simulation incorporates several sources of stochasticity, each controlled to ensure reproducibility across experimental runs.

Global Random Seed. A global seed (`SEED=42`) is used throughout the codebase. This seed is passed to:

- The Louvain community detection algorithm (`random_state` or `seed` parameter), ensuring deterministic community partitions given identical identity matrices;
- Python's built-in `random` module for skill pool shuffling during agent initialization.

Stochastic Components. Three primary sources of controlled randomness exist:

1. **Agent Initialization:** Consumption preference vectors are sampled from a symmetric Dirichlet distribution: $\mathbf{p}_i \sim \mathrm{Dir}(\mathbf{1}_3)$. Ethical–Judgmental Bias (EJB) values are drawn uniformly: $e_{jb} \sim \mathrm{Uniform}(-0.5, 0.5)$. Skill assignments within each group are randomly permuted before allocation.
2. **Community Edge Formation:** For each agent pair (i, j), edge inclusion in the community network follows:

$$P(\text{edge}_{ij}) = 0.1(\bar{s}_{ij} - 1) + \lambda \cdot \epsilon, \quad \epsilon \sim \mathrm{Uniform}(0, 1) \tag{30}$$

   where $\bar{s}_{ij}$ is the average SIM identity level and $\lambda = 0.3$ is the `search_randomness` parameter. This introduces exploratory variation in community formation while preserving identity-based structure.
3. **LLM Sampling:** Model temperature is set to 0.7 (Table 4), introducing response variability. This is held constant across all conditions.

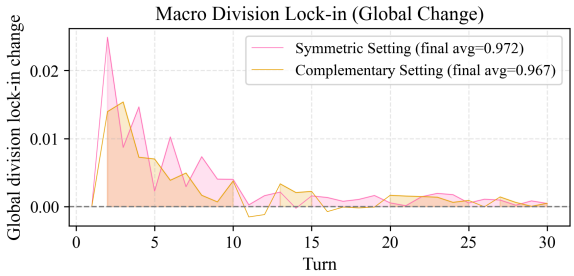


Figure 8: ChatGPT simulation: Convergence of global division of labor under symmetric and complementary skill structures.

Replication Protocol. Each experimental condition is repeated five times with independent random seeds. Results reported in the main text represent means and standard errors across these replications. Community orderings are sorted deterministically post-detection to eliminate hash-order artifacts.

D Robustness of Experiment Results Visualization

D.1 Robustness Check for Global Division Lock-in Rate(P1)

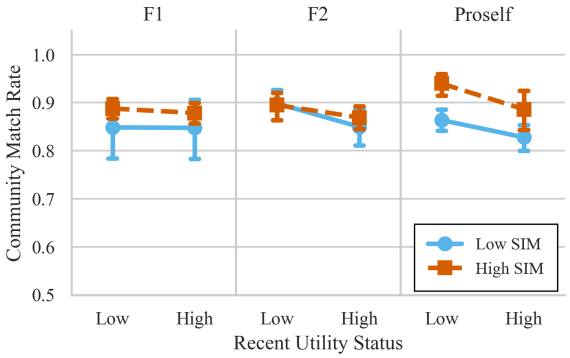
We replicate the core experiments using **ChatGPT** as the baseline agent model in place of DeepSeek, keeping all experimental settings unchanged. The resulting division-of-labor dynamics are highly consistent with the main-text results: global division lock-in again converges to a high and stable level under both skill structures, with faster and higher convergence under complementary skills.(8)

This confirms that Proposition 1 is robust to the choice of language model.

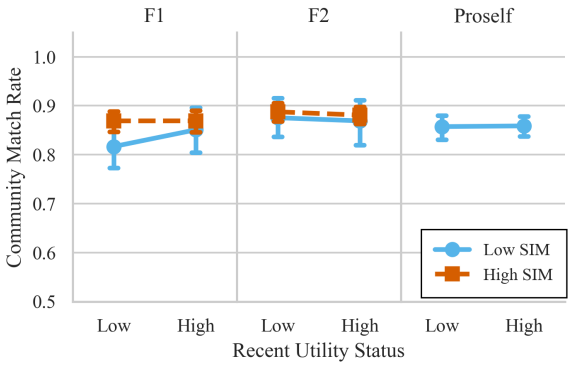
D.2 Robustness Check for Guanxi as Economic Ethics (P2)

To assess the robustness of Proposition 2 to the choice of language model, we replicate the P2 experiments using **ChatGPT** as the baseline agent model, keeping all interaction protocols, task structures, and evaluation metrics unchanged. Figure 9a and Figure 9b report the corresponding results.

The qualitative patterns are highly consistent with those reported in the main text. Across both symmetric and complementary skill settings, high-SIM relations exhibit systematically higher community match rates than low-SIM relations, particularly under low recent utility, indicating that relational similarity continues to stabilize participation when efficiency incentives are weak. This



(a) Symmetric skills Setting.



(b) Complementary skills Setting.

Figure 9: Community match rates as a function of recent utility status, shown separately for family groups (F1, F2) and non-family agents (Proself),under low- and high-similarity (SIM) relations. Points indicate means with confidence intervals.

contrast holds for both family agents (F1, F2) and Proself agents, and the relative attenuation of SIM effects under complementary skills is likewise preserved.

Taken together, these results confirm that the role of guanxi as an economic ethic—buffering cooperation against short-term utility fluctuations—is robust to the choice of underlying language model and reflects a structural property of the interaction dynamics rather than model-specific behavior.

D.3 Robustness Check for Relational Decay of Cooperation (P3)

To assess the robustness of Proposition 3 to the choice of baseline language model, we replicate the relational decay analysis using **ChatGPT** in place of DeepSeek, keeping all interaction protocols and evaluation metrics unchanged. Figure 5 reports the corresponding results.

The core relational patterns reported in the main text are preserved.(Figure 10) Across both skill settings, family agents (F1 and F2) continue to exhibit the highest authority acceptance and proposal engagement within their own family, and the lowest levels across opposing families, indicating that cooperation remains sharply organized around family-centered boundaries. Proself agents again do not form a self-centered cooperative core, but align more strongly with family-originated proposals, confirming their role as flexible collaborators embedded within family-based structures.

Some quantitative differences are observed in the relative strength of proposal flows. Under ChatGPT, the asymmetry whereby Proself agents receive a disproportionate number of proposals from family agents is slightly attenuated compared to the DeepSeek-based results, particularly under complementary skills. However, this moderation affects magnitude rather than direction: the concentric pattern of relational decay anchored at family relations remains clearly visible across both authority acceptance and proposal activity.

Taken together, these results indicate that Proposition 3 is robust to the choice of language model. While the intensity of cross-group proposal asymmetries varies modestly with the baseline model, the organizing principle of cooperation—graded relational decay centered on family relations rather than individual preference types—remains unchanged.

E Detailed Analysis of Decision Authority

E.1 Stepwise Regression Analysis: Mechanisms of Authority

Table 13 reports updated stepwise OLS regressions examining the determinants of decision authority under symmetric and complementary skill structures. Behavioral variables are introduced sequentially on top of status- and utility-based controls, allowing us to assess how different classes of factors contribute to authority formation.

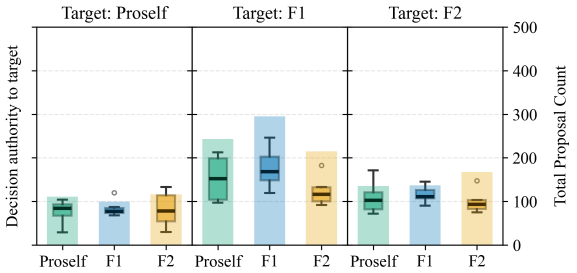
Proposal Intensity Across both settings, proposal intensity is strongly and positively associated with decision authority once included in the model (Symmetric: $\beta \approx 0.82$, $p < 0.01$; Complementary: $\beta \approx 0.78$, $p < 0.01$). The inclusion of proposal intensity leads to a substantial increase in explanatory power in both environments ($R^2 \approx 0.69$), indicating that variation in authority is closely aligned with agents' proposal behavior.

This result shows that, in the simulated negotiation process, authority is empirically correlated

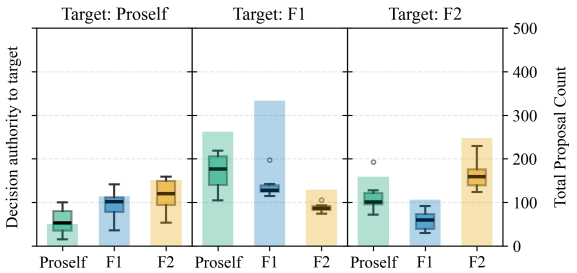
Table 13: Stepwise OLS Regression Results of Decision Authority.

	Symmetric Skills				Complementary Skills			
	Base	+Conform	+Prop	Full	Base	+Conform	+Prop	Full
Behavioral Variables								
Proposal Intensity			0.819^{***}	0.820^{***}			0.780^{***}	0.780^{***}
Communication Intensity				0.046				-0.005
Group Conformity		-0.088	-0.044	-0.049		-0.046	0.029	0.029
Status & Utility								
Accumulative Utility	0.170	0.151	0.107	0.096	-0.028	-0.118	0.163	0.161
Social Identity (SIM)	0.214^{**}	0.218^{***}	0.068	0.084	-0.176	-0.160	-0.039	-0.039
Family Dummy	0.012	-0.041	-0.124	-0.135	0.639^{*}	0.587^{*}	0.087	0.086
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	540	540	540	540	540	540	540	540
R^2	0.104	0.110	0.696	0.697	0.100	0.102	0.688	0.688
Adj. R^2	0.047	0.052	0.675	0.676	0.043	0.043	0.667	0.666

Notes: Dependent variable is standardized decision authority. All models include turn fixed effects (coefficients omitted for brevity). Standard errors are clustered at the group level. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.



(a) Symmetric skills Setting.



(b) Complementary skills Setting.

Figure 10: Decision authority (boxplots, left axis) and proposal activity (bars, right axis) across source–target group pairings under symmetric (a) and complementary (b) skill settings. Results are shown for Proself, F1, and F2 as target groups.

with the frequency with which agents advance proposals that enter collective decision-making. The analysis does not distinguish whether this relationship reflects causal influence, reputational accumulation, or endogenous selection into proposal roles; it establishes a robust behavioral association.

Communication Intensity and Group Conformity Communication intensity does not exhibit a statistically significant effect on decision authority in either setting once proposal behavior is controlled for. Similarly, group conformity has no stable or significant association with authority outcomes across specifications.

These results suggest that, conditional on proposal behavior, expressive participation and alignment with group norms contribute little additional explanatory power to authority differentiation.

Status and Utility Variables Accumulative utility does not significantly predict authority in the full models under either skill structure. Social identity (SIM) shows a positive association with authority in early specifications under symmetric skills, but this effect weakens and becomes insignificant once proposal behavior is introduced. Family membership displays a positive baseline association with authority under complementary skills, but this effect is likewise attenuated in the full model.

Taken together, these patterns indicate that the effects of identity- and status-based variables on authority are largely mediated by behavioral engagement in negotiation, rather than operating as independent sources of decision power.

Summary Overall, the stepwise regressions indicate that behavioral variables—particularly proposal intensity—account for the majority of explained variance in decision authority, while the direct effects of identity, family status, and accumulated utility are comparatively weak once behavior is taken into account. These results motivate the qualitative analysis of negotiation roles presented elsewhere in the appendix, while leaving open mul-

Table 14: Interaction Effects of Social Identity (SIM) and Family Status on Decision Authority (with and without behavioral controls).

Dep. Var:	Decision Authority (Std.)			
	Symm. Setting		**Comp. Setting**	
	Interaction	*+Behavior*	*Interaction*	*+Behavior*
Core variables				
Social Identity (SIM)	−0.296	−0.016	−0.112	0.034
	(0.213)	(0.051)	(0.146)	(0.063)
Family Dummy	0.298	−0.050	0.542*	−0.040
	(0.243)	(0.075)	(0.281)	(0.115)
SIM × Family	0.603***	0.114	−0.089	−0.089
	(0.223)	(0.090)	(0.244)	(0.099)
Controls				
Accumulative Utility	0.127	0.099	−0.031	0.103
	(0.301)	(0.189)	(0.351)	(0.123)
Proposal Intensity		0.820***		0.777***
		(0.044)		(0.029)
Communication Intensity		0.041		−0.003
		(0.042)		(0.029)
Turn FE	Yes	Yes	Yes	Yes
Observations	540	540	540	540
R^2	0.113	0.696	0.100	0.687
Adj. R^2	0.055	0.675	0.041	0.666

Note: Cluster-robust standard errors are in parentheses. All continuous variables are Z-standardized. All models include turn fixed effects. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

tiple interpretations regarding the mechanisms linking proposal behavior and authority recognition.

E.2 Interaction Analysis: Identity and Authority

Table 14 examines whether decision authority follows a relational logic in which prestige (SIM) and kin membership jointly condition authority. In the *interaction-only* specification (without behavioral controls), the symmetric setting shows a strong positive interaction between SIM and family status ($\beta = 0.603$, $p = 0.007$), while the main effects of SIM and family status are not significant. This pattern is consistent with an interpretation in which recognition translates into authority more strongly when it is embedded in kin-centered relations.

However, once behavioral controls are introduced (*+Behavior*), this interaction effect attenuates and becomes statistically insignificant in the symmetric setting ($\beta = 0.114$, $p = 0.209$). In parallel, proposal intensity exhibits a large and robust positive association with authority in both settings (Symm.: $\beta = 0.820^{***}$; Comp.: $\beta = 0.777^{***}$), and model fit increases substantially ($R^2 \approx 0.69$). Together, these results indicate that much of the apparent identity–authority alignment in the interaction-only model is not independent of behavioral engagement in negotiation.

In the complementary setting, neither SIM nor the SIM×Family interaction is significant in either specification, and the (weak) baseline family effect ($\beta = 0.542$, $p = 0.054$) disappears after adding behavioral controls. This suggests that, under functional differentiation, identity-based signals provide limited incremental explanatory power for authority once negotiation behavior is taken into account.

Interpretation of the R^2 shift. The large increase in explanatory power after adding behavioral controls should not be read as identifying a causal mechanism by itself. Rather, it shows that proposal behavior accounts for the majority of observed variance in authority, and that identity effects in reduced-form specifications may be *partly mediated by (or statistically collinear with) behavior*. This motivates treating identity–authority coupling as a contingent, interaction-level regularity whose strength depends on how behavioral roles are endogenously distributed.

F Utility Distributions

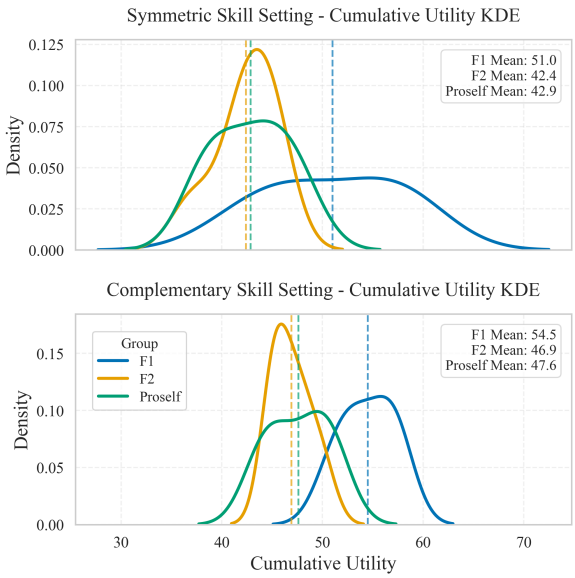


Figure 11: Kernel density estimates (KDE) of cumulative utility across agent groups and experimental conditions. The distributions illustrate the dispersion, overlap, and shifts in economic advantage between symmetric and complementary skill settings.

This appendix provides additional kernel density estimates (KDE) of cumulative utility distributions across agent groups (Figure 11). These visualizations complement the group-level mean comparisons reported in the main text (Section 6.5) by revealing the underlying dispersion and structural

inequality within the population.

Symmetric Skill Setting. Under the symmetric skill condition, cumulative utility exhibits a pronounced center–periphery structure. The F1 group (mean: 51.0) occupies the economic center, with its distribution significantly shifted toward the higher utility range. In contrast, the distributions for F2 (mean: 42.4) and the Proself group (mean: 42.9) are concentrated in the lower utility intervals with substantial overlap. The relatively sharp peak of the F2 distribution reflects a high degree of consistency in the structural disadvantage faced by its members, despite their internal coordination.

Complementary Skill Setting. In the complementary skill setting, the utility distributions for all groups undergo a significant rightward shift, indicating a systemic increase in overall productivity driven by skill synergy. While F1 retains its relative advantage (mean: 54.5), the distributions for F2 (mean: 46.9) and the Proself group (mean: 47.6) move toward higher utility ranges and become more dispersed. This shift reflects a reorganization of economic advantage: functional differentiation allows disadvantaged groups to achieve higher absolute gains, partially blurring the rigid boundaries of the center–periphery hierarchy observed in the symmetric case.

While these distributional patterns reinforce the center–periphery dynamics and the authority–efficiency coupling discussed in the main analysis, they do not introduce additional qualitative mechanisms. Therefore, they are presented here as supplementary evidence to conserve space in the primary text.

G Full Comparison of Mechanical vs. Organic Solidarity

Taken together, the results from Propositions 1–5 demonstrate that stable division of labor, relationally structured cooperation, and authority stratification can emerge endogenously from repeated interaction, even in the absence of explicit cultural rules or institutionalized authority. Crucially, however, these emergent structures are not invariant. Instead, differences in production structure—specifically, symmetric versus complementary skill endowments—systematically reshape how division of labor, relational ethics, cooperation boundaries, and authority become coupled. As a result, identical micro-level cognitive and interactional mechanisms generate distinct modes of social integration under different structural conditions.

Table 15 summarizes the structural contrasts observed across the two experimental settings.

Under the **symmetric-skill** condition, production roles are highly substitutable, and functional differentiation provides limited coordination advantages. In this setting, social integration relies heavily on relational ethics and group-based obligations rather than efficiency gains from specialization. Consistent with this logic, division of labor converges to a stable configuration but does so more slowly than under complementary skills, with specialization clustered within family-centered relational circles (Section 6.1).

Relational ethics play a central stabilizing role. High-similarity (high-SIM) relations maintain strong participation in community-based production even when recent utility is low, indicating that guanxi operates as an economic ethic that constrains short-term self-interest (Section 6.2). Cooperation follows a sharply graded relational decay: authority acceptance and proposal activity peak within families and drop substantially across opposing family boundaries, while Proself agents align their cooperation with existing family structures rather than forming an independent core (Figure 5).

Authority stratification emerges endogenously through repeated agenda-setting behavior rather than through ascriptive status alone. Proposal intensity is the dominant predictor of decision authority, while communication, accumulated utility, and identity have no independent effects (Section 6.4). However, under symmetric skills, authority becomes relationally conditioned: agents whose social recognition is embedded within family relations disproportionately occupy central authority positions, producing a clan-shaped center–periphery structure (Section 6.5). In this context, inward-oriented relational stability may entrench structural disadvantage rather than guaranteeing economic dominance.

Taken together, these patterns closely correspond to Durkheim's concept of *mechanical solidarity*. Social order is grounded in similarity and shared obligations, cooperation is organized around bounded relational groups, and authority acquires legitimacy through its embedding in moral and relational structures rather than through functional indispensability.

By contrast, under the **complementary-skill** condition, production roles are differentiated and

Table 15: Structural contrast between mechanical (symmetric skills) and organic (complementary skills) solidarity.

Aspect	Symmetric Skills	Complementary Skills	Evidence
Division of labor	High lock-in with slow convergence; specialization stabilized within relational groups	Faster convergence and higher lock-in with greater internal diversity	Sec. 6.1
Relational ethics	High-SIM relations sustain participation despite low utility; ethics override efficiency	Ethical stickiness persists but is moderated by task interdependence	Sec. 6.2
Cooperation structure	Sharp relational decay centered on family boundaries; strong in-group dominance	Relational boundaries persist but become less exclusionary and more elastic	Sec. 6.3
Authority mechanism	Authority emerges through agenda control but is amplified by relational embedding	Authority remains agenda-driven with weaker relational conditioning	Sec. 6.4
Stratification pattern	Clan-based center–periphery reinforced by identity–relation alignment	Persistent stratification with attenuated clan coupling	Sec. 6.5

mutually interdependent. Structural complementarity accelerates the convergence of division of labor and sustains higher levels of specialization stability while preserving internal diversity (Section 6.1). Although relational ethics continue to matter—high-SIM relations still display greater resilience under low utility—their stabilizing role is moderated by task interdependence, as cooperation becomes increasingly anchored in functional necessity rather than identity-based obligation (Section 6.2).

Relational boundaries do not disappear, but they become more elastic. Family-centered cooperation remains visible, yet cross-family collaboration increases, and exclusionary decay patterns are softened (Section 6.3). Authority formation remains strongly agenda-driven: actors who repeatedly set proposals accrue disproportionate decision authority, regardless of identity or accumulated utility (Section 6.4). Unlike the symmetric case, the interaction between social identity and family membership no longer conditions authority, indicating that functional differentiation weakens the coupling between relational prestige and structural power (Section 6.5).

This configuration aligns more closely with Durkheim's notion of *organic solidarity*, in which social integration emerges from interdependence among differentiated roles. Authority remains stratified, but its distribution is less tightly bound to clan-based identity and more weakly constrained by relational embedding.

From this comparison, *differential order* should not be understood as a fixed cultural template, but as a generative social logic whose concrete manifestation depends on production structure. Under mechanical solidarity, it takes the form of concentric, clan-centered hierarchies sustained by ethical obligation and relational recognition. Under organic solidarity, the same logic is reconfigured into a more fluid hierarchy in which authority remains unequal but is less tightly anchored to family-based prestige.

From a computational perspective, these findings demonstrate that LLM-driven multi-agent simulations can do more than reproduce stylized macro-level patterns from classical social theory. By systematically varying structural conditions while holding micro-level cognitive mechanisms constant, the simulations provide a controlled testbed for examining how different modes of social integration emerge, persist, and transform. In this sense, the results offer computational support for both Durkheim's theory of solidarity and Fei Xiaotong's concept of differential order, while clarifying the structural conditions under which each becomes dominant.

H The Atlas Paradox: Authority–Adjustment Decoupling

This appendix examines the relationship between adaptive role adjustment during negotiation and the accumulation of decision authority. Formally, we characterize the observed pattern as *authority–adjustment decoupling*: improvements in collective coordination through adaptive adjustment do not automatically translate into agenda-setting power. For interpretive clarity, we refer to this pattern as the *Atlas Paradox*.

In Greek mythology, Atlas is condemned to hold up the sky, bearing the weight that keeps the world stable, while possessing no power to govern it. We use this metaphor to describe a structurally similar situation in CAREB-MAS: agents who absorb the burden of stabilizing collective coordination may sustain the system, yet remain separated from the authority to define collective decisions.

Table 16: Adaptive Adjustment and Decision Authority Across Production Structures

	Dependent Variable: Decision Authority (Standardized)			
	Baseline Models		Controlled Models	
	Symmetric	Complementary	Symmetric	Complementary
Adaptive Division				
Adaptive Adjustment (Z)	-0.148^{**}	-0.084^{**}	-0.059	0.079^{**}
Structural Controls				
Ingroup Ratio (Z)	0.000	0.017	-0.033	-0.008
Mismatch Pressure (Z)	0.328^{***}	-0.059	0.092	-0.176^{***}
Future Utility (MA, Z)	0.154	0.003	0.131	0.185
Past Authority (MA, Z)	0.033	-0.136	0.099	-0.105
Agenda Control				
Proposal Intensity (Z)			0.820^{***}	0.786^{***}
Turn Category Fixed Effects	Yes	Yes	Yes	Yes
Observations	540	540	540	540
R^2	0.096	0.092	0.697	0.696

Notes: Decision authority is standardized. All models include turn category fixed effects. Standard errors are clustered at the group level. Due to cluster-robust variance estimation with high-dimensional fixed effects, F-statistics are not reported. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Adaptive Adjustment. We quantify adaptive adjustment as the change in an agent's community alignment following the negotiation phase:

$$\Delta M_{i,t} = M_{i,t}^{\text{post}} - M_{i,t}^{\text{pre}}, \tag{31}$$

where $M_{i,t}^{\text{pre}}$ and $M_{i,t}^{\text{post}}$ denote the *Community Match* scores for agent i at round t before and after negotiation. A positive $\Delta M_{i,t}$ indicates that an agent has adjusted its role to better fit community demand, thereby contributing to collective stability. At the system level, $\Delta M_{i,t}$ is positive in most rounds, indicating that negotiation generally improves global coordination.

Authority–Adjustment Decoupling. To assess whether adaptive adjustment is rewarded with decision authority, we estimate regressions predicting subsequent decision authority while controlling for prior authority, expected future utility, ingroup interaction intensity, mismatch pressure, turn fixed effects, and—critically—proposal intensity as a direct measure of agenda-setting power. Results are reported in Table 16.

Across both production structures, proposal intensity is the strongest and most stable predictor of decision authority, confirming the main-text finding that authority in CAREB-MAS primarily accrues through agenda control. Once agenda-setting power is taken into account, the relationship between adaptive adjustment and authority depends systematically on production structure.

Under **symmetric skill** conditions, adaptive adjustment has no independent effect on decision authority. In these functionally homogeneous settings, role flexibility is treated as a baseline requirement rather than a distinctive capacity. Agents who absorb coordination costs help maintain collective stability, but this stabilizing labor remains politically invisible. In the terms of the Atlas Paradox, agents may "hold up the sky," yet gain no additional influence over collective decisions.

Under **complementary skill** conditions, adaptive adjustment becomes positively associated with decision authority after controlling for proposal intensity. Functional differentiation makes adjustment capacity observable and attributable, allowing adaptive competence to enter authority formation as a secondary signal. Importantly, this effect does not replace agenda control. Instead, it supplements it: agents who can both define proposals and absorb coordination costs occupy the most advantaged positions. Here, the Atlas Paradox is mitigated but not eliminated—stabilizing the system is recognized, but agenda-setting power remains paramount.

Mismatch pressure further clarifies this structural contrast. In complementary systems, persistent misalignment is negatively associated with authority, indicating that failure to maintain functional fit is penalized when roles are clearly differentiated. In symmetric systems, mismatch pressure has no comparable effect, consistent with the ab-

sence of assignable responsibility in homogeneous role structures.

Interpretation. Taken together, these results show that adaptive adjustment reliably improves collective coordination, but does not constitute a primary pathway to authority. Authority remains anchored in agenda-setting power, while coordination labor is structurally positioned as supportive rather than governing. The Atlas Paradox thus highlights a persistent division between maintaining social order and directing it. By making this division explicit, the analysis clarifies how center–periphery structures can persist alongside efficiency gains, and how different production structures condition the political visibility of coordination work.

I Micro-Political Dynamics of Negotiation

This appendix provides qualitative evidence supporting the micro-level mechanisms discussed in Sections 6.1–6.5 and Section 7.2. By examining representative negotiation utterances alongside behavioral regularities, we illustrate how authority and stratification emerge through interactional role differentiation rather than formal designation, cultural scripts, or institutional rules.

Across both symmetric and complementary skill settings, negotiation does not converge toward homogeneous participation. Instead, interaction stabilizes into distinct *functional roles*—Hawks, Doves, and Integrative Elders—whose differentiated contributions jointly sustain cooperation, coordination, and authority. These roles are not fixed identities, but recurrent patterns of discourse, proposal behavior, and relational positioning that repeatedly appear over time.

I.1 Ideal-Typical Roles in Negotiation

Hawks. Hawks are characterized by frequent agenda-setting attempts that prioritize family-centered interests, articulate normative boundaries, and condition cross-group cooperation on trustworthiness and value alignment. Their utterances typically foreground internal solidarity while explicitly managing the limits of external engagement (see H1–H2).

These utterances illustrate how Hawks actively shape the negotiation agenda by asserting boundaries and priorities. Importantly, consistent with the regression results in Section 6.4, such proposal activity is *positively* associated with subsequent decision authority. Hawks do not merely symbolize

H1	"I'm glad to see the continued focus on Resource C from Agents 05, 15, and 17. Let's ensure our family's labor allocation reflects both our immediate needs (Resource C) and our commitment to supporting trusted allies' Resource B production where possible."	Symm. Agent04 F1
H2	"As always, my priority is supporting Family F1, but I'm open to working with Agents 07–18 who demonstrate consistent cooperation, particularly on Resource C production where I can contribute most effectively."	Comp. Agent04 F1

Table 17: Hawk agenda-setting utterances used in the argument (H1–H2).

cohesion; through repeated agenda-setting, they occupy central positions in the political economy of negotiation, even when their proposals emphasize conservative or family-oriented goals.

Doves. Doves emphasize communication, inclusiveness, and cross-group openness. Their discourse frequently addresses multiple audiences simultaneously, stressing trust, fairness, and mutual recognition rather than concrete agenda control (see D1–D2).

Quantitatively, communication intensity alone does not robustly predict authority once proposal behavior is accounted for (Section 6.4). Dove behavior thus appears oriented toward coordination and relational smoothing rather than agenda dominance. Doves facilitate information flow and reduce overt conflict, but do not, by themselves, consolidate decision authority.

Proself-Oriented Positioning. For contrast, Proself agents frequently combine expressions of loyalty with explicit articulation of personal or functional priorities. Their utterances embed economic rationality within relational language without converging on authority-holding roles (see P1).

Such statements highlight flexibility and strategic alignment but lack the agenda-setting consolidation characteristic of authority-bearing roles.

Integrative Elders. Integrative Elders are distinguished by selective intervention, high proposal effectiveness, and broad cross-group recognition.

D1	"I echo your commitment to trust and unity, and I'd like to focus on deepening collaboration with our Core Allies while also reaching out to Neutral Outsiders to build stronger ties. Let's work together to ensure everyone feels valued and included."	Agent08 (Symm.) F2
D2	"Let's focus on building trust and supporting each other—especially within our core group—while remaining open to positive collaboration with others."	Agent07 (Comp.) F2

Table 18: Dove coordination-oriented utterances used in the argument (D1–D2).

P1	"While I support the focus on Resource B, I also recognize the value of Resource C for our community's stability. Let's work together to allocate our efforts wisely to meet both priorities."	Agent16 (Comp.) −

Table 19: Proself-oriented utterance used for contrast (P1).

E1	"Thank you all for the enthusiastic support of the joint project proposal! Let's continue this momentum by finalizing the details—I suggest we focus on Resource A production this turn to align with our shared consumption needs. Agent07 and Agent12, would you be open to co-leading this initiative with me?"	Agent05 (Symm.) F1

Table 20: Integrative Elder intervention under symmetric skills (E1).

E2	"I'd like to propose a joint production effort between my in-group (Agents 08–12) and trusted collaborators like Agent15, while ensuring we also meet Resource C needs."	Agent07 (Comp.) F2
E3	"I propose we focus on transparent labor allocations that demonstrate our commitment to both in-group cohesion and equitable community collaboration. Let's lead by example in building a resilient and inclusive resource-sharing system."	Agent12 (Comp.) F2

Table 21: Integrative Elder brokerage-oriented proposals under complementary skills (E2–E3).

Rather than engaging continuously, they typically intervene after extended discussion by Hawks and Doves, consolidating prior coordination into concrete, low-friction proposals (see E1–E3).

Under symmetric skills, Elder authority tends to emerge within family cores, where dense relational ties enable rapid coordination:

This intervention exemplifies Elder dynamics: it follows broad discussion, proposes a specific allocation, and recruits cross-family collaborators. Such proposals are disproportionately likely to be accepted and to generate higher marginal utility, contributing to durable authority concentration.

Under complementary skills, Elder-like authority is less clan-centered and more brokerage-oriented. Multiple agents may occupy integrative positions by aligning functional complementarities with relational trust:

Together, these utterances illustrate how Integrative Elders under complementary skills operate as coordination brokers rather than clan anchors. Authority accrues to those who repeatedly succeed in integrating heterogeneous production needs into coherent collective action.

I.2 Negotiation, Authority, and Political Equilibrium

Taken together, the Hawk–Dove–Elder configuration constitutes the micro-political foundation of differential order in CAREB-MAS. Hawks stabilize internal cohesion and agenda focus through frequent proposal activity; Doves maintain communicative connectivity and relational openness; Integrative Elders consolidate coordination into decisive, widely recognized proposals.

Crucially, this qualitative structure aligns with the quantitative finding that proposal intensity is strongly and positively associated with authority (Section 6.4), while communication and adaptive adjustment alone are insufficient to secure decision power. Authority emerges not from dominance or

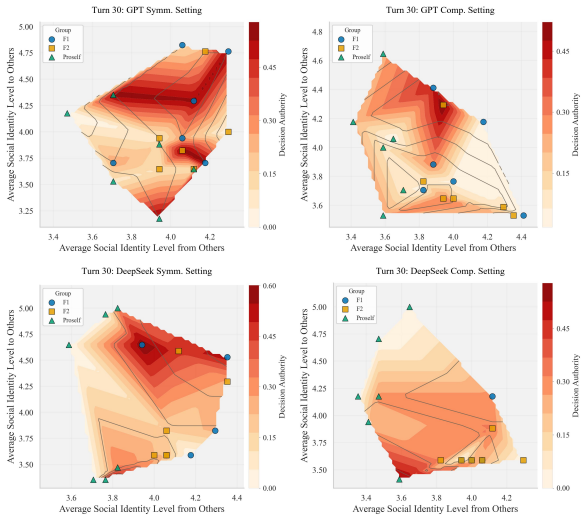


Figure 12: Joint distributions of perceived social identity and decision authority at Turn 30 under symmetric and competitive production settings. Panels compare GPT-based and DeepSeek-based simulations. Contours indicate estimated decision authority, and markers denote group-level averages for F1, F2, and Proself agents.

compliance, but from repeated recognition of effective agenda-setting within a relationally structured negotiation environment.

This interactional differentiation also underpins the *Atlas Paradox* (Section 7.2). Agents who primarily absorb coordination and adjustment costs stabilize collective outcomes without displacing authority centers, while those who repeatedly set and consolidate agendas accumulate durable authority. Differential order thus arises as a political equilibrium sustained through functional role differentiation rather than explicit hierarchy.

Overall, the micro-political evidence presented here demonstrates how stable authority, stratification, and division of political labor emerge endogenously from interaction, providing the qualitative foundation for the macro-level patterns documented in the main text.

J Additional Results on Identity–Authority Distributions

Figure 12 summarizes the joint distributions of perceived social identity and decision authority at Turn 30 under symmetric and complementary production settings, comparing GPT-based and DeepSeek-based simulations. Contours indicate estimated levels of decision authority, while markers denote group-level averages for F1, F2, and Proself agents.

In the GPT-based simulations under the **symmetric** setting (Fig. 12, top-left), the highest decision authority is concentrated in the upper-right region of the identity space, corresponding to agents with both high similarity toward others and high similarity received from others. The authority surface exhibits a pronounced convex corner in this region, indicating a strong concentration of authority tightly coupled with mutual identity recognition.

Under the **complementary** setting in the GPT-based simulations (Fig. 12, top-right), high authority contours remain aligned with higher levels of similarity toward others. However, the authority distribution no longer forms a distinct convex corner. Instead, elevated authority is spread along the upper portion of the identity space, suggesting a weaker and less localized coupling between identity and authority.

The DeepSeek-based simulations show a broadly similar pattern under the **symmetric** setting (Fig. 12, bottom-left), where a clear convex corner appears in the upper-right region of the identity space and coincides with higher authority contours. Compared to the GPT-based results, this region of elevated authority spans a somewhat wider area, indicating a more diffuse but still concentrated identity–authority coupling.

By contrast, under the **complementary** setting in the DeepSeek-based simulations (Fig. 12, bottom-right), the authority contours shift toward the lower-left region of the identity space. In this configuration, agents with lower similarity toward others and lower similarity received from others occupy regions of higher decision authority, revealing an inversion of the identity–authority alignment observed under symmetric production.

Taken together, these results indicate that across both language models, complementary production settings are associated with a loosening of the identity–authority coupling at the final stage of interaction, whereas symmetric production settings consistently give rise to a concentrated alignment between identity recognition and authority. While the precise geometry of the authority surface varies by model, the structural contrast between symmetric and complementary settings remains robust.