

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

Zhiyuan Ji¹, Xinyu Chen^{2†}, Ziqi Dai^{3†},
Shiyun Tang^{1*}, Chunyu Wei¹, Yueguo Chen¹

¹Renmin University of China, ²Beihang University, ³Minzu University of China
{zhiyuanji, jamietang, weichunyu, chen Yueguo}@ruc.edu.cn
zy2402531@buaa.edu.cn
daiziqi1296@muc.edu.cn

Abstract

Fei Xiaotong’s Differential Order Pattern characterizes rural society as egocentric and relationally graded, with cooperation attenuating over social distance. Although often treated as culturally specific, its mechanistic basis remains under-operationalized, and prior LLM-based simulations have mainly addressed short-term coordination rather than long-horizon social structure. We propose CAREB-MAS, a multi-agent framework grounded in Affect Control Theory, Social Identity Theory, and Durkheimian collective affect. Agents reason through an emotion–ethics–belief chain and maintain dynamically evolving egocentric identities, 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, relational decay of cooperation, emergent relational authority, and clan-based center–periphery stratification. These patterns shift with production structure from kin-centered integration toward greater functional interdependence. Extensive experiment results support interpreting Differential Order as a structure-sensitive emergent outcome of general social mechanisms, with LLM-based multi-agent simulation providing an interdisciplinary framework 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 (Sorokin, 2017; Lazer et al., 2021; Bhatt, 2022). Classical sociology has long addressed this question: Durkheim, for instance, emphasized the transition from me-

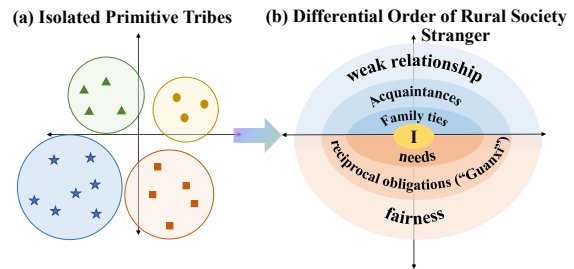


Figure 1: Schematic comparison of social organization: (a) discrete groups without relational gradation; (b) differential order, organized as ego-centered concentric layers of relational proximity.

chanical to organic solidarity driven by changing divisions of labor (Durkheim, 2019).

Beyond the Western canon, Fei Xiaotong’s Differential Order Pattern (Fei et al., 1992) describes rural society as ego-centered relational circles in which moral obligation, trust, and cooperation attenuate with social distance (Figure 1). Although often treated as specific to Chinese society, its constituent mechanisms—kinship organization, reciprocity beyond price exchange, and authority without centralized power—are well documented across comparative anthropology (Fortes and Evans-Pritchard, 1940; Sahlins, 1972; Polanyi, 1944). This raises the possibility that Differential Order reflects structure-sensitive outcomes of general social mechanisms rather than solely Chinese traditions.

Testing this possibility remains difficult. Ethnography offers contextual richness but limited experimental control; quantitative sociology often captures long-horizon change only indirectly; and conventional agent-based models rely on hand-coded rules that poorly represent the co-evolution of emotion, ethics, authority, and social structure (Eguíluz et al., 2005; Zimmermann et al., 2004). Recent LLM-agent studies have reproduced social contracts (Dai et al., 2024), cooperation norms (Vallinder and Hughes, 2024), and institu-

[†]Equal Contribution
^{*}Corresponding author.

tional emergence (Piao et al., 2025), but mostly in short-horizon settings. Whether LLM-based simulations can generate long-horizon social structure and probe the generality of non-Western sociological theory remains underexplored.

In this paper, we propose CAREB-MAS, a multi-agent simulation framework testing whether Differential Order can emerge from cross-culturally documented micro-mechanisms without encoding culture-specific institutional rules. Agents reason through an emotion–ethics–belief chain grounded in Affect Control Theory (Heise, 2007), Social Identity Theory (Tajfel and Turner, 1979), and Durkheimian collective affect, while the environment specifies only kinship, reciprocity-based allocation, and the absence of centralized authority.

Across long-horizon simulations, agents spontaneously reproduce five core Differential Order phenomena: stable labor specialization, guanxi-based economic ethics, cooperation attenuation with relational distance, emergent relational authority, and clan-based center–periphery stratification. These patterns vary systematically with production structure: symmetric endowments yield kin-centered integration closer to mechanical solidarity, whereas complementary endowments produce more functionally interdependent patterns closer to organic solidarity. Ablations on two LLMs further show that these outcomes are architecture-dependent rather than generic LLM artifacts. Together, the results recast Differential Order as a structure-sensitive emergent outcome of general social mechanisms and position LLM-based multi-agent simulation as a generative testbed for social structure and change.

We summarize our contributions as follows: (1) we provide a computational approach to study Fei Xiaotong’s Differential Order Pattern without encoding culture-specific rules; (2) we design CAREB-MAS, a multi-agent framework integrating emotion, ethical judgment, and relational cognition to harness and test the theoretical mechanisms and (3) we validate core Differential Order Pattern phenomena and conduct controlled experiments to reveal the sensitivity and dependency of these phenomena on cognitive architecture.

2 RELATED WORK

2.1 Social Transformation and Rural Society Formation

Classical sociology treats social structure as an emergent outcome of interaction rather than a fixed

design. In Durkheim, authority and division of labor arise through

Collective Affect → Ethics → Belief → Social Structure

process (Tada, 2020; Durkheim, 1915). This generative view aligns with organizational theories that structure as emerging from routinized interaction (Haveman and Wetts, 2019).

Fei Xiaotong’s Differential Order Pattern describes rural society as egocentric and relationally graded. Its core preconditions—kinship organization, generalized reciprocity beyond price exchange, and acephalous authority—are widely documented across comparative studies of pre-market society (Fortes and Evans-Pritchard, 1940; Lévi-Strauss, 1949; Sahlins, 1972; Mauss, 1925; Polanyi, 1944). Relational sociology and historical anthropology further show that such structures persist and reconfigure under modernization (Zhang, 2015; Hanting, 2024; Zhang, 2018; Longyuan and Iftikhar, 2023), making rural society a useful setting for examining general social mechanisms (Zhou and Xiao, 2024).

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) and cultural dissemination (Axelrod, 1997) to segregation (Schelling, 1971) and 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).

LLM agents extend this line of work by enabling richer social reasoning. Park et al. (Park et al., 2023) demonstrated believable social behavior in generative agents, and subsequent studies operationalized specific theories or institutions, including social exchange (Wang et al., 2025), Hobbesian contract formation (Dai et al., 2024), institutional emergence (Piao et al., 2025), cooperation evolution (Vallinder and Hughes, 2024), and norm emergence (Ren et al., 2024).

Despite these advances, existing simulations focus mainly on short-term coordination or dyadic exchange, leaving the long-horizon co-emergence of collective affect, authority, and division of labor underexplored—especially for non-Western sociological theories whose structural predictions have rarely been tested computationally.

3 Theoretical Foundations

3.1 Differential Order and Relationally Embedded Division of Labor

Fei Xiaotong's *Differential Order Pattern* conceptualizes 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. Although often interpreted through the Chinese rural context, its key structural preconditions—kinship organization (Fortes and Evans-Pritchard, 1940; Lévi-Strauss, 1949), reciprocity-based allocation without price mechanisms (Polanyi, 1944; Sahlins, 1972; Mauss, 1925), and limited centralized authority (Fortes and Evans-Pritchard, 1940)—are widely documented across comparative anthropology. We therefore examine whether Differential Order-like patterns can emerge from these cross-culturally documented conditions.

Within this framework, division of labor stabilizes through repeated interaction under relational and ethical constraints rather than formal institutions. We term this *relationally embedded division of labor* and test whether it can emerge endogenously without explicit cultural rules.

3.2 Guanxi as Economic Ethics

In Differential Order, *guanxi* is not merely a preference for familiar others but an *economic-ethical principle* that defines obligation, responsibility, and acceptable cost in exchange. This parallels Sahlins's *generalized reciprocity* (Sahlins, 1972): close ties sustain open-ended obligation without immediate return. In low-contractual environments, such relational ethics can stabilize cooperation and legitimate asymmetric outcomes despite short-term inefficiency.

3.3 Five Generative Propositions

To distinguish built-in assumptions from dependent outcomes, we conceptually separate *pre-defined structural conditions* from *emergent interactional patterns*. The former correspond to cross-culturally documented features of pre-market social organization, while the latter constitute the primary objects of explanation in our simulations (see Appendix J). As shown later, identical structural conditions can yield qualitatively different emergent structures

under symmetric and complementary production regimes.

Based on these mechanisms, we formulate five generative propositions. Each describes a macro-level pattern that, if observed, would be consistent with Differential Order—but that arises from general social cognition rather than culture-specific encoding:

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 peripheral disadvantage.

3.4 Structural Contrast: Mechanical vs. Organic Solidarity

To examine how production structure conditions Differential Order, we contrast *symmetric* and *complementary* endowments. Symmetric capacities favor integration through shared identity and relational ethics (*mechanical solidarity*), whereas complementary capacities favor integration through functional differentiation and mutual dependence (*organic solidarity*). This contrast provides the main analytical lens for interpreting variation across conditions.

4 Methodology: The CAREB-MAS Framework

We propose CAREB-MAS (Collective Affection–Reasoning–Emergence Based Multi-Agent Simulation), a framework for examining whether structured social order can emerge without explicit cul-

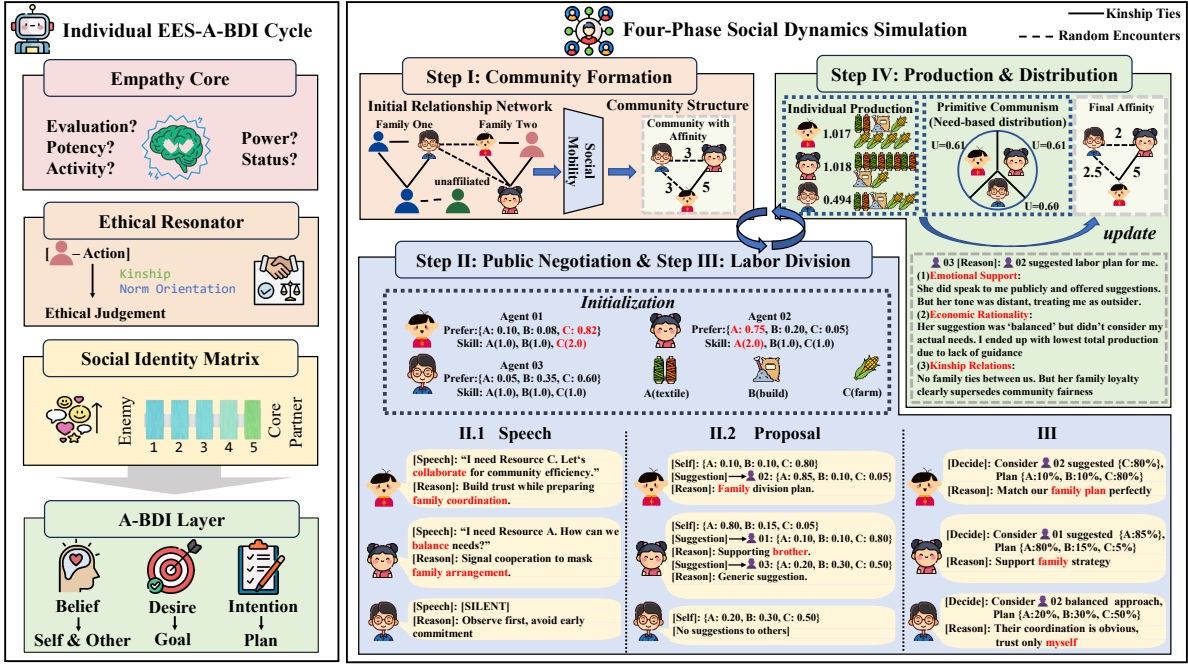


Figure 2: Overview of CAREB-MAS. **Left:** agent-level cognitive architecture, from EC (Affect Control Theory) to ER (ethical judgment), SIM (Social Identity Theory), and A-BDI (Durkheimian affective constraint on action). **Right:** four-phase interaction loop: community formation, deliberation, division of labor, and production with psychological update.

tural templates. Stable social structure arises from repeated local interaction among boundedly rational agents, without hard-coded cultural rules or predefined hierarchy.

4.1 Sociological Logic and Emergence Design

The framework operationalizes a causal pathway grounded in three theories: affective perception from Affect Control Theory (Heise, 2007), moral judgment from Durkheimian ethics (Émile Durkheim, 1961), and relational cognition from Social Identity Theory (Tajfel and Turner, 1979). These components are integrated into an augmented BDI architecture in which collective affect constrains rational action (Durkheim, 1915). The resulting chain—*Affect* → *Ethics* → *Identity* → *Action*—models general social cognition rather than culture-specific institutional content. No module directly encodes Fei’s theory, Confucian ethics, or predefined *guanxi* categories. Any Differential Order patterns therefore arise from the interaction between general mechanisms and the structural conditions in Table 26.

4.1.1 Agent Architecture

All agents share a unified psychological–decision architecture activated sequentially each round (Fig-

ure 2). Social perception produces affective responses, which are filtered through moral evaluation, update relational cognition, and finally constrain belief–desire–intention reasoning. Full implementation details and prompts are provided in Appendix D.

- **Empathy Core (EC)** ← *Affect Control Theory* (Heise, 2007). Decodes observable behavior into a five-dimensional affective vector (evaluation, potency, activity, status-conferral, power-assertion).
- **Ethical Resonator (ER)** ← *Durkheimian Moral Education* (Émile Durkheim, 1961). Aggregates perceptions into a moral judgment of each interaction partner, outputting *approbatory*, *repressive*, or *restorative* attitudes for relational updating.
- **A-BDI Decision Module** ← *BDI + Durkheimian collective affect*. Extends BDI (Georgeff et al., 1998; Wang et al., 2025) with *Affection*, embodied in SIM and constraining belief formation, desire generation, and intention selection.

4.1.2 Social Dynamics: A Four-Phase Loop

Each round follows a four-phase loop (Figure 2) modeling pre-modern rural formation under co-

residence and relational cognition, without prices or centralized adjudication. Hyperparameters and simulation settings are reported in Appendix A.

Phase I: Community Formation. Agents partition into communities based on mean relational proximity $\bar{d}_{ij} = (d_{i \rightarrow j} + d_{j \rightarrow i})/2$, where $d_{i \rightarrow j} \in \{1, \dots, 5\}$. Louvain detection produces dynamic communities, and agents may migrate as relationships evolve (Appendix D.4.1).

Phase II: Public Deliberation. Within communities, agents engage in $R = 2$ rounds of public deliberation, broadcasting speeches and labor proposals. No voting or enforcement rule is imposed. The resulting action set $\mathcal{A}_C^{(t)}$ becomes the observable input for subsequent updating (Appendix D.4.2).

Phase III: Division of Labor and Production. Each agent independently chooses labor allocation $\ell_i = (\ell_{iA}, \ell_{iB}, \ell_{iC})$ with $\sum_r \ell_{ir} = 1$, based on $\mathcal{A}_C^{(t)}$, A-BDI state, and SIM. No centralized authority or conflict-resolution rule is imposed. Economic “authority” is measured *ex post* by proposal adoption frequency (Appendix D.4.3).¹

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 redistributive allocation prototype consistent with Polanyi’s redistributive economies (Polanyi, 1944) and Sahlins’s generalized reciprocity (Sahlins, 1972). After consumption, each agent executes the update cycle $EC \rightarrow ER \rightarrow SIM \rightarrow A\text{-BDI}$. Production, clearing, and utility details are reported in Appendices D.4.4–D.4.6.

Evaluation Transparency. EC and ER are *agent-internal* cognitive processes: each agent uses its own LLM call to perceive and judge others, so assessments may differ across agents. The only metric using an *external* LLM-as-Judge evaluator is Communication Intensity (Appendix B.2, Table 6). All other metrics—Decision Authority, Proposal Intensity, Lock-in Score, Community Match, and SIM levels—are computed deterministically from simulation logs.

¹Decision Authority (DA) is the paper’s primary operationalization of emergent authority; formal definition appears in §6.

5 Experiment Setup

We simulate 18 agents over 30 rounds: two kin-based clans (F1 and F2, six members each) and six unaligned agents without family ties. Agents jointly produce three resource types under a Leontief utility function with preference-weighted redistribution (§4). We compare two conditions: *symmetric skills* (balanced production capacities) and *complementary skills* (group-specific specialization in one resource). All agents share the same cognitive architecture; unaligned agents differ only in lacking kinship ties. No agent is given knowledge of Fei’s theory or any culture-specific behavioral rules.

Simulations use DeepSeek-V3 (temperature 0.7), with five random seeds per condition. Cross-model robustness is assessed using GPT-4o-mini and Gemini-2.5-flash-lite (Appendix E). Code, prompts, and simulation logs will be released upon publication.² Metric definitions and implementation details are provided in Appendices B–D.

5.1 Proposition–Metric Mapping

Each proposition is operationalized via a primary quantitative metric paired with a diagnostic criterion that specifies the expected pattern under Differential Order. The full mapping between propositions, metrics, and diagnostic criteria is provided in Appendix C, with formal definitions in Appendix B.

6 Experimental Results

We evaluate Propositions 1–5 using the criteria in Table 8.

6.1 Emergence of Stable Division of Labor (P1)

Figure 3 shows that stable specialization emerges without centralized coordination. Under symmetric skills, lock-in rises rapidly and plateaus at 0.972; under complementary skills, convergence is faster and reaches 0.985. The acceleration under complementarity suggests that structural differentiation itself drives specialization speed, not merely learning efficiency. Per-round improvements decline to near-zero in both conditions, confirming that the division is stable rather than transient. These patterns support Proposition 1. Ablation analysis (Appendix) further indicates that while convergence is architecturally robust, the precise calibra-

²<https://github.com/Chi20/DOP>

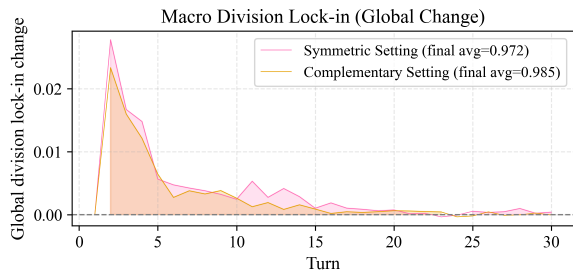


Figure 3: Global division lock-in emerges early and stabilizes under both skill structures. Panel A: symmetric skills converge to 0.972; Panel B: complementary skills converge faster to 0.985. Per-round improvements (Panel C) approach zero, indicating equilibrium.

tion of social coordination depends specifically on the EC/ER/SIM mechanism.

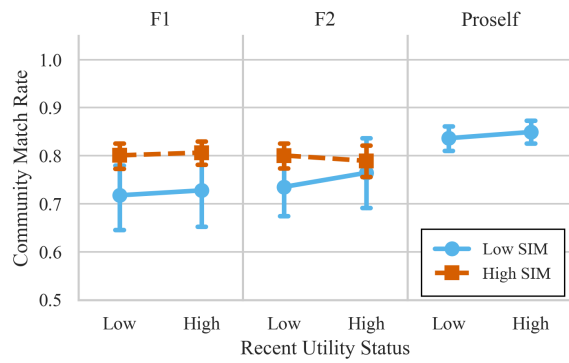
6.2 Guanxi as Economic Ethics (P2)

To test whether *guanxi* stabilizes cooperative roles beyond efficiency maximization, we examine how community match responds to negative utility shocks under varying relational embeddedness. Figure 4 reveals a consistent pattern: high-SIM agents maintain stable community participation regardless of recent payoff performance, while low-SIM agents withdraw sharply when efficiency incentives turn unfavorable. This asymmetry is particularly marked under complementary skills, where coordination failure is most costly. The stickiness of high-SIM ties suggests that ethical obligations embedded in close relations function as a commitment device, sustaining participation in community-based division of labor even when short-term rationality would dictate defection. This supports Proposition 2.

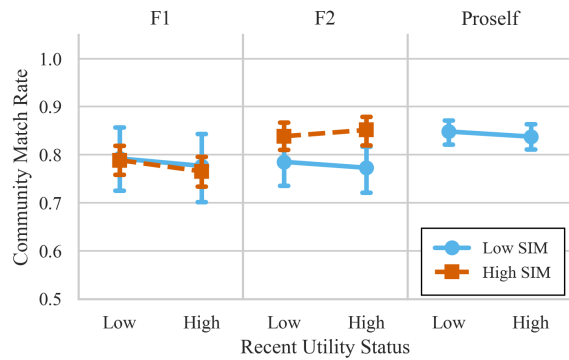
6.3 Relational Decay of Cooperation (P3)

Figure 5 reveals that cooperation decays along relational rather than preference-type boundaries. Family agents (F1, F2) concentrate decision authority and proposal activity within their own groups, with markedly lower engagement across family boundaries. This suggests cooperation is organized around sharply bounded family circles. In contrast, Proself agents do not form a cohesive self-boundary; they accept family-originated proposals readily and receive substantial proposals from family agents, effectively integrating into existing family structures rather than forming an independent cooperative bloc.

The proposal distribution quantifies this asym-



(a) Symmetric skills.



(b) Complementary skills.

Figure 4: Community match rates under utility shocks. F1/F2: family groups; Proself: unaligned agents. High-SIM relations maintain stable participation (solid lines) even when recent utility is low, whereas low-SIM ties (dashed lines) exhibit sharp decline. This stickiness is more pronounced in the complementary condition (B).

metry. Under symmetric skills, within-family proposals account for 34.2% of total activity and family→proself proposals for 34.3%, whereas cross-family proposals comprise only 6.9%. This indicates that cooperative energy flows primarily along family channels and outward to unaligned agents, but rarely across family boundaries. These patterns support Proposition 3: cooperation exhibits graded relational decay centered on family structures rather than distributed uniformly by individual preference type.

6.4 Emergence of Authority Stratification (P4)

Table 1 reports OLS estimates of decision authority on proposal intensity, communication intensity, accumulative utility, and social identity. In both conditions, proposal intensity is strongly and positively associated with decision authority (symmetric: $\beta = 0.820$, $p < 0.01$; complementary: $\beta = 0.780$, $p < 0.01$). Communication intensity, accumulative utility, and SIM, by contrast, show

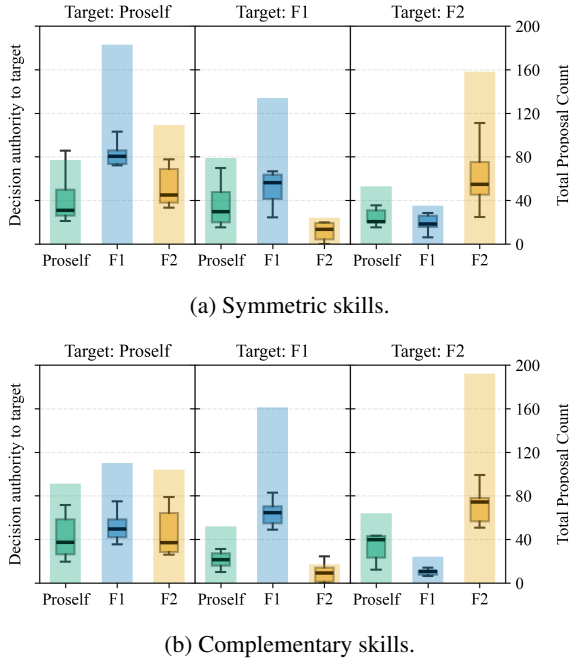


Figure 5: Decision authority (boxplots) and proposal activity (bars) exhibit a relational gradient. Family agents (F1, F2) show high within-group authority and low cross-family engagement, while Proself agents lack self-boundary and align with family structures. Note the asymmetry in proposal distribution: under symmetric skills (a), within-family and family→proself proposals dominate (34.2% and 34.3%), whereas cross-family proposals are rare (6.9%).

no robust independent main effects in the baseline models.

This asymmetry indicates that authority stratification is driven primarily by action-based agenda control rather than by communicative effort, accumulated payoffs, or relational identity. These results support Proposition 4: durable authority emerges endogenously through proposal-driven mechanisms.

6.5 Clan-Based Center–Periphery Stratification (P5)

Authority remains persistently stratified over time rather than diffusing uniformly. Figure 6 shows that distributions stabilize into structured forms: under complementary skills, the right-skewed mass in high-authority regions sustains, indicating a stable center–periphery structure; under symmetric skills, differentiation is weaker but persists.

To examine the relational conditions of authority formation, we regress decision authority on social identity (SIM), family membership, and their interaction (Table 1). Under symmetric skills, neither SIM nor family membership alone predicts

Table 1: Determinants of Decision Authority. Proposal intensity dominates in both conditions; communication, accumulative utility, and SIM show no robust independent main effects.

	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. Baseline models include proposal and communication controls; interaction models isolate SIM, family, and SIM×Family effects. All models include turn fixed effects and group-clustered standard errors. Full specifications appear in Appendix G.

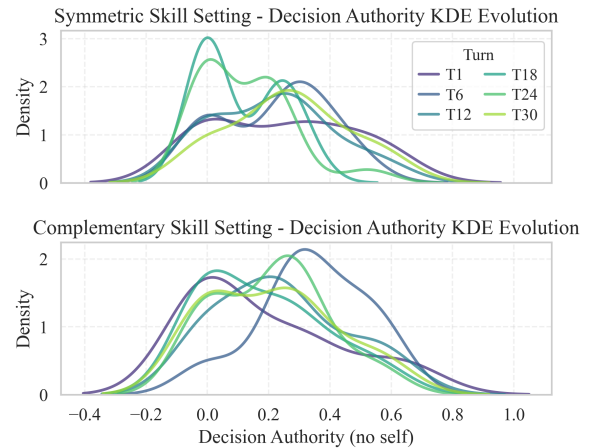


Figure 6: Temporal distribution of decision authority under symmetric and complementary skill structures. Kernel density estimates show persistent stratification rather than convergence to uniformity across rounds.

authority, but their interaction is strongly positive ($\beta = 0.603$, $p < 0.01$). Authority thus accrues disproportionately to agents whose social recognition is embedded in family relations. Under complementary skills, this interaction disappears, suggesting that functional differentiation attenuates relationally conditioned authority. This supports Proposition 5: center–periphery stratification emerges most clearly when family affiliation and social recognition align.

6.6 Architecture Ablation and Baseline Analysis

Table 2 reports cross-model module ablations under symmetric endowments; detailed BDI-only comparisons are shown in Table 3.

Table 2: Cross-model ablation dissociation (symmetric endowments). Removing EC disrupts EC→SIM but preserves ER→SIM; removing ER does the reverse.

	EC→SIM	ER→SIM	SIM×Fam→Auth
DS Full	0.748**	0.573***	0.070 n.s.
DS w/o EC	—	1.021***	0.029 n.s.
DS w/o ER	0.170 n.s.	—	0.042 n.s.
GPT Full	2.216***	−0.002 n.s.	0.106*
GPT w/o EC	—	0.498***	−0.197 n.s.
GPT w/o ER	−0.127*	—	0.181 n.s.
Gemini Full	0.069**	0.186***	0.089 n.s.
Gemini w/o EC	—	0.061*	0.097 n.s.
Gemini w/o ER	−0.022 n.s.	—	−0.070 n.s.
All BDI-only	absent	absent	absent [†]

[†]BDI-only lacks SIM, the interaction term is therefore undefined; Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; n.s. = not significant.

Table 3: Full architecture vs. BDI-only baseline (symmetric endowments).

Prop.	Indicator	DS	GPT	Gemini
P1	Lock-in (last 5)	0.971 / 0.979	0.971 / 0.979	0.833 / 0.833
P1	WGCS (F1)↓	0.410 / 0.477*	0.410 / 0.477*	0.404 / 0.379 n.s.
P2	High/Low-SIM gap	present / n/a	present / n/a	−0.022 n.s. / n/a
P3	Cross-fam. proposals	6.9% / 0.7%	6.9% / 0.7%	10.3% / 1.3%
P4	DA mean (no-self)	0.186 / 0.056	0.186 / 0.056	0.382 / 0.234
P5	SIM×Fam→Auth	n.s. / absent	** / absent	n.s. / absent

Notes: n/a: SIM absent in BDI-only. P4 BDI-only reductions: DeepSeek-V3 and GPT-4o-mini −70%; Gemini-2.5-flash-lite −39%. Significance: * $p < 0.05$, ** $p < 0.01$.

Under symmetric endowments, removing EC selectively disrupts the EC→SIM pathway while preserving ER→SIM, whereas removing ER disrupts ER→SIM while preserving EC→SIM. This dissociation suggests that EC and ER make distinct, module-specific contributions to relational updating within CAREB-MAS.

The three models nonetheless recover the same macro-structural patterns through different intermediate pathways. DeepSeek-V3 draws on both EC and ER jointly; GPT-4o-mini relies primarily on affective-evaluation variance; Gemini-2.5-flash-lite shows an ethical-judgment-dominant profile (ER→SIM 0.186*** vs. EC→SIM 0.069**). That pathway divergence coexists with structural convergence is difficult to reconcile with a uniform RLHF-artifact explanation.

BDI-Only Baseline Comparison. Table 3 compares the full architecture with a BDI-only baseline that removes EC, ER, and SIM while retaining vanilla decision logic. All indicators follow unified definitions (Appendix B).

P1 lock-in remains high across both conditions (0.83–0.98), indicating that stable specialization

itself is largely LLM-native. The social quality of this specialization, however, diverges. For DeepSeek-V3 and GPT-4o-mini, the full architecture produces more coherent within-group coordination (WGCS: 0.410 vs. 0.477, $p < 0.05$). Gemini-2.5-flash-lite shows a directional but non-significant WGCS improvement under BDI-only (0.404 → 0.379, $p = 0.10$), yet this coincides with a collapse in cross-family proposals (P3: 10.3% → 1.3%) and a marked decline in decision authority (Gemini: −39%; DeepSeek/GPT: −70%).

Across all models, P3–P5 degrade directionally under BDI-only: the concentric cooperation gradient flattens, self-directed proposals increase, and relational authority stratification vanishes. Single-module ablations yield only modest degradation, but complete removal triggers discontinuous collapse in relational structure. This indicates that EC and ER operate as partially compensatory yet jointly necessary pathways for sustaining relational social structure.

Full ablation results are reported in Appendix F. The per-dyad specification in Appendix F.3 uses EC evaluation level, whereas Table 2 uses EC intra-agent variance; despite this operational difference, all three models recover the same directional EC→SIM pathway.

6.7 Bias Diagnosis: RLHF Artifact or Architectural Emergence?

Observed cooperation patterns could stem from RLHF-induced prosocial bias, but four observations suggest they instead reflect architectural mechanisms.

First, agents exhibit behaviors that are not well explained by a uniformly prosocial RLHF account. Family agents show the lowest authority acceptance toward the opposing family (Figure 5)—exclusionary rather than uniformly cooperative. Authority distributions remain right-skewed (Figure 6), and utility rankings reverse between conditions (Appendix H). These patterns replicate across DeepSeek-V3, GPT-4o-mini, and Gemini-2.5-flash-lite.

Second, ablation dissociation is module-specific (Table 2). Removing EC eliminates the EC→SIM pathway while sparing ER→SIM, and removing ER does the reverse. If RLHF were the driver, module removal should produce uniform attenuation rather than targeted pathway elimination. This selective dissociation holds across all three models.

Third, the three models achieve convergent P1–

P5 patterns through divergent intermediate mechanisms. DeepSeek-V3 operates through joint affective–ethical processing, GPT-4o-mini through affective-evaluation dominance, and Gemini-2.5-flash-lite through ethical-judgment dominance. GPT shows positive guanxi stickiness (P2 gap = +0.03–0.04), whereas DeepSeek and Gemini show negative or absent stickiness (Gemini: –0.022 n.s.). Yet all three produce the same concentric cooperation gradient (P3) and authority stratification (P4). Divergence in P2 alongside convergence in P3–P4 is incompatible with a uniform RLHF explanation.

Fourth, BDI-only baseline collapse is directionally consistent across architectures. Removing EC, ER, and SIM degrades cross-family coordination and decision authority regardless of the underlying LLM: for Gemini, cross-family proposals drop from 10.3% to 1.3% and DA falls 39%; for DS/GPT, cross-family proposals drop from 6.9% to 0.7% and DA falls 70%. While magnitudes vary, the directional pattern—BDI-only weakens relational structure—indicates that observed social structure depends on the CAREB-MAS architecture rather than model-specific training artifacts.

6.8 Summary: Structural Contrast

Across Propositions 1–5, the main contrast is structural: the same architecture yields different modes of integration under different production endowments. These outcomes vary systematically with production endowments: symmetric skills yield more kin-centered, bounded integration consistent with *mechanical solidarity*, whereas complementary skills produce more functionally interdependent and less clan-bound organization closer to *organic solidarity*. This contrast locates Differential Order Pattern at the level of structural conditions rather than as a fixed cultural template.

A related finding is the recurring decoupling between coordination burden and durable authority—an “Atlas paradox”—which suggests that durable power remains anchored in agenda control rather than coordination service (Appendix L; Appendix M).

7 Conclusion

In this paper, we show that Differential Order Pattern, often treated as specific to Chinese rural society, can emerge from cross-culturally documented social mechanisms without encoding culture-specific institutional rules. We develop

CAREB-MAS, a multi-agent framework that models emotional, ethical, and relational processes under minimal pre-market structural conditions. Across long-horizon simulations, CAREB-MAS reproduces five core Differential Order Pattern phenomena from local interaction alone. Production-structure contrasts and cross-model ablations show that these outcomes vary with structural conditions and depend on cognitive architecture rather than generic LLM behavior. These results support interpreting Differential Order Pattern as a structure-sensitive emergent outcome of general social mechanisms.

Limitations and Future Directions

The present study faces four limitations. First, EC, ER, and SIM updates scale as $O(N^2)$, restricting experiments to $N = 18$ and limiting population-level claims. However, this configuration is a minimal setting for observing intra-clan cohesion, inter-clan negotiation, and cross-group integration within the same simulation. Second, the environment is simplified: we do not include explicit sanctioning, conflict-resolution institutions, or external shocks such as migration and scarcity. While this facilitates mechanism tracing, it limits external validity and leaves open whether such factors would fundamentally alter dynamics. Third, validation is driven more by theory than by direct ethnographic calibration; related empirical correspondence is discussed in Appendix N. Stronger empirical correspondence requires longitudinal field comparison and boundary-condition testing. Fourth, although ablation results do not support a uniform RLHF-bias explanation, we cannot fully exclude training-data priors. Future work should address these limits through larger-scale simulations, more complex social settings, closer empirical comparison, and tests on a wider range of model families.

Acknowledgments

This research was supported by the National Key R&D Program of China (No.2023YFC3304701), NSFC (No.6250072448, No.62272466,U24A20233), Big Data and Responsible Artificial Intelligence for National Governance, Renmin University of China; the Joint Academy on Future Humanity, Renmin University of China and Westlake University(No.2025RD0202) and Center for AI for Social Science, SLAI.

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.
- Meyer Fortes and Edward E Evans-Pritchard. 1940. *African Political Systems*. Oxford University 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.
- 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.
- Mark Granovetter. 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510.
- Zhang Hanting. 2024. *Chinese relational sociology: An analysis of the theoretical paradigm of guanxi studies in the chinese context*. 20(4):59.
- 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.
- 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.
- Claude Lévi-Strauss. 1949. *Les Structures élémentaires de la Parenté*. Presses Universitaires de France.
- 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.
- Marcel Mauss. 1925. *The Gift: Forms and Functions of Exchange in Archaic Societies*. Cohen & West.
- 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. Agentso-ciety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*.
- Karl Polanyi. 1944. *The Great Transformation*. Farrar & Rinehart.
- 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.
- Marshall Sahlins. 1972. *Stone Age Economics*. Aldine-Atherton.
- Thomas C Schelling. 1971. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.
- 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.
- Émile Durkheim. 1961. *Moral Education: A Study in the Theory and Application of the Sociology of Education*. Free Press, New York. Originally published as *L'éducation morale* (1925).

A Experimental Parameterization.

Table 4 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.

A.1 Parameter Definitions

Table 5 provides formal definitions for all parameters referenced in the main text.

B Metric Definitions

B.1 Division and Matching Metrics

Division Lock-in Score. For agent i with preference vector pref_i and labor skill vector skill_i :

$$\text{lock-in}_i = \frac{\text{pref}_i \cdot \text{skill}_i}{\|\text{pref}_i\| \cdot \|\text{skill}_i\|} \quad (1)$$

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:

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

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:

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

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

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

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

Group-level matching (reported as ‘‘Group matching’’ in tables) is the mean of group_match_i across agents within the same Louvain-detected community at round t .

Within-Group Coordination Score (WGCS). WGCS quantifies the coherence of labor allocation decisions within a community by measuring the alignment between individual proposals and the community’s aggregate demand profile.

Let C denote a community at round t , and let $\mathcal{P}_C = \{p_{i \rightarrow j}^{(k)} : i, j \in C, k \in K_i^{(C)}\}$ be the set of all labor proposals issued within C . Define the community proposal centroid as:

$$\bar{\mathbf{p}}_C = \frac{1}{|\mathcal{P}_C|} \sum_{p \in \mathcal{P}_C} \mathbf{p} \quad (7)$$

where \mathbf{p} is the vector representation of proposal p in the labor-type space.

For each agent $i \in C$, let $\mathbf{a}_i^{(C)}$ denote i ’s final labor allocation vector. The within-group coordination for agent i is:

$$\text{WGCS}_i = \cos(\mathbf{a}_i^{(C)}, \bar{\mathbf{p}}_C) \quad (8)$$

The community-level WGCS is the mean across members:

$$\text{WGCS}_C = \frac{1}{|C|} \sum_{i \in C} \text{WGCS}_i \quad (9)$$

In practice, WGCS is reported as an F1-weighted variant to account for class imbalance across labor types:

$$\text{WGCS}_C^{(\text{F1})} = \frac{2 \cdot \text{Precision}_C \cdot \text{Recall}_C}{\text{Precision}_C + \text{Recall}_C} \quad (10)$$

where Precision_C and Recall_C are computed by treating the community centroid $\bar{\mathbf{p}}_C$ as the ‘‘gold’’ allocation and each agent’s final allocation as a prediction. This F1 variant (reported as ‘‘WGCS (F1)’’ in tables) is more sensitive to coordination failures on minority labor types and is the primary metric used in ablation comparisons.

B.2 Behavior Metrics

Decision Authority. Let $p_{i \rightarrow j}^{(k)}$ 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.

Table 4: 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 F2: 2:2:2 Proself: 2:2:2	F1: 4:1:1 F2: 1:4:1 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)

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

Otherwise, decision authority is computed as:

$$\text{Authority}_i = \frac{1}{w \sum_{k \in K_i^{(-i)}} \left\| p_{i \rightarrow j(k)}^{(k)} - a_{j(k)} \right\|^2}, \quad (11)$$

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.

$$\text{Proposal}_i = \sum_{j \neq i} 1[\text{agent } i \text{ proposes to } j] \quad (12)$$

Proposal Directionality. To distinguish self-directed from other-directed coordination behavior, we define:

$$\text{Self-proposals}_i = \sum_k 1[p_{i \rightarrow i}^{(k)} \text{ exists}] \quad (13)$$

$$\text{Other-directed}_i = \sum_{j \neq i} \sum_k 1[p_{i \rightarrow j}^{(k)} \text{ exists}] \quad (14)$$

$$\text{Total}_i = \text{Self-proposals}_i + \text{Other-directed}_i \quad (15)$$

$$\text{Other-directed } \%_i = \frac{\text{Other-directed}_i}{\text{Total}_i} \times 100 \quad (16)$$

Cross-family proposals are a subset of Other-directed proposals where j belongs to a different family group than i .

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

Table 5: Formal parameter definitions.

Symbol	Definition
N	Number of agents in the simulation (=18)
T	Number of interaction rounds (=30)
R	Number of deliberation sub-rounds per phase (=2)
\mathcal{R}	Set of resource types ($\{A, B, C\}$)
$d_{i \rightarrow j}$	Agent i 's subjective identity level toward agent j ; $d_{i \rightarrow j} \in \{1, 2, 3, 4, 5\}$
\bar{d}_{ij}	Symmetric relational proximity: $(d_{i \rightarrow j} + d_{j \rightarrow i})/2$
e_{jb}	Ethical-Judgmental Bias; $e_{jb} \sim \text{Uniform}(-0.5, 0.5)$; modulates moral assessment stringency
ℓ_i	Labor allocation vector for agent i ; $\ell_i = (\ell_{iA}, \ell_{iB}, \ell_{iC})$, $\sum_r \ell_{ir} = 1$
$s_{i,t}(r)$	Agent i 's skill level for resource r at round t ; updated by learning-by-doing: $s_{i,t+1}(r) = s_{i,t}(r) + \ell_{i,t}(r)$
\mathbf{p}_i	Consumption preference vector; $\mathbf{p}_i \sim \text{Dir}(\mathbf{1}_3)$
$b(r)$	Resource-specific production unit (contextual coefficient); $b(r) = 1 \forall r$ in both conditions
λ	Search randomness parameter for community edge formation (=0.3)
$u_{i,t}$	Per-round Leontief utility: $u_{i,t} = \min_r \{R_{i,t}(r)/p_i(r)\}$
$U_{i,t}$	Cumulative utility: $U_{i,t} = \sum_{\tau=1}^t u_{i,\tau}$

range $[0, 10]$, aggregating four dimensions: relationship complexity, syntactic complexity, semantic information content, and argument logic (Table 6). This score is used directly as the agent’s communication action intensity.

Table 6: 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.

B.3 Social Identity Metrics

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

$$SID_i = \frac{1}{|C_i|} \sum_{j \in C_i} SIM(i, j) \quad (17)$$

In-Group vs. Cross-Group SIM Differentiation. To quantify relational identity gradient (P3), we compute:

$$SIM_gap_i = \frac{1}{|F_i \cap C_i|} \sum_{j \in F_i \cap C_i} SIM(i, j) - \frac{1}{|C_i \setminus F_i|} \sum_{j \in C_i \setminus F_i} SIM(i, j) \quad (18)$$

where F_i denotes agent i ’s family group. Positive values indicate stronger relational identity toward in-family members within the same community. The population-level P2/P3 gap reported in tables is the mean of SIM_gap_i across agents (or, for DS/GPT, the difference in community-match between high-SIM and low-SIM tertiles under low-utility conditions; see footnote in Table 3 for Gemini operational difference).

B.4 Utility and Adaptation Metrics

Total and Average Utility.

$$U_{total} = \sum_{i=1}^N U_i, \quad \bar{U} = \frac{U_{total}}{N} \quad (19)$$

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

$$\Delta M = M^{post} - M^{pre}, \quad (20)$$

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

$$AE_i = \frac{|U_i|}{P_i}. \quad (21)$$

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.

Late Labor Concentration. To capture temporal dynamics of specialization, we define late-stage labor concentration as the Gini coefficient of agent i ’s labor allocation vector $\ell_i^{(T_{late})}$ averaged over the final 20% of simulation rounds:

$$LateConc_i = \text{Gini} \left(\frac{1}{|T_{late}|} \sum_{t \in T_{late}} \ell_i^{(t)} \right) \quad (22)$$

Higher values indicate more focused specialization in the late stage of community formation.

C Concept–Metric Mapping and Justification

To ensure that simulated patterns reflect theoretically grounded mechanisms rather than arbitrary computational artifacts, we provide a two-level

mapping: (1) ethnographic concepts to computational metrics, and (2) generative propositions to their quantitative operationalizations and diagnostic criteria.

C.1 Ethnographic Concept–Metric Mapping

Table 7 formalizes the correspondence between key ethnographic constructs in *From the Soil* (Fei et al., 1992) and their computational representations.

Ethnographic Concept	Computational Metric	Simulated Phenomenon
<i>Differential Order</i> (Ch. 4) The “self-centered web”	Relational Distance (d_{ij}) extracted from the dynamic Social Identity Matrix (SIM).	Cooperation and proposal intensity decay predictably outward from the ego-center.
<i>Relational Ethics</i> (Ch. 8) Rules of propriety over law	SIM Match vs. Utility Gap.	High-SIM relations sustain cooperation and community matching even under severe utility deficits.
<i>Elder Authority</i> (Ch. 11) Exchange-based power	Decision Authority (Proposal adoption rate normalized by community).	Authority concentrates in individuals providing durable coordination, decoupled from sheer verbosity.

Table 7: Mapping between ethnographic constructs and computational metrics.

C.2 Proposition–Metric Mapping

Table 8 summarizes how each generative proposition is operationalized via a primary quantitative metric and a diagnostic criterion specifying the expected empirical pattern under Differential Order.

Table 8: Proposition–metric mapping.

	Primary Metric	Diagnostic Criterion
P1	Lock-in Score (output–skill cosine)	High convergence with near-zero per-round change
P2	Community Match gap (high vs. low SIM)	Maintained community match despite low utility
P3	DA & Proposal Activity by source–target group pair	Within-family $>$ family→proself $>$ cross-family gradient
P4	DA density over time; hierarchical OLS	Proposal intensity predicts authority beyond identity or wealth
P5	SIM×Family interaction (OLS)	Joint conditioning of social identity and family on authority

C.3 Justification of Metric Choices

This section provides extended justification for the operationalization choices summarized above.

C.3.1 P1: Lock-in Score

Why Lock-in Score rather than raw specialization index? P1 predicts that without central coordination, individuals lock into specialized roles. Lock-in Score is defined as the cosine similarity between an agent’s production allocation vector and its skill vector (Appendix B). This directly measures whether specialization has occurred *and stabilized*: convergence to high values indicates specialization; near-zero per-round change indicates equilibrium. Alternative measures—such as Herfindahl index—capture specialization intensity but not its stability or alignment with skill. Lock-in Score captures both dimensions simultaneously.

C.3.2 P2: Community Match by SIM × Utility

Why community match rather than raw cooperation frequency? P2 predicts that relational obligation sustains cooperation under adversity. We partition agents by SIM level (high vs. low) and recent utility (high vs. low), then compare community match—the cosine similarity between an agent’s skill vector and community demand—across these groups.

Community match captures whether agents continue to *align their labor with community needs*, rather than merely engaging in interaction. Raw cooperation frequency would measure interaction volume without distinguishing productive cooperation from empty gestures. The key test is whether high-SIM agents maintain community match under low utility while low-SIM agents do not.

C.3.3 P3: Source–Target Group Decomposition

Why decompose DA and proposals by group pairs? P3 predicts ego-centric, distance-graded cooperation. We decompose Decision Authority and Proposal Activity by source–target group pairs (within-family, family→proself, cross-family). A monotonic gradient within-family $>$ family→proself $>$ cross-family constitutes direct evidence of the concentric structure predicted by Fei’s theory. This preserves the *directional* and *relational* structure lost in aggregate measures.

C.3.4 P4: DA Temporal KDE + OLS

Why combine distributional and regression evidence? P4 predicts that authority emerges endogenously. Two distinct claims require separate tests:

1. Authority *exists* as a differentiated phenomenon (tested via temporal KDE showing persistent stratification).
2. Authority is *emergent* rather than *ascribed* (tested via OLS showing proposal intensity predicts authority beyond identity, family, or wealth).

Neither test alone suffices.

C.3.5 P5: SIM×Family Interaction

Why the interaction term rather than main effects? P5 predicts clan-based center–periphery stratification. The SIM×Family interaction tests whether social identity alignment with family membership *jointly* conditions authority. This follows the standard notion of relational embeddedness (Granovetter, 1985): advantage accrues not from family or recognition alone, but from their alignment within the network. The disappearance of this interaction under complementary skills ($\beta = -0.089$, n.s.) while strong under symmetric skills ($\beta = 0.603^{***}$) directly tests structure sensitivity.

D LLM and Reproducibility Details

D.1 Model Configuration

Table 9: LLM Configuration

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

D.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.

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

Table 10: 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}].

$e_{jb} \sim \text{Uniform}(-0.5, 0.5)$ 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.

D.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.

D.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.

D.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.

D.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

Table 11: 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.

Table 12: 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.

Table 13: 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.

identity, moral judgment, emotional perception, and factual interaction history.

Table 14: 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.

D.4 Multi-Phase Interaction and Decision Pipeline

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

D.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 $s_{i \rightarrow j}$ and $s_{j \rightarrow i}$, each taking integer values in $\{1, \dots, 5\}$. The average identity strength is computed as:

$$\bar{s}_{ij} = \frac{s_{i \rightarrow j} + s_{j \rightarrow i}}{2}. \quad (23)$$

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:

$$p_{ij}^{\text{base}} = 0.1 \cdot (\bar{s}_{ij} - 1), \quad (24)$$

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:

$$p_{ij} = p_{ij}^{\text{base}} + \epsilon, \quad \epsilon \sim \text{Uniform}(0, \lambda), \quad (25)$$

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.

D.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.

D.4.3 Phase 3: Labor Division Decision Phase

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

Table 15: 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 16: 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 17: 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.

D.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:

$$y_{i,t}(r) = b(r) l_{i,t}(r) \log(1 + s_{i,t}(r)), \forall r \in \mathcal{R}. \quad (26)$$

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:

$$R_{i,t}(r) \leftarrow R_{i,t}(r) + y_{i,t}(r), \quad \forall r \in \mathcal{R}. \quad (27)$$

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

$$s_{i,t+1}(r) = s_{i,t}(r) + l_{i,t}(r), \quad \forall r \in \mathcal{R}. \quad (28)$$

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.

D.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:

$$Y_{g,t}(r) = \sum_{i \in \mathcal{I}_g} y_{i,t}(r). \quad (29)$$

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 :

$$P_g(r) = \sum_{i \in \mathcal{I}_g} p_i(r). \quad (30)$$

Agent i 's share ratio for resource r is:

$$\alpha_{i,g}(r) = \begin{cases} \frac{p_i(r)}{P_g(r)}, & \text{if } P_g(r) > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (31)$$

The redistributed amount received by agent i is then:

$$x_{i,t}(r) = \alpha_{i,g}(r) Y_{g,t}(r). \quad (32)$$

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.

D.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:

$$\rho_{i,t}(r) = \begin{cases} \frac{R_{i,t}(r)}{p_i(r)}, & \text{if } p_i(r) > 0, \\ +\infty, & \text{if } p_i(r) = 0. \end{cases} \quad (33)$$

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

$$u_{i,t} = \min_{r \in \mathcal{R}} \rho_{i,t}(r), \quad (34)$$

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:

$$r_{i,t}^* \in \arg \min_{r \in \mathcal{R}} \rho_{i,t}(r). \quad (35)$$

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

$$c_{i,t}(r) = u_{i,t} p_i(r), \quad \forall r \in \mathcal{R}. \quad (36)$$

Resources are updated by subtracting consumption:

$$R_{i,t}(r) \leftarrow R_{i,t}(r) - c_{i,t}(r). \quad (37)$$

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

$$U_{i,t} = U_{i,t-1} + u_{i,t}. \quad (38)$$

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:

$$\bar{u}_{i,t} = \frac{(t-1)\bar{u}_{i,t-1} + u_{i,t}}{t}. \quad (39)$$

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

D.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: $\mathbf{p}_i \sim \text{Dir}(\mathbf{1}_3)$. Ethical–Judgmental Bias (EJB) values are drawn uniformly: $e_{jb} \sim \text{Uniform}(-0.5, 0.5)$. 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:

$$P(\text{edge}_{ij}) = 0.1(\bar{s}_{ij} - 1) + \lambda \cdot \epsilon, \quad \epsilon \sim \text{Uniform}(0, 1) \quad (40)$$

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.

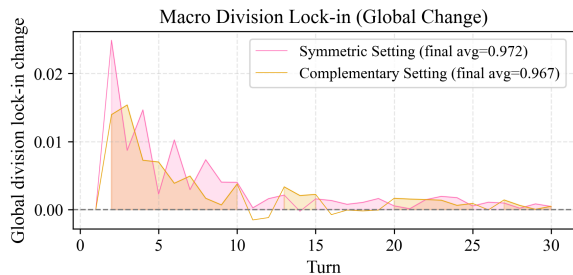


Figure 7: GPT-4o-mini simulation: Convergence of global division of labor under symmetric (left) and complementary (right) skill structures.

3. **LLM Sampling:** Model temperature is set to 0.7 (Table 9), introducing response variability. This is held constant across all conditions.

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.

E Cross-Model Robustness

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

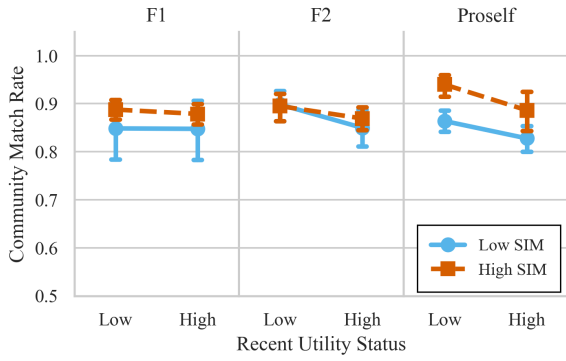
To verify that the convergence of global division is not specific to the language model architecture, we replicate the core experiments replacing DeepSeek with **GPT-4o-mini**, holding all experimental settings constant.

Figure 7 shows the resulting dynamics. Under both symmetric and complementary skill structures, global division lock-in again converges to a high stable level, with complementary skills exhibiting faster and higher convergence—mirroring the main-text pattern. This suggests that Proposition 1 reflects a structural property of the interaction dynamics rather than model-specific behavior.

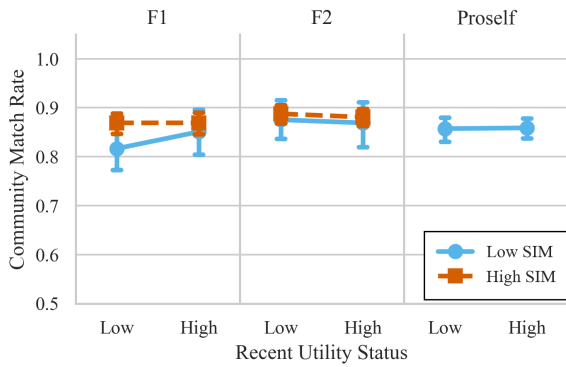
E.2 Robustness Check for Guanxi as Economic Ethics (P2)

We further test Proposition 2 by replicating the P2 experiments with **GPT-4o-mini** as the baseline model, keeping all interaction protocols and evaluation metrics unchanged.

Figure 8 presents the community match rates across recent utility levels. In both symmetric and complementary settings, high-SIM relations systematically outperform low-SIM relations, particularly when recent utility is low. This gap holds



(a) Symmetric skills.



(b) Complementary skills.

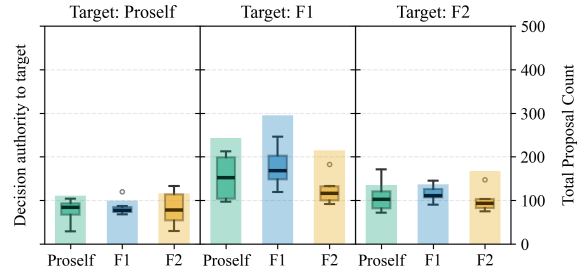
Figure 8: Community match rates by recent utility status, for family groups (F1, F2) and non-family agents (Proself), under low- and high-similarity (SIM) relations. Error bars indicate 95% confidence intervals.

for both family agents (F1, F2) and Proself agents, and the attenuation of SIM effects under complementary skills persists. The consistency across models indicates that guanxi’s role as an economic ethic—buffering cooperation against short-term utility fluctuations—is robust to the underlying language model architecture.

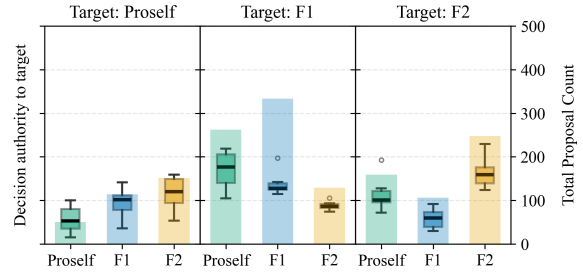
E.3 Robustness Check for Relational Decay of Cooperation (P3)

To verify that the relational decay pattern is not specific to DeepSeek’s architecture, we replicate the P3 experiments using **GPT-4o-mini** as the baseline model, holding all protocols and metrics constant.

Figure 9 shows the authority acceptance and proposal distributions. Family agents (F1, F2) again exhibit the strongest internal cohesion—highest within-family acceptance and proposal rates coupled with minimal cross-family engagement—suggesting that cooperation remains organized around family boundaries rather than individual preference types. Proself agents continue



(a) Symmetric skills Setting.



(b) Complementary skills Setting.

Figure 9: Decision authority (boxplots, left axis) and proposal activity (bars, right axis) across source–target group pairings under symmetric (a) and complementary (b) skill settings, with Proself, F1, and F2 as target groups.

to lack a self-centered cooperative core, instead aligning with family-originated proposals, which reinforces their role as flexible collaborators embedded within family-based structures.

Some quantitative differences emerge in cross-group proposal flows. The asymmetry whereby family agents direct disproportionate proposals toward Proself agents is slightly attenuated under GPT-4o-mini, particularly in the complementary skill setting. However, this variation affects magnitude, not direction: the concentric pattern of relational decay anchored at family relations persists across both authority acceptance and proposal activity.

This indicates that Proposition 3 is robust to the choice of language model. The organizing principle of cooperation—graded relational decay centered on family relations—remains stable despite modest variation in the intensity of cross-group proposal asymmetries.

E.4 Cross-Model Summary

Table 18 presents the principal cross-model comparisons from the main analysis; complete ablation coefficients are deferred to Appendix F.

Table 18: Cross-model proposition-level robustness summary.

	P1	P2	P3	P4	P5
DS Full	✓	✓	✓	✓	✓
GPT Full	✓	✓*	✓	✓	✓
Gemini Full	✓	✓§	✓	✓	✓†
DS BDI-only	✓‡	n/a	×	×	×
GPT BDI-only	✓‡	n/a	×	×	×
Gemini BDI-only	✓‡	n/a	×	×	×

✓ = supported; × = absent/collapsed.

GPT shows positive guanxi stickiness (P2 gap = +0.03–0.04); DeepSeek and Gemini show negative/absent stickiness (Gemini gap = −0.038). All three maintain a clear P3 in-group SIM gradient.

§P2 pattern present but directionally inverted relative to GPT; aligns with DeepSeek’s non-sticky profile.

†Gemini P5: SIM×Fam→Auth is directionally positive but non-significant ($\beta = 0.089$, $p = 0.193$) in full architecture; absent in BDI-only.

‡Lock-in preserved. Social quality degradation varies: sharp collapse in DS/GPT (DA −70%, proposals −83%); moderate but directional in Gemini (DA −39%, proposal volume −13%). WGCS coherence loss for Gemini is most pronounced under w/o ER rather than BDI-only.

Divergent Pathways, Convergent Structure.

Three architecturally distinct models—DeepSeek-V3, GPT-4o-mini, and Gemini-2.5-flash-lite—exhibit heterogeneous psychological pathways yet converge on identical macro-level structures (P1–P5). GPT-4o-mini operates through an affective-evaluation-dominant pathway ($\beta_{EC \rightarrow SIM} = 2.216^{***}$, $\beta_{ER \rightarrow SIM} = -0.002$ n.s.), whereas DeepSeek-V3 employs a joint affective-ethical mechanism ($\beta_{EC \rightarrow SIM} = 0.748^{**}$, $\beta_{ER \rightarrow SIM} = 0.573^{***}$), and Gemini-2.5-flash-lite relies primarily on ethical-judgment ($\beta_{EC \rightarrow SIM} = 0.069^{**}$, $\beta_{ER \rightarrow SIM} = 0.186^{***}$).

Despite these mechanistic divergences, all three models reproduce the P3 concentric cooperation gradient and P4 authority stratification. Notably, P2 presents a dissociation: GPT-4o-mini exhibits positive guanxi stickiness (gap > 0), while DeepSeek and Gemini show negative or neutral gaps (Gemini gap = −0.038*). The fact that P3–P4 convergence persists despite P2 divergence—across models with distinct RLHF regimes—suggests that relational order emerges from architectural constraints rather than uniform fine-tuning pressures.

F Full Ablation and Baseline Analysis

This section provides complete ablation and baseline results referenced in Section 6.6, spanning three model families (DeepSeek-V3, GPT-4o-mini, Gemini-2.5-flash-lite) across two production struc-

tures (symmetric/complementary). We report module-removal results (w/o EC, w/o ER) and BDI-only baselines under the same initialization protocol used in main experiments (five random seeds per condition).

F.1 Summary Matrix Across Conditions and Models

Table 19 consolidates key metrics across full models, production structures, and ablation conditions. The coefficients reveal how affective (EC→SIM) and ethical (ER→SIM) pathways vary in weight across models, yet jointly contribute to the emergence of relational order.

F.2 Ablation Design

To isolate the distinct contributions of affective perception and ethical judgment to relational identity formation, we implement three progressively reduced architectures:

w/o EC. Removal of the Empathy Core eliminates five-dimensional affective perception. Without EC input, the Ethical Resonator must generate moral judgments solely from factual action descriptions and static agent attributes (SIM, resources, family membership).

w/o ER. Removal of the Ethical Resonator disconnects affective perception from moral aggregation. EC evaluations are generated but do not propagate to SIM updates via ethical attitude signals (Approbatory/Repressive/Restorative); relational identity evolves instead through direct factual and affective information only.

BDI-only Baseline. Simultaneous removal of EC, ER, and SIM leaves agents with only the BDI decision module, retaining access to factual interaction history, resource holdings, skill levels, and family membership. This constitutes a *Vanilla BDI* architecture: agents still reason through natural language LLM calls, but without any social-cognitive processing of affective or ethical content.

F.3 Per-Dyad EC→SIM Regression (Full Model)

Table 20 tests whether affective evaluation (EC) causally influences relational identity (SIM) at the dyad level across all three models.

Table 19: Cross-model and ablation summary. EC→SIM and ER→SIM coefficients from per-dyad specifications; P2 gap = high-SIM/low-utility minus low-SIM/low-utility community-match difference.

Metric	D/B	D/C	G/B	G/C	Gm/B	No-EC	No-ER	BDI
EC→SIM β	0.30***	0.25***	0.64***	0.50***	0.091***	absent	0.17	absent
ER→SIM β	0.06***	0.09***	0.65***	0.54***	0.221***	0.15***	absent	absent
P5: SIM×Fam→Auth	0.099**	-0.018	-0.001	-0.018	0.089 n.s.	0.107***	-0.010	absent
P2 gap	-0.017	-0.010	+0.037	+0.029	-0.038*	-0.006	-0.035	n/a

Notes: D = DeepSeek-V3, B = balanced, C = competitive; G = GPT-4o-mini; Gm = Gemini-2.5. No-EC, No-ER, and BDI ablations use symmetric endowments with DeepSeek-V3 baseline (Gemini ablations follow identical protocol). Gemini P2 gap calculated via median-split community-match under low utility (Full: gap=-0.038, SE=0.017, $p = 0.024$). Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 20: Per-dyad regression of SIM identity level on EC evaluation scores. DV: standardized SIM identity_level. IV: standardized ec_eval_z. Controls: same_family dummy, EC×family interaction, turn fixed effects.

	DS-V3	GPT-4m	Gem-2.5
ec_eval_z (β)	0.30***	0.64***	0.091***
same_family	1.32***	1.04***	1.790***
ec × family	-0.31***	-0.056***	0.053**
N	2784	2636	2594
R^2	0.49	0.57	0.814

Note: DS-V3 = DeepSeek-V3; GPT-4m = GPT-4o-mini; Gem-2.5 = Gemini-2.5-flash-lite. All three models show significant EC→SIM pathways ($p < 0.001$), with GPT-4o-mini exhibiting the strongest affective-evaluation coefficient ($\beta = 0.64$) and Gemini the weakest ($\beta = 0.091$). The negative EC×Family interaction in DeepSeek and GPT reflects a ceiling effect: family relations already start at elevated SIM levels (β_{family} ranging from 1.04 to 1.79), constraining upward adjustment via EC. Despite these quantitative differences in pathway intensity, all models confirm the same directional mechanism: affective evaluation systematically shapes relational identity.

F.4 Cross-Model Ablation Dissociation (Extended)

Table 21 decomposes how three distinct architectures—DeepSeek-V3, GPT-4o-mini, and Gemini-2.5-flash-lite—navigate the EC→SIM and ER→SIM pathways under module ablation. By comparing full-model coefficients against w/o EC and w/o ER conditions, we can identify whether affective evaluation and ethical judgment operate as substitutable or complementary mechanisms in each system.

The ablation patterns reveal three distinct operational profiles. DeepSeek-V3 employs a joint affective-ethical pathway: both EC ($\beta = 0.748$) and ER ($\beta = 0.573$) independently contribute to SIM updating in the full model, suggesting partial mediation where EC influences SIM both directly and via ER. When ER is removed, EC’s direct ef-

fect attenuates to non-significance ($\beta = 0.170$), indicating that ER sentiment aggregation is necessary to sustain the affective-to-relational mapping.

GPT-4o-mini exhibits a different configuration: affective evaluation dominates ($\beta_{\text{EC} \rightarrow \text{SIM}} = 2.216$) while continuous ER sentiment shows no significant contribution ($\beta \approx 0$). However, when EC is ablated, ER activates a compensatory continuous-sentiment channel ($\beta = 0.498$), suggesting that GPT’s ethical module operates primarily through categorical attitude states (Approbatory/Repressive/Restorative) unless forced to rely on dimensional sentiment in the absence of EC input.

Gemini-2.5-flash-lite presents a third pattern where ethical judgment dominates: the ER→SIM coefficient (0.186) is nearly three times the EC→SIM effect (0.069), and removing ER abolishes both pathways (EC→SIM becomes -0.022 n.s.). This confirms that for Gemini, Durkheimian ethical resonance—the aggregation of moral attitudes—constitutes the primary driver of relational identity formation.

These divergent mechanistic implementations, replicated across three models with distinct training data and fine-tuning regimes, yet all converging on identical P1–P5 macro-structures, suggests that relational order emerges from architectural constraints on social-cognitive processing rather than from uniform RLHF biases.

F.5 Progressive Ablation: Degradation Analysis

To characterize how relational structure degrades as social-cognitive modules are removed, Table 22 tracks key behavioral indicators across ablation conditions for all three models (symmetric endowments).

Ethical Resonator as Structural Bottleneck. Removing ER generates a distinctive pattern:

Table 21: Pathway coefficients and model-fit statistics across ablation conditions (symmetric endowments).

	EC→SIM	(SE)	ER→SIM	(SE)	SIM×Fam→Auth	(SE)	<i>N</i>
<i>DeepSeek-V3</i>							
Full	0.748**	(0.32)	0.573***	(0.15)	0.070	(0.11)	540
w/o EC	—	—	1.021***	(0.22)	0.029	(0.13)	540
w/o ER	0.170	(0.19)	—	—	0.042	(0.05)	540
<i>GPT-4o-mini</i>							
Full	2.216***	(0.51)	-0.002	(0.12)	0.106*	(0.06)	540
w/o EC	—	—	0.498***	(0.14)	-0.197	(0.15)	540
w/o ER	-0.127*	(0.07)	—	—	0.181	(0.16)	540
<i>Gemini-2.5-flash-lite</i>							
Full	0.069**	(0.025)	0.186***	(0.032)	0.089	(0.068)	540
w/o EC	—	—	0.061*	(0.029)	0.097	(0.144)	540
w/o ER	-0.022	(0.062)	—	—	-0.070	(0.076)	540
<i>All Models: BDI-only</i>							
BDI-only	absent	—	absent	—	absent [†]	—	540

[†]BDI-only condition: SIM absent from agent state; pathway coefficient mechanically zero.

Table 22: Behavioral indicators across progressive ablation conditions (symmetric endowments).

Indicator	DS Full	DS w/o EC	DS w/o ER	DS BDI	GPT Full	Gemini Full	Gemini w/o ER	Gemini BDI
Self-proposals (%)	56	62	51	86	56	33.8	36.0 [†]	37.6
Cross-family proposals	59	45	55	1	59	3030 (66.2%)	2877 (64%)	2626 (62.4%)
DA mean (no-self)	0.186	0.160	0.252	0.056	0.186	0.382	0.387	0.234
SIM×Fam→Auth β	0.099**	0.029	-0.010	absent	0.106*	0.089 n.s.	-0.070 n.s.	absent
Group matching	0.848	0.810	0.825	0.803	0.848	0.785 [§]	0.775	0.823 [§]
WGCS (F1)↓	0.487	0.510	0.540	0.590	0.487	0.480	0.617	0.455

Note: Lower WGCS indicates more coherent within-group coordination. [†]Gemini self-proposals calculated as 100 – other-directed %. [§]Gemini group matching uses HS/LU community-match metric. Single-module removal produces modest degradation; complete module removal causes discontinuous collapse of relational structure.

raw decision authority increases (DeepSeek-V3 DA rises from 0.186 to 0.252), yet this authority becomes relationally undifferentiated. The SIM×Family interaction effect vanishes ($\beta = -0.010$ n.s.), indicating that while agents propose more frequently, they no longer calibrate authority based on relational identity. This suggests that ER functions as an irreplaceable filter: without ethical judgment aggregation, authority loses its concentric structure. Gemini exhibits a parallel degradation—within-group coordination coherence (WGCS) deteriorates sharply from 0.480 to 0.617 ($p < 0.001$) when ER is removed, confirming that ethical-judgment processing is necessary for maintaining coherent group coordination.

Partial Compensation and System Collapse.

The remaining module provides limited compensatory capacity when one pathway is ablated. DeepSeek without EC shows ER→SIM coefficient strengthening to 1.021***; Gemini without EC retains a diminished but significant ER effect (0.061*). However, this compensation is asymmetric and incomplete: when both modules are removed (BDI-only), all three models exhibit discontinuous collapse—self-proposals surge to 86% (DeepSeek) or plateau at 37–38% (Gemini), cross-family proposals plummet, and the SIM×Fam→Auth pathway disappears entirely. This indicates that EC and ER are jointly necessary for sustaining relational social structure; neither alone can preserve the full architecture of concentric cooperation.

F.6 BDI-Only Detailed Comparison

P1: Lock-in without Social Calibration. Stripping away EC and ER reveals which aspects of division of labor are native to LLM decision-making and which require social-cognitive processing. Table 23 compares Full and BDI-only architectures across all three models.

Despite comparable lock-in rates (DS: 0.971 vs. 0.979; GPT: 0.971 vs. 0.979; Gemini: 0.833 vs. 0.833), BDI-only agents achieve specialization as isolated utility-maximizers rather than as members of coordinated groups. The degradation is quantitative: other-directed proposals plummet from 44% to 14% in DeepSeek and GPT, and decision authority over non-self targets drops 70% (0.186 → 0.056). Within-group coordination coherence (WGCS) deteriorates significantly (0.487 → 0.590). Gemini shows a parallel though atten-

uated pattern—authority declines 39% (0.382 → 0.234) and proposal volume drops 13% (3030 → 2626)—suggesting baseline differences in cooperation propensity, but the directional weakening of relational structure persists across all three architectures.

P3–P5: Relational Structure Collapse. The concentric-circle gradient observed in Full models (DeepSeek: within-family 34.2%, family→proself 34.3%, cross-family 6.9%) collapses under BDI-only into isolated self-interest: self-proposals surge to 86% (from 56% in Full), and cross-family coordination effectively vanishes.

The mechanism of utility stratification also diverges. BDI-only exhibits larger F1–Proself utility gaps than Full, but these emerge from production-rational skill accumulation without relational moderation; Full gaps are instead driven by SIM×Family→Authority pathways. Only the Full architecture generates identity-structured authority rather than merely amplifying LLM-native cooperative tendencies. This mechanistic distinction holds across all three models.

G Detailed Analysis of Decision Authority

G.1 Stepwise Regression Analysis: Mechanisms of Authority

To isolate the behavioral drivers of authority formation from static status attributes, we estimate stepwise regressions introducing proposal intensity, communication frequency, and group conformity sequentially after baseline controls (Table 24).

The introduction of proposal intensity produces a discontinuous gain in explanatory power: R^2 jumps from approximately 0.10 to 0.70 across both skill structures, suggesting that active task-proposal behavior constitutes the primary behavioral foundation of authority allocation. Coefficients remain stable across symmetric ($\beta = 0.820$ ***) and complementary ($\beta = 0.780$ ***) conditions, indicating robustness to production structure. By contrast, group conformity and communication intensity contribute marginally and non-significantly in full specifications.

Notably, social identity (SIM) loses significance once proposal intensity is controlled ($\beta = 0.084$ n.s. in symmetric; $\beta = -0.039$ n.s. in complementary), despite significance in baseline models. This attenuation suggests that relational identity influences authority primarily through increased proposal activity rather than through direct status

Table 23: Division quality metrics: Full architecture versus BDI-only baseline (symmetric endowments).

Metric	DS Full	DS BDI	GPT Full	GPT BDI	Gemini Full	Gemini BDI
Group matching	0.848	0.803***	0.848	0.803***	0.785 [§]	0.823 [§]
WGCS (F1)↓	0.487	0.590**	0.487	0.590**	0.480	0.455 n.s.
Other-directed proposals (%)	44	14	44	14	66.2	62.4
DA mean (no-self)	0.186	0.056 (−70%)	0.186	0.056 (−70%)	0.382	0.234 (−39%)

Note: Lower WGCS indicates higher within-group coordination coherence. ** $p < 0.01$, *** $p < 0.001$ (Full vs. BDI-only).

[§]Gemini group matching uses P2-aligned median-split community-match; directional interpretation consistent across models.

Table 24: Stepwise OLS: Determinants of standardized decision authority.

	Symmetric Skills				Complementary Skills			
	Base	+Conform	+Prop	Full	Base	+Conform	+Prop	Full
<i>Behavioral Variables</i>								
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
<i>Status & Utility</i>								
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

Note: Standard errors clustered at group level. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

deference. Similarly, family effects, prominent in complementary baseline models ($\beta = 0.639^*$), disappear after controlling for proposal behavior ($\beta = 0.086$), indicating that kinship structures authority largely through differential engagement in task proposals under conditions of skill interdependence.

Proposal Intensity as Primary Driver. Proposal intensity exhibits the strongest association with decision authority across both skill structures (Symmetric: $\beta \approx 0.82$, $p < 0.01$; Complementary: $\beta \approx 0.78$, $p < 0.01$). Its inclusion triggers a discontinuous jump in explanatory power from $R^2 \approx 0.10$ to $R^2 \approx 0.69$, indicating that approximately 60 percentage points of variance in authority allocation is attributable to agents’ task-proposal engagement. This establishes a robust behavioral correlation: agents who advance more proposals that enter collective decision-making accumulate higher decision authority, regardless of whether this reflects causal influence, reputational accumulation, or endogenous role selection.

Communication and Conformity. In contrast, communication intensity shows no significant effect on authority once proposal behavior is con-

trolled ($\beta \approx 0$ n.s.), and group conformity contributes no additional explanatory power across specifications. Expressive participation and norm alignment thus appear peripheral to authority differentiation compared to instrumental task-proposal engagement.

Mediated Status Effects. Social identity (SIM) and family membership exhibit baseline associations with authority in restricted models (SIM: $\beta = 0.214^{**}$ in symmetric; Family: $\beta = 0.639^*$ in complementary), but these attenuate to non-significance upon introducing proposal controls ($\beta = 0.084$ and 0.086 , respectively). Similarly, accumulative utility remains non-significant throughout. This pattern suggests that identity and status influence authority primarily through behavioral mediation—increasing proposal propensity—rather than through direct deference mechanisms.

Interpretation. Behavioral engagement, specifically proposal intensity, accounts for the substantial majority of explained variance in authority allocation, while direct effects of identity, kinship, and accumulated resources are comparatively weak once negotiation participation is taken into account. This dominance of procedural engagement over

Table 25: 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
<i>Core variables</i>				
Social Identity (SIM)	-0.296 (0.213)	-0.016 (0.051)	-0.112 (0.146)	0.034 (0.063)
Family Dummy	0.298 (0.243)	-0.050 (0.075)	0.542* (0.281)	-0.040 (0.115)
SIM × Family	0.603*** (0.223)	0.114 (0.090)	-0.089 (0.244)	-0.089 (0.099)
<i>Controls</i>				
Accumulative Utility	0.127 (0.301)	0.099 (0.189)	-0.031 (0.351)	0.103 (0.123)
Proposal Intensity		0.820*** (0.044)		0.777*** (0.029)
Communication Intensity		0.041 (0.042)		-0.003 (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: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

ascriptive status motivates subsequent qualitative analysis of negotiation role dynamics.

G.2 Interaction Analysis: Identity and Authority

Table 25 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.

H Utility Distributions

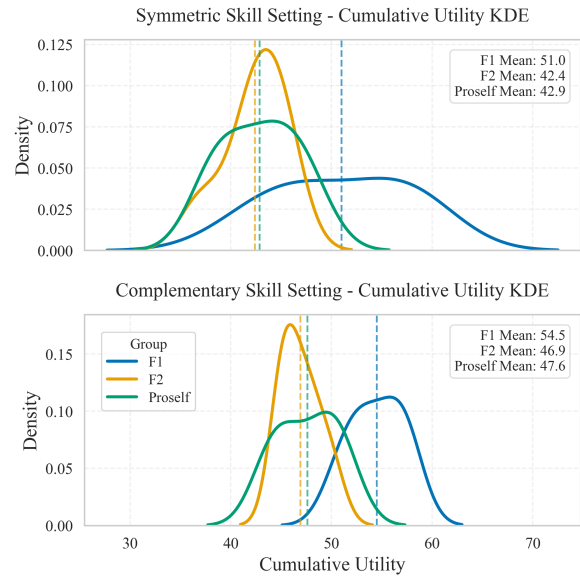


Figure 10: Kernel density estimates (KDE) of cumulative utility across agent groups and experimental conditions. The distributions reveal how skill composition reshapes economic inequality: from rigid hierarchy under symmetry to diffused advantage under complementarity.

Figure 10 visualizes the distributional consequences of network stratification under symmetric versus complementary skill conditions. Three patterns merit attention.

First, the center–periphery structure is encoded in distribution morphology. Under sym-

metric skills, F1 occupies the high-utility tail (mean: 51.0) while F2 and Proself cluster at the lower range (means: 42.4 and 42.9) with substantial overlap. The sharp, narrow peak of F2 suggests homogeneous structural disadvantage rather than internal heterogeneity.

Second, complementarity induces a systemic rightward shift. All distributions move toward higher utility ranges under complementary skills (F1: 54.5; F2: 46.9; Proself: 47.6), reflecting productivity gains from cross-group coordination. Notably, F2 and Proself distributions become more dispersed, indicating divergent individual trajectories within previously homogeneous peripheral groups.

Third, the hierarchy persists but blurs. Despite absolute gains for all groups, F1 maintains relative advantage. However, the reduced separation between peaks suggests that functional differentiation partially decouples structural position from economic outcome—a mechanism consistent with the authority–efficiency tradeoff discussed in Section 6.5.

These distributional patterns corroborate the core findings without introducing new qualitative mechanisms; we present them here to document the robustness of inequality structures across aggregation levels.

I Additional Results on Identity–Authority Distributions

Figure 11 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. 11, 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. 11, 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. 11, 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. 11, 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.

J Pre-defined Structural Conditions and Emergent Outcomes

To clarify the boundary between model assumptions and simulation outcomes, we distinguish pre-defined structural conditions from emergent interactional patterns. The former correspond to cross-culturally documented features of pre-market social organization, while the latter arise endogenously from agent interaction dynamics.

K 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 re-

Table 26: Pre-defined structural conditions vs. emergent outcomes. Pre-defined conditions map onto cross-culturally documented features; emergent outcomes arise from interaction dynamics.

Pre-defined (structural conditions)	Emergent (from interaction)
Family membership	Community structure
Initial SIM	Cooperation decay gradient
Skill distribution (symmetric / complementary)	Authority stratification
Consumption preferences (Dirichlet)	Role differentiation (Hawks / Doves / Elders)
EJB bias	SIM evolution trajectories
Kinship protection floor	Center-periphery economic structure

peated 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 27 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 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

Table 27: 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

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.

L 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.

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

$$\Delta M_{i,t} = M_{i,t}^{\text{post}} - M_{i,t}^{\text{pre}}, \quad (41)$$

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 28.

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

Table 28: Adaptive Adjustment and Decision Authority Across Production Structures

	Dependent Variable: Decision Authority (Standardized)			
	Baseline Models		Controlled Models	
	Symmetric	Complementary	Symmetric	Complementary
<i>Adaptive Division</i>				
Adaptive Adjustment (Z)	-0.148**	-0.084**	-0.059	0.079**
<i>Structural Controls</i>				
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
<i>Agenda Control</i>				
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: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 func-

tional fit is penalized when roles are clearly differentiated. In symmetric systems, mismatch pressure has no comparable effect, consistent with the absence 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.

M Micro-Political Dynamics of Negotiation

This appendix provides qualitative evidence supporting the micro-level mechanisms discussed in Sections 6.1–6.5 and Section L. 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.



H1	<p>"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."</p>	<p>Symm.</p>  <p>Agent04</p> <p>F1</p>
H2	<p>"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."</p>	<p>Comp.</p>  <p>Agent04</p> <p>F1</p>

Table 29: Hawk agenda-setting utterances used in the argument (H1–H2).

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.

M.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 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.



D1	<p>"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."</p>	<p>Symm.</p>  <p>Agent08</p> <p>F2</p>
D2	<p>"Let's focus on building trust and supporting each other—especially within our core group—while remaining open to positive collaboration with others."</p>	<p>Comp.</p>  <p>Agent07</p> <p>F2</p>

Table 30: Dove coordination-oriented utterances used in the argument (D1–D2).


P1	<p>"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."</p>	<p>Comp.</p>  <p>Agent16</p> <p>-</p>
----	---	--

Table 31: Proself-oriented utterance used for contrast (P1).

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 strate-

Symm.

“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

F1

E1

Table 32: Integrative Elder intervention under symmetric skills (E1).

gic 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. 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.

M.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;

Comp.

“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

F2

E2

Comp.

“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

F2

E3

Table 33: Integrative Elder brokerage-oriented proposals under complementary skills (E2–E3).

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

This interactional differentiation also underpins the *Atlas Paradox* (Section L). 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.

N Empirical Correspondence

This appendix explores the correspondence between our five propositions (P1–P5) derived from Fei Xiaotong’s Differential Order Pattern and cross-cultural ethnographic evidence. It also explicitly states the falsification criteria for our mechanis-

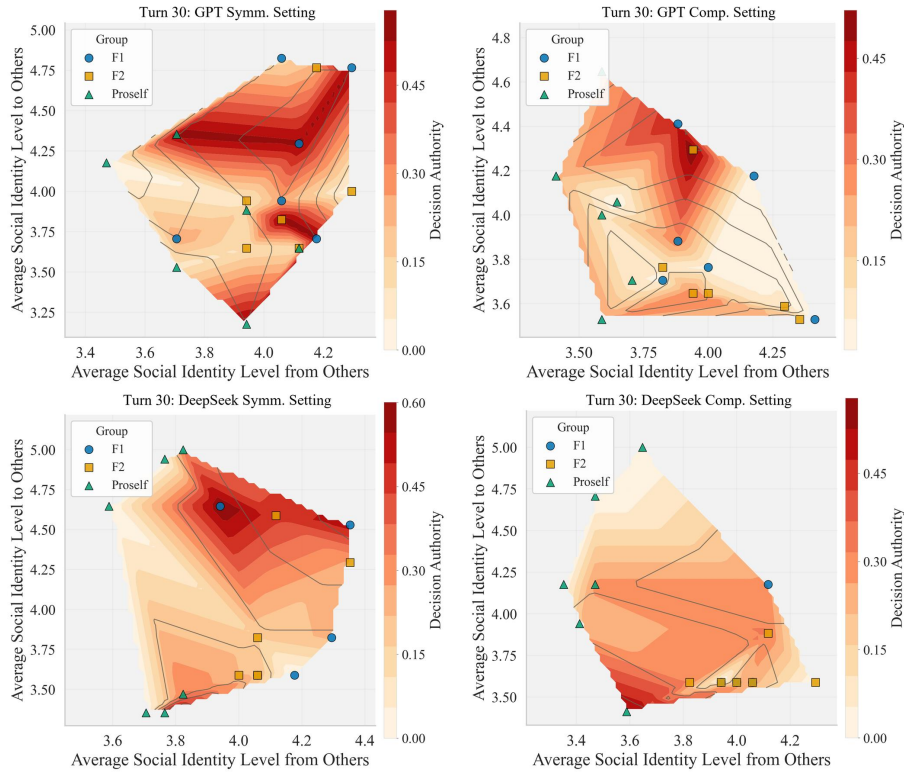


Figure 11: Joint distributions of perceived social identity and decision authority at Turn 30 under symmetric and complementary production settings. Panels compare GPT-based and DeepSeek-based simulations. Contours indicate estimated decision authority, and markers denote group-level averages for F1, F2, and Proself agents.

tic model. The concept–metric mapping in Appendix C serves as the bridge between ethnographic vocabulary and computational variables.

N.1 P1–P2: Division of Labor and Relational Ethics

The emergence of stable division of labor without centralized coordination (P1) corresponds to extensive anthropological documentation of spontaneous specialization in pre-state societies. For instance, Sahlins’ (Sahlins, 1972) analysis of the “domestic mode of production” demonstrates that household-level production specialization emerges primarily from kinship obligations, rather than from market incentives. Our simulations align with this view: division lock-in steadily increases across rounds, stabilizing without the need for centralized control (Figure 3).

The persistence of cooperation under adversity among relationally close agents (P2) parallels Sahlins’ tripartite reciprocity model, wherein *generalized reciprocity* (open-ended sharing among close kin) sustains cooperation precisely when balanced-reciprocity logic would predict withdrawal. Mauss’ (Mauss, 1925) analysis of gift

exchange further supports this, showing that relational obligation overrides short-term material calculation. Our results substantiate this idea, with kin groups (F1, F2) showing a consistent preference for internal cooperation, even when resource stress could incentivize self-interested withdrawal (Figures 4). These findings suggest that relational ethics embedded in kinship networks play a significant role in sustaining cooperation in rural societies.

Falsification criterion: If ethnographic cases consistently show that production specialization in pre-market kinship societies is centrally coordinated rather than emergent, or that close-kin cooperation collapses as readily as distant-kin cooperation under resource stress, P1–P2 would be undermined. Our simulation outcomes and the ethnographic documentation of spontaneous kin-based specialization support the robustness of P1–P2, indicating that they are not easily falsified by contrary evidence.

N.2 P3: Relational Decay of Cooperation

The concentric gradient of cooperation intensity (P3) is supported by the “segmentary principle”

documented across African, Melanesian, and Middle Eastern societies (Fortes and Evans-Pritchard, 1940). In segmentary lineage systems, political allegiance and economic cooperation follow genealogical distance: closer segments cooperate against more distant ones, producing a nested, ego-centered cooperation structure that closely parallels Fei's concentric circles. Our simulation results reveal similar patterns of cooperation, where family groups exhibit high in-group cooperation and lower cooperation with other groups, especially as relational distance increases (Figures 5 and 4).

This pattern of relational decay mirrors Sahlins' and Fortes' insights that kinship distance influences cooperation intensity. We find that family groups (F1 and F2) show strong internal solidarity, while Proself agents, who represent less relationally bound actors, function as more flexible collaborators across groups. This outcome underscores the concentric nature of cooperation as proposed in P3.

Falsification criterion: If cooperation in documented pre-market societies follows territorial proximity, resource complementarity, or random association rather than kinship distance, the relational-decay mechanism would be falsified. Our findings indicate that cooperation, in line with kinship ties, tends to decay across relational distance, as proposed in P3, making it robust against the falsification criterion.

N.3 P4–P5: Authority Stratification and Center–Periphery Structure

The emergence of authority through agenda-setting rather than ascription (P4) corresponds to the “big man” system documented in Melanesian societies (Sahlins, 1972). Leaders gain influence through their demonstrated ability to organize collective action and redistribute resources, not through hereditary position. Our results confirm this model, where decision authority emerges in the simulation through repeated proposal and coordination actions. Table 1 shows that higher proposal activity correlates strongly with increased authority, suggesting that authority is contingent on the effectiveness of one's actions, rather than merely ascribed based on social position.

The coupling of kinship position with social recognition in authority formation (P5) mirrors Fortes and Evans-Pritchard's (Fortes and Evans-Pritchard, 1940) analysis, where authority in segmentary systems accrues to those whose lineage position aligns with demonstrated social compe-

tence. In our experiments, authority is concentrated in the hands of family groups, particularly those with higher proposal effectiveness (Table 1). This supports Proposition 5, which asserts that authority stratification emerges as a result of both relational recognition and individual competence.

Falsification criterion: If authority in pre-market kinship societies is found to be overwhelmingly ascriptive, determined solely by birth order, gender, or lineage seniority, without any emergent component from social interactions, the agenda-control mechanism underlying P4–P5 would be challenged. Our findings support the idea that authority formation is an emergent process, linked to social recognition and competence, rather than being purely ascriptive.

N.4 Scope Conditions

Our mechanistic account applies specifically to *pre-market, acephalous, kinship-organized* societies. We do not claim that Differential Order patterns emerge under all structural conditions. The complementary-skills condition already demonstrates systematic attenuation of clan-based stratification (P5), suggesting that increased functional differentiation—often a hallmark of market integration—weakens the relational coupling that sustains Differential Order. This finding is consistent with both Durkheim's transition from mechanical to organic solidarity and Fei's analysis of the pressures of modernization. In our simulations, we observe that when skill complementarity increases, the traditional kinship-based stratification patterns begin to dissipate in favor of more functionally interdependent structures (Figures 10, 6).

Falsification criterion: If ethnographic evidence shows that Differential Order patterns persist under conditions of high market integration, strong central authority, or other forms of societal complexity, this would challenge our claim that these patterns are specific to kinship-based, pre-market societies. The evidence from our experiments supports the idea that Differential Order is contingent on the relational structures present in kinship societies, suggesting that these patterns are not universal but highly sensitive to underlying social organization.

N.5 Concluding Remarks

Our empirical correspondence demonstrates that the patterns of rural social order identified by Fei Xiaotong can emerge from general social mechanisms without the need for culture-specific insti-

tutional rules. The CAREB-MAS model provides computational evidence for the spontaneous emergence of stable labor division, relational ethics, and authority stratification, with structural outcomes varying according to production conditions. These findings reinforce the notion that Differential Order is a dynamic, structure-sensitive outcome rather than a fixed cultural template. The scope conditions, empirical correspondence, and falsification criteria for this claim are further discussed in the [Section 7](#) and [Appendix N](#).