The Ratchet Effect in Silico through Interaction-Driven Cumulative Intelligence in Large Language Models

Ren Zhuang (venomrko97@gmail.com, 2022112011050@stu.hznu.edu.cn)[1]
[1]School of Information Science and Technology, Hangzhou Normal University

Abstract

Human intelligence scales through cumulative cultural evolution (CCE), a ratchet process in which innovations are retained against entropic drift. Large language model training, by contrast, still depends primarily on static corpora and parameter growth, leaving little room for endogenous accumulation through interaction. We present POLIS (Population Orchestrated Learning and Inference Society), a framework in which heterogeneous agents generate solutions, verify one another's outputs, retain validated artifacts in shared cultural memory, and internalize them through parameter updates. On mathematical reasoning benchmarks, populations of 1–4B-parameter models achieved average gains of 8.8–18.9 points over base models and narrowed the gap to 70B+ monoliths. Mechanistic ablations identify peer verification as the main ratchet operator and show that internalization sustains accumulation across rounds, providing computational evidence that epistemic vigilance organizes durable knowledge growth. These results position structured social interaction as a scaling lever orthogonal to parameter count.

Keywords cumulative cultural evolution; ratchet effect; collective intelligence; epistemic vigilance; multi-agent learning

Introduction

Human cognition scales through cumulative cultural evolution (CCE), and collective competence progressively exceeds the innovative capacity of any solitary individual (Mesoudi & Thornton, 2018). The ratchet effect that enables this depends on sociocognitive mechanisms ensuring innovations are selectively retained and transmitted with sufficient fidelity to prevent regression (Dean et al., 2012; Tennie et al., 2009). Unlike other primates exhibiting transient social learning (Bandura, 1977; Whiten et al., 1999), humans stabilize knowledge against entropic drift, transforming ephemeral interactions into durable cultural artifacts (Boyd et al., 2011; Henrich, 2015). This scaling-by-organization is naturally framed as an emergent property of complex adaptive systems (Anderson, 1972; Lansing, 2003).

Modern LLM training, by contrast, remains largely individualistic. Capability scales through parameters and static data (Hoffmann et al., 2022; Kaplan et al., 2020), with alignment imported from outside the agent's ecology (Ouyang et al., 2022). Despite emergent reasoning abilities (Wei, Tay, et al., 2022), current models resemble isolated learners that exploit existing information but cannot sustain an autonomous ratchet effect, lacking the high-fidelity social transmission that CCE requires (Dean et al., 2012).

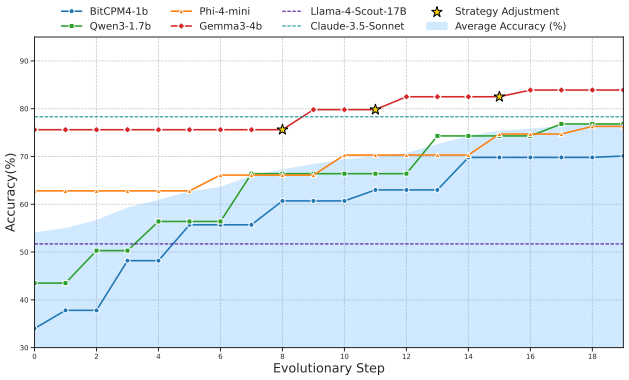

Figure 1: **Interaction supports cumulative gains.** POLIS improves over rounds, consistent with ratchet-like accumulation of verified strategies that reduces the gap to much larger monolithic models.

We introduce **POLIS** (Population Orchestrated Learning and Inference Society), a computational framework for artificial sociogenesis that maps CCE mechanisms to language model training (Tomasello, 1999). POLIS organizes a heterogeneous micro-society in which agents generate solutions independently, verify one another's outputs, retain validated artifacts in shared cultural memory, and internalize them through parameter updates. This perspective aligns with calls to build machines that learn and think with people rather than scaling isolated learners alone (Collins et al., 2024), and connects to traditions on collective intelligence (Leimeister, 2010). Our results show that structured social interaction can function as a scaling lever orthogonal to parameter count, as illustrated in Figure 1, where small interacting models close the performance gap to much larger monoliths over rounds of CCE.

Our key contributions are

- We formalize the ratchet cycle as an explicit loop over variation, social selection, and retention, implemented with modular operators for generation, verification, memory, and internalization.
- We introduce a high-precision social learning architecture in which unanimous peer verification and preference-based quality selection admit transmissible artifacts into shared cultural memory.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

- We show across four reasoning benchmarks that this closed loop yields cumulative gains for heterogeneous populations, with verification and internalization governing sustained accumulation.

Related Work

Cumulative Cultural Evolution and the Ratchet

CCE explains open-ended cognitive scaling as a variation, selection, and inheritance cycle (Mesoudi & Thornton, 2018; Tomasello, 1999) in which socially transmitted innovations are retained against loss (Dean et al., 2012). The cognitive-gadgets hypothesis frames many reasoning procedures as culturally acquired software rather than fixed priors (C. Heyes, 2019b). Computational models of CCE often track transmission among agents or populations, whereas POLIS closes the loop between external artifacts and internal parameter updates. It implements the ratchet as an explicit training cycle in which candidate solutions are externalized, selectively retained through verification, and internalized into agent parameters so that gains persist across episodes.

Epistemic Vigilance and Social Verification

Interactionist accounts emphasize that evaluation is often more reliable than production (Mercier & Sperber, 2011), motivating epistemic vigilance mechanisms that filter misinformation (Csibra & Gergely, 2009). Collective decision-making improves accuracy when aggregation preserves independence rather than amplifying correlated noise (Bahrami et al., 2010; Woolley et al., 2010). Existing multi-agent LLM systems often aggregate outputs through majority voting or mixture-of-experts (EBRAHIMI & ASUDEH, 2025). POLIS uses unanimous peer verification to decide what enters cultural memory, turning aggregation into a social selection process that prioritizes independent confirmation and suppresses correlated error.

Multi-Agent Learning and LLM Self-Improvement

LLM self-improvement methods such as STaR (Zelikman et al., 2022) and related approaches (Gao et al., 2025; C. Li et al., 2023; M. Li et al., 2023; Tong et al., 2024; Xu et al., 2024) improve a single agent by generating and learning from its own solutions, so the selection signal remains local to that agent. Multi-agent finetuning introduces interaction for cross-validation (Chen et al., 2024), and its effectiveness depends on preserving independence; homogeneous populations and low-precision aggregation can amplify correlated errors (Hong & Page, 2004; Lorenz et al., 2011). POLIS combines heterogeneous generation, peer verification, shared cultural memory, and parameter internalization in a single cross-round learning loop.

Modeling Artificial Sociogenesis

To realize a cumulative-cultural ratchet in artificial populations, a system must repeatedly generate novelty, select against error, and retain validated innovations with sufficient fidelity to prevent drift (Dean et al., 2012; Mesoudi & Thornton, 2018). We formulate POLIS as a computational model of artificial sociogenesis that instantiates this cycle through adaptive scaffolding, decentralized verification, and parametric internalization.

Notation. We denote the heterogeneous population as $\mathcal{P} = \{\pi_{\theta_i}\}_{i=1}^{N}$, parameterized by weights θ_i. The variation module comprises a stochastic generator $\mathcal{G}$, an adaptive scaffolding function $\mathcal{S}$, and an external anchor Ω. Selection is mediated by an epistemic filter $\mathcal{F}$ and a quality preference $\mathcal{Q}$. Retention relies on a cultural memory $\mathcal{M}$ and an internalization process updating θ. Where needed, we denote the internalization step size by λ_θ to avoid ambiguity with LoRA hyperparameters.

Variation Dynamics

Evolution requires continuous variation (Boyd & Richerson, 1988; Hong & Page, 2004). We model it through three coupled mechanisms for generation, adaptation, and constraint.

Stochastic Generator ($\mathcal{G}$). To prevent mode collapse, problems are synthesized via a stochastic generator $\mathcal{G}$. Let $z \sim \mathcal{N}(0, I)$ be a latent noise vector. The problem set Q_t is sampled conditioned on the current difficulty state d_t as follows.

$$q \sim \mathcal{G}(z \mid d_t) \quad \forall q \in Q_t. \tag{1}$$

Adaptive Scaffolding ($\mathcal{S}$). To maintain the population within the *zone of proximal development* (ZPD) (Van de Pol et al., 2010), the difficulty d_t adapts homeostatically, closely related to curriculum learning in machine learning (Bengio et al., 2009). Let $\bar{r}_t$ be the population average pass rate. Difficulty evolves as follows.

$$d'_{t+1} \leftarrow d_t + \eta \cdot (\bar{r}_t - r_{\text{target}}), \tag{2}$$

where η is the adaptation rate.

External Anchor (Ω). To prevent drift from external norms (Lorenz et al., 2011), we gate difficulty updates on a held-out generalization gap Δ_{gen} as follows.

$$d_{t+1} = \begin{cases} d_t - \delta & \text{if } \Delta_{\text{gen}} \geq \Omega \\ d_t + \eta(\bar{r}_t - r_{\text{target}}) & \text{otherwise} \end{cases} \tag{3}$$

Selection Dynamics

The ratchet effect requires high-fidelity transmission (Dean et al., 2012). We implement this via a two-stage filter.

Epistemic Filter ($\mathcal{F}$). First, we apply a decentralized truth-seeking mechanism. For a candidate solution a, a subset of peers K acts as verifiers. Drawing on the asymmetry of argumentation (Mercier & Sperber, 2011), we require strict consensus, defined as

$$\mathcal{F}(a) = \mathbb{I}\left[\sum_{k \in K} \mathbb{I}(\text{Verify}_k(q, a)) = |K|\right]. \tag{4}$$

Solutions failing this check ($\mathcal{F}(a) = 0$) are discarded. We set K to all peers except the author, which imposes a conservative criterion and increases precision for suppressing hallucinations.

Quality Preference (Q). Among verified solutions, evolution favors those that are communicable. Agents assign a quality score $Q(a)$ (e.g., clarity), operationalized through peer preference judgments aggregated into a stable ranking signal such as TrueSkill (Herbrich et al., 2006). The selected artifact a^* is optimized for transmission as follows.

$$a^* = \arg \max_{a \text{ s.t. } \mathcal{F}(a)=1} Q(a \mid q). \tag{5}$$

Retention Dynamics

Finally, selected traits must be preserved to enable accumulation.

Cultural Memory ($\mathcal{M}$). We maintain an externalized buffer $\mathcal{M}_t$ representing the elite history of the society. The memory is updated via a priority queue mechanism as follows.

$$\mathcal{M}_{t+1} \leftarrow \text{TopK}\left(\mathcal{M}_t \cup \{(q, a^*)\}, Q\right). \tag{6}$$

This buffer functions as an externalized scaffold (Kirsh, 1995).

Internalization (θ). The cycle closes when public artifacts are converted into private competencies. Consistent with the theory of cognitive gadgets (C. Heyes, 2019a), agents update parameters θ via LoRA as follows.

$$\theta_{t+1} \leftarrow \theta_t - \lambda_\theta \nabla_\theta \mathbb{E}_{(q,a)\sim\mathcal{M}_{t+1}}[\log \pi_\theta(a \mid q)]. \tag{7}$$

Artificial Sociogenesis in POLIS

Each interaction round follows a three-phase loop.

1. **Variation.** Generate challenges q near the population's current competence in a ZPD-style regime (Vygotsky, 1978) using a stochastic generator $\mathcal{G}$. Adapt difficulty with scaffolding $\mathcal{S}$ and gate updates with an external regulator Ω.
2. **Selection.** Agents propose solutions independently. Peers verify candidates via the epistemic filter $\mathcal{F}$ (Mercier & Sperber, 2011), and verified artifacts are ranked by quality preference Q.
3. **Retention.** Selected artifacts enter cultural memory $\mathcal{M}$ and are internalized into each agent's parameters θ via lightweight updates (LoRA) (Hu et al., 2022). This converts public solutions into private competence.

Experiments

We evaluated whether POLIS yields cumulative, ratchet-like improvements in mathematical reasoning and whether these gains depend on the specific sociocognitive operators defined in our model. Beyond final benchmark accuracy, we analyzed learning dynamics over interaction rounds and performed mechanistic ablations to identify which components are necessary to prevent drift.

Experimental Setup

Models. We form a heterogeneous population of four LLMs ranging from 1B to 4B parameters, including BitCPM4 (M. Team et al., 2025), Qwen3 (Yang et al., 2025), Phi-4 (Abouelenin et al., 2025), and Gemma3 (G. Team et al., 2025), Llama3.2 (Dubey et al., 2024).

Benchmarks. We evaluate on MATH500 (Hendrycks et al., 2021), GSM8K (Cobbe et al., 2021), GPQA Diamond (Rein et al., 2024), and AIME24 (MAA, 2024) using Chain-of-Thought (CoT) prompts (Wei, Wang, et al., 2022). We focus on mathematical reasoning as its clear ground truth enables unsupervised decentralized verification, and its multi-step structure renders cumulative strategy acquisition directly observable. This setting makes social filtering and cross-round retention directly measurable within the learning loop.

Baselines. We compare POLIS against isolated self-improvement methods, including STaR (Zelikman et al., 2022), SGDS (M. Li et al., 2023), MAGPIE (Xu et al., 2024), GRA (Gao et al., 2025), MuggleMath (C. Li et al., 2023), DART (Tong et al., 2024), along with monolithic systems such as GPT-4o (Hurst et al., 2024), Claude 3.5 Sonnet (Anthropic, 2024), Qwen2.5-72B (Q. Team, 2024), and Llama-4-Scout-17B (MAA, 2025).

Implementation. Internalization updates parameters via LoRA (r=16, α=32, dropout 0.05) (Hu et al., 2022) after 1024 verified artifacts accumulate. Adaptive scaffolding initializes at d_0=1.0 with update rate 0.05; generation uses temperature 0.6 and top-p 0.95.

Main Results

Social Organization as a Scaling Axis. Table 1 presents the primary comparative results across four benchmarks. POLIS consistently improves each base model and outperforms strong isolated self-improvement baselines. The baseline set spans local self-training, strategic data generation, and larger monolithic models, positioning POLIS within a broad map of established improvement regimes. Averaged across tasks, POLIS yields gains of 8.8–18.9 points over the base models and 2.9–6.5 points over the strongest isolated baseline among those tested. Interaction structure thus emerges as a scaling axis alongside parameter count. MAGPIE slightly outperforms POLIS on GPQA with smaller models; we attribute this to its emphasis on diverse instruction generation, which aligns well with the benchmark's graduate-level question variety. Nevertheless, POLIS surpasses MAGPIE on GPQA with larger models and achieves higher average accuracy across all benchmarks.

Verifying the ratchet effect. Figure 2a shows that population performance accumulates steadily over rounds, a hallmark of CCE (Dean et al., 2012). The curve exhibits a consistent

Table 1: **Social organization is a scaling axis.** POLIS consistently improves over both base models and isolated self-improvement baselines across benchmarks, supporting interaction structure as a complement to parameter scaling.

Method	MATH500	GSM8K	GPQA	AIME24	Avg.	MATH500	GSM8K	GPQA	AIME24	Avg.
Model	BitCPM4-1B					Qwen3-1.7B				
Baseline	34.0	60.8	28.5	3.0	31.6	43.5	68.5	32.0	3.8	37.0
STaR	40.2	70.3	29.1	3.1	35.7	51.3	76.8	32.8	4.0	41.2
SGDS	54.5	78.1	37.8	3.8	43.6	64.0	84.2	41.5	5.0	48.7
MAGPIE	39.8	72.4	**40.1**	3.0	38.8	49.5	78.0	43.2	3.9	43.7
GRA	51.0	76.5	32.5	4.2	41.1	61.8	82.5	36.5	5.5	46.6
MuggleMath	46.5	79.5	30.2	3.4	39.9	58.5	85.1	34.5	4.4	45.6
DART	57.1	81.2	34.5	**5.0**	44.5	66.2	87.4	38.0	6.1	49.4
POLIS	**70.1**	**85.5**	33.7	4.9	**47.4**	**76.8**	**90.3**	**49.1**	**7.5**	**55.9**
Model	Phi-4-mini					Gemma3-4B				
Baseline	62.8	75.0	38.5	5.5	45.5	75.6	82.5	42.0	7.0	51.8
STaR	66.5	80.5	39.0	5.6	47.9	77.2	85.0	42.4	7.1	52.9
SGDS	70.8	85.5	44.0	6.2	51.6	80.1	87.0	46.5	7.9	55.4
MAGPIE	65.5	82.1	45.5	5.5	49.7	76.5	84.5	48.1	7.0	54.0
GRA	70.1	84.0	41.0	6.5	50.4	79.5	86.8	44.0	8.0	54.6
MuggleMath	68.0	84.8	40.1	5.8	49.7	77.8	86.5	42.8	7.2	53.6
DART	71.5	86.5	42.5	7.0	51.9	80.5	88.2	45.0	8.5	55.6
POLIS	**76.3**	**89.2**	**47.4**	**8.1**	**55.3**	**83.9**	**91.8**	**56.5**	**10.2**	**60.6**

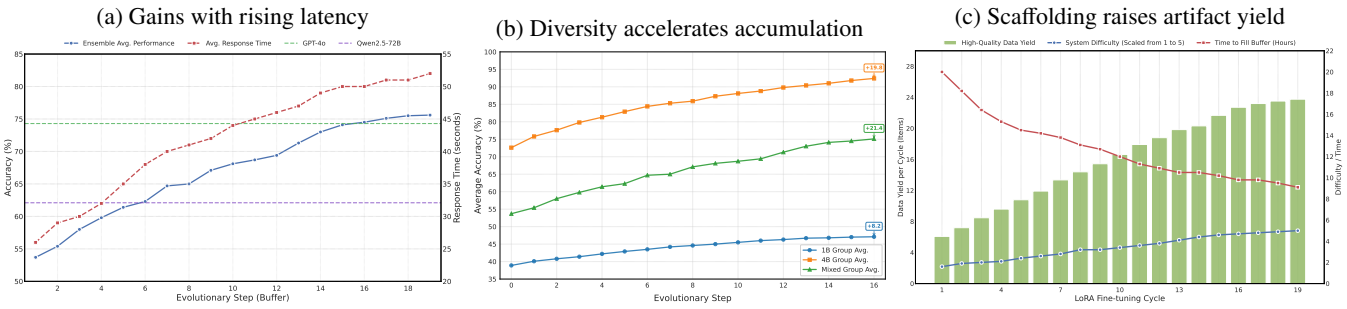

Figure 2: **Population dynamics and trade-offs.** POLIS exhibits ratchet-like accumulation across rounds, and this accumulation is accelerated by diversity and supported by adaptive scaffolding that increases the yield of verified training artifacts while improving efficiency.

upward trajectory rather than transient fluctuations, while average response time increases with difficulty as agents internalize longer strategies.

Cognitive Diversity and Collective Intelligence. We investigate whether gains stem from population structure rather than agent count alone. Figure 2b contrasts the Mixed Group against scale-homogeneous controls. The heterogeneous society improves faster and reaches a higher plateau, consistent with the diversity prediction theorem (Hong & Page, 2004) and with heterogeneity reducing correlated error in homogeneous settings (Lorenz et al., 2011). More broadly, these dynamics match classic accounts of collective intelligence as an organizational resource rather than a property of any single agent (Hutchins, 1995; Leimeister, 2010).

Efficiency dynamics. Figure 2c tracks system metabolism, defined as artifact yield relative to problem difficulty. Even as difficulty rises, the yield of verified artifacts increases, indicating that the population internalizes strategies rather than merely memorizing answers.

Analysis

Ablation Study. Cumulative evolution depends on selective retention of variants against noise. Ablation studies of our model's sociocognitive mechanisms, as illustrated in Table 2, show that removing the Epistemic Filter causes the largest performance drop, consistent with argumentative theory (Mercier & Sperber, 2011). The filter governs which artifacts enter cultural memory and therefore which reasoning traces are internalized across rounds. Ablating the Quality Preference also impairs performance, revealing that the Epistemic Filter enforces correctness while the Quality Preference enhances communicability. The system is more impaired by a lack of neural plasticity than by a lack of historical records, supporting the view that external artifacts serve as tempo-

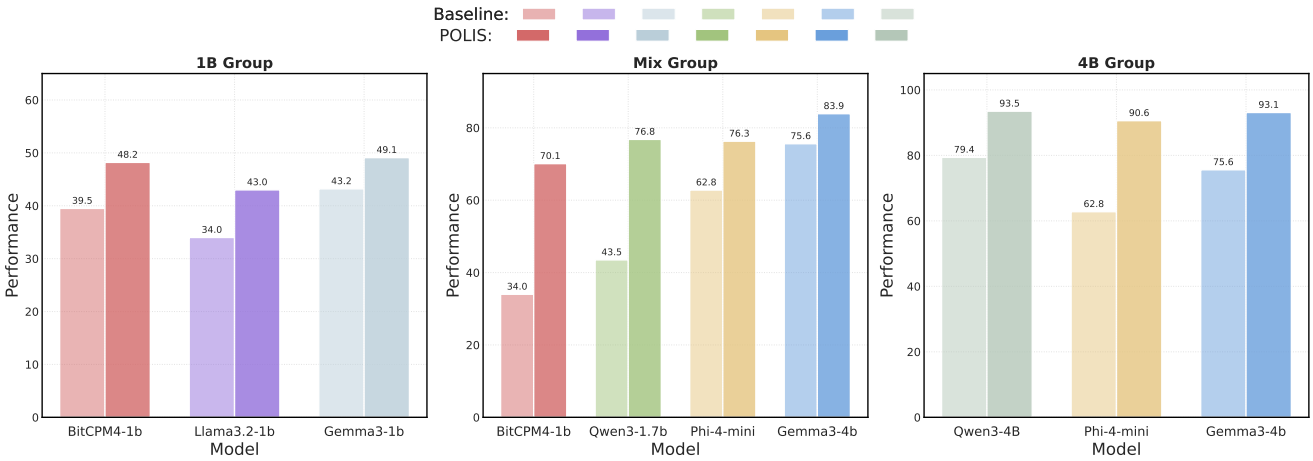

Figure 3: **Diversity redistributes gains without sacrificing strong models.** POLIS improves every member of the population, with the largest gains accruing to weaker models in heterogeneous societies while stronger models do not regress.

Table 2: **Verification is the main ratchet operator.** Removing the Epistemic Filter yields the largest drop in average accuracy, while internalization and quality selection also contribute to sustained accumulation.

Method Variant	Avg. Acc
POLIS	54.8
- Epistemic Filter $\mathcal{F}$	37.8
- Quality Preference $\mathcal{Q}$	47.8
- Internalization θ	45.5
- Cultural Memory $\mathcal{M}$	49.6
- Adaptive Scaffolding $\mathcal{S}$	53.6
- Stochastic Generator $\mathcal{G}$	52.8
- External Anchor Ω	53.3

rary scaffolds whose durable effect arises through parameter updates (C. Heyes, 2019b).

The Dividend of Cognitive Diversity. Figure 3 isolates the effect of population structure by comparing models trained in homogeneous groups versus the heterogeneous POLIS society. Diversity functions as a non-rivalrous resource: weaker agents benefit most, stronger agents retain performance, and the group preserves complementary strengths under shared verification. This pattern matches the diversity prediction theorem (Hong & Page, 2004) and aligns with distributed cognition accounts of collective performance (Hutchins, 1995; Lorenz et al., 2011).

Comparing Social and Solitary Generalization. Figure 4 contrasts the social trajectory against isolated learning across four distinct cognitive environments. Across all benchmarks, POLIS outperforms both the base models and the isolated-learning control. On AIME24, for example, BitCPM4-1b improves by 63% relative to baseline, whereas isolated learning yields only a modest gain. The gap reflects a structural property of the learning loop: peer verification governs which artifacts are retained and internalized, improving generalization across regimes.

Discussion

Interaction as a Scaling Law

POLIS shows that capability can scale through interaction structure as well as parameter count, consistent with distributed and situated accounts of cognition (Brown et al., 1989; Hollan et al., 2000; Hutchins, 1995). In our framework, decentralized verification establishes an endogenous epistemic norm by granting candidate artifacts knowledge status only after independent peer checking, aligning the production-evaluation asymmetry in interactionist theories of reasoning (Mercier & Sperber, 2011) with metacognitive monitoring (Flavell, 1979). Ablation results reveal a functional hierarchy among operators. Removing the Epistemic Filter causes the greatest performance decline, followed by internalization and Quality Preference. Variation supplies candidates; social filtering determines which ones persist (Tennie et al., 2009).

The Cultural Ratchet

The gains in POLIS depend on both selection and retention. Unanimous verification and quality-based preference preserve artifacts that are both correct and communicable, while internalization converts these public artifacts into private competence, aligning with the idea that many reasoning procedures function as culturally learned cognitive gadgets (C. Heyes, 2019b) and with the broader cultural intelligence hypothesis (Herrmann et al., 2007; Tomasello et al., 2005). This sequence of externalization and internalization reflects how social interaction scaffolds individual development (Vygotsky, 1978). In POLIS, the cultural memory $\mathcal{M}$ serves as a zone of proximal development by providing exemplars just beyond each agent's current competence. The observation that weaker

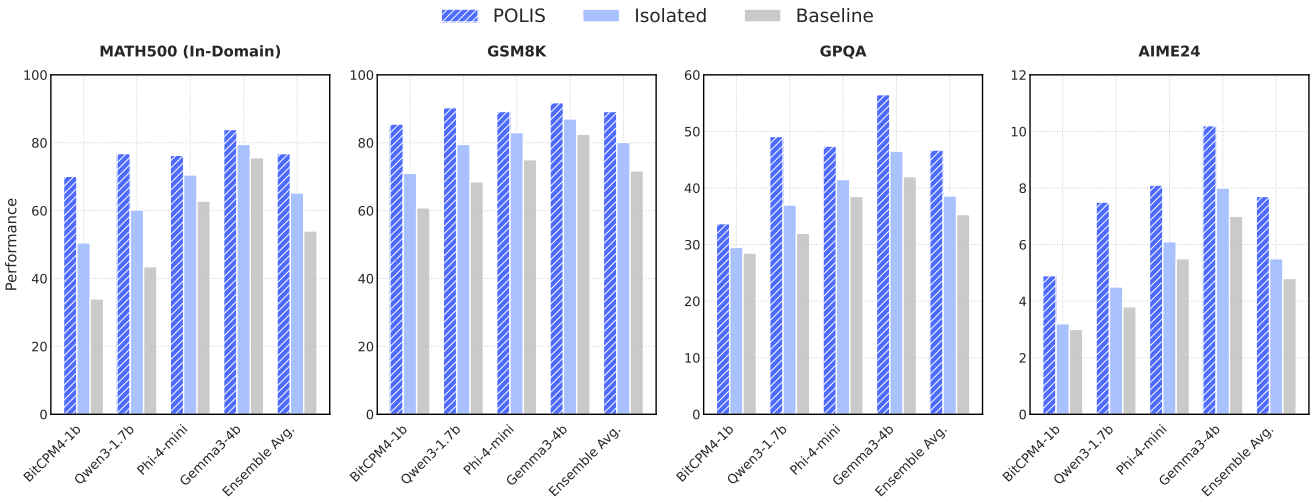

Figure 4: **Interaction yields robust gains beyond solitary learning.** POLIS improves over both base models and an isolated-learning control, consistent with peer verification reducing errors that escape single-agent verification horizons.

models benefit disproportionately in heterogeneous societies, as shown in Figure 3, supports this scaffolding interpretation.

Implications for Human CCE

POLIS offers a computational testbed for probing the foundations of cumulative cultural evolution and, more specifically, a micro-institution for cumulative intelligence. Our results isolate three ingredients for this process, namely mechanisms that generate diversity, high-precision social filters that reject errors, and pathways that internalize innovations into stable competence. The prominence of verification aligns with anthropological evidence that cumulative cultures depend strongly on population connectivity and error-correction norms (Derex & Boyd, 2016; Powell et al., 2009). Human epistemic vigilance remains richer, drawing on mental-state reasoning (Frith & Frith, 2006; C. M. Heyes & Frith, 2014), whereas POLIS agents rely on outcome-based consensus. This contrast makes POLIS useful for cognitive science in a second sense. It separates innovation, social evaluation, memory, and internalization into distinct operators that can be manipulated experimentally. Population structure, verifier independence, and memory policy can therefore be studied as organizational variables rather than treated as background conditions. Extending this framework toward richer social inference is a natural next step for computational studies of CCE.

Scalability and Trade-offs

POLIS highlights critical design tensions in multi-agent societies. Verification improves information quality but raises interaction costs at scale, making verifier allocation and communication topology central variables. Internalization, conversely, risks converging agents toward local optima and eroding diversity. Effective societies therefore require sufficient structural commonality to stabilize transmission, paired with enough heterogeneity to preserve complementary strengths. These dynamics mirror organizational learning: institutions standardize evaluation for reliability while sustaining distributed perspectives for adaptability. POLIS similarly clarifies the cultural ratchet effect, where cumulative progress hinges on retaining error correction without depleting the generative diversity needed for novel variants. Within the framework, this balance is governed by verifier independence, memory selectivity, and update frequency.

By operationalizing these tensions, POLIS shifts the research focus from static accuracy benchmarks to investigate which organizational structures maximize long-term knowledge accumulation under fixed compute and communication budgets. Treating institutions rather than isolated agents as the primary unit of analysis bridges cognitive science and machine learning. Furthermore, the framework facilitates controlled ablation studies that systematically compare centralized and decentralized verification protocols, evaluate sparse and dense communication topologies, and examine varying memory horizons, enabling a rigorous exploration of social architecture under constrained resources.

Conclusion

The dominant approach to scaling intelligence trains ever larger models on expanding datasets. POLIS shows how populations of bounded reasoners accumulate capability through social interaction. Experiments identify verification as the operator that governs what the society retains, while internalization turns shared artifacts into durable competence. This result aligns with cognitive science accounts in which epistemic vigilance stabilizes cumulative learning. Viewing intelligence as scalable through organization shifts the question from model size to connectivity. POLIS provides a testbed for this scaling axis and for institutional variation in cumulative intelligence.

Acknowledgments

I am grateful to Professor Ben Wang and Professor Shuifa Sun from Hangzhou Normal University for their guidance, encouragement, and continued support throughout the course of this work.

References

Abouelenin, A., Ashfaq, A., Atkinson, A., Awadalla, H., Bach, N., Bao, J., Benhaim, A., Cai, M., Chaudhary, V., Chen, C., et al. (2025). Phi-4-mini technical report: Compact yet powerful multimodal language models via mixture-of-loras. *arXiv preprint arXiv:2503.01743*.

Anderson, P. W. (1972). More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, *177*(4047), 393–396.

Anthropic, A. (2024). The claude 3 model family: Opus, sonnet, haiku. *Claude-3 Model Card*, *1*, 1.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081–1085.

Bandura, A. (1977). Social learning theory. *Englewood Cliffs, NJ: Prentice Hall*.

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *Proceedings of the 26th annual international conference on machine learning*, 41–48.

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago press.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(supplement_2), 10918–10925.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *1989*, *18*(1), 32–42.

Chen, H., Sun, Q., Dai, Q., & Wang, J. (2024). Multiagent finetuning: Self improvement with diverse reasoning chains. *arXiv preprint arXiv:2410.14228*.

Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., et al. (2021). Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*.

Collins, K. M., Sucholutsky, I., Bhatt, U., Chandra, K., Wong, L., Lee, M., Zhang, C. E., Zhi-Xuan, T., Ho, M., Mansinghka, V., et al. (2024). Building machines that learn and think with people. *Nature human behaviour*, *8*(10), 1851–1863.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148–153.

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science*, *335*(6072), 1114–1118.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987.

Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al. (2024). The llama 3 herd of models. *arXiv e-prints*, arXiv–2407.

EBRAHIMI, S., & ASUDEH, A. (2025). A survey on reliability, transparency, accountability, and fairness in llm-based multi-agent systems through the responsibility lens.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

Gao, X., Pei, Q., Tang, Z., Li, Y., Lin, H., Wu, J., Wu, L., & He, C. (2025). A strategic coordination framework of small llms matches large llms in data synthesis. *arXiv preprint arXiv:2504.12322*.

Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., & Steinhardt, J. (2021). Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Herbrich, R., Minka, T., & Graepel, T. (2006). Trueskill™: A bayesian skill rating system. *Advances in neural information processing systems*, *19*.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *science*, *317*(5843), 1360–1366.

Heyes, C. (2019a). Cognition blindness and cognitive gadgets. *Behavioral and Brain Sciences*, *42*.

Heyes, C. (2019b). Précis of cognitive gadgets: The cultural evolution of thinking. *Behavioral and Brain Sciences*, *42*, e169.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190), 1243091.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., et al. (2022). Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(2), 174–196.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385–16389.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. (2022). Lora: Low-rank adaptation of large language models. *ICLR*, *1*(2), 3.

Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Rad-

ford, A., et al. (2024). Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Kirsh, D. (1995). The intelligent use of space. *Artificial intelligence*, *73*(1-2), 31–68.

Lansing, J. S. (2003). Complex adaptive systems. *Annual review of anthropology*, *32*(1), 183–204.

Leimeister, J. M. (2010). Collective intelligence. *Business & Information Systems Engineering*, *2*, 245–248.

Li, C., Yuan, Z., Yuan, H., Dong, G., Lu, K., Wu, J., Tan, C., Wang, X., & Zhou, C. (2023). Mugglemath: Assessing the impact of query and response augmentation on math reasoning. *arXiv preprint arXiv:2310.05506*.

Li, M., Zhang, Y., Li, Z., Chen, J., Chen, L., Cheng, N., Wang, J., Zhou, T., & Xiao, J. (2023). From quantity to quality: Boosting llm performance with self-guided data selection for instruction tuning. *arXiv preprint arXiv:2308.12032*.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, *108*(22), 9020–9025.

MAA. (2024). American invitational mathematics examination - aime. *American Invitational Mathematics Examination - AIME*. https://huggingface.co/datasets/math-ai/aime24

MAA. (2025). The llama 4 herd: The beginning of a new era of natively multimodal ai innovation. https://ai.meta.com/blog/llama-4-multimodal-intelligence/

Mercier, H., & Sperber, D. (2011). Why do humans reason? arguments for an argumentative theory. *Behavioral and brain sciences*, *34*(2), 57–74.

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B*, *285*(1880), 20180712.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, *35*, 27730–27744.

Powell, A., Shennan, S., & Thomas, M. G. (2009). Late pleistocene demography and the appearance of modern human behavior. *Science*, *324*(5932), 1298–1301.

Rein, D., Hou, B. L., Stickland, A. C., Petty, J., Pang, R. Y., Dirani, J., Michael, J., & Bowman, S. R. (2024). Gpqa: A graduate-level google-proof q&a benchmark. *First Conference on Language Modeling*.

Team, G., Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., Perrin, S., Matejovicova, T., Ramé, A., Rivière, M., et al. (2025). Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*.

Team, M., Xiao, C., Li, Y., Han, X., Bai, Y., Cai, J., Chen, H., Chen, W., Cong, X., Cui, G., et al. (2025). Minicpm4: Ultra-efficient llms on end devices. *arXiv preprint arXiv:2506.07900*.

Team, Q. (2024). Qwen2 technical report. *arXiv preprint arXiv:2412.15115*.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2405–2415.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675–691.

Tong, Y., Zhang, X., Wang, R., Wu, R., & He, J. (2024). Dart-math: Difficulty-aware rejection tuning for mathematical problem-solving. *Advances in Neural Information Processing Systems*, *37*, 7821–7846.

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational psychology review*, *22*(3), 271–296.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., et al. (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824–24837.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682–685.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, *330*(6004), 686–688.

Xu, Z., Jiang, F., Niu, L., Deng, Y., Poovendran, R., Choi, Y., & Lin, B. Y. (2024). Magpie: Alignment data synthesis from scratch by prompting aligned llms with nothing. *arXiv preprint arXiv:2406.08464*.

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al. (2025). Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Zelikman, E., Wu, Y., Mu, J., & Goodman, N. (2022). Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, *35*, 15476–15488.