IECGA 2025 Conference

The 1st International Electronic Conference on Games

15–16 October 2025 | Online

The Power of Stories: Narrative Priming in Networked Multi-Agent LLM Interactions

Gerrit Großmann[1], Larisa Ivanova[1,3], Sai Leela Poduru[1,2], Mohaddeseh Tabrizian[1,2], Islam Mesabah[1], David A. Selby[1], Sebastian J. Vollmer[1,2]

[1] German Research Center for Artificial Intelligence (DFKI), [2] Department of Computer Science, University of Kaiserslautern–Landau (RPTU),
[3] Department of Language Science and Technology, Saarland University
larisa.ivanova@dfki.de

BACKGROUND

Why Narratives?

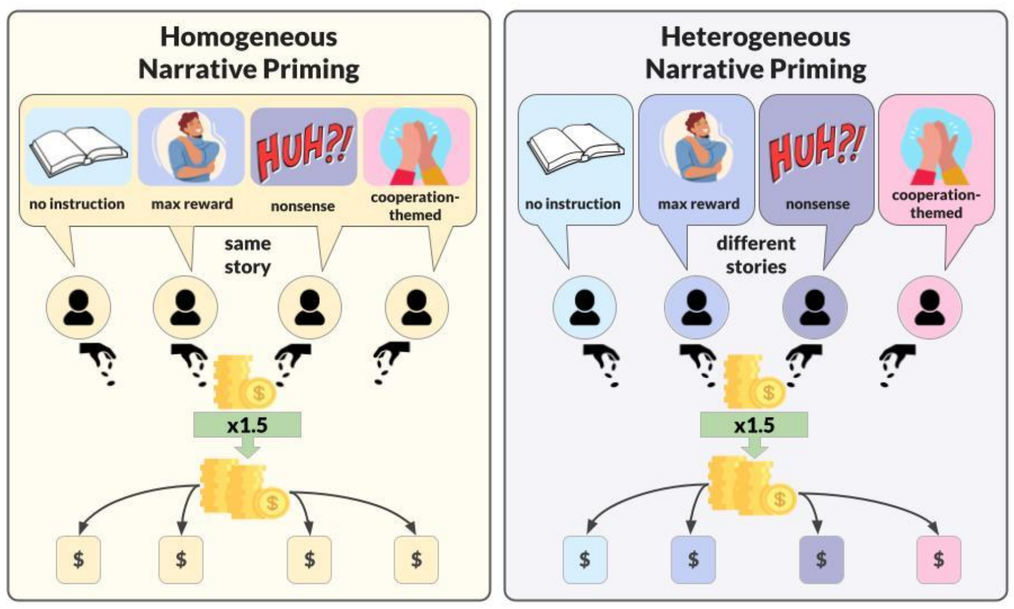
In humans, shared stories enable large-scale cooperation [1,2]
We ask: can narrative priming also shape cooperation among artificial agents?

Why LLM Agents?

LLMs exhibit social-like behaviors: advances in reasoning, communication, and coordination across tasks, LLM multi-agent simulations [3,4,5,6,7]
→ controlled testbed for social simulations: transparent context, full logging

Research Questions

1. How do narratives influence negotiation behavior?
2. What differs when agents share the same story versus different ones?
3. What happens when the agent numbers grow?
4. Are agents resilient against self-serving participants?



METHOD

Repeated Public Goods Game (PGG) – classic social dilemma paradigm:

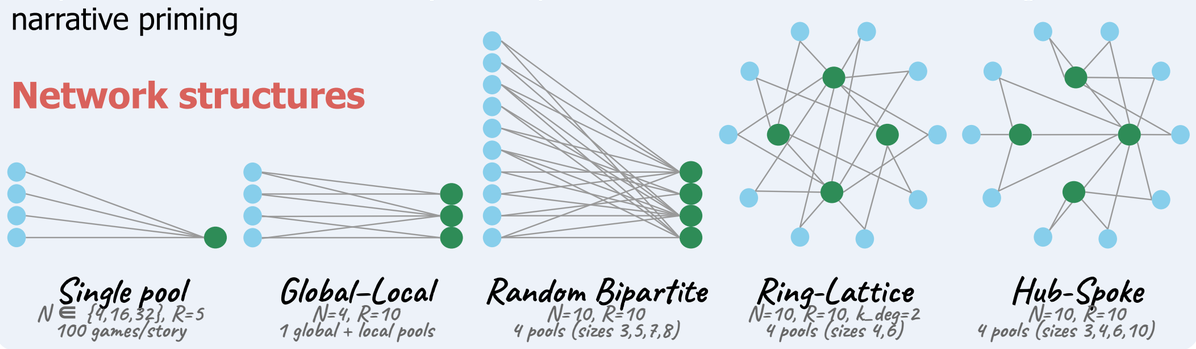
- **Endowment**: tokens each agent receives per round *R*, *e*=10
- **Multiplier**: efficiency factor applied to pooled contributions, *m*=1.5
- **Payoff**: agent's return per round = kept tokens + (m × Σ contrib / pool size)
- **Collaboration score**: share of total tokens collectively invested in pools across all rounds ([0,1])

Procedure

➪ **Setup**: define condition (same- vs different-story) and network topology
↳ **Play**: each round, agents allocate integer contributions *k* ≤ remaining tokens to each pool they belong to
↳ **Feedback**: after every round, agents receive only per-pool contribution summaries (from others in shared pools) and their own payoffs
↳ **Metrics**: compute per-round & cumulative payoffs; derive collab score
↳ **Repetition**: 10 rounds × 100 games per story per topology → robust averages

Implementation: LLM agents (LLaMA-3.3-70B-Instruct) are independently prompted; each agent's system message embeds one bedtime-style story (8 cooperation-themed, 4 controls) + complete PGG rules → consistent framing but varied narrative priming

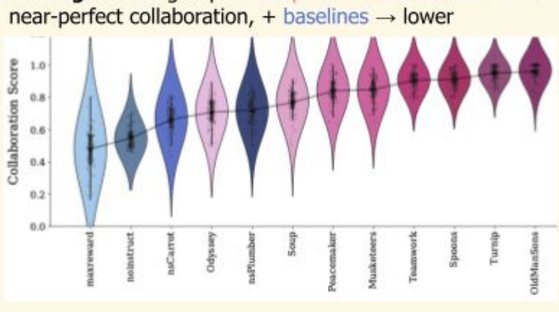
Network structures



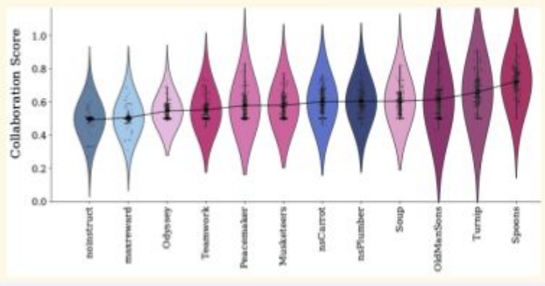
RESULTS

Condition	Observation	Implication
Homogeneous Cooperative Stories	Near-perfect collaboration, highest cumulative payoffs	*Shared* narratives *align* decision heuristics
Heterogeneous Stories	Cooperation collapses, self-interested narratives outperform	*Misaligned* narratives *drive exploitation*
Network Topology	Effect persists, topology shifts *where* cooperation flows: local clusters in ring, hub dominance in hub-spoke	Narrative priming generalizes beyond structure
Adaptation Across Rounds	Declining/increasing contributions depending on partners' behavior	Implicit, round-level learning

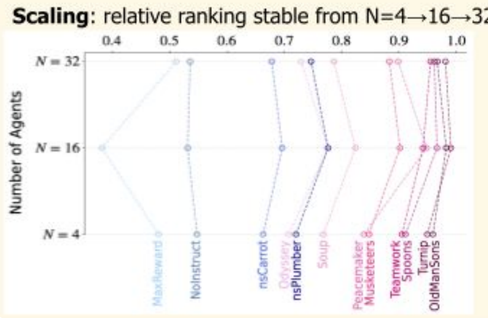
Homogeneous groups + cooperation-themed stories → near-perfect collaboration, + baselines → lower



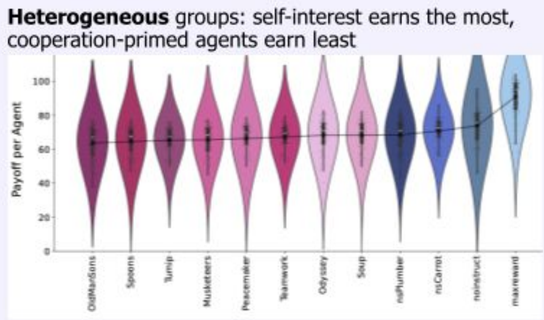
Robustness: adding a free-rider lowers overall cooperation but preserves story-ranking trend



Scaling: relative ranking stable from N=4→16→32



Heterogeneous groups: self-interest earns the most, cooperation-primed agents earn least



CONTRIBUTIONS

- **Narrative coherence reliably promotes cooperation**: aligned story framing produces stable cooperative equilibria and higher payoffs
- **Divergent narratives destabilize coordination**: mixed or conflicting frames shift behavior toward self-interest across network types
- **Topology shapes allocation, not direction**: structure (single-/multi-pool: global-local, random bipartite, geometric ring-lattice, star hub-spoke) alters where cooperation flows but not whether it emerges
- **Emergent adaptation**: round-to-round behavioral shifts reflect reinforcement by textual framing and feedback, not explicit communication
- **Design implication**: maintaining coherent narrative framing can systematically modulate cooperation in LLM multi-agent systems

LIMITATIONS & FUTURE WORK

- Mechanistic interpretability: narrative semantics vs structure vs RLHF?
- Larger *N* and varied topologies
- Adversarial narratives
- Cross-model replication
- Measure decay over time

REFERENCES

[1] Boyd, R., Richerson, P.J., 2009: Culture and the Evolution of Human Cooperation
[2] Harari, Y.N.: Sapiens, 2014: A Brief History of Humankind. Random House
[3] Aher et al., 2023: Using LLMs to Simulate Multiple Humans and Replicate Human Subject Studies
[4] Bianchi et al., 2024: How Well Can LLMs Negotiate? NegotiationArena Platform and Analysis
[5] Park et al., 2023: Generative Agents: Interactive Simulacra of Human Behavior
[6] Piatti et al., 2024: Cooperate or Collapse: Emergence of Sustainable Cooperation in a Society of LLM Agents
[7] Zhou et al., 2025: The PIMMUR Principles: Ensuring Validity in Collective Behavior of LLM Societies