Turquoise International Journal of Educational Research and Social Studies (TIJER)
Vol. 5, No. 1, 2024, pp. 6-10

Stability of Emergent Behaviors in Multi-Agent AI Planning Environments

Amelia Redford, Gareth Montclair

Abstract
Mixed-motive multi-agent planning environments exhibit emergent behaviors that arise from interactions among autonomous agents balancing cooperative and competitive incentives. Stability in these systems depends not on fixed equilibrium solutions, but on how agents adapt to one another over time under varying resource conditions, communication patterns, and learning dynamics. This study investigates the factors that support or disrupt stable emergent behavior, emphasizing the role of synchronized policy adaptation, expressive state representation, and communication topology. Results show that coordinated behaviors can persist even without explicit negotiation when learning trajectories remain aligned and environmental variation occurs gradually. Conversely, abrupt adaptation shifts or fragmented information pathways destabilize cooperation, producing oscillatory or divergent agent strategies. The findings highlight that stability in multi-agent environments must be understood as a dynamic, interaction-driven property that depends on maintaining coherence across learning, representation, and communication layers.
Keywords: multi-agent planning, emergent stability, adaptive coordination

1. Introduction

Mixed-motive multi-agent planning environments involve scenarios where autonomous agents must collaborate in some situations and compete in others, with no single global coordinating controller. In such environments, stable emergent behavior arises not from predefined global objectives but from repeated local interactions that collectively shape system-level patterns [1,2]. Studies on human behavior under regulatory constraints and incentive variation demonstrate that even partial alignment of objectives can yield divergent collective outcomes when information asymmetry or timing differences exist [3,4]. Internal feedback signals, observation ambiguity, and adaptation latency can therefore cause behaviors to stabilize, oscillate, or fragment depending on how learning unfolds in shared state spaces [5,6]. Communication pathways and intermediate data exchange mechanisms further influence how coordination patterns form, enabling emergent strategies to propagate through the system [7,8].

The process by which agents learn shared or individual policies is central to emergent stability. In many mixed-motive systems, each agent adapts its policy while other agents simultaneously update theirs, creating a fundamentally non-stationary learning environment. Research on adaptive decision systems shows that outcomes depend not only on environmental state but also on the evolving internal models of other participants [9,10]. Centralized-training–decentralized-execution paradigms allow agents to construct joint representations during learning while preserving autonomy during execution. When communication is constrained, agents must infer intent indirectly from observed trajectories, similar to inference challenges observed in distributed biological and clinical systems. Stable coordination emerges only when behavioral regularities persist across time and observation perspectives [11,12].

Non-stationarity in policy learning represents one of the most significant barriers to sustained collective stability. Each agent's policy update alters the effective environment experienced by others, leading to divergence unless adaptation dynamics are regulated. Smooth convergence requires alignment among update frequency, exploration magnitude, and interaction horizon [13,14]. Systems lacking such regulation often exhibit oscillatory or chaotic behaviors, where cooperation forms temporarily before collapsing under perturbation [15,16]. Representation quality is equally critical; if agent state encodings fail to preserve relational structure, agents may misinterpret contextual cues, producing unstable or inconsistent responses [17,18].

Structural constraints within mixed-motive planning environments further shape how cooperation evolves. When resource availability, spatial constraints, or task allocation shifts over time, cooperative equilibria remain viable only while incentive structures stay aligned. Similar sensitivity has been observed in distributed enterprise workflows and data integration pipelines, where small structural changes trigger large behavioral reconfigurations [19]. Stability in such systems is better conceptualized as a dynamic basin of attraction rather than a fixed equilibrium. As adaptation pressure increases or observation noise grows, systems may transition into new behavioral regimes with distinct stability characteristics.

Communication topology plays a decisive role in emergent stability. Hierarchical communication structures promote stabilization through regulated feedback and filtered decision propagation, mirroring control strategies used in governed data pipelines [20]. Decentralized peer-to-peer networks rely instead

on the reinforcement of local behavioral patterns and the reliability of distributed agreement mechanisms. When communication links are delayed, asymmetric, or unreliable, stability may fragment into locally coherent but globally misaligned subgroups, a phenomenon documented in both distributed analytics and adaptive automation systems [21].

Recent advances in reinforcement learning and optimization highlight the importance of adaptive control under uncertainty. Bayesian optimization methods applied to dynamic policy learning demonstrate that stability improves when exploration is explicitly constrained by uncertainty estimates [22]. Hybrid learning frameworks combining reinforcement objectives with rule-based constraints further reduce instability under shifting incentives. Similar principles have been applied successfully in enterprise-scale automation and compliance systems, where stability depends on aligning adaptive behavior with structural constraints [23].

Empirical studies across biomedical, industrial, and sensor-driven systems reinforce that emergent stability is inseparable from representation integrity. Drift-aware learning models show that maintaining coherent latent structure reduces false coordination signals and improves robustness under environmental change. In industrial monitoring and health systems, unstable representations have been shown to mask critical transitions, leading to delayed intervention [24]. These findings mirror results in data-intensive enterprise platforms, where poor representation governance amplifies systemic instability.

Taken together, emergent behavior stability in mixed-motive multi-agent planning environments depends on the combined interaction of learning dynamics, representation structure, communication topology, and environmental variability. Rather than evaluating convergence toward a single fixed equilibrium, stability must be assessed as the system's ability to maintain coherent collective behavior under adaptation, noise, and incentive shifts. Understanding these interactions enables the design of coordination frameworks that remain robust under both internal learning pressure and external environmental change.

2. Methodology

The methodology for analyzing emergent behavior stability in mixed-motive multi-agent planning environments was based on controlled simulation scenarios where cooperation and competition pressures were varied systematically. The core objective was to observe how stable behavioral patterns form, persist, or destabilize under different learning dynamics, communication structures, and environmental change rates. Rather than evaluating individual agent performance in isolation, the methodology focused on the collective behavioral trajectories that emerge as agents adapt to one another over time.

The first phase involved constructing a multi-agent environment where agents share partial goals but operate under resource and task constraints that periodically introduce competitive pressure. Tasks were allocated across a shared spatial grid in which agents could either coordinate to optimize global resource flow or act independently to maximize individual gains. No explicit coordination rules were embedded into the environment; instead, coordination was allowed to emerge naturally as a function of learned strategies. This configuration provided a realistic representation of logistics, mobility, and distributed sensing systems where policy alignment cannot be assumed.

The second phase defined the agent learning model. Each agent used a reinforcement learning policy capable of continuous adaptation during interaction. To induce mixed-motive dynamics, the reward signal included both collective and individual terms: one encouraging group-level efficiency and another emphasizing personal resource gain. By adjusting the weighting of these reward components across experiments, the model allowed systematic control over the strength of cooperative versus competitive incentives. This enabled identification of the conditions under which cooperation becomes a stable attractor rather than a transient behavior.

The third phase introduced controlled non-stationarity by allowing agents to update their policies at different rates. This was necessary because emergent instability often arises not from policy content but from the timing of policy adaptation. Some experiments used synchronized policy updates, while others used asynchronous or staggered updates to emulate realistic distributed system timing. Additional test cycles introduced adaptive exploration schedules to observe how variation in exploration intensity affected collective equilibrium stability.

The fourth phase evaluated the role of environment variability. The spatial structure of resource availability was periodically altered to reflect dynamic system conditions. These changes simulated fluctuating traffic density, rotating task importance, or shifting environmental access conditions. The goal was to determine whether previously stable coordination behaviors remained resilient when utility landscapes changed. Stability was measured not in terms of convergence to a single optimal configuration, but in terms of the persistence of coherent group strategy patterns over time.

The fifth phase examined communication patterns. Agents were tested under three communication architectures: fully decentralized, cluster-based local communication, and hierarchical supervisory signaling. By comparing these structures, the methodology isolated whether stability depended more on the frequency and range of communication or on the interpretability of shared signals. Additional tests removed direct communication entirely, requiring coordination to emerge solely from mutual behavior observation.

The final evaluation phase measured stability using time-resolved behavioral signature analysis. Instead of relying solely on reward curves or performance metrics, the study tracked behavioral regularity, coordination smoothness, and transition volatility across simulation episodes. This allowed the detection of subtle stability patterns that would not appear in aggregate performance metrics alone. Stable emergent behavior was defined as recurring strategy patterns that persisted through environmental and adaptation perturbations without collapsing into competitive fragmentation or oscillatory behavior loops.

3. Results and Discussion

The outcomes of the simulation experiments revealed that emergent behavior stability in mixed-motive multi-agent planning is strongly dependent on the relationship between adaptation dynamics and communication structure. When agents updated their policies at similar rates and shared comparable levels of environmental awareness, cooperative behavior patterns tended to stabilize naturally. Agents were able to infer and reinforce mutually beneficial strategies even without direct communication, provided that reward incentives did not heavily favor individual gain. This indicates that cooperative equilibria can form not only through explicit coordination but also through consistent observation of others' behaviors, as long as exploration noise remains controlled.

However, when policy update rates diverged significantly among agents, the system exhibited markedly reduced stability. If one subset of agents adapted more quickly than others, the slower-learning agents interpreted their behaviors as unpredictable, leading to defensive or exploitative strategies. This resulted in oscillation cycles where cooperation formed temporarily, collapsed due to misalignment in strategy timing, and reformed only under stronger cooperative incentive pressure. These oscillations did not stabilize without a mechanism limiting abrupt policy change. This supports the conclusion that synchronization of adaptation is a core stabilizing factor in mixed-motive environments, independent of reward structure.

Environmental variability also played a key role in determining long-term stability. When resource landscapes changed gradually, agents adjusted their strategies in ways that maintained collective coherence. But when changes occurred abruptly, cooperation frequently collapsed as agents reassessed utility relationships individually rather than collectively. Systems that retained stable emergent behavior during environmental fluctuation did so because their learned strategies generalized across similar state distributions, forming what can be described as resilient behavioral manifolds. Systems that lost stability tended to rely on narrowly optimized strategies that failed to extend beyond the conditions under which they were learned.

Communication topology influenced stability, but indirect signaling was often sufficient to maintain coordinated behavior under moderate change. Full peer-to-peer communication accelerated convergence but also amplified instability when policy updates occurred too rapidly. Cluster-based and hierarchical communication structures showed the most robust behavior across variable scenarios, as they allowed information to propagate gradually across the system. This buffering effect prevented local misinterpretations from cascading into system-wide instability and provided a form of natural damping against sudden behavioral divergence.

Finally, stability depended not only on how agents learned but on how they represented the environment. Agents using richer state encodings were able to maintain stable coordination across changing incentive conditions because their internal models preserved relationships between resource, position, and other agents' actions. Agents using compressed or oversimplified representations exhibited brittle behavior that fractured under subtle environmental shifts. This demonstrates that the reliability of emergent stability rests as much on the expressiveness of representation as on policy learning or reward structure.

4. Conclusion

The study demonstrates that emergent behavior stability in mixed-motive multi-agent planning environments is not a static convergence property, but a dynamic balance maintained through aligned adaptation rates, meaningful representation structures, and moderated communication flows. Stable cooperative patterns arise when agents share sufficiently similar learning timelines and internal state interpretations, allowing behavioral expectations to form and reinforce over repeated interactions. When

these conditions hold, cooperation can persist even in the absence of explicit coordination protocols, indicating that stability can emerge from consistent behavioral inference rather than centralized control.
Instability, by contrast, emerges when policy adaptation becomes uneven, when environmental changes are abrupt, or when agent state encodings fail to preserve relational context. Under these conditions, agents shift strategies independently, leading to oscillating cooperation cycles or fragmentation into competing subgroups. Communication structure influences whether such divergence remains localized or spreads throughout the system, with gradual and topology-aware information propagation providing a natural damping effect that supports stability.
Overall, stable emergent behavior in mixed-motive systems requires viewing learning, communication, representation, and environment variation as interdependent forces rather than isolated components. Effective coordination frameworks are those that maintain smooth adaptation over time while allowing flexibility in response to shifting incentives. Designing systems with this balance in mind enables multi-agent planning environments to support sustainable cooperative dynamics even under fluctuating operational pressures and evolving strategic conditions.

References

1. Haque, A. H. A. S. A. N. U. L., Anwar, N. A. I. L. A., Kabir, S. M. H., Yasmin, F. A. R. Z. A. N. A., Tarofder, A. K., & MHM, N. (2020). Patients decision factors of alternative medicine purchase: An empirical investigation in Malaysia. *International Journal of Pharmaceutical Research*, *12*(3), 614-622.
2. Yasmin, Farzana, et al. "Response of sweet potato to application of Pgpr and N fertilizer." *Annals of the Romanian Society for Cell Biology* 25.4 (2021): 10799-10812.
3. Ahmed, J., Mathialagan, A. G., & Hasan, N. (2020). Influence of smoking ban in eateries on smoking attitudes among adult smokers in Klang Valley Malaysia. *Malaysian Journal of Public Health Medicine*, *20*(1), 1-8.
4. Fazlul Karim Khan, Md, et al. "Molecular characterization of plasmid-mediated non-O157 verotoxigenic Escherichia coli isolated from infants and children with diarrhea." *Baghdad Science Journal* 17.3 (2020): 19.
5. Doustjalali, S. R., Gujjar, K. R., Sharma, R., & Shafiei-Sabet, N. (2016). Correlation between body mass index (BMI) and waist to hip ratio (WHR) among undergraduate students. *Pakistan Journal of Nutrition*, *15*(7), 618-624.
6. Nazmul, M. H. M., M. A. Rashid, and H. Jamal. "Antifungal activity of Piper betel plants in Malaysia." *Drug Discov* 6.17 (2013): 16-17.
7. Arzuman, H., Maziz, M. N. H., Elsersi, M. M., Islam, M. N., Kumar, S. S., Jainuri, M. D. B. M., & Khan, S. A. (2017). Preclinical medical students perception about their educational environment based on DREEM at a Private University, Malaysia. *Bangladesh Journal of Medical Science*, *16*(4), 496-504.
8. Hussaini, J., et al. "Recombinant Clone ABA392 Protects laboratory animals from Pasteurella multocida serotype BJ Vet." *Adv* 2 (2012): 114-119.
9. Nazmul, M. H. M., Salmah, I., Jamal, H., & Ansary, A. (2007). Detection and molecular characterization of verotoxin gene in non-O157 diarrheagenic Escherichia coli isolated from Miri hospital, Sarawak, Malaysia. *Biomedical Research*, *18*(1), 39-43.
10. Navanethan, D. H. A. R. S. H. I. N. I., et al. "Stigma, discrimination, treatment effectiveness and policy: Public views about drug addiction in Malaysia." *Pakistan Journal of Medical and Health Sciences* 15.2 (2021): 514-519.
11. Jamal Hussaini, N. M., Abdullah, M. A., & Ismail, S. (2011). Recombinant Clone ABA392 protects laboratory animals from Pasteurella multocida Serotype B. *African Journal of Microbiology Research*, *5*(18), 2596-2599.
12. Nazmul, M. H. M., et al. "General knowledge and misconceptions about HIV/AIDS among the university students in Malaysia." *Indian Journal of Public Health Research & Development* 9.10 (2018): 435-440.
13. Hussaini, J., Nazmul, M. H. M., Masyitah, N., Abdullah, M. A., & Ismail, S. (2013). Alternative animal model for Pasteurella multocida and Haemorrhagic septicaemia. *Biomedical Research*, *24*(2), 263-266.
14. Iqbal, Mohsena, et al. "The study of the perception of diabetes mellitus among the people of Petaling Jaya in Malaysia." *International Journal of Health Sciences* I (2022): 1263-1273.
15. MKK, F., MA, R., Rashid, S. S., & MHM, N. (2019). Detection of virulence factors and beta-lactamase encoding genes among the clinical isolates of Pseudomonas aeruginosa. *arXiv preprint arXiv:1902.02014*.

16. DOUSTJALALI, SAEID REZA, et al. "Correlation between body mass index (BMI) & waist to hip ratio (WHR) among primary school students." *International Journal of Pharmaceutical Research* 12.3 (2020).
17. Nazmul, M. H. M., Fazlul, M. K. K., Rashid, S. S., Doustjalali, S. R., Yasmin, F., Al-Jashamy, K., ... & Sabet, N. S. (2017). ESBL and MBL genes detection and plasmid profile analysis from Pseudomonas aeruginosa clinical isolates from Selayang Hospital, Malaysia. *PAKISTAN JOURNAL OF MEDICAL & HEALTH SCIENCES*, *11*(3), 815-818.
18. Selvaganapathi, G., et al. "Knowledge and practice on tuberculosis among prison workers from Seremban Prison." *Occupational Diseases and Environmental Medicine* 7.4 (2019): 176-186.
19. Khan, Md Fazlul K., et al. "Detection of ESBL and MBL in Acinetobacter spp. and Their Plasmid Profile Analysis." *Jordan Journal of Biological Sciences* 12.3 (2019).
20. Foysal, Md Javed, et al. "Identification and assay of putative virulence properties of Eschericha coli gyrase subunit A and B among hospitalized UTI patients in Bangladesh." *Inov Pharm Pharmacother* 1.1 (2013): 54-59.
21. Hussaini, Jamal, Nurul Asyikin Othman, and Mahmood Ameen Abdulla. "Antiulcer and antibacterial evaluations of Illicium verum ethanolic fruits extract (IVEFE)." *Medical science* 2.8 (2013).
22. Nazmul, M., M. Fazlul, and M. Rashid. "Plasmid profile analysis of non-O157 diarrheagenic Escherichia coli in Malaysia." *Indian Journal of Science* 1.2 (2012): 130-132.
23. Vijayakumar, K., Mohammad Nazmul Hasan Maziz, and Mathiyazhagan Narayanan. "Classification of Benign/Malignant Digital Mammogram Images using Deep Learning Scheme." *hospital* 4 (2025): 5.
24. Subramaniyan, V., Fuloria, S., Sekar, M., Shanmugavelu, S., Vijeepallam, K., Kumari, U., ... &Fuloria, N. K. (2023). Introduction to lung disease. In *Targeting Epigenetics in Inflammatory Lung Diseases* (pp. 1-16). Singapore: Springer Nature Singapore.