2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Structured Diversity Control: A Dual-Level Framework for Group-Aware Multi-Agent Coordination

Shuocun Yang[1], Huawen Hu[1], Xuan Liu[1],Yincheng Yao[1],Enze Shi[1], Shu Zhang[1,*]

Abstract— **Controlling the behavioral diversity is a pivotal challenge in multi-agent reinforcement learning (MARL), particularly in complex collaborative scenarios. While existing methods attempt to regulate behavioral diversity by directly differentiating across all agents, they lack deep characterization and learning of multi-agent composition structures. This limitation leads to suboptimal performance or coordination failures when facing more complex or challenging tasks. To bridge this gap, we introduce Structured Diversity Control (SDC), a framework that redefines the system-wide diversity metric as a weighted combination of intra-group diversity, which is minimized for cohesion and inter-group diversity, which is maximized for specialization. The trade-off is governed by a pre-set Diversity Structure Factor (DSF), allowing for fine-grained, group-aware control over the collective strategy. Our method directly constrains the policy architecture without altering reward functions. This structural definition of diversity enables SDC to deliver substantial performance gains across various experiments, including increasing average rewards by up to 47.1% in multi-target pursuit and reducing episode lengths by 12.82% in complex neutralization scenarios. The proposed method offers a novel analytical perspective on the problem of cooperation in group-aware multi-agent systems.**

I. INTRODUCTION

Multi-agent reinforcement learning (MARL) has become a powerful paradigm for solving complex cooperative tasks [1], [2]. A central challenge in MARL is fostering effective coordination, where agents must balance individual behaviors with collective goals [3]. While foundational techniques like parameter sharing improve sample efficiency, they often lead to policy homogenization, limiting the system's adaptability and performance in scenarios requiring diverse roles [4], [5]. This highlights the critical need for explicit behavioral diversity control — a mechanism to regulate not just the overall level of diversity, but its underlying structure.

Behavioral diversity control has emerged as a critical research area in multi-agent coordination, garnering increasing attention in recent years. This growing interest has led to the development of various approaches for quantifying and regulating agent behavioral differences. Hu et al. have explored methods to measure diversity by learning latent representations of policies to compute a Multi-Agent Policy Distance (MAPD) [6]. Bettini et al. introduced System Neural Diversity (SND) [7], a Wasserstein distance-based metric that aggregates pairwise agent distances to measure system-level behavioral heterogeneity while maintaining agent number invariance. Building upon SND, Bettini et al. further proposed DiCo [8], which achieves precise diversity control by representing agent policies as combinations of shared homogeneous and agent-specific heterogeneous components with architectural constraints. Zhang et al. developed Intrinsic Action-tendency Consistency (IAM) [6], integrating intrinsic rewards with Centralized Training with Decentralized Execution (CTDE) frameworks to address action-tendency disagreements through prediction-based incentive mechanisms.

Despite these advances, existing diversity control methods still face significant limitations in structured multi-agent scenarios with groupings. These pioneering methods, including both the metrics and the control frameworks built upon them, share a common limitation: their formulation is monolithic. Current approaches either promote diversity among individual agents [9]–[12] or enforce system-wide uniformity [13], [14], treating all agents as a single population without differentiating between desired low diversity within groups versus high diversity between groups. This monolithic nature leaves the critical challenge of controlling diversity structure unaddressed, leading to suboptimal performance in complex coordination tasks where different groups require distinct diversity characteristics—some needing tight cooperation while others requiring specialized differentiation. To address this gap, we introduce Structured Diversity Control (SDC), a novel framework that extends the principle of architectural constraint-based control to structured multi-agent systems. SDC introduces a mechanism to explicitly manage the structure of diversity by redefining the total diversity metric. It is calculated from three key components:

(1) **Intra-Group Diversity**, a measure of policy dissimilarity within each group, which is typically minimized to facilitate tight cooperation;

(2) **Inter-Group Diversity**, a measure of policy dissimilarity between groups, which is encouraged to enable an effective division of labor;

(3) **Diversity Structure Factor (DSF)**, which governs the balance between these two components, allowing practitioners to inject prior knowledge about the task's coordination requirements.

SDC offers a generalizable solution compatible with any actor-critic framework. Our core contributions are threefold:

- We propose a novel dual-level framework for explicit behavioral diversity control in grouped MARL scenarios, bridging a critical gap left by monolithic control methods.
- We introduce the Diversity Structure Factor, a mecha-

*Corresponding author: Shu Zhang, shu.zhang@nwpu.edu.cn.
[1] Northwestern Polytechnical University, Xi'an 710072, China.

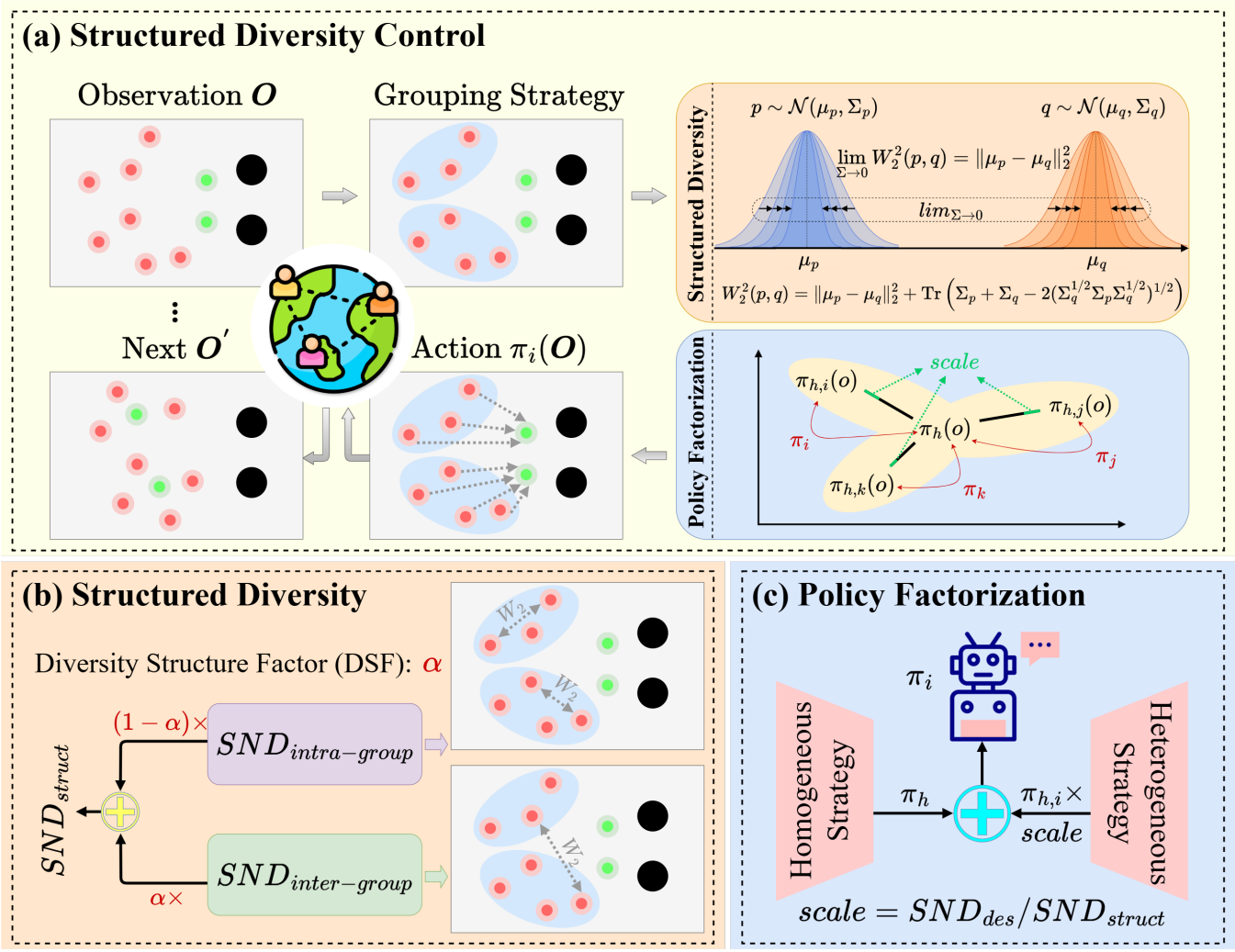

Fig. 1. Overview of the Structured Diversity Control (SDC) framework within a MARL training loop: **(a) Structured Diversity Control:** Within the SDC framework, a structured diversity constraint is imposed on the agent policies. **(b) Structured Diversity:** At each training step, the framework first computes the structured system diversity $\mathrm{SND}_{\mathrm{struct}}$ from the heterogeneous policy components. **(c) Policy Factorization:** Each agent's final policy is represented as the sum of a parameter-shared, homogeneous component and a rescaled, per-agent heterogeneous component. The computed $\mathrm{SND}_{\mathrm{struct}}$ is then compared to a desired target value $\mathrm{SND}_{\mathrm{des}}$ to generate a scale factor. This factor is used to rescale the heterogeneous components, resulting in the final updated policies.

nism for fine-grained tuning of group-based strategies to match diverse task demands.

- We provide extensive results showing that by enabling more granular control over behavioral diversity, SDC achieves significantly superior task performance and more effective specialization compared to its highly competitive counterpart, DiCo, especially in complex cooperative tasks requiring structured team coordination.

II. Related Work

Behavioral diversity control in multi-agent reinforcement learning is an emerging but promising research area [15]. Early studies have focused on single-agent scenarios, encouraging exploration by introducing entropy regularization or diversity rewards [16], [17]. However, it is often difficult to effectively coordinate the relationship between individuals and the whole when these approaches are directly extended to multi-agent systems [18]–[22].

A. Diversity via Internal Mechanisms

Such approaches directly motivate agents to explore diverse strategies and avoid behavioral convergence by designing reward mechanisms, learning frameworks, or strategy optimization methods within the agents. By designing an inherent reward mechanism, Ben Eysenbach et al. [23] propose a method to maximize mutual information, encourage agents to explore different strategies and avoid behavioral convergence. Through self-course learning, Igor Mordatch et al. [24] have promoted agents to develop a variety of

tool use strategies in interaction. Max Jadeberg et al. [25] use population methods such as evolutionary algorithms or MAP-Elites. They advocate maintaining a diverse population of agents and promoting diversity through parallel training and strategy exchange.

B. Diversity via External Interaction

This type of approach motivates agents to develop diverse strategies in their interactions with the environment or other agents by introducing external interaction mechanisms. Marcin Andrychowicz et al. [26] introduce a competition mechanism to promote agents to develop diversified strategies in confrontation. Peng Peng et al. [27] introduce opponent modeling modules in the Actor-Critic framework to further enhance diversity. These efforts have solved the problem of strategy convergence in MARL from different angles, provided theoretical support and practical verification for the design and optimization of complex multi-agent systems, and promoted the application of MARL in open environments, dynamic tasks, and complex interaction scenarios.

Although existing methods have achieved significant results in controlling behavioral diversity, they primarily focus on the diversity of individual agents or the overall system [28], [29], lacking a systematic analysis of behavioral diversity from a grouping perspective [30]–[35]. Specifically, existing methods do not fully consider the need for behavioral diversity in grouped tasks and lack fine-grained control over intra-group cooperation and inter-group task allocation, making it difficult to achieve efficient task distribution and collaboration in complex scenarios [36]. Therefore, we propose SDC, which provides a new solution for task allocation and collaboration in complex multi-agent systems by introducing two key concepts, intra-group diversity and inter-group diversity, and constraining the behavior of agents from the grouping perspective.

III. METHOD

In this section, we present Structured Diversity Control (SDC), a novel framework Fig. 1 designed to control behavioral diversity in multi-agent systems through explicit group-aware coordination. Our approach extends architectural constraint-based methods to address both intra-group cohesion and inter-group specialization simultaneously. We first formalize the problem setting and then elaborate on the key components of our proposed methodology.

A. Problem Formulation: Grouped POMGs

We formulate the multi-agent cooperative task as a Partially Observable Markov Game (POMG) [37], defined by the tuple $\langle \mathcal{A}, \mathcal{S}, \{\mathcal{O}_i\}, \{\mathcal{A}_i\}, P, R, \gamma \rangle$. Here, $\mathcal{A}$ is the set of N agents. A key characteristic of our setting is that the set of N agents, $\mathcal{A}$, is partitioned into K disjoint groups, denoted by $\mathcal{G} = \{G_1, G_2, \ldots, G_K\}$. Each group $G_k \in \mathcal{G}$ contains n_k agents, satisfying the conditions $\bigcup_{k=1}^{K} G_k = \mathcal{A}$ and $G_i \cap G_j = \emptyset$ for $i \neq j$. Each agent i learns a policy π_i that outputs a continuous action distribution, typically a multivariate Gaussian $\mathcal{N}(\mu_i, \Sigma_i)$.

Our goal is to learn policies that not only maximize the shared reward but also adhere to a structured behavioral diversity constraint defined over these groups.

B. Dual-Level Diversity Formulation

Our core idea is to create a diversity measure that is aware of the group structure by decomposing the standard system-wide diversity metric into two semantically meaningful components. Our formulation is built upon the System Neural Diversity (SND) metric [7], which provides a principled way to quantify behavioral diversity using the 2-Wasserstein distance (W_2) as its underlying pairwise measure. In our work, we leverage this same foundational metric but decompose it to apply in a structured, dual-level manner, capturing the nuanced coordination requirements of grouped systems.

1) Intra-Group Diversity: The first component of our formulation measures the cohesion within a group. For each group $G_k \in \mathcal{G}$, we define its intra-group diversity as the average pairwise SND among its members. In this regard, we follow the standard method used by DiCo [8] but apply it at the sub-group level. The intra-group diversity of the heterogeneous policy components, $\text{SND}_{\text{intra}}(G_k)$, is calculated as:

$$\text{SND}_{\text{intra}}(G_k) = \frac{2}{n_k(n_k-1)} \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} W_2(\pi_{h,i}(o), \pi_{h,j}(o)) \tag{1}$$

A lower $\text{SND}_{\text{intra}}$ value reflects stronger intra-group cohesion, leading to more effective collaborative behavior.

2) Inter-Group Diversity: The second component explicitly measures the specialization between groups. For any two groups, G_k and G_l, we define their inter-group diversity as the average pairwise SND between their respective members. The inter-group diversity of the heterogeneous components, $\text{SND}_{\text{inter}}(G_k, G_l)$, is:

$$\text{SND}_{\text{inter}}(G_k, G_l) = \frac{1}{n_k n_l} \sum_{i \in G_k} \sum_{j \in G_l} W_2(\pi_{h,i}(o), \pi_{h,j}(o)) \tag{2}$$

A higher $\text{SND}_{\text{inter}}$ value corresponds to a clearer division of labor between groups, as they adopt distinct, specialized policies.

C. Structured Diversity and Policy Factorization

SDC computes the total structured diversity, $\text{SND}_{\text{struct}}$, by structurally combining the dual-level components using the Diversity Structure Factor (DSF) α. For a general system with K groups, the total structured diversity is a weighted sum:

$$\begin{aligned} \text{SND}_{\text{struct}} = (1-\alpha) \cdot \left(\frac{1}{K} \sum_{k=1}^{K} \text{SND}_{\text{intra}}(G_k) \right) + \\ \alpha \cdot \left(\frac{1}{\binom{K}{2}} \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} \text{SND}_{\text{inter}}(G_k, G_l) \right) \end{aligned} \tag{3}$$

where $\binom{K}{2}$ is the binomial coefficient, representing the total number of unique group pairs. A high α prioritizes inter-group specialization, whereas a low α emphasizes intra-group cohesion. This formulation is valid for any $K \geq 2$.

Inspired by prior approaches that emphasize structural constraints on policy architectures [8], we represent each agent's final policy π_i as the sum of a shared component π_h and a rescaled per-agent deviation $\pi_{h,i}$. To enforce the desired overall diversity level, SND_{des}, we use a scaling factor, scale:

$$\pi_i(o) = \pi_h(o) + \text{scale} \cdot \pi_{h,i}(o) \tag{4}$$

where the scaling factor is computed as:

$$\text{scale} = \frac{\text{SND}_{\text{des}}}{\text{SND}_{\text{struct}}} \tag{5}$$

This mechanism, leveraging the properties of the Wasserstein distance, guarantees that the diversity of the final policies, when measured using the same structural formulation as Eq. (3), will match the target value, SND_{des}, without altering the learning objective.

IV. EXPERIMENTS

We conduct a series of experiments designed to comprehensively evaluate our proposed SDC framework. Our evaluation aims to answer three key questions: (1) To what extent can SDC achieve precise regulation of structured system diversity, ensuring alignment with desired targets? (2) How does the DSF modulate the balance between intra-group cohesion and inter-group specialization in the learned policies? (3) Does structured, group-aware diversity control yield measurable performance gains over monolithic control, particularly in tasks requiring coordinated multi-agent behaviors?

A. Experimental Setup

1) Environments: We use two challenging multi-agent environments designed to test structured cooperative strategies. As shown in Fig 1(a), the green and red circles denote the pursuers and escapees, respectively, while the black shapes act as static obstacles.

a) Multi-Pursuer Tag: Our first environment extends the simple tag scenario from the Multi-Agent Particle Environment (MPE). To study the impact of scale and task complexity, we establish a three-tiered difficulty system: 5 pursuers vs. 2 escapees (difficult), 6 vs. 2 (medium), and 7 vs. 2 (easy). Crucially, we implement a group collaboration mechanism where pursuers are partitioned into two teams, each tasked with capturing one of the escapees. This setup necessitates both tight intra-group cohesion and clear inter-group specialization . The evaluation metric is the team's average reward.

b) Shielded Tag: To further test coordinated target neutralization, we introduce a variant of simple tag. In this environment, each of the N escapee is assigned a shield attribute. A target is only neutralized when its shield value is reduced to zero by the pursuers. The episode terminates when all targets are neutralized. The primary evaluation metric is the episode length; a shorter episode indicates a more efficient strategy. This task explicitly rewards "focus fire" where high intra-group cohesion and high inter-group specialization are paramount for success.

2) Implementation Details: In our experiments, we partition the pursuer agents into distinct groups, with the number of groups set to match the number of escapees. All agents are trained using the IPPO algorithm [38]. For experiments, we train for at least 12 million environment steps across 600 parallel environments, using at least 4 distinct random seeds to ensure reliability and reproducibility. Our primary baseline for comparison is DiCo [8], the highly competitive method for architectural diversity control.

B. Verification of Diversity Control Capability

First, we verify that SDC can accurately control the total system diversity $\text{SND}(\{\pi_i\})$ to a desired level, SND_{des}. To this end, we fix the DSF to a representative value of $\alpha = 0.9$ and train a series of models with varying target diversity levels, $\text{SND}_{\text{des}} \in \{0.3, 0.4, \ldots, 0.9\}$. The results are presented in Fig. 2.

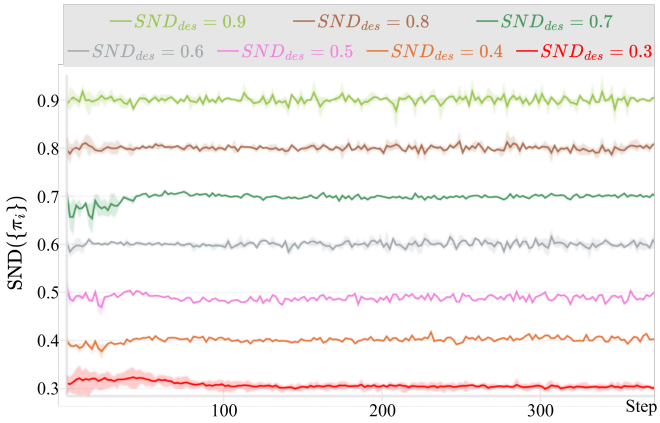

Fig. 2. Verification of SDC's diversity control capability. Each colored line represents a separate training run with a different target diversity level, SND_{des}, ranging from 0.3 to 0.9, at a fixed DSF of $\alpha = 0.9$.

As the Fig. 2 clearly illustrates, for each specified target value, the measured system diversity of the final policies rapidly converges to and then stably oscillates around its target level throughout the training process. For instance, the run with $\text{SND}_{\text{des}} = 0.3$ (red line) quickly settles near the 0.3 level, while the run with $\text{SND}_{\text{des}} = 0.9$ (light green line) consistently maintains its diversity around 0.9. This provides strong evidence that our structural redefinition of the SND metric does not compromise the fundamental control capability of the architectural constraint approach, validating SDC as a reliable method for precise diversity control.

C. Analysis of the Diversity Structure Factor (DSF)

In this section, we analyze the effect of the DSF(α), which is the core of our structural control mechanism. We

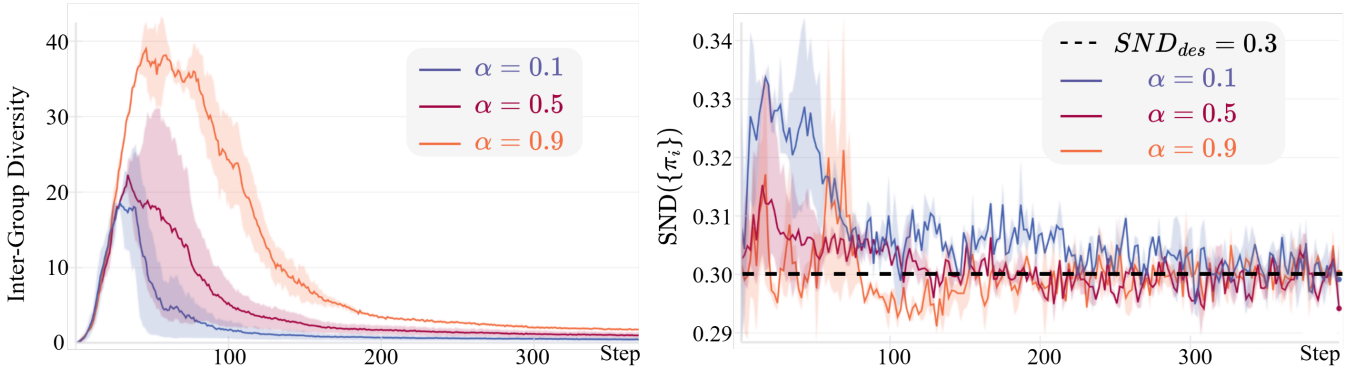

Fig. 3. Analysis of the Diversity Structure Factor (DSF, α). (**Left**) The effect of α on the inter-group diversity of the heterogeneous components. Higher α values, which place more weight on inter-group diversity, lead to a correspondingly higher and more sustained level of specialization between groups. (**Right**) Verification of total diversity control under different DSF settings. For a fixed target diversity of $\text{SND}_{\text{des}} = 0.3$, the measured system diversity $\text{SND}(\{\pi_i\})$ remains stable and accurate across all tested α values, demonstrating the robustness of our control mechanism.

conduct ablation studies with representative DSF values of $\alpha \in \{0.1, 0.5, 0.9\}$. Across these settings, the system maintains robust control over the total diversity. The results, presented in Fig. 3, confirm that the DSF acts as an intuitive and effective control knob for tuning the collective strategy. The left panel of Fig. 3 shows the inter-group diversity of the heterogeneous components throughout training. As hypothesized, a higher value of α directly correlates with a higher and more sustained level of inter-group diversity. This demonstrates that by increasing the weight on the inter-group component in our structured diversity metric, SDC successfully encourages greater specialization and a more pronounced division of labor between the agent groups.

Crucially, this fine-grained structural control does not compromise the overall stability of the system. The right panel of Fig. 3 shows the measured total system diversity for a fixed target of $\text{SND}_{\text{des}} = 0.3$ across the different α settings. The plot clearly indicates that SDC maintains robust and accurate control over the total diversity, regardless of how the internal diversity structure is weighted. This validates that the DSF provides a direct and interpretable mechanism for practitioners to inject task-specific priors about the desired balance between cohesion and specialization, without sacrificing the precision of the overall diversity control.

D. Performance Comparison in Group-Based Tasks

Having validated SDC's control mechanisms, we now evaluate its performance in tasks that benefit from structured coordination.

1) Results in Shielded Tag: In the Shielded Tag environment, where the need for structured coordination is even more critical, SDC's advantage becomes more pronounced. For this experiment, we set a target diversity of $\text{SND}_{\text{des}} = 0.3$ for both methods. As illustrated in the left panel of Fig. 4, SDC agents learn significantly more efficient strategies, achieving consistently shorter episode lengths compared to DiCo agents. This indicates a faster neutralization of all targets. The strategic superiority of SDC stems from its ability to foster an optimal "focus fire" strategy, which this task explicitly rewards. SDC's structured control mechanism, governed by the DSF, can be configured to simultaneously encourage high intra-group cohesion and high inter-group specialization. In contrast, DiCo's monolithic constraint struggles to find this delicate balance. It often results in agents spreading their attacks too thinly across multiple targets, which is a highly inefficient way to overcome the shield mechanic, thus leading to longer episodes.

Crucially, the right panel of Fig. 4 provides a critical piece of evidence: the observed performance gap is not due to a failure in diversity control. Both SDC and DiCo successfully steer their total system diversity to the specified target of 0.3 and maintain it throughout training. This isolates the source of the performance difference, providing compelling evidence that SDC's superiority is a direct result of its ability to organize diversity into a more effective, group-aware structure, which is essential for solving complex, structured cooperative tasks.

2) Results in Multi-Pursuer Tag: In the Multi-Pursuer Tag environment, as shown by the reward curves in Fig. 5, SDC consistently achieves higher average rewards across all three difficulty settings (5vs2, 6vs2, and 7vs2). SDC demonstrates a significant performance advantage over the monolithic control of DiCo.

The underlying reason for this superiority is revealed through qualitative analysis of the learned behaviors. The trajectory visualizations in Fig. 6 (5vs2), Fig. 7 (6vs2), and Fig. 8 (7vs2) consistently show a stark contrast in strategy. DiCo's agents, governed by a single global diversity constraint, frequently exhibit a "collapsing" behavior, where all pursuers myopically chase a single escapee. This leaves the second target completely uncontested, resulting in a highly suboptimal team strategy.

In stark contrast, SDC agents successfully learn to execute a structured, two-group strategy. As seen in the visualizations, they effectively partition themselves to simultaneously

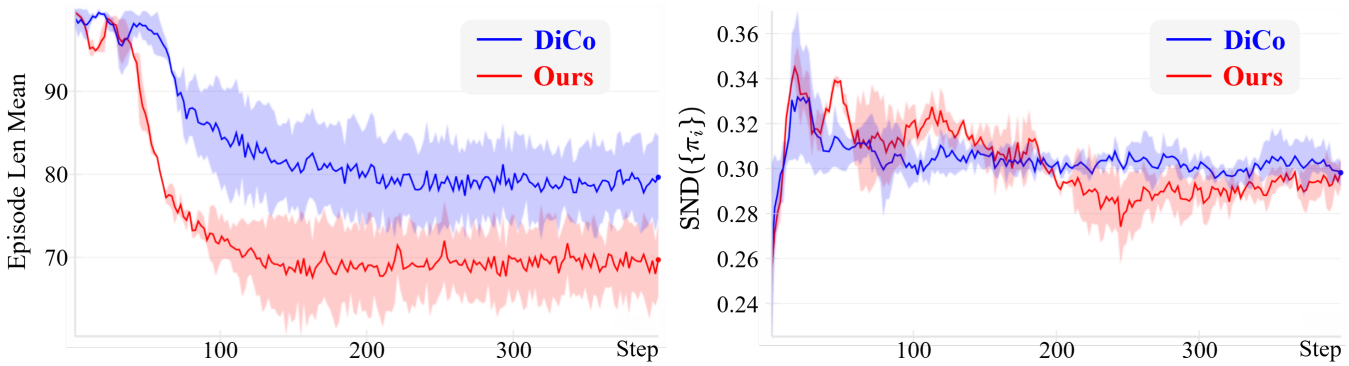

Fig. 4. Performance and diversity comparison in the Shielded Tag environment (7vs2 scenario) with a target diversity of $\text{SND}_{\text{des}} = 0.3$. (**Left**) Mean episode length over training. SDC (red) learns a superior strategy, completing the task significantly faster than DiCo (blue). (**Right**) Measured total system diversity, $\text{SND}(\{\pi_i\})$. Both methods successfully converge to and maintain the target diversity level of 0.3.

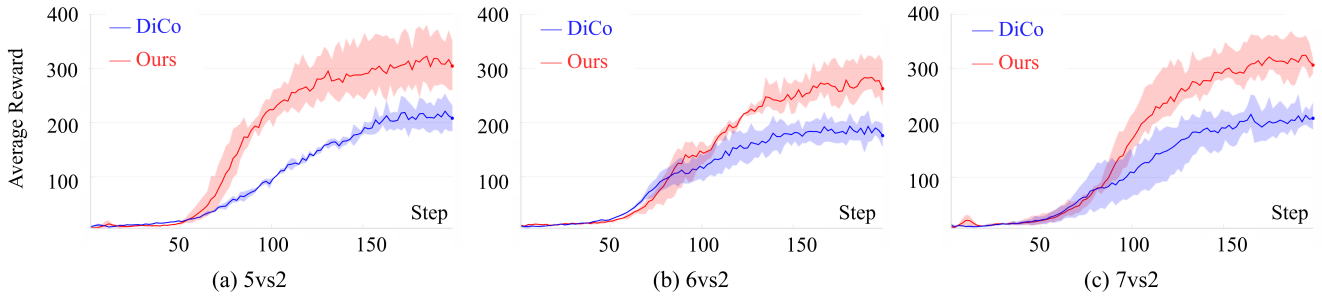

Fig. 5. Comparison of average rewards between DiCo and SDC across scenarios with a target diversity of $\text{SND}_{\text{des}} = 0.3$. (a) Comparison of Average Rewards for DiCo and SDC in "5 chasing 2" scenario. (b) Comparison of Average Rewards for DiCo and SDC in "6 chasing 2" scenario. (c) Comparison of Average Rewards for DiCo and SDC in "7 chasing 2" scenario.

Fig. 6. DiCo and SDC visualization process in "5 chasing 2" scenario. The four frames correspond to 0%, 25%, 75%, and 100% of the episode's total duration, respectively.

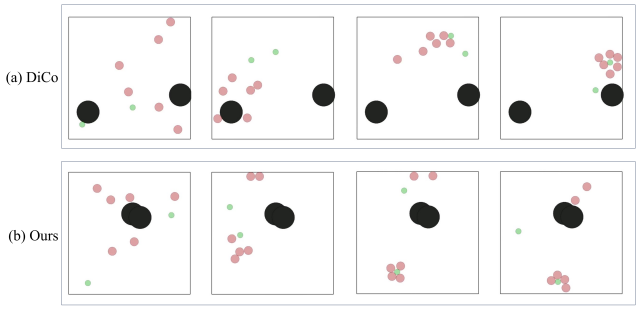

Fig. 7. DiCo and SDC visualization process in "6 chasing 2" scenario. The four frames correspond to 0%, 25%, 75%, and 100% of the episode's total duration, respectively.

pursue both targets. This emergent division of labor is a direct result of our structured diversity control. To further validate this, we plot the inter-group diversity of the heterogeneous components in Fig. 9. The results confirm that SDC maintains a significantly higher level of inter-group diversity compared to DiCo. By enabling this structured control that encourages inter-group specialization, SDC facilitates a more effective division of labor, which is critical for success in this multi-target scenario. This strategic advantage becomes even more pronounced as the number of pursuers increases (Fig. 8), where SDC's ability to organize agents into cohesive pursuit groups prevents redundant chasing and maximizes spatial coverage, leading to better performance.

To provide a quantitative basis for this strategic difference, we further analyze the number of collisions per episode, which in this context represent successful captures

Fig. 8. DiCo and SDC visualization process in "7 chasing 2" scenario. The four frames correspond to 0%, 25%, 75%, and 100% of the episode's total duration, respectively.

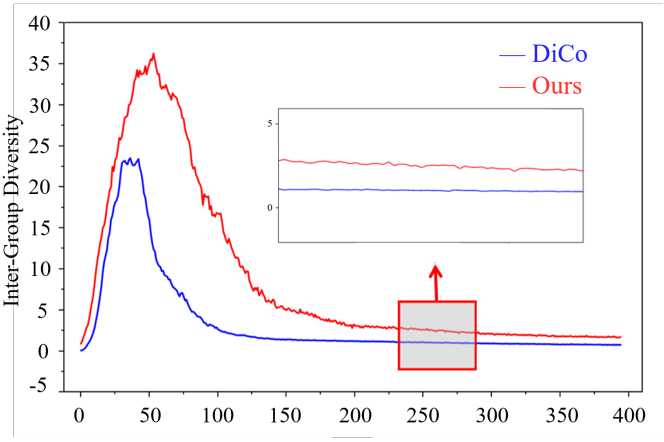

Fig. 9. Comparison of inter-group diversity between DiCo and SDC with the same target diversity.

of the escapees. As illustrated in Fig. 10, the violin plots show the distribution of total collisions per episode for SDC (red) and DiCo (blue). SDC achieves a significantly higher number of collisions than DiCo across all scenarios, increasing the average collision count from approximately 41.7 to 61.3, which corresponds to a 47.0% improvement. This discrepancy stems directly from the observed strategies: SDC's effective division of labor allows its two groups to engage both targets in parallel, creating more opportunities for captures within the same timeframe, directly explaining the higher average rewards reported in Fig. 5.

V. CONCLUSION

In this paper, we introduced Structured Diversity Control (SDC), a novel framework for structured behavioral diversity control in grouped multi-agent reinforcement learning that simultaneously achieves cohesion and specialization. SDC decomposes system-wide diversity into intra-group and inter-group components through architectural constraints controlled by the Diversity Structure Factor, enabling fine-grained control over group coordination without modifying learning objectives.

Extensive experimental evaluation across multiple cooperative scenarios demonstrates that SDC achieves superior performance compared to strong baselines while maintaining

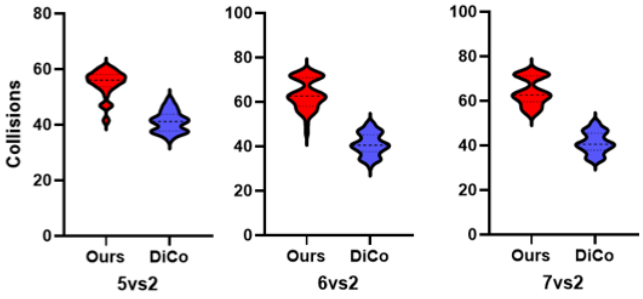

Fig. 10. Quantitative comparison of successful captures in the Multi-Pursuer Tag scenarios.

precise diversity control and interpretable group structures. These results validate the importance of structured diversity control for complex multi-agent coordination tasks. SDC holds significant promise for applications in scenarios requiring group-level team coordination, including multi-robot logistics, autonomous drone swarms, intelligent manufacturing, and distributed sensing systems. The structured diversity control principles validated in this work provide a promising pathway toward solving large-scale, heterogeneous coordination problems, potentially enabling breakthrough applications across diverse domains and facilitating the transition of multi-agent systems from theoretical foundations to practical engineering deployment.

For future work, we will focus on exploring dynamic DSF adaptation mechanisms, designing data-driven grouping strategies, and validating the proposed framework in real-world scenarios such as multi-robot coordination and UAV swarms. We believe that the structural diversity control principle established in this work provides a solid foundation for advancing more sophisticated and effective multi-agent strategies.

VI. ACKNOWLEDGMENT

This work was supported by National Natural Science Foun dation of China (62376219); Faculty Construction Project (Grant No. 25SH02010044).

REFERENCES

[1] S. Yan, L. König, and W. Burgard, "Agent-agnostic centralized training for decentralized multi-agent cooperative driving," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1002–1009, IEEE, 2024.

[2] E. Candela, L. Parada, L. Marques, T.-A. Georgescu, Y. Demiris, and P. Angeloudis, "Transferring multi-agent reinforcement learning policies for autonomous driving using sim-to-real," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8814–8820, IEEE, 2022.

[3] H. Hu, E. Shi, C. Yue, S. Yang, Z. Wu, Y. Li, T. Zhong, T. Zhang, T. Liu, and S. Zhang, "Harp: Human-assisted regrouping with permutation invariant critic for multi-agent reinforcement learning," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4287–4293, IEEE, 2025.

[4] A. Mahajan, T. Rashid, M. Samvelyan, and S. Whiteson, "Maven: Multi-agent variational exploration,"

[5] T. Wang, H. Dong, V. Lesser, and C. Zhang, "Roma: Multi-agent reinforcement learning with emergent roles," *International Conference on Machine Learning,International Conference on Machine Learning*, Jul 2020.

[6] J. Zhang, Y. Zhang, X. S. Zhang, Y. Zang, and J. Cheng, "Intrinsic action tendency consistency for cooperative multi-agent reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 17600–17608, 2024.

[7] M. Bettini, A. Shankar, and A. Prorok, "System neural diversity: Measuring behavioral heterogeneity in multi-agent learning," *arXiv preprint arXiv:2305.02128*, 2023.

[8] M. Bettini, R. Kortvelesy, and A. Prorok, "Controlling behavioral diversity in multi-agent reinforcement learning," *arXiv preprint arXiv:2405.15054*, 2024.

[9] I.-J. Liu, Z. Ren, R. A. Yeh, and A. G. Schwing, "Semantic tracklets: An object-centric representation for visual multi-agent reinforcement learning," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5603–5610, IEEE, 2021.

[10] N. Jaques, A. Lazaridou, E. Hughes, C. Gulcehre, P. Ortega, D. Strouse, J. Leibo, and N. Freitas, "Social influence as intrinsic motivation for multi-agent deep reinforcement learning," *arXiv: Learning,arXiv: Learning*, Oct 2018.

[11] T. Wang, J. Wang, Y. Wu, and C. Zhang, "Influence-based multi-agent exploration," *arXiv: Learning,arXiv: Learning*, Oct 2019.

[12] C. Li, T. Wang, C. Wu, Q. Zhao, J. Yang, and C. Zhang, "Celebrating diversity in shared multi-agent reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3991–4002, 2021.

[13] Z. Hu, D. Shishika, X. Xiao, and X. Wang, "Bi-cl: A reinforcement learning framework for robots coordination through bi-level optimization," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 581–586, IEEE, 2024.

[14] P. Feng, J. Liang, S. Wang, X. Yu, X. Ji, Y. Chen, K. Zhang, R. Shi, and W. Wu, "Hierarchical consensus-based multi-agent reinforcement learning for multi-robot cooperation tasks," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 642–649, IEEE, 2024.

[15] S. Nikkhoo, Z. Li, A. Samanta, Y. Li, and C. Liu, "Pimbot: Policy and incentive manipulation for multi-robot reinforcement learning in social dilemmas," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5630–5636, IEEE, 2023.

[16] M. Shen and J. P. How, "Safe adaptation in multiagent competition," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 12441–12447, IEEE, 2022.

[17] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 3, pp. 1086–1095, 2019.

[18] J. Bloom, P. Paliwal, A. Mukherjee, and C. Pinciroli, "Decentralized multi-agent reinforcement learning with global state prediction," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8854–8861, IEEE, 2023.

[19] Y. Chen, J. Mao, Y. Zhang, D. Ma, L. Xia, J. Fan, D. Shi, Z. Cheng, S. Gu, and D. Yin, "Ma4div: Multi-agent reinforcement learning for search result diversification," *arXiv preprint arXiv:2403.17421*, 2024.

[20] A. Esmaeili, N. Mozayani, M. R. J. Motlagh, and E. T. Matson, "The impact of diversity on performance of holonic multi-agent systems," *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 186–201, 2016.

[21] N. Suryanarayanan and I. Hitoshi, "Diversifying experiences in multi agent reinforcement learning," in *2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA)*, pp. 47–52, IEEE, 2019.

[22] Z. Wang, S. Feng, D. Wang, K. Song, G. Wu, Y. Zhang, H. Zhao, and G. Yu, "Diversity-enhanced conversational recommendation via multi-agent reinforcement learning," 2024.

[23] B. Eysenbach, A. Gupta, J. Ibarz, and S. Levine, "Diversity is all you need: Learning skills without a reward function," *arXiv preprint arXiv:1802.06070*, 2018.

[24] B. Baker, I. Kanitscheider, T. Markov, Y. Wu, G. Powell, B. McGrew, and I. Mordatch, "Emergent tool use from multi-agent autocurricula," in *International conference on learning representations*, 2019.

[25] M. Jaderberg, V. Dalibard, S. Osindero, W. M. Czarnecki, J. Donahue, A. Razavi, O. Vinyals, T. Green, I. Dunning, K. Simonyan, *et al.*, "Population based training of neural networks," *arXiv preprint arXiv:1711.09846*, 2017.

[26] T. Bansal, J. Pachocki, S. Sidor, I. Sutskever, and I. Mordatch, "Emergent complexity via multi-agent competition," *arXiv preprint arXiv:1710.03748*, 2017.

[27] H. Ryu, H. Shin, and J. Park, "Multi-agent actor-critic with hierarchical graph attention network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 7236–7243, 2020.

[28] Z. Song, R. Zhang, and X. Cheng, "Helsa: Hierarchical reinforcement learning with spatiotemporal abstraction for large-scale multi-agent path finding," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7318–7325, IEEE, 2023.

[29] J. Bloom, P. Paliwal, A. Mukherjee, and C. Pinciroli, "Decentralized multi-agent reinforcement learning with global state prediction," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8854–8861, IEEE, 2023.

[30] R. Makar, S. Mahadevan, and M. Ghavamzadeh, "Hierarchical multi-agent reinforcement learning," in *Proceedings of the fifth international conference on Autonomous agents*, pp. 246–253, 2001.

[31] M. Meindl, F. Molinari, D. Lehmann, and T. Seel, "Collective iterative learning control: Exploiting diversity in multi-agent systems for reference tracking tasks," *IEEE Transactions on Control Systems Technology*, vol. 30, no. 4, pp. 1390–1402, 2021.

[32] T. Sarkar, "Cpper: A controlled partial prioritized experience replay for reinforcement learning in its multi-agent extension," in *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, vol. 1, pp. 1–6, IEEE, 2023.

[33] S. Sun and K. Xu, "Temporal inconsistency-based intrinsic reward for multi-agent reinforcement learning," in *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, IEEE, 2023.

[34] J. Zhang and M. Cao, "Strategy competition dynamics of multi-agent systems in the framework of evolutionary game theory," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 1, pp. 152–156, 2019.

[35] Y. Jwa, M. Gwak, J. Kwak, C. W. Ahn, and P. Park, "Scalable robust multi-agent reinforcement learning for model uncertainty," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 3402–3407, IEEE, 2023.

[36] L. Canese, G. C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, M. Re, and S. Spanò, "Multi-agent reinforcement learning: A review of challenges and applications," *Applied Sciences*, vol. 11, no. 11, p. 4948, 2021.

[37] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of markov decision processes," *Mathematics of operations research*, vol. 27, no. 4, pp. 819–840, 2002.

[38] C. S. De Witt, T. Gupta, D. Makoviichuk, V. Makoviychuk, P. H. Torr, M. Sun, and S. Whiteson, "Is independent learning all you need in the starcraft multi-agent challenge?," *arXiv preprint arXiv:2011.09533*, 2020.