# Coordination Connectivity: Shared Initialization Shapes the Joint-Policy Landscape in MARL

**Yixiang Fan**[1,2,3] **Dongmin Li**[1,3] **Weihao Zhang**[1,3] **Hao Ma**[1,3]
Yinxing Dai[1,3] Shijie Zheng[1,2,3] Jingjing Huang[1,3] Zhiqiang Pu[1,3*]

[1]National Key Laboratory of Cognition and Decision Intelligence for Complex Systems, Institute of Automation, Chinese Academy of Sciences
[2]School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences [3]National Key Laboratory of Information Systems Engineering

## Abstract

Mode connectivity studies reveal that high-performing neural networks are often connected by low-loss paths in parameter space. We investigate the analogue of this phenomenon in cooperative multi-agent reinforcement learning (MARL), where a solution is not a single model but a team of decentralized policies. We introduce *coordination connectivity*, a diagnostic evaluating whether linear interpolation between the actor parameters of two trained policy teams preserves high team return. The resulting *coordination barrier* measures whether two successful teams are separated by a performance cliff along the joint-policy path. We evaluate this diagnostic on the StarCraft Multi-Agent Challenge (SMAC) across five controlled training histories, isolating the effects of shared initialization, continued training, parameter decoupling, and HAPPO-style separate-actor specialization. Across three SMAC maps and four seeds, policy teams derived from a shared MAPPO checkpoint exhibit near-zero interpolation barriers, even after non-shared specialization. In contrast, interpolation paths involving independently trained HAPPO teams exhibit substantially larger barriers. Mixed-team assembly, single-agent hot-swap, clone-all, role-swap, and weight-space probes further indicate that shared-initialized agents specialize into distinct roles while remaining within a low-barrier coordination component. These results suggest that shared initialization implicitly structures the joint-policy space of cooperative MARL, providing a mode-connectivity-inspired lens for analyzing coordination, specialization, and same-source cross-version team compatibility.

## 1 Introduction

Deep neural networks are optimized through highly non-convex objectives, yet stochastic gradient methods often find high-performing solutions in practice. This tension has motivated the study of *mode connectivity*: whether independently trained neural networks occupy isolated optima, or whether they are connected by low-loss paths in parameter space [Garipov et al., 2018, Draxler et al., 2018]. A robust body of literature reveals that successful networks can often be connected by simple curves or even linear paths under appropriate training histories or alignments [Frankle et al., 2020, Kuditipudi et al., 2019, Benton et al., 2021, Entezari et al., 2022, Ainsworth et al., 2023]. Beyond its theoretical significance for understanding loss landscapes [Li et al., 2018, Fort et al., 2019], this perspective has catalyzed practical advances in model ensembling, weight merging, and permutation alignment [Wortsman et al., 2022, Ilharco et al., 2023, Matena and Raffel, 2022, Yadav et al., 2023, Jordan et al., 2023, Singh and Jaggi, 2020].

In cooperative multi-agent reinforcement learning (MARL), an analogous geometric question remains largely unexplored. During execution, a MARL solution is not a single centralized controller, but

*Corresponding author: zhiqiang.pu@ia.ac.cn

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

a *team* of decentralized policies whose success depends on how agent behaviors interlock. Two trained teams may both achieve high return in the same environment while relying on substantially different role assignments, timing patterns, target priorities, or other learned coordination conventions. Interpolating their actor parameters may therefore produce intermediate teams that implement neither convention, leading to sharp drops in team return. This suggests a multi-agent analogue to mode connectivity: are high-performing MARL teams isolated coordination solutions, or are they connected by low-barrier paths in joint-policy space?

To investigate this, we introduce *coordination connectivity*, a diagnostic framework for evaluating the joint-policy landscape of MARL teams. Given two trained policy teams with aligned actor architectures, we linearly interpolate their actor parameters agent by agent:

$$\theta_i(\alpha) = (1-\alpha)\theta_i^A + \alpha\theta_i^B, \quad i \in \{1, \ldots, N\}.$$

Here, $\theta_i^A$ and $\theta_i^B$ denote the actor parameters of agent $i$ in teams A and B, respectively; $N$ is the number of agents; and $\alpha \in [0, 1]$ is the interpolation coefficient. We then evaluate the team return along the resulting joint-policy path. The induced *coordination barrier* measures the largest performance drop along this path relative to the weaker endpoint. A near-zero barrier indicates that two teams are connected by a high-return interpolation path, whereas a substantial barrier exposes a coordination cliff. Unlike standard mode connectivity, the object here is a team of policies, and the evaluated metric is team return rather than supervised loss.

We deploy this diagnostic to analyze a central tension in cooperative MARL: parameter sharing versus heterogeneous specialization. Parameter sharing is ubiquitous for its sample efficiency and stability [Gupta et al., 2017, Terry et al., 2020, Yu et al., 2022], while non-shared policies and role-specialized methods encourage agent-specific behaviors [Kuba et al., 2022, Wang et al., 2020, 2021b, Li et al., 2021, 2024]. Historically, these paradigms are compared via endpoint performance. We ask a complementary, geometric question: when agents specialize into distinct roles after a shared initialization, do they remain connected to the shared source checkpoint, or do they become separated by substantial coordination barriers? By answering this question, we evaluate *same-source cross-version compatibility*: compatibility among policies that share an ancestral initialization but diverge in later checkpoints, parameter-sharing modes, or continuation algorithms. We use coordination connectivity as a post-hoc diagnostic to reveal geometric structure between trained teams that endpoint performance alone cannot expose.

To answer this, we conduct a controlled empirical study on the StarCraft Multi-Agent Challenge (SMAC) [Samvelyan et al., 2019]. We compare five training histories: shared MAPPO, scratch separate-actor HAPPO, continued shared MAPPO, MAPPO-initialized separate-actor MAPPO, and MAPPO-initialized separate-actor HAPPO (M2H). Across three maps and four seeds, teams derived from a shared MAPPO initialization exhibit near-zero interpolation barriers, even after separate-actor specialization. In contrast, paths involving independently trained HAPPO teams exhibit substantially larger barriers.

To determine whether low-barrier connectivity coexists with agent specialization, we conduct behavioral and structural probes. Mixed-team evaluations show that same-source combinations retain high win rates, while cross-history mixtures involving HAPPO-Full drop substantially. Hot-swap evaluations confirm that single-agent replacement by the MAPPO-Shared actor preserves high performance for same-source specialists. Clone-all and role-swap probes further demonstrate that specialized agents are not interchangeable copies. Together with weight-space analysis, these probes indicate that shared-initialized agents specialize while remaining within a low-barrier coordination component.

Our contributions are as follows:

1. We formulate *coordination connectivity* and the *coordination barrier* as a mode-connectivity-inspired diagnostic framework for cooperative MARL teams.
2. We demonstrate empirically across three SMAC maps, four seeds, and seven targeted interpolation paths that shared-initialized histories form low-barrier paths ($B \leq 0.13$), whereas cross-history paths involving independently trained HAPPO exhibit substantially larger barriers ($B \geq 0.42$).
3. We provide behavioral and structural evidence—via mixed-team assembly, single-agent hot-swap, clone-all, role-swap, and weight-space probes—that low-barrier connectivity coexists with role-sensitive specialization.

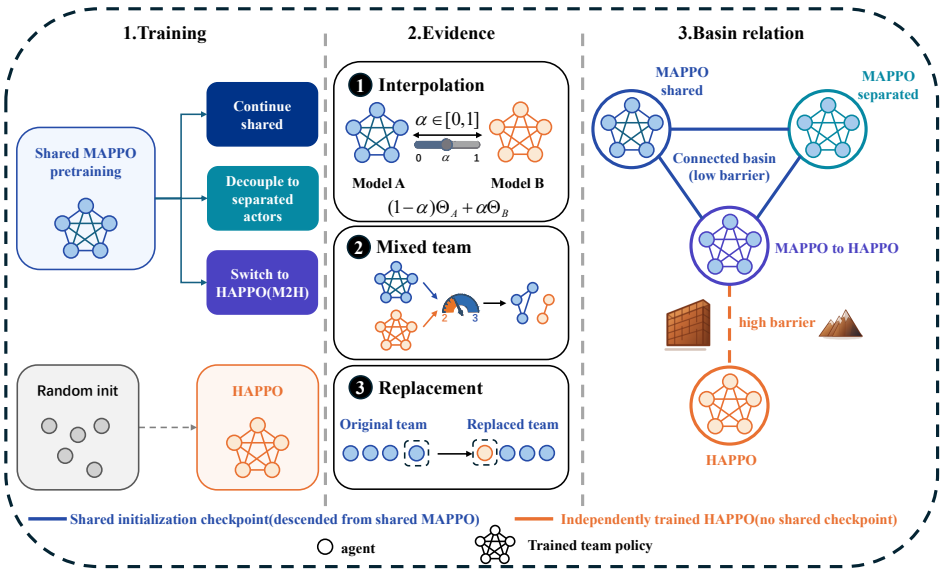



Figure 1: Overview of coordination connectivity. **Left**: Five training histories branch from two initialization sources—a shared MAPPO checkpoint and random initialization. **Center**: Post-hoc diagnostic protocols probe the joint-policy landscape: actor-weight interpolation, mixed-team assembly, and single-agent replacement. **Right**: Shared-initialized endpoints form a low-barrier coordination component, while independently trained HAPPO endpoints are separated by high coordination barriers.

4. We connect parameter sharing, heterogeneous specialization, and same-source teammate compatibility through a joint-policy landscape perspective, providing a diagnostic lens for MARL training dynamics.

## 2 Related work

**Mode connectivity and model merging.** Mode connectivity studies whether high-performing neural networks are connected by low-loss paths in parameter space [Garipov et al., 2018, Draxler et al., 2018]. Subsequent work has shown that such paths can be linear under shared early training trajectories, stability conditions, or appropriate parameter alignments [Frankle et al., 2020, Kuditipudi et al., 2019, Benton et al., 2021, Entezari et al., 2022, Ainsworth et al., 2023]. These insights have enabled model soups, task arithmetic, Fisher merging, TIES merging, and related fusion methods [Wortsman et al., 2022, Ilharco et al., 2023, Matena and Raffel, 2022, Yadav et al., 2023, Jordan et al., 2023, Singh and Jaggi, 2020]. Our work adopts the interpolation-based diagnostic from this literature but changes both the object and the metric: instead of evaluating the loss or accuracy of a single network, we measure the team performance of a joint policy formed by multiple decentralized actors.

**Parameter sharing and heterogeneity in cooperative MARL.** Many cooperative MARL algorithms operate within centralized training with decentralized execution (CTDE), including actor-critic methods, value factorization, and PPO-style policy optimization [Lowe et al., 2017, Foerster et al., 2018, Sunehag et al., 2018, Rashid et al., 2018, Wang et al., 2021a, Yu et al., 2022]. Parameter sharing improves sample efficiency and exploits exchangeability among homogeneous agents [Gupta et al., 2017, Terry et al., 2020, Christianos et al., 2021]. Heterogeneous methods instead facilitate agent-specific behaviors through non-shared optimization, role discovery, diversity objectives, or learnable sharing masks [Kuba et al., 2022, Zhong et al., 2024, Wang et al., 2020, 2021b, Li et al., 2021, 2024]. These works primarily evaluate endpoint performance. We instead study geometric relations *between* trained teams.

**Zero-shot coordination and ad-hoc teamwork.** Zero-shot coordination (ZSC) and ad-hoc teamwork focus on cooperation with novel partners not encountered during training [Stone et al., 2010,

Mirsky et al., 2022]. A central obstacle is convention mismatch: multiple incompatible policies may solve the same cooperative task but suffer large cross-play drops. Existing approaches mitigate this through symmetry-aware training, population diversity, belief modeling, or partner-generalization objectives [Hu et al., 2020, Lupu et al., 2021, Lucas and Allen, 2022, Hu et al., 2021a, Treutlein et al., 2021, Strouse et al., 2021]. We study policies derived from a shared source checkpoint and measure whether they remain connected under interpolation, mixing, and replacement.

**Landscape diagnostics for multi-agent coordination.** Our work treats coordination connectivity as a post-hoc landscape diagnostic for trained MARL teams. Standard MARL evaluation asks whether a final team performs well. Coordination connectivity asks a different question: whether two trained teams can be connected by a path of high team performance after training has ended. This distinction is important because cooperative MARL can admit multiple successful but mutually incompatible conventions. A team may achieve high win rate by adopting a particular assignment of roles, timing patterns, or target priorities; another team may solve the same task with a different convention. Endpoint evaluation alone cannot reveal whether these conventions are compatible under interpolation or replacement. Our diagnostic is therefore complementary to endpoint-performance comparisons, partner-generalization benchmarks, and post-hoc model-merging analyses.

**Relation to role discovery and diversity.** Prior work on roles and diversity explicitly encourages heterogeneous behaviors, often through latent role variables, diversity regularization, or task decomposition [Wang et al., 2020, 2021b, Li et al., 2021, 2024]. Our study asks whether low-barrier connectivity persists after agents become non-shared and behaviorally specialized. Clone-all and role-swap probes test whether the non-shared endpoints exhibit role-sensitive specialization while remaining compatible with a shared ancestral policy.

## 3 Method

We formalize *coordination connectivity* as a multi-agent analogue of linear mode connectivity. Rather than interpolating a single neural network, we interpolate a team of decentralized actor policies and evaluate a scalar team-performance metric along the resulting joint-policy path.

### 3.1 Cooperative MARL setup

We consider a finite-horizon cooperative partially observable Markov game with $N$ agents. At time step $t$, each agent $i$ receives a local action-observation history $\tau_t^i$ and selects an action $a_t^i$ using a decentralized actor policy $\pi_{\theta_i}(a_t^i \mid \tau_t^i)$. The agents execute the joint action $\mathbf{a}_t = (a_t^1, \ldots, a_t^N)$ and receive a shared reward $r_t$. A policy team is parameterized by $\Theta = (\theta_1, \ldots, \theta_N)$, where actor parameters may be agent-specific or shared across agents. Let $\boldsymbol{\tau}_t = (\tau_t^1, \ldots, \tau_t^N)$ denote the joint collection of local histories. Under decentralized execution, the joint action distribution factorizes as:

$$\Pi_\Theta(\mathbf{a}_t \mid \boldsymbol{\tau}_t) = \prod_{i=1}^{N} \pi_{\theta_i}(a_t^i \mid \tau_t^i). \tag{1}$$

We use $J(\Theta)$ to denote the scalar team-performance metric used by the diagnostic. A canonical choice in Markov games is the expected finite-horizon return:

$$J_{\text{return}}(\Theta) = \mathbb{E}_{\rho_0, P, \Pi_\Theta}\left[\sum_{t=0}^{H-1} r_t\right], \tag{2}$$

where $H$ is the episode horizon, $\rho_0$ is the initial-state distribution, $P$ denotes the environment dynamics, and the expectation is over trajectories induced by $\Pi_\Theta$. In our SMAC experiments, we instantiate $J(\Theta)$ as test win rate, normalized to $[0, 1]$ for barrier computation.

### 3.2 Agent-wise actor interpolation

Given two trained policy teams

$$\Theta_A = (\theta_1^A, \ldots, \theta_N^A), \qquad \Theta_B = (\theta_1^B, \ldots, \theta_N^B),$$

with identical actor architectures and input-output interfaces, including the same observation preprocessing, action space, recurrent hidden size, and MLP dimensions, we align actors by agent index and define the interpolated team at coefficient $\alpha \in [0, 1]$ as:

$$\Theta(\alpha) = (\theta_1(\alpha), \ldots, \theta_N(\alpha)), \qquad \theta_i(\alpha) = (1 - \alpha)\theta_i^A + \alpha\theta_i^B. \tag{3}$$

Interpolation is applied only to floating-point actor parameters. Critics, optimizer states, replay buffers, and training statistics are not interpolated because decentralized evaluation relies solely on actors. For recurrent actors, hidden states are reset at episode start and then evolved by the interpolated actor under the standard evaluation protocol; hidden states themselves are not interpolated. When one endpoint originates from a parameter-shared algorithm, we replicate its shared actor weights across all agent indices before applying Eq. (3). For separate-actor endpoints, actors are aligned by native agent index. This fixed alignment matches the execution semantics of the trained teams and avoids a post-hoc assignment search that would conflate coordination compatibility with permutation optimization. We examine sensitivity to agent-index changes through role-swap probes in Section 4.4.

### 3.3 The coordination barrier

We define the *coordination barrier* between two policy teams as the worst performance drop along the interpolation path, measured relative to the weaker endpoint:

$$B(\Theta_A, \Theta_B) = \max\left\{0,\ \min\left\{J(\Theta_A), J(\Theta_B)\right\} - \min_{\alpha\in[0,1]} J(\Theta(\alpha))\right\}. \tag{4}$$

The outer maximum ensures that the barrier is non-negative; if the interpolation path outperforms the weaker endpoint, we treat the barrier as zero rather than negative. We say two teams are $\epsilon$-*coordination-connected* if $B(\Theta_A, \Theta_B) \leq \epsilon$. In practice, we report a sampled estimate of $B$ on the grid $\alpha \in \{0.0, 0.1, \ldots, 1.0\}$. Appendix B gives an informal behavioral intuition for low barriers.

### 3.4 Diagnostic scope and design controls

Coordination connectivity is a post-hoc diagnostic of trained teams under fixed agent-index linear interpolation. A low barrier indicates that the sampled path preserves team performance for the aligned pair of trained teams. Compatibility under arbitrary agent permutations or with independently trained partners lies outside this diagnostic. A high barrier identifies a low-performing region along the interpolation path relative to the endpoint performance.

This distinction is central to the experimental design. The five histories in Table 1 isolate different possible explanations. MAPPO-Shared controls for additional training after the shared MAPPO checkpoint. MAPPO-Sep controls for parameter decoupling without changing to HAPPO updates. M2H tests whether HAPPO-style separate-actor specialization after a shared source preserves connectivity. HAPPO-Full tests a separately initialized heterogeneous endpoint with no shared MAPPO ancestor. MAPPO-Full provides the corresponding fully trained shared-policy endpoint from random initialization. Together, these controls isolate the key comparison between same-source specialization and independently trained heterogeneous specialization.

## 4 Experiments

We evaluate whether shared initialization induces low-barrier coordination paths in cooperative MARL. The experiments proceed from interpolation, to pairwise barrier structure, to functional compatibility, and finally to specialization, structural, and behavioral probes.

### 4.1 Setup: controlled training histories and evaluation protocol

**Environments and training histories.** We evaluate on three SMAC maps [Samvelyan et al., 2019]: `3s5z`, `corridor`, and `MMM2`. To isolate the effect of shared initialization, we compare five controlled training histories, summarized in Table 1. Throughout, *separate actors* denotes agent-specific actor networks with no parameter sharing; all methods are trained under the corresponding CTDE procedure and are not independent learners.

Table 1: Controlled training histories. **MAPPO-Pre** denotes the shared MAPPO checkpoint used as initialization for the three continued-training methods.

| Name | Initialization | Continued training | Role |
|---|---|---|---|
| MAPPO-Full | random | MAPPO, shared actor | shared-policy endpoint |
| HAPPO-Full | random | HAPPO, separate actors | independent heterogeneous endpoint |
| MAPPO-Shared | MAPPO-Pre | MAPPO, shared actor | extra-training control |
| MAPPO-Sep | MAPPO-Pre | MAPPO, separate actors | decoupling control |
| M2H | MAPPO-Pre | HAPPO, separate actors | specialization path |

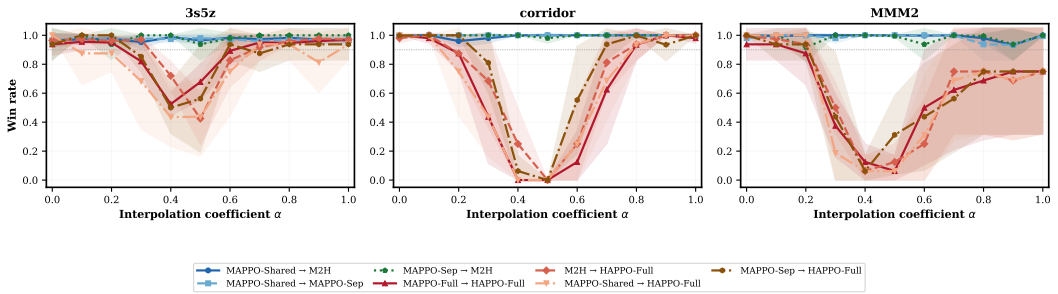



Figure 2: Interpolation win-rate curves for seven paths across three SMAC maps. Same-source paths (P1, P2, P6; blue/green) maintain high win rate along the interpolation range. Cross-history paths involving HAPPO-Full (P3–P5, P7; red/brown) exhibit mid-path coordination cliffs. Shaded regions indicate $\pm 1$ cross-seed standard deviation over four seeds.

All methods use identical actor architectures within each map. Experiments are repeated over four random seeds per map. Training uses MAPPO [Yu et al., 2022] and HAPPO [Kuba et al., 2022] implementations from HARL [Zhong et al., 2024]. For interpolation, we evaluate actor-only parameter interpolation on $\alpha \in \{0.0, 0.1, \ldots, 1.0\}$ with greedy action selection and 100 evaluation episodes for each map, seed, and interpolation point. We evaluate seven targeted paths: three same-source paths (P1, MAPPO-Shared $\rightarrow$ M2H; P2, MAPPO-Shared $\rightarrow$ MAPPO-Sep; P6, MAPPO-Sep $\rightarrow$ M2H) and four cross-history paths involving HAPPO-Full (P3, MAPPO-Full $\rightarrow$ HAPPO-Full; P4, M2H $\rightarrow$ HAPPO-Full; P5, MAPPO-Shared $\rightarrow$ HAPPO-Full; P7, MAPPO-Sep $\rightarrow$ HAPPO-Full).

**Evaluation philosophy.** All interpolation and compatibility probes are applied after training, without additional fine-tuning, so that they measure compatibility already present between the learned teams. We keep agent-index alignment fixed, use the same decentralized actor interfaces, and evaluate only actor parameters. Under this definition, teams that require post-hoc permutation search, additional optimization, or retraining fall outside fixed-index linear connectivity.

### 4.2 Shared initialization preserves low-barrier interpolation

Figure 2 shows interpolation win-rate curves for all seven paths across three maps. All same-source paths (P1, P2, P6) remain close to endpoint-level performance across the interpolation range. In contrast, cross-history paths involving HAPPO-Full (P3–P5, P7) exhibit pronounced mid-path cliffs. The effect is strongest on `corridor`, where every cross-history path collapses to near-zero win rate around $\alpha \approx 0.4$–$0.5$, even though the endpoint teams are high-performing on this map.

Table 2 summarizes the corresponding coordination barriers. Same-source paths have near-zero barriers across all maps ($B \leq 0.13$), while all cross-history paths exhibit substantially larger barriers. On `corridor`, cross-history barriers reach $B \approx 1.0$, meaning near-total coordination collapse at mid-path. On `MMM2`, larger cross-seed variance suggests that exact barrier magnitudes should be interpreted with caution. The repeated $0.69$ mean for HAPPO-Full-related paths on `MMM2` arises from the high seed-level variance of the shared HAPPO-Full endpoint and is explained in Appendix C.1.

Table 2: Coordination barriers $B(\Theta_A, \Theta_B)$ across three SMAC maps, reported as mean $\pm$ cross-seed standard deviation over four seeds. Barriers use win rate normalized to $[0, 1]$. The identical `MMM2` means for the HAPPO-Full-related paths reflect seed-level endpoint instability; see Appendix C.1.

| Path | Description | `3s5z` | `corridor` | `MMM2` |
|---|---|---|---|---|
| *Same-source paths* | | | | |
| P1 | MAPPO-Shared $\rightarrow$ M2H | $0.01 \pm 0.02$ | $0.06 \pm 0.03$ | $0.08 \pm 0.10$ |
| P2 | MAPPO-Shared $\rightarrow$ MAPPO-Sep | $0.02 \pm 0.01$ | $0.00 \pm 0.00$ | $0.11 \pm 0.11$ |
| P6 | MAPPO-Sep $\rightarrow$ M2H | $0.12 \pm 0.12$ | $0.02 \pm 0.03$ | $0.13 \pm 0.12$ |
| *Cross-history paths* | | | | |
| P3 | MAPPO-Full $\rightarrow$ HAPPO-Full | $0.42 \pm 0.02$ | $0.98 \pm 0.03$ | $0.69 \pm 0.41$ |
| P4 | M2H $\rightarrow$ HAPPO-Full | $0.53 \pm 0.19$ | $0.98 \pm 0.03$ | $0.69 \pm 0.41$ |
| P5 | MAPPO-Shared $\rightarrow$ HAPPO-Full | $0.69 \pm 0.20$ | $1.00 \pm 0.00$ | $0.69 \pm 0.41$ |
| P7 | MAPPO-Sep $\rightarrow$ HAPPO-Full | $0.56 \pm 0.27$ | $1.00 \pm 0.00$ | $0.69 \pm 0.41$ |

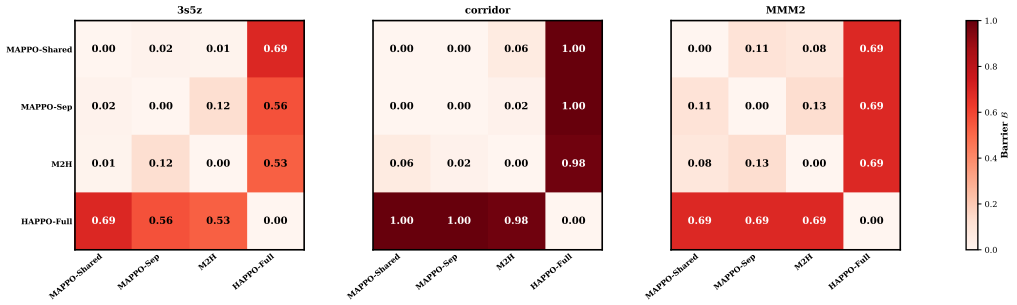



Figure 3: Pairwise coordination barrier matrix across SMAC maps. Each cell reports $B(\Theta_A, \Theta_B)$ averaged over four seeds. Shared-initialized histories form a low-barrier block, while paths involving HAPPO-Full have substantially larger barriers.

### 4.3 Pairwise barrier structure

The interpolation curves and barrier table describe targeted paths. We next summarize the same phenomenon as a pairwise barrier structure among MAPPO-Shared, MAPPO-Sep, M2H, and HAPPO-Full. Figure 3 shows that the three histories derived from MAPPO-Pre form a low-barrier block, while entries involving HAPPO-Full are consistently high. This matrix view shows that low barriers recur across multiple same-source paths, with the shared-initialized histories forming a small connected component under the diagnostic.

The matrix also clarifies the role of HAPPO-Full. M2H and MAPPO-Sep also use separate actors while remaining low-barrier with MAPPO-Shared and with each other. This comparison points to training ancestry as the main distinction. Same-source separate actors remain connected to the MAPPO-Pre-derived component, whereas independently trained HAPPO-Full often lies outside it. This supports the interpretation that shared initialization constrains subsequent specialization into a compatible region of joint-policy space.

### 4.4 Functional compatibility and specialization probes

We next test whether low-barrier interpolation corresponds to functional compatibility under discrete team composition changes, and whether this compatibility coexists with specialization. Table 3 reports representative functional-compatibility results. Full tables and visualizations are provided in Appendix C.

Mixed-team assembly combines half the agents from one team with the remaining agents from another, aligned by index. Same-source mixtures maintain high win rates, whereas cross-history mixtures involving HAPPO-Full drop substantially. Single-agent hot-swap yields the same pattern: replacing one M2H or MAPPO-Sep specialist with the MAPPO-Shared actor preserves high performance, while replacement into HAPPO-Full produces larger drops and higher variance.

Table 3: Representative functional-compatibility probes. Values are win rate (%) mean ± standard deviation over four seeds. Full results are reported in Appendix C.4.

| Relation | Team composition | 3s5z | corridor | MMM2 |
|---|---|---|---|---|
| *Mixed-team assembly* | | | | |
| Same-source | M2H + MAPPO-Shared | 93.8±10.8 | 93.8±10.8 | 87.5±12.5 |
| Cross-history | HAPPO-Full + MAPPO-Shared | 62.3±27.6 | 31.2±41.0 | 50.0±50.0 |
| *Single-agent hot-swap* | | | | |
| Same-source | M2H ← MAPPO-Shared | 96.2±2.7 | 100.0±0.0 | 95.2±9.5 |
| Cross-history | HAPPO-Full ← MAPPO-Shared | 74.9±15.6 | 50.0±37.5 | 53.7±38.5 |

The specialization probes show that same-source compatibility coexists with role-sensitive specialization. Clone-all copies a single agent's actor to all positions; it dramatically degrades non-shared teams, showing that the agents carry role- or position-specific information. Role-swap further shows nonzero position sensitivity for same-source specialists, with the largest drops for independently trained HAPPO-Full. Thus, shared-initialized agents specialize, yet remain compatible with the shared-source actor. This also supports the fixed agent-index alignment used in our interpolation diagnostic: misaligning agent indices would introduce role-mismatch artifacts that could artificially inflate barriers.

Finally, weight-space probes provide a structural correlate of the interpolation results. Same-source specialists remain close to the MAPPO-Pre checkpoint in source-checkpoint cosine, whereas HAPPO-Full is substantially less aligned with this reference. These metrics provide supporting structural evidence that low-barrier paths are associated with shared-source parameter alignment.

### 4.5 Interpreting the probe suite

The probe suite clarifies what the barrier structure represents. Endpoint performance alone does not determine connectivity: MAPPO-Full and HAPPO-Full can both reach high endpoint performance on several maps and seeds, yet interpolation between them can still collapse. Clone-all and role-swap show that low-barrier same-source teams contain role- or position-specific policies, establishing that connectivity coexists with specialization. Mixed-team and hot-swap evaluations provide a functional counterpart by testing actual hybrid teams under decentralized execution.

Together, these probes show that the same-source relation is expressed across interpolation, specialization, and execution-time compatibility. Parameter sharing appears to establish a source convention that can survive later decoupling and specialization. The separate actors in MAPPO-Sep and M2H develop distinct specialized behaviors while preserving compatibility with the shared-source component. In contrast, independently trained HAPPO-Full can discover a different convention that works internally but is not linearly connected to the MAPPO-Pre-derived histories. This distinction is the main empirical claim of the paper. A complementary question is whether the same separation is also visible at the policy level.

### 4.6 Behavioral differences along interpolation paths

We therefore compare policy behavior along the same interpolation paths. For each corridor path and seed, we collect a fixed set of local policy inputs from rollouts of the two endpoint teams and evaluate the interpolated policies and both endpoints on those inputs. We report individual total variation (TV) and joint symmetric KL divergence as complementary measures of policy-level difference, averaged over the two endpoints. Appendix D provides the full protocol and additional midpoint results on 3s5z.

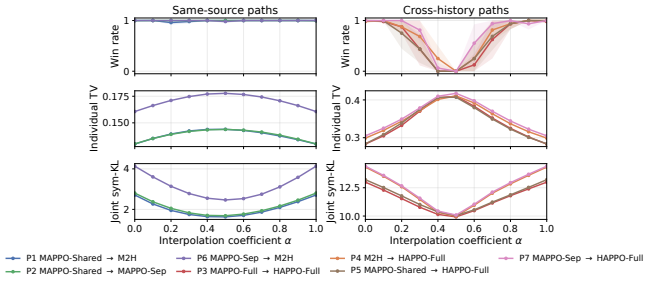



Figure 4: Behavioral differences along interpolation paths on `corridor`, averaged over four seeds. The rows show win rate, individual TV, and joint symmetric KL. Same-source paths show smaller behavioral differences than cross-history paths.

Figure 4 shows the same qualitative separation as the return barriers. Along same-source paths, individual TV stays below $0.18$, and mid-path joint symmetric KL is between $1.6$ and $2.5$ nats. Cross-history paths show larger behavioral differences: individual TV reaches $0.41$, and joint symmetric KL stays above about $10$ nats. These larger differences occur in the same region of the interpolation path as the sharp win-rate drops. The nonzero differences along same-source paths show that distinct endpoint behaviors can coexist with low-barrier connectivity. The behavioral measurements therefore provide a policy-level correlate of the coordination-barrier structure.

## 5 Discussion and conclusion

**Main findings.** Coordination connectivity reveals a clear difference between teams that share the same training history and teams trained independently. Across the three SMAC maps, all same-source paths remain low-barrier, with $B \leq 0.13$, whereas all cross-history paths involving independently trained HAPPO-Full have substantially larger barriers, with $B \geq 0.42$ and values close to $1.0$ on `corridor`. This pattern persists even after the actors are separated and specialized. The mixed-team and hot-swap experiments show that same-source teams also retain higher compatibility under direct team composition changes. Clone-all and role-swap show that the actors have developed role- or position-specific behaviors, while the behavioral measurements in Figure 4 show smaller policy differences along same-source paths. Together, these results show that a shared training history can preserve both linear connectivity and team compatibility after specialization.

The controlled training histories further show that this relation persists across different continuation choices. MAPPO-Sep and M2H both depart from the shared MAPPO checkpoint and use separate actors, and M2H additionally switches to HAPPO-style updates, yet both remain low-barrier to other same-source descendants. By contrast, independently trained HAPPO-Full is separated from the MAPPO-Pre-derived histories by much larger barriers. The effect of the shared training history therefore remains visible after later training has changed both the individual actors and, in M2H, the continuation algorithm.

**Why coordination connectivity matters.** First, the results extend the geometric perspective of mode connectivity from a single neural network to a team of decentralized policies. A MARL solution contains multiple actors whose behaviors must remain coordinated, and these actors can specialize after they separate. The finding that same-source teams remain connected by high-return linear paths shows that this team-level connection can survive such specialization. Independently trained teams, in contrast, can reach high endpoint performance while remaining separated by sharp coordination cliffs.

Second, the geometric relation has a direct functional counterpart. Same-source agents can be mixed or replaced with much smaller performance loss than cross-history agents in our experiments. Thus, the low-barrier relation is not confined to interpolation itself: the same training-history structure is also reflected when team composition changes. This gives a functional interpretation to the same-source compatibility studied by coordination connectivity.

Third, the results suggest several practical uses when policies continue to evolve after training. High-return interpolation paths may provide useful intermediate team policies without additional

training. The mixed-team and hot-swap results also motivate partial team updates, agent replacement, and checkpoint or version selection when the policies involved share the same training history. In this setting, coordination connectivity can help identify groups of policy versions that remain usable together after separate updates.

**Scope and limitations.** Our primary experiments are conducted on three SMAC maps, with additional four-seed results on two cooperative MPE tasks in Appendix E. Appendix E also reports a single-seed VDN experiment, so broader validation across value-decomposition methods and other MARL benchmarks remains for future work. The current diagnostic uses fixed agent-index alignment and linear interpolation; nonlinear paths or permutation-optimized interpolation may reveal additional connections between trained teams. We have not yet identified when during shared pre-training the low-barrier structure first emerges. The behavioral analysis in Appendix D covers `corridor` and `3s5z`. Finally, `MMM2` exhibits substantial seed-level variance, so exact barrier magnitudes on this map should be interpreted together with the reported cross-seed standard deviations.

**Conclusion.** We introduced *coordination connectivity*, a diagnostic for testing whether trained MARL teams are connected by high-return interpolation paths in joint-policy space. Across five controlled training histories, seven targeted interpolation paths, and three SMAC maps, teams derived from the same MAPPO initialization remain low-barrier, whereas paths involving independently trained HAPPO exhibit large coordination cliffs. Mixed-team assembly, hot-swap, clone-all, role-swap, weight-space probes, and policy-divergence measurements further show that the same-source teams can develop distinct roles while retaining both linear connectivity and team compatibility. Overall, our results show that shared initialization can leave a persistent structure among later specialized teams, linking the geometry of the joint-policy space to how well those teams continue to coordinate.

**Code availability.** Code is available at `https://github.com/FanYixiang2000/Coordination-Connectivity`.

## Acknowledgments and Disclosure of Funding

**Funding.** This work was supported by the National Natural Science Foundation of China under Grant 62322316; the CAS Project for Young Scientists in Basic Research (Grant No. YSBR-123); the National Natural Science Foundation of China under Grant No. 62503472; and the Young Scientists Foundation of CSAA (Guidance Navigation and Control, GNC) under Grant CSAA-YSF2025-GNC-08.

**Competing interests.** The authors declare no competing interests.

## References

S. Ainsworth, J. Hayase, and S. Srinivasa. Git re-basin: Merging models modulo permutation symmetries. In *ICLR*, 2023.

G. Benton, W. Maddox, S. Lotfi, and A.G. Wilson. Loss surface simplexes for mode connecting volumes and fast ensembling. In *ICML*, 2021.

F. Christianos, G. Papoudakis, and S.V. Albrecht. Scaling multi-agent reinforcement learning with selective parameter sharing. In *ICML*, 2021.

F. Draxler, K. Veschgini, M. Salmhofer, and F. Hamprecht. Essentially no barriers in neural network energy landscape. In *ICML*, 2018.

R. Entezari, H. Sedghi, O. Saukh, and B. Neyshabur. The role of permutation invariance in linear mode connectivity of neural networks. In *ICLR*, 2022.

J. Foerster, G. Farquhar, et al. Counterfactual multi-agent policy gradients. In *AAAI*, 2018.

S. Fort, H. Hu, and B. Lakshminarayanan. Deep ensembles: A loss landscape perspective. *arXiv:1912.02757*, 2019.

J. Frankle, G.K. Dziugaite, D. Roy, and M. Carbin. Linear mode connectivity and the lottery ticket hypothesis. In *ICML*, 2020.

W. Fu, C. Yu, et al. Revisiting some common practices in cooperative multi-agent reinforcement learning. In *ICML*, 2022.

T. Garipov, P. Izmailov, D. Podoprikhin, D. Vetrov, and A.G. Wilson. Loss surfaces, mode connectivity, and fast ensembling of DNNs. In *NeurIPS*, 2018.

J.K. Gupta, M. Egorov, and M. Kochenderfer. Cooperative multi-agent control using deep reinforcement learning. In *AAMAS ALA Workshop*, 2017.

H. Hu, A. Lerer, A. Peysakhovich, and J. Foerster. "Other-play" for zero-shot coordination. In *ICML*, 2020.

H. Hu, A. Lerer, B. Cui, L. Pineda, N. Brown, and J. Foerster. Off-belief learning. In *ICML*, 2021a.

J. Hu, S. Jiang, et al. Rethinking the implementation tricks and monotonicity constraint in cooperative multi-agent reinforcement learning. *arXiv:2102.03479*, 2021b.

G. Ilharco, M.T. Ribeiro, M. Wortsman, et al. Editing models with task arithmetic. In *ICLR*, 2023.

K. Jordan, H. Sedghi, O. Saukh, et al. REPAIR: Renormalizing permuted activations for interpolation repair. In *ICLR*, 2023.

J.G. Kuba, R. Chen, M. Wen, et al. Trust region policy optimisation in multi-agent reinforcement learning. In *ICLR*, 2022.

R. Kuditipudi, X. Wang, H. Lee, et al. Explaining landscape connectivity of low-cost solutions for multilayer nets. In *NeurIPS*, 2019.

C. Li, T. Wang, et al. Celebrating diversity in shared multi-agent reinforcement learning. In *NeurIPS*, 2021.

H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein. Visualizing the loss landscape of neural nets. In *NeurIPS*, 2018.

X. Li, L. Pan, and J. Zhang. Kaleidoscope: Learnable masks for heterogeneous multi-agent reinforcement learning. In *NeurIPS*, 2024.

R. Lowe, Y.I. Wu, A. Tamar, et al. Multi-agent actor-critic for mixed cooperative-competitive environments. In *NeurIPS*, 2017.

K. Lucas and R.E. Allen. Any-Play: An intrinsic augmentation for zero-shot coordination. In *AAMAS*, pages 853–861, 2022.

A. Lupu, B. Cui, H. Hu, and J. Foerster. Trajectory diversity for zero-shot coordination. In *ICML*, 2021.

M. Matena and C. Raffel. Merging models with Fisher-weighted averaging. In *NeurIPS*, 2022.

R. Mirsky, I. Carlucho, et al. A survey of ad hoc teamwork: Definitions, methods, and open problems. *arXiv:2202.10450*, 2022.

T. Rashid, M. Samvelyan, et al. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *ICML*, 2018.

M. Samvelyan, T. Rashid, et al. The StarCraft multi-agent challenge. *CoRR abs/1902.04043*, 2019.

S.P. Singh and M. Jaggi. Model fusion via optimal transport. In *NeurIPS*, 2020.

P. Stone, G.A. Kaminka, S. Kraus, and J.S. Rosenschein. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *AAAI*, 2010.

D.J. Strouse, K.R. McKee, et al. Collaborating with humans without human data. In *NeurIPS*, 2021.

P. Sunehag, G. Lever, et al. Value-decomposition networks for cooperative multi-agent learning based on team reward. In *AAMAS*, 2018.

J.K. Terry, N. Grammel, S. Son, B. Black, and A. Agrawal. Revisiting parameter sharing in multi-agent deep reinforcement learning. *arXiv:2005.13625*, 2020.

J.K. Terry, B. Black, et al. PettingZoo: Gym for multi-agent reinforcement learning. In *NeurIPS*, 2021.

J. Treutlein, M. Dennis, C. Oesterheld, and J. Foerster. A new formalism, method, and open issues for zero-shot coordination. In *ICML*, 2021.

J. Wang, Z. Ren, et al. QPLEX: Duplex dueling multi-agent Q-learning. In *ICLR*, 2021a.

T. Wang, H. Dong, V. Lesser, and C. Zhang. ROMA: Multi-agent reinforcement learning with emergent roles. In *ICML*, 2020.

T. Wang, T. Gupta, A. Mahajan, B. Peng, S. Whiteson, and C. Zhang. RODE: Learning roles to decompose multi-agent tasks. In *ICLR*, 2021b.

M. Wortsman, G. Ilharco, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *ICML*, 2022.

P. Yadav, D. Tam, L. Choshen, C. Raffel, and M. Bansal. TIES-merging: Resolving interference when merging models. In *NeurIPS*, 2023.

C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu. The surprising effectiveness of PPO in cooperative multi-agent games. In *NeurIPS Datasets and Benchmarks*, 2022.

Y. Zhong, J.G. Kuba, X. Feng, et al. Heterogeneous-agent reinforcement learning. *JMLR*, 2024.

# A Experimental details

**Hyperparameters.** All training uses the HARL framework [Zhong et al., 2024], following established practices for cooperative MARL [Fu et al., 2022, Hu et al., 2021b]. The shared hyperparameters are: learning rate $5 \times 10^{-4}$ for both actor and critic, GAE $\lambda = 0.95$, PPO clip ratio $\epsilon = 0.2$, and entropy coefficient $0.01$. The discount factor is $\gamma = 0.99$ for `3s5z` and `corridor`, and $\gamma = 0.95$ for `MMM2`, following the per-map HARL tuned defaults. Both actor and critic use a two-layer MLP encoder of width 128; `3s5z` additionally employs a 128-dimensional GRU recurrent layer, while `corridor` and `MMM2` use feedforward-only (non-recurrent) architectures, consistent with the HARL tuned configurations. Each update collects episode_length $\times$ 20 parallel rollout threads timesteps, yielding batch sizes of 3,000 (`3s5z`), 3,600 (`MMM2`), and 8,000 (`corridor`). For MAPPO-Pre training, we train with shared parameters for approximately $2/3$ of the total training budget per map ($6.66 \times 10^6$ steps for `3s5z`; $1.33 \times 10^7$ steps for `corridor` and `MMM2`). All subsequent training histories continue from the converged checkpoint. The total training budget is $10^7$ steps for `3s5z` and $2 \times 10^7$ steps for `corridor` and `MMM2`. HAPPO-Full and MAPPO-Full are trained from random initialization for the same total budget.

**Map characteristics.** `3s5z` features a symmetric composition of 3 Stalkers and 5 Zealots against an identical enemy army. `corridor` places 6 Zealots in a narrow passageway, requiring tight spatial coordination. `MMM2` involves a mixed unit composition (1 Medivac, 2 Marauders, 7 Marines) against a stronger enemy army, demanding both micromanagement and healing coordination. All maps are drawn from SMAC v1 [Samvelyan et al., 2019].

# B Behavioral intuition for low coordination barriers

This section provides an informal behavioral intuition for why shared-source teams may exhibit low coordination barriers. Suppose the interpolated joint policy $\Theta(\alpha)$ remains behaviorally close—in total variation (TV) distance—to a high-performing reference team $\Theta_0$ throughout the interpolation path. For bounded rewards, standard finite-horizon coupling arguments imply that the return discrepancy is controlled by the per-step joint-policy deviation. Informally, if the per-step joint TV deviation is at most $\delta$, then the resulting return difference scales as $\mathcal{O}(H^2 R_{\max} \delta)$.

For decentralized policies, the joint policy factorizes as $\Pi_\Theta = \prod_i \pi_{\theta_i}$, yielding

$$D_{\mathrm{TV}}\left(\prod_{i=1}^{N} \pi_i, \prod_{i=1}^{N} \pi_i'\right) \leq \sum_{i=1}^{N} D_{\mathrm{TV}}(\pi_i, \pi_i'). \tag{5}$$

This factorization suggests one possible mechanism for low barriers: if each agent's interpolation path induces only small behavioral drift from a shared-source reference family, then the aggregate joint-policy deviation can remain small enough to preserve team return. Shared initialization may help by anchoring all agents near a common behavioral and parametric region. Section 4.6 and Appendix D directly measure these behavioral differences along interpolation paths. Same-source paths show smaller differences than cross-history paths, consistent with this intuition.

# C Additional coordination-connectivity results

This section provides the full visual evidence behind the compact probe summary in the main text. The key comparison is between same-source histories derived from MAPPO-Pre and cross-history combinations involving independently trained HAPPO-Full.

## C.1 Why the `MMM2` cross-history barriers repeat

In Table 2 and Figure 3, the HAPPO-Full-related cross-history barriers on `MMM2` all have the same mean value, $0.69$ after rounding. The repeated value arises from the shared unstable HAPPO-Full endpoint on `MMM2`, whose original-team win rate is $75.0\% \pm 43.3\%$ over four seeds (Table 5). Under the barrier definition in Eq. (4), a seed where the HAPPO-Full endpoint itself fails gives $\min\{J(\Theta_A), J(\Theta_B)\} = 0$ and therefore contributes zero barrier, even if intermediate policies also fail. In the remaining seeds, HAPPO-Full reaches high endpoint performance, but interpolation

through cross-history paths collapses at intermediate $\alpha$, producing large per-seed barriers. Averaging these seed-level outcomes yields $0.6875$, displayed as $0.69$, for all four HAPPO-Full-related paths. Thus, the repeated value reflects endpoint instability and seed-level averaging on `MMM2`; the qualitative conclusion is the separation between same-source paths and HAPPO-Full-related cross-history paths, not the exact equality of the rounded means.

### C.2 Functional compatibility visualizations

Figure 5 and Figure 6 show the execution-time compatibility probes.

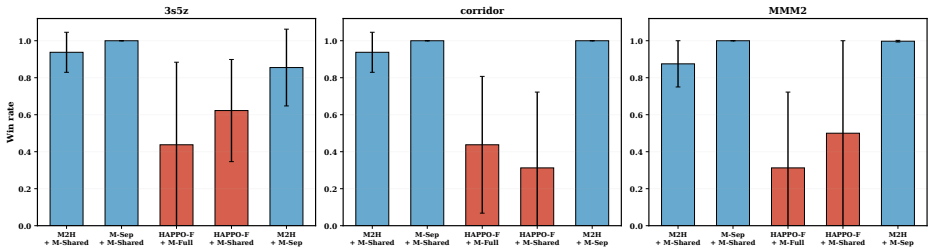



Figure 5: **Mixed-team assembly.** Same-source combinations retain higher win rates across maps, whereas cross-history mixtures involving HAPPO-Full show larger performance drops and higher seed-level variance.

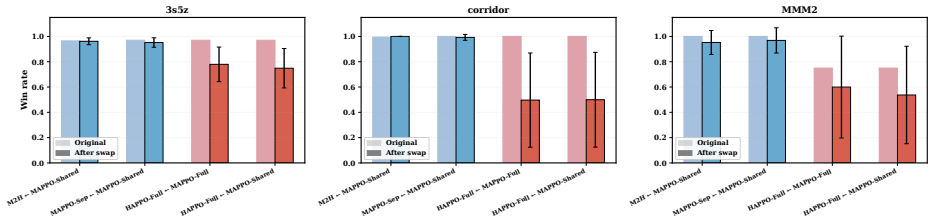



Figure 6: **Single-agent hot-swap.** Replacing one actor with the MAPPO-Shared actor is more stable for same-source receiving teams than for independently trained HAPPO-Full.

### C.3 Specialization and weight-space visualizations

Figure 7 and Figure 8 report the specialization and weight-space probes.

### C.4 Full functional-compatibility tables

Tables 4 and 5 report the complete numerical results for mixed-team assembly and single-agent hot-swap. Mixed-team assembly constructs hybrid teams by taking part of the agents from one trained team and the remaining agents from another trained team, while preserving agent-index alignment. Single-agent hot-swap replaces one actor in a receiving team with a donor actor and evaluates the resulting hybrid team without additional training.

### C.5 Specialization and weight-space tables

As shown in Table 6, the clone-all probe copies one agent's actor to all positions. If a non-shared team survived clone-all, then its agents could be interpreted as nearly interchangeable copies. Instead, clone-all collapses non-shared teams, showing that their agents carry position- or role-specific information. Furthermore, Table 7 reports the weight-space probes, which provide a structural correlate of retained same-source alignment.

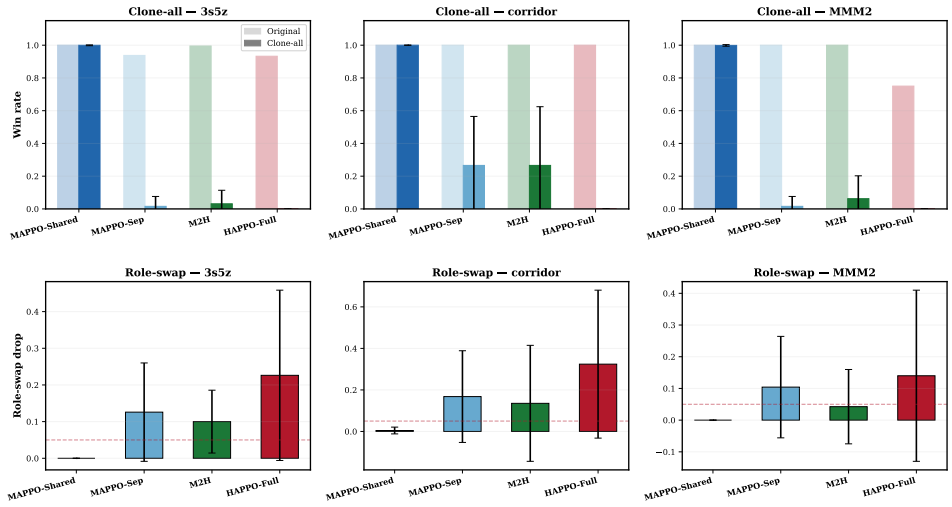



Figure 7: **Clone-all and role-swap specialization probes.** Clone-all and role-swap results show that non-shared agents are not interchangeable copies, demonstrating that specialization coexists with same-source compatibility.

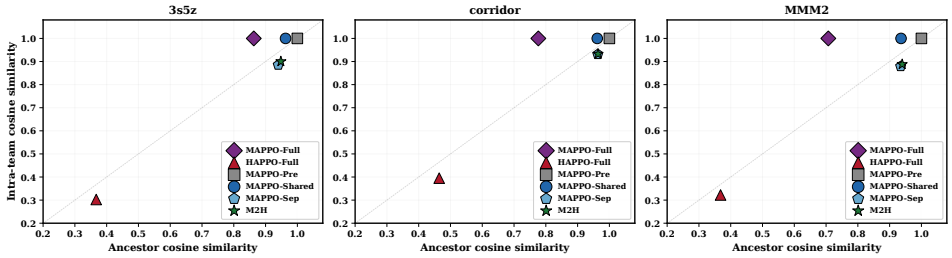



Figure 8: **Weight-space probes.** Same-source methods remain closer to the MAPPO-Pre ancestor than independently trained HAPPO-Full. This structural pattern is consistent with the interpolation and compatibility results.

Table 4: **Mixed-team assembly results.** Values are win rate (%) mean ± standard deviation over four seeds. Teams are formed by combining agents from two training histories using fixed agent-index alignment.

| Team composition | `3s5z` | `corridor` | `MMM2` |
|---|---|---|---|
| *Same-source mixtures* | | | |
| M2H + MAPPO-Shared | $93.8 \pm 10.8$ | $93.8 \pm 10.8$ | $87.5 \pm 12.5$ |
| MAPPO-Sep + MAPPO-Shared | $100.0 \pm 0.0$ | $100.0 \pm 0.0$ | $100.0 \pm 0.0$ |
| M2H + MAPPO-Sep | $85.5 \pm 20.8$ | $100.0 \pm 0.0$ | $99.8 \pm 0.4$ |
| *Cross-history mixtures* | | | |
| HAPPO-Full + MAPPO-Full | $43.8 \pm 44.6$ | $43.8 \pm 37.0$ | $31.2 \pm 41.0$ |
| HAPPO-Full + MAPPO-Shared | $62.3 \pm 27.6$ | $31.2 \pm 41.0$ | $50.0 \pm 50.0$ |

Table 5: **Single-agent hot-swap results.** Values are team win rate (%) after replacing one agent with a donor actor, averaged over replacement positions and four seeds.

| Receiving team ← Donor | 3s5z | corridor | MMM2 |
|---|---|---|---|
| *Same-source replacement by MAPPO-Shared actor* | | | |
| M2H ← MAPPO-Shared | $96.2 \pm 2.7$ | $100.0 \pm 0.0$ | $95.2 \pm 9.5$ |
| MAPPO-Sep ← MAPPO-Shared | $95.2 \pm 3.7$ | $99.2 \pm 2.3$ | $96.9 \pm 10.0$ |
| *Cross-history replacement into HAPPO-Full* | | | |
| HAPPO-Full ← MAPPO-Full | $77.9 \pm 13.7$ | $49.7 \pm 37.2$ | $60.0 \pm 40.3$ |
| HAPPO-Full ← MAPPO-Shared | $74.9 \pm 15.6$ | $50.0 \pm 37.5$ | $53.7 \pm 38.5$ |
| *Original team, no replacement* | | | |
| M2H | $96.5 \pm 3.9$ | $99.5 \pm 0.9$ | $100.0 \pm 0.0$ |
| MAPPO-Sep | $97.0 \pm 2.4$ | $100.0 \pm 0.0$ | $100.0 \pm 0.0$ |
| HAPPO-Full | $97.0 \pm 1.6$ | $100.0 \pm 0.0$ | $75.0 \pm 43.3$ |

Table 6: **Clone-all probe results.** Values are win rate (%). Original denotes the unmodified team. Clone denotes the team obtained by copying one agent's actor to all positions, averaged over agents and four seeds.

| Method | 3s5z | | corridor | | MMM2 | |
|---|---|---|---|---|---|---|
| | Original | Clone | Original | Clone | Original | Clone |
| MAPPO-Shared | 100.0 | 99.9 | 100.0 | 100.0 | 100.0 | 99.9 |
| MAPPO-Sep | 93.8 | 1.6 | 100.0 | 26.6 | 100.0 | 1.6 |
| M2H | 99.5 | 3.1 | 100.0 | 26.6 | 100.0 | 6.2 |
| HAPPO-Full | 93.2 | 0.0 | 100.0 | 0.0 | 75.0 | 0.0 |

Table 7: **Weight-space probes across three SMAC maps.** Source-checkpoint cosine measures similarity to MAPPO-Pre; intra-team cosine measures pairwise similarity among agents in the same team.

| Method | 3s5z | | corridor | | MMM2 | |
|---|---|---|---|---|---|---|
| | Source | Intra | Source | Intra | Source | Intra |
| MAPPO-Shared | 0.963 | 1.000 | 0.962 | 1.000 | 0.936 | 1.000 |
| MAPPO-Sep | 0.940 | 0.885 | 0.963 | 0.931 | 0.935 | 0.879 |
| M2H | 0.948 | 0.900 | 0.964 | 0.933 | 0.940 | 0.888 |
| MAPPO-Full | 0.863 | 1.000 | 0.776 | 1.000 | 0.707 | 1.000 |
| HAPPO-Full | 0.367 | 0.303 | 0.464 | 0.395 | 0.368 | 0.322 |

# D Behavioral difference measurement details

This section gives the details for the measurements in Section 4.6.

**Comparison with the endpoints on the same policy inputs (corridor, full $\alpha$ grid).** For each of the seven interpolation paths and four seeds, both endpoint teams are run for two deterministic evaluation episodes. At every step, each agent's local observation and available-action mask are recorded. This gives a fixed set of local policy inputs from both endpoint teams. For every $\alpha \in \{0.0, 0.1, \ldots, 1.0\}$, the interpolated actors and the endpoint actors are evaluated on the same inputs. We compute two measures. Individual TV is the total-variation distance between the action distributions of the interpolated actor and an endpoint actor, averaged over agents, samples, and the two endpoints. For each agent, symmetric KL is $\frac{1}{2}\big[D_{\mathrm{KL}}(p\|q) + D_{\mathrm{KL}}(q\|p)\big]$; joint symmetric KL is the sum of these per-agent terms. Figure 4 shows the resulting curves. Since corridor uses feedforward actors, no recurrent state is needed for these comparisons.

**Direct comparison between endpoint teams on interpolated-team rollouts (3s5z and corridor, midpoint).** The first measurement compares each interpolated policy with its endpoints. A second measurement directly compares the two endpoint teams on the histories encountered by the

interpolated team. For six paths (same-source P1, P2, P6; cross-history P3, P5, P7), each seed, and $\alpha \in \{0.0, 0.5, 1.0\}$, the interpolated team is run for 20 deterministic episodes. Along these rollouts, the action distributions of the two endpoint policies are compared using total variation, Jensen-Shannon divergence, and greedy-action disagreement. We also record the normalized $\ell_2$ parameter distance between the endpoint actors, defined for each aligned agent pair as $\|w_a - w_b\|_2 / (\|w_a\|_2 + \|w_b\|_2)$ and then averaged over agents. Table 8 reports the group means at $\alpha = 0.5$ over the $3 \times 4 = 12$ path-seed combinations in each group.

Table 8: Midpoint behavioral and parameter differences between endpoint teams. Values are means over three paths and four seeds per group. The behavioral measures compare the two endpoints on histories from the mid-path interpolated team. Cross-history pairs show larger differences than same-source pairs, together with much larger barriers.

| Map | Group | Barrier $B$ | TV | JS | Greedy disagr. | Param. dist. | Mid-path WR |
|---|---|---|---|---|---|---|---|
| `3s5z` | Same-source | 0.05 | 0.26 | 0.12 | 0.29 | 0.22 | 0.95 |
| `3s5z` | Cross-history | 0.56 | 0.44 | 0.26 | 0.46 | 0.57 | 0.53 |
| `corridor` | Same-source | 0.03 | 0.25 | 0.12 | 0.25 | 0.18 | 1.00 |
| `corridor` | Cross-history | 0.99 | 0.43 | 0.27 | 0.43 | 0.52 | 0.00 |

# E Additional results on MPE and VDN

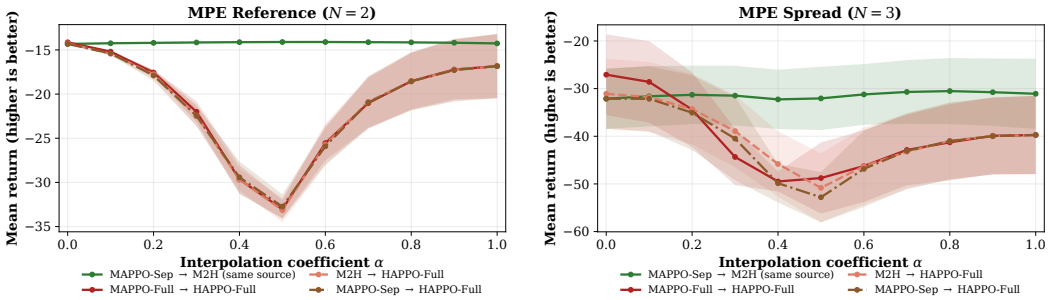
This section reports additional coordination-connectivity results on cooperative MPE tasks and VDN.

**MPE experiments.** We use two cooperative MPE tasks implemented with PettingZoo [Terry et al., 2021], Simple Reference ($N$=2 agents) and Simple Spread ($N$=3 agents), with episode length 25 and a total budget of $10^7$ environment steps for each training history. The actor networks use hidden sizes $[128, 128]$. Following the same design as the SMAC experiments, shared MAPPO pre-training uses $2/3$ of the budget. MAPPO-Sep and M2H then continue with separate actors for the remaining $1/3$, while MAPPO-Full and HAPPO-Full are trained from random initialization for the full budget. We use four seeds, the same interpolation grid $\alpha \in \{0.0, 0.1, \ldots, 1.0\}$, deterministic evaluation, and 50 evaluation episodes per interpolation point. Because these tasks do not use win rate, $J(\Theta)$ is the mean evaluation return and the barriers are reported on the raw return scale. Table 9 reports the results.

Table 9: Coordination barriers on the two cooperative MPE tasks, reported as mean $\pm$ cross-seed standard deviation over four seeds on the raw return scale.

| Type | Path | Simple Reference | Simple Spread |
|---|---|---|---|
| Same-source | MAPPO-Sep $\rightarrow$ M2H | $0.0 \pm 0.0$ | $1.1 \pm 1.9$ |
| Cross-history | MAPPO-Full $\rightarrow$ HAPPO-Full | $16.2 \pm 3.3$ | $9.5 \pm 2.1$ |
| Cross-history | M2H $\rightarrow$ HAPPO-Full | $16.3 \pm 3.3$ | $9.6 \pm 3.9$ |
| Cross-history | MAPPO-Sep $\rightarrow$ HAPPO-Full | $15.9 \pm 2.9$ | $11.9 \pm 1.3$ |

**Single-seed VDN experiment.** We also report a single-seed VDN experiment on Simple Reference. Each agent has a two-layer utility network, and the team value is the sum of the individual utilities [Sunehag et al., 2018]. A source team is trained for 2,000 episodes. Two branches A and B then continue from this checkpoint for 1,000 episodes each with different exploration seeds ($\epsilon = 0.1$), while a scratch team is trained from random initialization for 3,000 episodes. The utility networks are interpolated agent-wise on the same $\alpha$ grid and evaluated for 50 episodes per point. As shown in Figure 10, the same-source path between branches A and B has a return barrier of $4.2$, while the path between branch A and the independently trained team has a barrier of $23.8$. The endpoint returns are approximately $-48$ for all three teams.




(a) Simple Reference (b) Simple Spread

Figure 9: Coordination connectivity on the two cooperative MPE tasks. Curves show mean evaluation return over four seeds, and shaded regions show $\pm 1$ cross-seed standard deviation. The same-source path stays nearly flat, while cross-history paths show clear drops at intermediate $\alpha$, matching the qualitative pattern observed on SMAC.

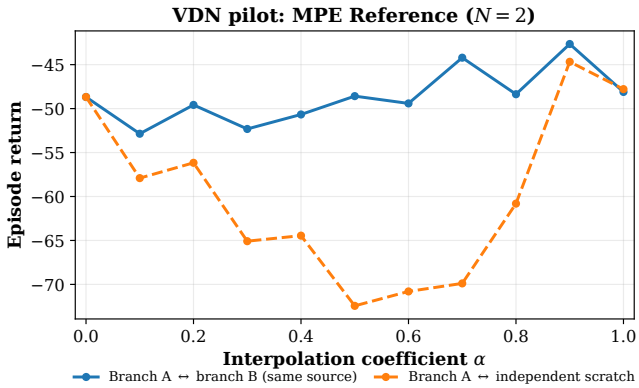



Figure 10: Single-seed VDN experiment on Simple Reference. The two teams branched from the same source checkpoint preserve return along the interpolation path, while the path to the independently trained team shows a clear mid-path drop.

# F Robustness to execution and observation perturbations

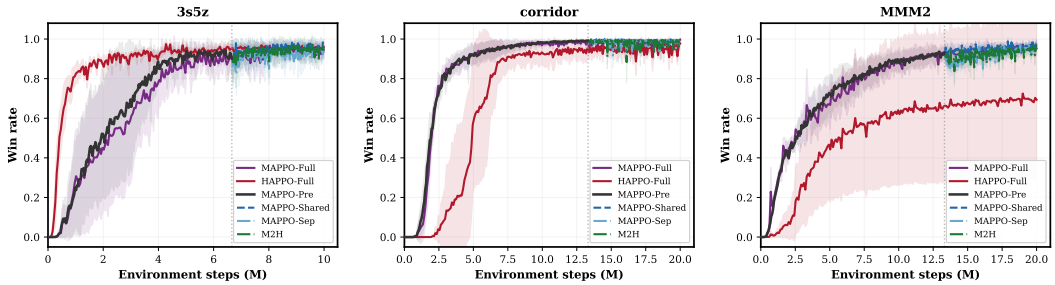
This section reports auxiliary robustness diagnostics using the same trained models from our coordination connectivity study. The experiments characterize endpoint behavior under controlled perturbations, including action delay, observation noise, and missing observations.

## F.1 Training curves

Figure 11 reports training curves for the five histories across the three SMAC maps, showing the endpoint performance reached before the interpolation and compatibility probes are applied.

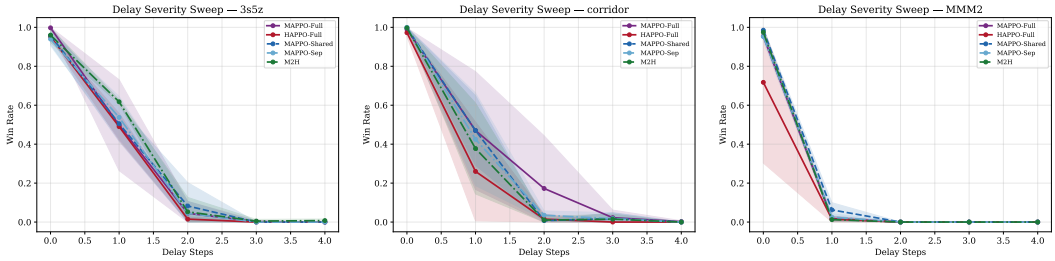
## F.2 Per-perturbation severity sweeps

Figures 12–14 report per-perturbation severity sweeps across the three SMAC maps.



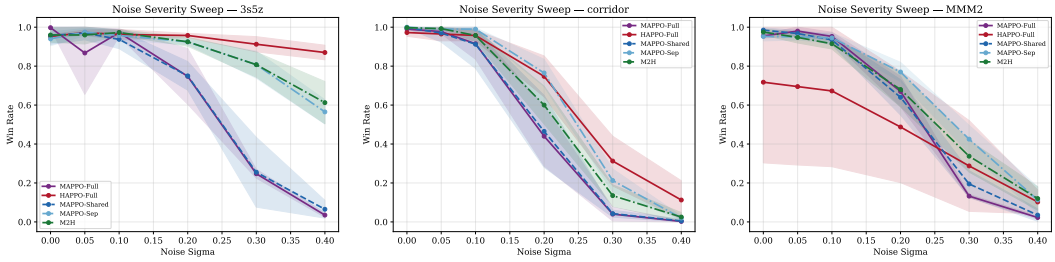

(a) `3s5z` (b) `corridor` (c) `MMM2`

Figure 11: **Training curves.** MAPPO-Full and HAPPO-Full are trained from random initialization; continued histories branch from the MAPPO pre-training checkpoint. The curves show endpoint performance before the interpolation and compatibility probes.



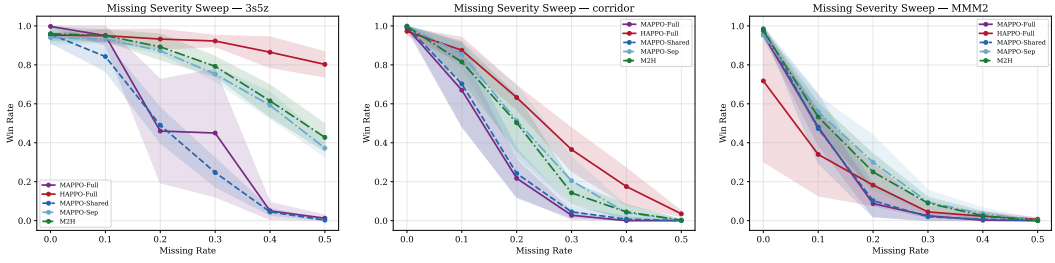

(a) `3s5z` (b) `corridor` (c) `MMM2`

Figure 12: **Action-delay severity sweeps.** Win rate is evaluated as the executed action is delayed by increasing numbers of timesteps.




(a) `3s5z` (b) `corridor` (c) `MMM2`

Figure 13: **Observation-noise severity sweeps.** Win rate is evaluated as Gaussian observation noise increases.




(a) `3s5z` (b) `corridor` (c) `MMM2`

Figure 14: **Missing-observation severity sweeps.** Win rate is evaluated as an increasing fraction of observation dimensions is masked.

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the coordination-connectivity diagnostic and the main SMAC findings, with claims restricted to the tested training histories and evaluation setting. The behavioral and MPE results provide additional supporting evidence without changing the main claim.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 discusses the main limitations, including benchmark coverage, when the low-barrier component emerges, nonlinear or permutation-optimized paths, and seed-level variance.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The paper does not present formal theoretical results. Appendix B provides only an informal behavioral intuition for the empirical observations.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Section 4 and Appendices A, D, and E describe the training and evaluation settings, principal hyperparameters, network architectures, and training budgets; the released code provides the remaining implementation details and exact configurations.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code is publicly available at `https://github.com/FanYixiang2000/Coordination-Connectivity`. The paper and appendices provide the corresponding experimental settings and evaluation protocols.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 4 and Appendices A, D, and E describe the training and evaluation settings, principal hyperparameters, network architectures, training budgets, and evaluation procedures; the released code provides the exact configurations and remaining implementation details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The main results report means and cross-seed standard deviations over four seeds, and interpolation plots show shaded $\pm 1$ cross-seed standard-deviation bands. The VDN experiment is explicitly identified as single-seed.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: The paper reports training budgets and batch sizes, but does not report the hardware type, memory, or wall-clock execution time for each experiment.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses simulated cooperative MARL benchmarks and does not involve human subjects or sensitive personal data. The research conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: Potential positive impacts include improved analysis and maintenance of multi-agent systems. Direct societal risks are limited in the present study because it is conducted in simulated environments and does not introduce a real-world deployment system or sensitive data.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The work does not release high-risk models or sensitive datasets and does not introduce a direct real-world deployment system requiring special release safeguards.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites the original SMAC, HARL, and PettingZoo sources. SMAC, HARL, and PettingZoo are distributed under the MIT License, and their license terms are respected.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The code released with this work is linked through the public repository, and the paper and appendices document the associated experimental and evaluation protocols.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The work does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The work does not involve crowdsourcing or research with human subjects, so IRB approval or an equivalent review is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core methodology does not involve LLMs as an important, original, or non-standard component.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.