Annals of Science and Inquiry
Vol. 2 No. 11 (2026)

Team Composition as a Test of Learned Cooperation

Spencer Butler[1] · Dylan Tucker[2] · Matthew Fisher[3]

[1] French Polynesia · [2] San Marino · [3] Algeria

Published: November 02, 2026 | Article 974

Abstract

A team can coordinate well with familiar partners while remaining fragile to a changed composition. Team-level generalization should distinguish policy quality from compatibility among participants.

Team Composition as a Test of Learned Cooperation

Abstract

A team can coordinate well with familiar partners while remaining fragile to a changed composition. Team-level generalization should distinguish policy quality from compatibility among participants.

Introduction

The cooperative setting of Guo et al. (2023) can be interpreted at the level of a particular assembled team or at the level of policies expected to work with other partners. Those are different targets. A review should identify whether the evidence varies team composition or only repeats trajectories from one composition. A favorable outcome in the latter case may establish useful joint behavior without describing broader compatibility. The relevant boundary concerns the relationships preserved among participants. Keeping partner familiarity visible helps separate a strong within-team result from a stronger claim about interchangeable cooperative competence.

The distribution behind a learning result

In the work of Henderson et al. (2017), it is clear that the distribution behind a successful reinforcement learning result is crucial. The authors emphasize that merely reporting episodes from one policy does not constitute a reliable evaluation. A selected policy, which represents what one training history produced, must be compared against a distribution of training histories and task instances. This implies that a learning procedure is not just a single outcome but an ensemble of potential outcomes.

Agarwal et al. (2021) further reinforce this point by highlighting the importance of defining both the primary summary and supporting distributional view in aggregate comparisons. They argue that pooling across tasks can mask underlying limitations, while reporting every task without a clear overall estimand makes comparisons arbitrary. To address these issues, it is necessary to treat training run variation, scenario variation, and measurement variation separately. This separation allows for a more nuanced evaluation of the robustness and reliability of learning procedures.

The distribution behind a successful reinforcement learning result must be carefully defined and evaluated, distinguishing between a selected policy and the broader distribution of outcomes. This approach ensures that claims are supported by a thorough analysis of various sources of variation, thereby enhancing the credibility of the results.

Foundations of collective motion

In the multi-agent actor-critic formulation, the distinctions between a participant's incentive, training information, and team performance outcome are crucial (Lowe et al., 2017). These elements should be transparent when interpreting collective motion scores, ensuring that the quality of coordination policy is not conflated with the compatibility among its members. For instance, a team might perform well with familiar partners but falter when new members join, highlighting the fragility of team-level generalization.

Rashid et al.'s QMIX method, on the other hand, focuses on linking individual decisions to a structured representation of team value (Rashid et al., 2018). This approach acknowledges that the representational framework is specific to the method rather than a universal characteristic of coordination. Evaluating whether task relationships can be expressed before attributing a score difference to a general advantage underscores the importance of methodological choices in reinforcement learning experiments.

Henderson et al. emphasize that implementation and evaluation choices significantly impact comparisons (Henderson et al., 2017). By distinguishing between successful training outcomes and the procedures that produce them, researchers can better assess learning reliability. In this context, repeated episodes of a selected policy are necessary to claim about learning, ensuring that independent training histories are considered when interpreting collective motion scores.

The objective is an operational definition

The objective of developing an operational definition for team coordination is to ensure that the measures used are directly applicable and not confounded by other factors. In a team setting, the translation of a task into a measurable quantity can be manipulated to allow one behavior to compensate for another, potentially leading to improved overall performance. However, this should not obscure the need to scrutinize these substitutions, as they might indicate changes in behavior that do not align with the intended coordination strategy. For instance, if a reward system is designed to favor one team member over another, it could lead to a situation where the optimal policy preserves a behavior that aligns with an individual's advantage rather than the collective success of the team.

Lowe et al. (2017) highlight the importance of examining incentives within teams, noting that shared success, individual advantage, and the criterion for retaining a policy need not be aligned. This distinction is crucial because it allows researchers to identify whether improvements in coordination are due to changes in the optimized surrogate, specific behavioral dimensions, or broader intended uses. By distinguishing these elements, an operational definition can provide clearer insights into team-level generalization, distinguishing policy quality from compatibility among participants.

Cooperative learning and policy updates

In the context of cooperative learning and policy updates, this section emphasizes how a team's performance with familiar partners contrasts with its adaptability when the composition changes. This shift underscores the importance of distinguishing between policy quality and compatibility among participants. Teams with familiar members tend to exhibit strong coordination but are more vulnerable to membership changes. In contrast, teams encountering new members might experience fluctuations, indicating that compatibility is a crucial factor for sustained success.

The analysis using a distributional view reveals whether gains in performance are broad or limited, suggesting that policy quality alone does not guarantee adaptability. This insight underscores the need to consider compatibility among team members, as a small set of strong individuals may dominate outcomes when the team composition changes. The leakage analysis further emphasizes this point by examining how information can cross the boundary between model development and evaluation. It highlights that even with a large dataset and sophisticated models, maintaining independence is essential for accurate assessment. This means that the decision-making process should remain unbiased to avoid selection bias, thereby ensuring that evaluations remain independent of the team's composition.

By distinguishing between explaining and predicting, the review clarifies the importance of an intended inferential goal. A model may predict well without necessarily explaining the underlying mechanisms. Therefore, the focus on policy quality alone is insufficient. Understanding compatibility among participants becomes crucial for evaluating how teams perform under different compositions. This distinction allows for a more nuanced evaluation that considers not just the outcomes but also the adaptability and robustness of team performance.

Who enters the evaluation population

The evaluation population is assembled through a systematic process rather than inferred from sample size alone, as Koh et al. (2020) suggest. This process can lead to differences between the reported, reviewed, retained cases and typical instances, highlighting that a larger selected sample might be more precise in describing the evaluated group without fully capturing the intended population. Shmueli (2010)'s distinction underscores the need to distinguish between prediction within the sample and selection criteria, which can influence the validity of the evaluation. By recording each inclusion stage and examining how earlier decisions affect later selections, researchers can better understand the selection process and clarify its impact on the assessment. This approach allows for a more explicit examination of the evaluation population, distinguishing it from ordinary cases and ensuring that the shared evaluation sample answers different questions than escalated cases within individual systems.

Evidence for comparative learning claims

The distributed behavioral model of flocking offers foundational insights into how local rules contribute to collective motion (Reynolds, 1987). This model underscores the importance of understanding what specific property a learned policy captures or preserves. In contrast, the comparison should clarify which collective property its score actually measures. For instance, if a team's coordination is evaluated based solely on how well it performs with familiar partners, this approach may not capture the team's potential to adapt and coordinate effectively when partners change.

Trust-region policy optimization provides a method for regulating policy improvement (Schulman et al., 2015). While its theoretical or algorithmic motivation remains distinct from evidence about one specific implementation, comparing update ideas can reveal how different strategies affect policy performance. A review should highlight matched information, tuning, and evaluation budgets necessary to claim an empirical advantage. For example, if a team's policy is evaluated with a trust-region approach, understanding the impact of varying the region size or other parameters can provide valuable insights into which strategies are most effective for maintaining coordination in different contexts.

Generalized advantage estimation addresses estimation used in policy-gradient learning (Schulman et al., 2015). Changing an estimator can affect variance, bias, and the experience generated by the resulting policy. Comparing these changes can help identify which estimators lead to more stable or adaptable policies, thereby supporting comparative learning claims about policy quality and compatibility among participants.

A score needs a construct

Measurement begins with a claim about meaning. As Shmueli (2010) suggests, a score designed to predict future success requires different justification compared to one intended to explain present behavior. In the context of team composition, this distinction is crucial because the performance of a team can vary significantly based on who is working together. For example, a team may coordinate well with familiar partners but struggle when new members are added. Guo et al. (2017) illustrates why model evaluation dimensions cannot be conflated into one notion of quality. Confidence, ranking ability, and correctness are distinct qualities that teams may exhibit differently. In the case of cooperative learning within a team, consensus can introduce another dimension that diverges from these measures. Future assessments should retain these separate quantities and examine cases where they disagree. This analysis could reveal a score that is repeatable but not useful, or one that is effective in certain populations but poorly calibrated in others. The resulting evidence would support a bounded interpretation of the score rather than a universal ranking of systems.

What the observation represents

The observation represented by one entry in a hypothetical chart can be interpreted at different levels of aggregation. At the first level, an individual team member's performance may be highlighted. This interpretation focuses on how well each team member contributes to overall coordination and policy quality. At a second level of aggregation, the composition of the team becomes significant. Here, the interpretation shifts to assess how the presence or absence of certain team members impacts the team’s ability to coordinate effectively and maintain policy quality. In software simulations, consider a team of idealized moving points representing individual team members. These points might move in a coordinated pattern when familiar team members are present, indicating strong compatibility. However, when a member is replaced by another, the pattern breaks, suggesting that the team’s coordination and policy quality may falter. This observation highlights the importance of understanding how team composition affects overall performance, distinguishing between the adaptability of individual team members and the robustness of the team as a whole.

Sorting records into comparable groups

Sorting records into comparable groups can obscure or highlight certain patterns by artificially constraining the data. For instance, if we group teams based on their initial performance, teams that performed well together might appear more cohesive when compared to those that struggled initially. This artificial grouping could exaggerate the observed association between team composition and coordination success. Conversely, if teams are sorted by their final performance, teams that improved over time might appear more cooperative than they were at the start. Such sorting can lead to an inflated or deflated appearance of the observed association, depending on how the groups are formed. In either case, it is crucial to understand that the grouping method itself can introduce bias into the analysis, potentially misleading conclusions about the nature of team coordination and compatibility.

Temporal order and the record

A cross-sectional record captures the current state of team dynamics at a single point in time, offering a snapshot of compatibility among team members. In contrast, a dated series of records tracks changes over time, allowing for a temporal analysis of how team composition affects coordination. In the cross-sectional record, we might observe that teams perform well with familiar partners but fail when new members are introduced. This suggests that team compatibility is a significant factor in performance. However, a dated series of records can reveal whether these initial observations were stable or if changes in team composition led to more variability in performance. For instance, if team coordination deteriorates over time despite the presence of familiar members, it indicates that team-level generalization might be more about policy quality than compatibility among participants. In this way, a temporal series provides a richer understanding of how team dynamics evolve and adapt, supporting a more nuanced evaluation of team composition's impact on performance.

Separating changes within and between groups

In an idealized scenario, consider a team composed of individuals who have learned to cooperate through repeated interactions. When the same members are reassembled into a new group, their performance tends to be consistent and predictable, suggesting that internal dynamics play a significant role in their coordination. Conversely, when the composition changes, the team's performance becomes more variable, indicating that external factors such as new team members' compatibility and individual strengths can significantly alter the outcome. This abstract pattern contrasts with the stable outcomes observed within the same group, highlighting the importance of both internal and external variables in shaping team performance. The variability introduced by changing group composition underscores the need to distinguish between policy quality and compatibility among participants, emphasizing the need for robust methods to evaluate collective behavior.

Conclusion

Therefore, distinguishing between policy quality and compatibility among participants becomes crucial. A team might perform well with familiar partners but fail with new members, indicating fragility of team-level generalization. This highlights the necessity to carefully define and evaluate the distribution behind successful learning results. An open question is how varying degrees of compatibility among team members affect the robustness of coordination policies and their generalization across different scenarios. An additional condition to explore is whether the effectiveness of QMIX or similar methods depends on the specific structure of team value representations and how this structure influences the team's overall performance and adaptability.

References

Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., et al. (2020). WILDS: A Benchmark of in-the-Wild Distribution Shifts. *arXiv preprint arXiv:2012.07421*. https://arxiv.org/abs/2012.07421

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques, 25-34*. https://doi.org/10.1145/37401.37406

Guo, C., Pleiss, G., Sun, Y., Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. *arXiv preprint arXiv:1706.04599*. https://arxiv.org/abs/1706.04599

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., Mordatch, I. (2017). Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. *arXiv preprint arXiv:1706.02275*. https://arxiv.org/abs/1706.02275

Guo, Y., Xie, X., Zhao, R., et al. (2023). Cooperation and Competition: Flocking with Evolutionary Multi-Agent Reinforcement Learning. *Lecture Notes in Computer Science*. doi:10.1007/978-3-031-30105-6_23.

Shmueli, G. (2010). To Explain or to Predict?. *Statistical Science, 25(3), 289-310*. https://doi.org/10.1214/10-STS330

Kapoor, S., Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns, 4(9), 100804*. https://doi.org/10.1016/j.patter.2023.100804

Agarwal, R., Schwarzer, M., Castro, P. S., Courville, A., Bellemare, M. G. (2021). Deep Reinforcement Learning at the Edge of the Statistical Precipice. *arXiv preprint arXiv:2108.13264*. https://arxiv.org/abs/2108.13264

Rashid, T., Samvelyan, M., de Witt, C. S., Farquhar, G., Foerster, J., Whiteson, S. (2018). QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:1803.11485*. https://arxiv.org/abs/1803.11485

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017). Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*. https://arxiv.org/abs/1707.06347

Schulman, J., Levine, S., Moritz, P., Jordan, M. I., Abbeel, P. (2015). Trust Region Policy Optimization. *arXiv preprint arXiv:1502.05477*. https://arxiv.org/abs/1502.05477

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., Meger, D. (2017). Deep Reinforcement Learning that Matters. *arXiv preprint arXiv:1709.06560*. https://arxiv.org/abs/1709.06560