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

Resource Costs Hidden by a Final Formation Score

Owen Hunter[1] · Isaac Henderson[2] · James Harrison[3]

[1] Tajikistan · [2] Angola · [3] Montserrat

Published: November 02, 2026 | Article 1048

Abstract

A final formation score leaves several resource expenditures invisible. Training, selection, communication, and execution should be distinguished when efficiency is part of the claim.

Resource Costs Hidden by a Final Formation Score

Abstract

A final formation score leaves several resource expenditures invisible. Training, selection, communication, and execution should be distinguished when efficiency is part of the claim.

Introduction

The method topic of Xu et al. (2023) can have different costs at different stages. A controller that is economical to execute may require extensive exploration or configuration search to obtain. A review should state which stage matters to the intended use and compare resources on that basis. Work per update, evaluation frequency, and information exchange can differ even when reported iteration counts match. The resulting efficiency statement should connect a defined expenditure to a defined outcome. This makes it possible to recognize a useful operational controller without confusing its execution footprint with the complete cost of producing and selecting it.

Average performance and consequential exceptions

A mean score offers a single summary of performance, but it does not capture all aspects of system reliability. For instance, if the goal is to ensure that the system performs well under a variety of conditions, focusing solely on average performance might overlook critical areas where the system could fail. As Agarwal et al. (2021) argue, evaluating the distribution of performance is crucial for understanding whether gains are broadly applicable or concentrated in easier scenarios. This distinction is particularly important when claiming dependability, as it allows us to assess not just typical behavior but also the robustness of the system under varied conditions.

The concept of a difficult subgroup versus a new population highlights another critical aspect: distinguishing between variability within a single population and shifts in conditions. As Koh et al. (2020) emphasize, performance changes across environments do not always translate to new populations. They might simply reflect different levels of difficulty within the same set of conditions. Therefore, when a system shows improved average performance, it is essential to investigate whether this improvement comes at the expense of robustness in less common scenarios. Failing to differentiate between these cases can lead to an undervaluation of the system's reliability. By identifying the environmental dimensions that matter and examining how they were sampled, we can better understand the true nature of the performance shifts and ensure that the intended claim of dependability is appropriately supported.

Collective motion and information structure

In the context of collective motion and information structure, QMIX’s method connects individual decisions with a structured representation of team value (Rashid et al., 2018). This connection is crucial for understanding how team dynamics influence performance. However, the representational restriction in QMIX does not universally apply to all coordination methods. An evaluation should therefore distinguish between these restrictions and common properties of coordination, ensuring that any observed differences are not attributed to a method's inherent limitations but rather to its specific design.

To address this issue, an evaluation analysis is needed to assess whether the task relationships under investigation can be adequately represented by QMIX’s structure (Agarwal et al., 2021). This analysis should clarify what aspects of coordination are captured and which are not. By defining the task population and summarizing relevant metrics, we can better interpret an overall ranking or score

difference. A distributional view is particularly useful here, as it allows us to determine if observed gains are consistent across different conditions or if they rely on a few exceptional cases.

The leakage analysis is also essential for ensuring that information does not leak from model development to evaluation (Kapoor & Narayanan, 2023). This analysis helps identify any biases that might arise from the interplay between data and model design. By examining which information was available at each decision point and whether the assessment remained independent of selection, we can ensure that evaluations are robust and not influenced by hidden resource expenditures or selection biases.

What crosses an environment boundary

In distinguishing between development and conclusion environments, it is crucial to recognize that the process of training, selection, communication, and execution may vary significantly. Training, for instance, involves the meticulous preparation and fine-tuning of agents or participants, often requiring substantial resource inputs. Selection processes might involve rigorous evaluations and assessments, also consuming time and resources. Communication, on the other hand, requires not just the transmission of information but also its effective reception and understanding by all participants. Execution, finally, encompasses the practical application of the learned skills in a real-world scenario, which can be fraught with unforeseen challenges that necessitate additional resource expenditures.

Kapoor and Narayanan (2023) further underscore this distinction by highlighting the importance of evaluating a destination versus adapting to it. If a method is repeatedly tested against target observations, the resulting success or failure provides evidence about how well the process has been adapted to those conditions. However, such evaluations do not inherently provide insights into whether the method could succeed in different or less familiar environments. This is because the outcomes observed may be heavily influenced by the specific tasks and roles assigned within the development environment, which might not generalize to other settings. Thus, while evidence from a successful demonstration can support a transport argument for certain conditions, it cannot be generalized without considering the unique aspects of the original environment.

Policy learning for coordination

In the context of policy learning for coordination, WILDS underscores the importance of distinguishing between training and evaluation environments. As Koh et al. (2020) note, conflating these stages can obscure resource expenditures such as training, selection, communication, and execution. These elements are distinct and should be separately evaluated to ensure that efficiency claims are not generalized across them.

For instance, a model of flocking behavior, which Reynolds (1987) elucidates, highlights the need for interpretable behavioral objectives. However, the evaluation of these objectives must also consider how a learned policy acquires and preserves them. This is critical because it allows us to understand whether the final formation score captures all necessary collective properties or if additional measures are needed.

Olfati-Saber's (2006) work on flocking through a mathematical dynamic-systems lens offers a complementary perspective, emphasizing the importance of understanding the dynamics at play rather than just the outcomes. By distinguishing between these approaches, we can better assess the robustness and adaptability of coordination strategies, ensuring that our policy learning is not only effective but also comprehensive.

Credit assignment and explanatory attribution

Credit assignment is crucial for understanding how individual contributions are perceived within a team, and this issue is compounded when the overall reward metric obscures resource expenditures and decision-making processes. Foerster et al. (2017) highlights that comparing an agent's performance to a baseline without considering alternative scenarios where different agents were present can lead to misattributing success. This suggests that training baselines alone are insufficient for explaining why a team succeeded, as they do not account for the comparative context needed to evaluate individual contributions.

In Rashid et al.'s (2018) factorized representation, the distinction between information available during training and during execution becomes even more critical. While this approach offers a method to connect team information with decisions, its effectiveness hinges on accurately representing the task at hand, including which interactions can be captured. Future evaluations should therefore focus on comparing teams that achieve similar returns but vary in their sensitivity to role substitution or changes in team composition. Such comparisons would reveal properties of the learned organization that are not fully captured by the overall reward metric alone. By retaining this distinction, we can better understand how the team uses available evidence and why certain configurations lead to success, rather than attributing it solely to broader observation capabilities.

Assessment of formation-learning evidence

In the emergent-flock analysis, the shared descriptive outcome of a final formation score may obscure resource expenditures such as training, selection, communication, and execution (Cucker & Smale, 2007). To address this issue, we should distinguish these components when evaluating efficiency in formation learning. For instance, training might involve the time and resources required to teach flock members their roles, while selection could refer to the process of choosing which individuals are part of the formation. Communication is another resource-intensive aspect that involves coordinating the actions of all members, and execution encompasses the practical application of these strategies during real-world scenarios.

Generalized advantage estimation, a method used in policy-gradient learning (Schulman et al., 2015), highlights the importance of choosing an appropriate estimator as it can significantly impact variance, bias, and the overall experience generated by the policy. This underscores the need for careful consideration of implementation choices, such as selecting the right estimator or adjusting the learning rate, which can affect the efficiency and effectiveness of formation learning (Engstrom et al., 2020).

The distribution behind a learning result

When evaluating a learning result, it is crucial to distinguish between the training process and the final formation score that emerges. In Agarwal et al.'s (2021) analysis, the distinction between these two elements becomes particularly important. A strong checkpoint or a successful population member might be highlighted in a paper, but if the procedure used to find such a system is unreliable, the conclusions drawn from it could be questionable. To address this issue, future research should clearly delineate the primary population and retain supporting distributions at lower levels of analysis. This would allow for a broader performance claim without conflating it with an overly specific or unreliable system, ensuring that the methodological analysis remains central to the scientific account.

Resources belong inside the comparison

Efficiency in a final formation score, as highlighted by Schulman et al. (2017), is contingent upon the route taken to achieve that score and the resources involved in it. The policy optimization framework presented by Schulman et al. (2017) allows for the separation of an update rule from any extra evaluation or selection opportunities that might be included around it. If a method leverages these additional opportunities, they should be counted as part of its procedure, rather than being considered separately.

Henderson et al. (2017) suggests retaining unsuccessful configurations and tuning efforts when describing the route to a final score. This approach acknowledges that efficiency is not solely about the final outcome but also about how resources are allocated and utilized throughout the process. A resource performance curve could provide a more comprehensive comparison by distinguishing between rapid early learning and longer-term improvements, as well as revealing whether extra computation primarily aids in policy selection or learning itself. Stopping decisions also need to be included in this account, as each observation that guides stopping necessitates another evaluation for claims about the selected output. In essence, both cost accounting and evidence accounting share a common boundary: they must include all decisions used to obtain the reported system.

What the observation represents

In a hypothetical simulation chart, one entry represents the total time spent on formation training across all agents. At a second level of aggregation, this same entry might represent the average training time per agent, highlighting how resource expenditures are distributed and averaged. The distinction between these two levels of aggregation is crucial for understanding the overall efficiency of the system, as it reveals whether resources are being evenly distributed or concentrated among certain agents. This analysis allows researchers to identify potential bottlenecks in the formation-learning process and adjust training strategies accordingly.

Sorting records into comparable groups

Sorting records into comparable groups could obscure or highlight resource expenditures by altering how we categorize and compare different instances of a formation score. For example, if we group formations based on their final scores alone, the training, selection, communication, and execution phases that lead to those scores might be overlooked. Consider a scenario where a team's final score is highly dependent on its ability to execute complex maneuvers under pressure. If we only look at the final score without sorting by training or selection criteria, we might miss the fact that some teams had better prepared and selected players, leading to higher scores. This grouping method could thus misrepresent the relative importance of each phase in the overall success of the formation. In contrast, if we sort formations into groups based on these phases, we can more accurately assess how resource allocation and management influenced the final outcome.

Temporal order and the record

To clarify the distinction between cross-sectional and dated series of records, consider a hypothetical scenario where a simulation model tracks a fleet of autonomous vehicles navigating through a complex urban environment. In a cross-sectional record, all vehicles' resource expenditures are aggregated at a single point in time, obscuring the temporal dynamics involved. Conversely, a dated series of records would detail resource usage over multiple time points, revealing how resource allocation changes with each new decision or task.

For instance, training costs might be visible in a cross-sectional record but not in a dated series, where only the final formation score is recorded. This means that within a single snapshot, the entire process of training and selection appears as a single cost, masking the varied stages of learning and preparation. In contrast, a dated series would highlight the incremental costs associated with each phase of training, selection, communication, and execution, providing a clearer picture of resource allocation over time.

The temporal order in a dated series thus enables a more nuanced understanding of resource utilization, distinguishing between distinct activities that might otherwise be conflated. This distinction is crucial for identifying inefficiencies or optimizable areas within the fleet's operations.

Conclusion

One downstream implication of this distinction is that interpreting the efficiency of a process requires careful consideration of the resource costs associated with training, selection, communication, and execution. A final formation score, which focuses on the success of the execution, may not fully account for the resource expenditures involved in the preparatory stages. Therefore, to gain a comprehensive understanding of the process, it is important to distinguish between training, selection, communication, and execution, and recognize that these stages can involve significant resource inputs. An open question remains as to whether the success demonstrated in a development environment can reliably be generalized to other settings, given the potential for different environments to require distinct resource allocations and challenges.

References

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

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

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

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

Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., et al. (2020). Implementation Matters in Deep Policy Gradients: A Case Study on PPO and TRPO. *arXiv preprint arXiv:2005.12729*. https://arxiv.org/abs/2005.12729

Xu, D., Guo, Y., Yu, Z., et al. (2023). PPO-Exp: Keeping Fixed-Wing UAV Formation with Deep Reinforcement Learning. *Drones, 7*(1), 28. doi:10.3390/drones7010028.

Olfati-Saber, R. (2006). Flocking for Multi-Agent Dynamic Systems: Algorithms and Theory. *IEEE Transactions on Automatic Control, 51(3), 401-420*. https://doi.org/10.1109/TAC.2005.864190

Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., Whiteson, S. (2017). Counterfactual Multi-Agent Policy Gradients. *arXiv preprint arXiv:1705.08926*. https://arxiv.org/abs/1705.08926

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

Cucker, F., Smale, S. (2007). Emergent Behavior in Flocks. *IEEE Transactions on Automatic Control, 52(5), 852-862*. https://doi.org/10.1109/TAC.2007.895842

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

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

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