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

Reward Sharing and Credit Assignment under Evolutionary Selection

Dylan Reynolds[1] · Lucas Morris[2] · Jason Reed[3]

[1] Saint Martin (French part) · [2] Germany · [3] Vietnam

Published: November 02, 2026 | Article 959

Abstract

A shared team reward and a policy-level fitness score need not express the same objective. Their relationship matters when cooperative behavior is selected through competition.

Reward Sharing and Credit Assignment under Evolutionary Selection

Abstract

A shared team reward and a policy-level fitness score need not express the same objective. Their relationship matters when cooperative behavior is selected through competition.

Introduction

The hybrid design named in Guo et al. (2023) joins within-team interaction to selection among policies. A team reward can distribute one evaluative signal across participants, while the fitness rule chooses which policies persist. These operations create different attribution questions. A favorable population trend does not reveal which participant behavior generated it, and a participant's local change does not by itself explain selection. A review should retain both levels in the account of improvement. Credit assignment becomes easier to discuss when the quantities driving interaction and those driving replacement are named separately, especially where their preferred behaviors could differ.

Selection can reward a lineage without explaining a role

In the context of evolutionary selection, rewards can be bestowed upon a lineage without uniquely explaining the specific role or interactions that led to their success. For example, a team’s cooperative behavior might thrive through competition, yet its survival could be due to an algorithmic reward scheme that does not directly correlate with individual contributions. Shared rewards within this framework can obscure the distinction between a participant’s actions and the collective environment created by their teammates. A hypothetical scenario would reveal persistent behavioral properties despite changes in team composition, thereby testing whether explanatory credit depends on alternative explanations rather than simply assessing algorithmic credit defined by training procedures.

Removing a component changes a system

In Lowe et al. (2017)'s study, removing a participant or module can indeed alter the dynamics of the entire system. This is evident when one considers the interaction setting, where changes in one component affect the experiences of all others. For instance, altering a participant's role might lead to shifts in how resources are allocated, affecting not just the specific participant but the entire group. Thus, removing a component can impact multiple processes simultaneously, illustrating that a single change can have far-reaching consequences.

Shmueli (2010) underscores the distinction between predictive usefulness and mechanistic explanation. To move toward understanding, it would be beneficial to replace the component with an alternative that maintains the same information, resources, and tuning opportunities but alters the proposed mechanism. By observing how the system behaves under this new configuration, we can gain insights into which parts of the original explanation were not essential. For example, if removing a component results in behavior that is similar to the original, it suggests that particular processes or mechanisms were not necessary for the observed outcomes. Conversely, if a simpler alternative performs similarly and ablation results are unfavorable, this could indicate that the component's mechanism was harmful rather than just an exception. Such evidence helps refine our understanding of which parts of the system contribute uniquely to its success.

Foundations of collective motion

QMIX, a method that connects individual decisions with a structured representation of team value (Rashid et al., 2018), acknowledges that this restriction is part of its scope rather than a universal property of coordination. This limitation highlights the importance of understanding how task relationships are expressed within the model to avoid misinterpreting differences in scores as indicative of broader advantages.

Henderson et al. (2017) underscore the significance of distinguishing between successful training outcomes and the procedures that produce them. When evaluating policies, it is crucial to consider whether repeated episodes characterize a selected policy and if independent training histories are necessary for claims about learning reliability. This distinction prevents conclusions about learning from being drawn too broadly based on single instances.

The study by Guo et al. (2017) further emphasizes the importance of distinguishing confidence from frequency of correctness for a defined outcome. Calibration depends on specific events and populations, meaning that a probability associated with one construct cannot be automatically inferred to apply to a different underlying concept merely because they are related. This highlights the need for careful interpretation when comparing scores or outcomes across different contexts.

Credit assignment and explanatory attribution

In Foerster et al. (2017), counterfactual policy gradients illustrate the need for distinguishing between systems that facilitate learning and those that provide explanatory attribution. A training signal can guide behavior effectively without being an accurate measure of individual contributions, especially when the system involves complex interactions or changes in team dynamics. To better understand contribution, it is crucial to consider what aspects are held constant during training, which actions are compared, and whether these comparisons reflect meaningful outcomes for the team. These considerations help clarify the interpretation of individual roles within a cooperative framework.

Rashid et al. (2018) further emphasizes the importance of methodological choices when analyzing team-level information. Their value factorization approach highlights how team performance is related to individual decisions, but it also underscores the need to scrutinize underlying assumptions and their relevance to the specific context. In cases where team success involves conflicts or complementary roles, the representation may not fully capture these nuances. This observation suggests that comparing different credit assignment methods requires reporting on representational limitations, available training data, and definitions of team success. Such a comprehensive comparison would reveal how learned cooperation is organized, rather than relying solely on a single ranking or measure.

Cooperative learning and policy updates

The leakage analysis in this context highlights that when evaluating cooperative learning models, it is crucial to ensure that the decision-making process remains independent of selection pressures (Kapoor & Narayanan, 2023). This means distinguishing between the information available at each step and how assessments are conducted. For instance, a large dataset and a sophisticated model might not adequately address boundary errors if they do not explicitly separate the training and evaluation phases.

WILDS provides a critical perspective on the challenges of moving from training to evaluation environments (Koh et al., 2020). It emphasizes the importance of naming the destination population when assessing robustness. Different factors such as changes in input distributions, labeling relationships, or

task definitions can all impact how well a model performs. For example, if a model is trained on a dataset with one set of labels and evaluated on another, it might not generalize as expected.

The distributed behavioral model of flocking offers an instructive example of local rules and collective motion (Reynolds, 1987). It underscores the need to define clear objectives for the behavior being modeled. However, this raises a separate question about how a learned policy acquires and preserves these objectives. For instance, a score derived from the model might not actually measure the collective property it is intended to capture. This highlights the importance of careful design and evaluation in cooperative learning systems.

Selection is part of the evidence pathway

Selection can influence outcomes by favoring certain behaviors or configurations over others, even when these are not explicitly rewarded in the system design. The process of selection itself-how choices are made and which outcomes are prioritized-becomes a significant aspect of research evidence. If a particular outcome is selected for publication, it may be interpreted as evidence for the very selection process that led to its prominence. This confusion can arise when researchers present evidence about the selection process itself as if it were evidence for the behavior or outcome it favored. By explicitly recording the decision pathway, however, these intermediate cases and their outcomes become meaningful in their own right. They provide insight into how the system operates and the criteria used to evaluate different behaviors, rather than being dismissed as mere alternatives that did not make the final cut.

Who enters the evaluation population

In Koh et al.'s (2020) work, the process of assembling the evaluation population is detailed rather than inferred from sample size alone. This approach acknowledges that what constitutes a representative sample can vary significantly from what actually enters the evaluation. For instance, cases that are reported, reviewed, retained, and easily adjudicated may not fully capture the broader set of scenarios that a system encounters in its actual operation.

Shmueli's (2010) distinction highlights the need to distinguish between prediction within the selected sample and understanding how the selection process itself was conducted. By examining each inclusion stage, future research could clarify whether earlier decisions within an agent workflow influence the final evaluation population. This distinction is crucial because a shared evaluation sample can only answer questions about that specific sample, not about the broader set of cases a system encounters in its actual operation.

Both approaches-evaluating within the selected sample and understanding the selection process-are valuable but must be interpreted separately. The shared evaluation sample provides insights into how one particular system was evaluated, whereas assessing escalated cases offers a broader view of what types of scenarios a system might face and how it performs under those conditions. By maintaining this separation in interpretation, researchers can avoid conflating questions about prediction with questions about selection bias.

Evidence for comparative learning claims

The flocking algorithms-and-theory paper, while providing a mathematical dynamic-systems perspective on collective behavior (Olfati-Saber, 2006), does not equate to empirical learning benchmarks. Instead, it highlights that the role of these algorithms differs from those used in empirical studies. Similarly, the implementation study underscores the importance of concrete implementation choices in policy-gradient comparisons (Engstrom et al., 2020). A baseline label alone cannot account for all relevant decisions made during a project. This deeper examination of policy gradients by Ilyas et al. (2018) reveals that internal diagnostics and final task outcomes can diverge, suggesting that selection processes in empirical studies may not fully align with the objectives they aim to achieve. Thus, evidence from these comparative learning claims suggests a need for more nuanced understanding of how collective behaviors are selected and evaluated.

A score needs a construct

In Shmueli's (2010) framework, explaining and predicting are distinct analytic goals, where a high score might predict future success without necessarily elucidating the underlying reasons for success. In Guo et al.'s (2017) analysis, scores can serve as useful predictors but may fail to explain the specific outcomes relevant to users. For instance, a system could rank cases accurately and provide plausible probabilities, yet still misclassify certain instances or underperform in critical areas. This highlights that a score's meaning must be carefully specified to avoid treating it as a universal measure of model quality or performance. Thus, a scoring framework should distinguish between discrimination (correctly classifying instances), calibration (assigning probabilities appropriately), and decision usefulness (relevance to the evaluation population). Each component should be examined independently before being interpreted together, ensuring that each aspect aligns with its intended purpose and relevance.

Time changes the information boundary

Event time, recording time, and availability time can indeed diverge, as evidenced by Kapoor and Narayanan (2023), who emphasize the importance of keeping these separate when evaluating decisions made prior to the final data collection. This temporal disjunction introduces a significant challenge, especially when considering the evolving nature of data and the environment.

Koh et al. (2020) further complicates the issue by highlighting that the world producing the observations can change over time. A later evaluation might reveal a new population, altered recording methods, or revised outcome definitions. To address these temporal shifts, future work should clearly distinguish between untouched assessments and adaptations informed by later information. Timestamps alone are insufficient. They typically capture only one phase of data collection. A scientifically valuable temporal evaluation must articulate what was known at the time of decision-making, what has changed, and which relationships remain valid, allowing earlier results to inform later decisions.

What the observation represents

In a hypothetical chart representing credit assignment within a cooperative system, one entry might denote an individual's contribution to the group's success. This unit-level representation suggests that each member's actions are independently evaluated for their impact on overall performance, which is a fundamental premise in understanding how individual contributions translate into collective outcomes.

At a second level of aggregation, however, the same chart might show the cumulative effect of these individual units forming a lineage or a group as a whole. Here, the focus shifts from assessing discrete actions to evaluating the success of the entire lineage or group against competition. This higher-level

analysis does not necessarily replicate the unit-level contributions. Instead, it assesses the collective fitness and survival of the lineage in the face of evolutionary selection.

The shift from individual units to the lineage as a whole highlights the complexity of credit assignment under evolutionary selection. While individual actions are crucial for understanding how credit should be distributed, the system's success is more accurately gauged by the collective performance of the lineage. This distinction underscores the need to consider both micro-level contributions and macro-level outcomes when evaluating cooperative behavior within competitive environments.

Conclusion

One consequence for interpretation is the need to distinguish between untouched assessments and adaptations informed by later information. This highlights the importance of temporal context in understanding credit assignment within cooperative systems. An open question is how the relationship between unit-level contributions and lineage-level success changes under different evolutionary pressures, and whether this relationship is consistent across various competitive scenarios.

References

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

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

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

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

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

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

Ilyas, A., Engstrom, L., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., et al. (2018). A Closer Look at Deep Policy Gradients. *arXiv preprint arXiv:1811.02553*. https://arxiv.org/abs/1811.02553

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

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

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.

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

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, 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