Statistical Science
2025, Vol. 40, No. 4, 515–516
https://doi.org/10.1214/25-STS1020

Introduction to the Special Issue on Reinforcement Learning

Nan Jiang and Ambuj Tewari

The year 2025 is a landmark for Reinforcement Learning (RL). Earlier this year, the 2024 ACM Turing Award went to RL pioneers Andrew G. Barto and Richard S. Sutton. Their foundational work, dating back to the 1980s, transformed RL from a speculative idea into one of the central paradigms of modern artificial intelligence (AI). RL powered the AlphaGo system that defeated Go world champion Lee Sedol in 2016 [2], and it now plays a key role in aligning large language models (LLMs) with human preferences via reinforcement learning from human feedback (RLHF) and in improving their reasoning abilities through reinforcement learning from verifiable rewards (RLVR). None of these achievements could have been foreseen as Barto and Sutton were laying the mathematical and algorithmic foundations of the field, at a time when learning was not yet regarded as central to AI. In addition to their many research contributions, they also wrote the definitive textbook *Reinforcement Learning: An Introduction* [3, 4], which has educated generations of students and continues to influence how the field is taught and understood today.

RL has moved from the periphery of artificial intelligence (AI) research to its very center. In hindsight, this evolution seems almost inevitable. As AI pioneer John McCarthy observed, "intelligence is the computational part of the ability to achieve goals in the world." RL concerns itself directly with this notion of goal-directed behavior through the lens of computationally bounded agents. A goal is formalized via a *reward function*, which assigns numerical values to desirable outcomes. An RL agent takes *actions* in an *environment* and observes its *state*, which evolves as a result of those actions. The environment may occasionally deliver rewards when certain states are reached (for instance, a "win" in a board game such as Chess or Go). A *policy* specifies which action to take in each state, and the central objective in RL is to learn a policy that maximizes the agent's long-term reward. The expected long-term reward that the agent receives starting from a given state or state-action pair is called the *value function*.

Although RL is traditionally viewed as a subfield of AI, it has always drawn strength from a rich interplay with adjacent disciplines. The early history of RL was shaped by ideas from trial-and-error learning in animal psychology, by dynamic programming in operations research, by optimal control in the control theory community, and by foundational contributions from statistics and probability theory. Classic results such as Blackwell's analysis of Markov decision processes and the Robbins–Munro theory of stochastic approximation provided the mathematical backbone for many later developments in RL. These early insights continue to influence research in the field to this day.

A major turning point in the development of RL was the transition from the so-called tabular setting, where both state and action spaces are small and knowledge about one state does not generalize to others, to problems involving high-dimensional or continuous spaces that require *function approximation*. The idea of using neural networks for function approximation in RL was already advocated in the influential book *Neuro-Dynamic Programming* [1], but most research at the time remained focused on the tabular case because of instability issues and limited computational power. Around 2013–2015, the rise of deep learning reignited this vision, giving birth to "deep RL," which integrated RL and neural networks into a unified and empirically powerful paradigm. This integration enabled landmark successes such as AlphaGo and catalyzed a surge of theoretical work on RL with function approximation, some of which is surveyed in the papers in this special issue.

Today, RL is a broad and rapidly developing area within AI, advancing both in theoretical depth and in the range of practical applications. When Editor Moulinath Banerjee invited us to guest edit a special issue of *Statistical Science* on RL, we knew that a comprehensive survey of the field would be impossible. We therefore chose to focus on a small set of topics, shaped by our own research interests, that we believe capture key frontiers of contemporary RL. We were fortunate that many leading researchers agreed to contribute their perspectives and share their insights for this special issue.

In our brief description of the RL problem above, we emphasized the "online" setting in which an agent actively interacts with its environment and takes actions in it. The first paper in this issue focuses on this standard

Nan Jiang is Associate Professor, Department of Computer Science, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA (e-mail: nanjiang@illinois.edu). Ambuj Tewari is Professor, Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: tewaria@umich.edu).

setup, particularly in the context of function approximation, where exact quantities of interest in RL, such as policies or value functions, are replaced by their approximations within a parametric or nonparametric function class. The paper also considers a more powerful scenario called *sample-based planning*, in which learning takes place in a simulator that allows the researcher to reset or query any desired state. A unifying theme across these settings is the need for active exploration, with *optimism in the face of uncertainty* serving as a key guiding principle for algorithm design and analysis.

- "Sample-based Planning and Learning with Function Approximation" by Tor Lattimore and Csaba Szepesvári

In many practical situations, however, the agent does not have the luxury of online interaction. Instead, it must learn from data that have already been collected, often by another agent or under a fixed policy. The next two papers provide a broad overview of this *offline* setting, which has emerged as one of the most active areas of recent RL research. Offline RL should be particularly appealing to statisticians because it emphasizes extracting as much reliable information as possible from limited and noisy data. In contrast to online RL, which promotes exploration of new regions of the state–action space, offline RL prioritizes robustness and guaranteed performance through the principle of *pessimism in the face of uncertainty*.

- "On the Statistical Complexity for Offline and Low-Adaptive Reinforcement Learning with Structures" by Ming Yin, Mengdi Wang, and Yu-Xiang Wang
- "Offline Reinforcement Learning in Large State Spaces: Algorithms and Guarantees" by Nan Jiang and Tengyang Xie

Many popular RL algorithms, such as Q-learning and temporal-difference (TD) methods, rely on solving regression problems, typically using the squared loss as the objective function. The following paper argues that alternative loss functions can provide both practical advantages and new theoretical insights.

- "The Central Role of the Loss Function in Reinforcement Learning" by Kaiwen Wang, Nathan Kallus, and Wen Sun

The statistical and computational challenges of RL become even more pronounced when the agent can only partially observe the underlying state of the environment. In such settings, the agent must simultaneously deal with the twin challenges of inferring hidden states and learning effective policies. The next paper surveys recent advances and open problems in this area, identifying structural conditions under which learning and control remain tractable.

- "Partially Observable RL: Benign Structures and Simple Generic Algorithms" by Qinghua Liu and Chi Jin

The next two papers explore the connections between RL and other well-established areas of learning and control. The first develops an online convex optimization framework for control. The second revisits the deep ties between RL and stochastic approximation, a classical topic that has provided essential tools for the analysis of learning algorithms.

- "The Theory of Online Control" by Elad Hazan and Karan Singh
- "Stochastic Approximation and Reinforcement Learning: The Interface and a Little Beyond" by Vivek Borkar

The final paper in this special issue turns to an important question of replicability that arises when RL algorithms are used in real-world decision-making, such as the selection of treatments in digital health intervention trials. Replicability ensures that independent researchers analyzing post-trial data from the same underlying population will reach similar conclusions. This paper develops a framework for achieving replicable decision rules in such settings, combining insights from contextual bandits, causal inference, and biostatistics.

- "Replicable Bandits for Digital Health Interventions" by Kelly W. Zhang, Nowell Closser, Anna L. Trella, and Susan A. Murphy

We are grateful to all the contributors for sharing their expertise and to the reviewers for their thoughtful feedback. We hope that this special issue will serve as both an accessible entry point for statisticians interested in RL and a valuable reference for experts in the field. The breadth of topics represented here illustrates the remarkable progress in RL and its growing connections to statistics, optimization, and control. We also hope that readers will find this collection as stimulating and enjoyable to read as we did while curating it.

REFERENCES

[1] Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA.

[2] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G. V. D., Schrittwieser, J., Antonoglou, I., Panneershelvam, V. et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* **529** 484–489. https://doi.org/10.1038/nature16961

[3] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning*: *An Introduction*, MIT Press, Cambridge, MA.

[4] Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning*: *An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3889951