As a research community, one of our primary goals is to develop algorithms that will allow us to solve challenging real-world problems. However, our existing algorithms are difficult to apply, and we do not understand when they will likely work. As scientists, it is our job to produce knowledge that will let us better solve both existing and new problems. Furthermore, RL is evaluated empirically, so the knowledge we gain can only be as insightful as the experiments we conduct. The most common reinforcement learning (RL) experiment is performance evaluation, which compares the performance of several algorithms to establish that one method is superior to the others. Unfortunately, these experiments are often improperly set up, leading to results that are not reproducible and consequential [Henderson et al., 2018]. While many have proposed improvements for performance evaluations [Whiteson et al., 2011] [Balduzzi et al., 2018] [Jordan et al., 2020] [Agarwal et al., 2021], performance evaluations suffer from an inescapable limitation: they can only show which algorithm(s) work well, but not why. To continue developing principled approaches to solve new problems, we must design experiments to improve our understanding of RL algorithms, not performance on benchmark problems.

This limitation of performance evaluation can also cause researchers to focus on developing new methods based on misunderstood concepts. For example, the Categorical DQN algorithm [Bellemare et al., 2017] was presented as a better method to approximate value functions, but the only experiments in the paper were performance evaluations. Based on claims of superior performance, researchers proposed other distributional approximations, all purporting to achieve superior performance [Dabney et al., 2018] [Barth-Maron et al., 2018] [Yang et al., 2019]. However, later works showed that the distributional value representation only benefits neural networks and stems from providing a richer set of features throughout learning, not an improved representation of value [Lyle et al., 2019] [Dabney et al., 2021]. These later experiments provide more value than any experiment showing improved performance because they give insights into what properties are essential for successful learning and can be generalized beyond a specific benchmark setting.

If we want our experiments to produce knowledge that can be generalized, we need to switch from asking competitive questions, e.g., does algorithm X outperform algorithm Y, to questions that are scientific [Hooker, 1995] in that they aim to further our understanding of how each algorithm works. This scientific style of experimentation should be the primary form of experimentation in papers. Performance evaluation should serve as a sanity check to confirm that the algorithms continue to perform well beyond carefully controlled experiments. Thankfully, there are many works performing experiments that further our understanding. For example, Tucker et al. [2018] conducted experiments to measure sources of variance and showed that action-dependent control variates produce little variance reduction over state-dependent ones, Linke et al. [2020] studied how the dynamics of different optimizers impact exploration when using intrinsic motivation, and Ghosh and Bellemare [2020] examined the representational properties for stable off-policy temporal difference learning. By measuring properties other than performance, these works produced valuable insights that go beyond a single algorithm and can be used to design new methods to meet specific challenges. When researchers conduct this type of experiment, designing algorithms for new problems becomes easier.

If we want to keep making progress and leverage the talents and creativity of our large community, we should stop competing to find the best algorithm and start working together to understand the necessary properties to solve sequential decision-making problems.

Preprint. Under review.
References


