Humans can outperform Q-learning in terms of learning efficiency

Holger Mohr (holger.mohr@tu-dresden.de)
Department of Psychology, Zellescher Weg 17
01069 Dresden, Germany

Katharina Zwosta (katharina.zwosta@tu-dresden.de)
Department of Psychology, Zellescher Weg 17
01069 Dresden, Germany

Dimitrije Markovic (dimitrije.markovic@tu-dresden.de)
Department of Psychology, Chemnitzer Str. 46b
01187 Dresden, Germany

Sebastian Bitzer (sebastian.bitzer@tu-dresden.de)
Department of Psychology, Chemnitzer Str. 46b
01187 Dresden, Germany

Uta Wolfensteller (uta.wolfensteller@tu-dresden.de)
Department of Psychology, Zellescher Weg 17
01069 Dresden, Germany

Hannes Ruge (hannes.ruge@tu-dresden.de)
Department of Psychology, Zellescher Weg 17
01069 Dresden, Germany

Abstract

Reinforcement learning algorithms like Q-learning maximize cumulative reward by strengthening or weakening stimulus-response associations based on feedback signals. It has been recently shown that Q-learning can be employed to achieve super-human performance in complex tasks (Mnih et al., 2015). However, this requires comprehensive training involving thousands of training episodes, indicating that humans might outperform Q-learning in terms of learning efficiency. Here we explicitly show this by analyzing human learning strategies on a simple stimulus-response learning task involving only four stimuli and four response options (Mohr et al., 2018). By comparing response data from N = 85 subjects with response data generated by the Q-learning algorithm, we show that humans explore the space of stimulus-response pairings more efficiently than Q-learning on the presented learning task. Moreover, using additional computational models, we also show that the subjects accomplished this by integrating implicit task structure into their learning strategies, with some subjects implementing specific response heuristics in order to maximize learning efficiency while keeping memory and computational resources bounded. We conclude that by engaging high-level cognitive processes, humans can minimize the number of errors during learning and can thus outperform Q-learning in terms of learning efficiency.

References

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529 - 533. doi: 10.1038/nature14236

Mohr, H., Zwosta, K., Markovic, D., Bitzer, S., Wolfensteller, U., & Ruge, H. (2018). Deterministic response strategies in trial-and-error learning. *bioRxiv*. doi: 10.1101/258459