

Human-level control through deep reinforcement learning

[Volodymyr Mnih](#), [Koray Kavukcuoglu](#), [David Silver](#), [Andrei A. Rusu](#), [Joel Veness](#), [Marc G. Bellemare](#), [Alex Graves](#), [Martin Riedmiller](#), [Andreas K. Fidjeland](#), [Georg Ostrovski](#), [Stig Petersen](#), [Charles Beattie](#), [Amir Sadik](#), [Ioannis Antonoglou](#), [Helen King](#), [Dharshan Kumaran](#), [Daan Wierstra](#), [Shane Legg](#) & [Demis Hassabis](#)

Nature 518, 529–533 (26 February 2015)

The theory of reinforcement learning provides a normative account¹, deeply rooted in psychological² and neuroscientific³ perspectives on animal behaviour, of how agents may optimize their control of an environment. To use reinforcement learning successfully in situations approaching real-world complexity, however, agents are confronted with a difficult task: they must derive efficient representations of the environment from high-dimensional sensory inputs, and use these to generalize past experience to new situations. Remarkably, humans and other animals seem to solve this problem through a harmonious combination of reinforcement learning and hierarchical sensory processing systems^{4,5}, the former evidenced by a wealth of neural data revealing notable parallels between the phasic signals emitted by dopaminergic neurons and temporal difference reinforcement learning algorithms³. While reinforcement learning agents have achieved some successes in a variety of domains^{6,7,8}, their applicability has previously been limited to domains in which useful features can be handcrafted, or to domains with fully observed, low-dimensional state spaces. Here we use recent advances in training deep neural networks^{9,10,11} to develop a novel artificial agent, termed a deep Q-network, that can learn successful policies directly from high-dimensional sensory inputs using end-to-end reinforcement learning. We tested this agent on the challenging domain of classic Atari 2600 games¹². We demonstrate that the deep Q-network agent, receiving only the pixels and the game score as inputs, was able to surpass the performance of all previous algorithms and achieve a level comparable to that of a professional human games tester across a set of 49 games, using the same algorithm, network architecture and hyperparameters. This work bridges the divide between high-dimensional sensory inputs and actions, resulting in the first artificial agent that is capable of learning to excel at a diverse array of challenging tasks.

<http://www.nature.com/nature/journal/v518/n7540/full/nature14236.html#figures>