

The Research Group  
**Artificial Intelligence Lab**

has the honor to invite you to the public defense of the PhD thesis of

**Denis Steckelmacher**

to obtain the degree of Doctor of Sciences

Title of the PhD thesis:

**Model-Free Reinforcement Learning for the Real World**

Promotor:

**Prof. dr. Ann Nowé**

The defense will take place on

**Friday, November 6, 2020 at 17h00**

The defense can be followed through a live stream: <https://youtu.be/UGRDV6UfUBY>

### Members of the jury

Prof. dr. Bernard Manderick (VUB, chair)

Prof. dr. Elisa Gonzalez Boix (VUB, secretary)

Prof. dr. ir. Bram Vanderborght (VUB)

Prof. dr. Manuela Veloso (Machine Learning Department, School of Computer Science, Carnegie-Mellon University, US)

Prof. dr. Robert Babuska (Department of Cognitive Robotics, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, NL)

### Curriculum vitae

Denis Steckelmacher graduated as a MSc in Computer Science at the Université Libre de Bruxelles in 2016. He started his PhD at the VUB Artificial Intelligence Lab in 2016. His research focuses on sample-efficient Reinforcement Learning, applied to robots. He also has experience in digital electronics, digital circuit design (VHDL) and high-performance computing. He (co-)authored 9 peer-reviewed papers in international conferences, presented his research at several international events and won the Best Demo Award at the BNAIC conference in 2017 and 2019. He taught classes at the VUB, EHB Brussels, a winter school in Delft, a summer school in Nieuwpoort and a company training at the VUB.

### Abstract of the PhD research

Reinforcement Learning allows an artificial agent to learn how to perform a task, given only sensory inputs and rewards to be maximized. A challenge of Reinforcement Learning, that prevents its wide-spread deployment in many real-world settings, is the need for the agent to learn. When no simulator for a task is available, the agent must learn on the physical system or the user-facing software. In this case, any mistake made by the agent while learning, due to its still incomplete knowledge, or the task, or its inherent need to explore, may have costly effects.

In this thesis, we consider real-world tasks that may benefit from Reinforcement Learning and for which there is no simulator available. We focus on sample-efficiency, that measures how little interactions with the environment the agent needs before learning the task. We believe that with high sample-efficiency (fast learning), few errors are made, leading to an acceptable cost of training the agent.

We follow a research line that starts with theoretical insights and algorithms and aims towards practical applications on physical platforms. We review several families of Reinforcement Learning algorithms, then introduce our first contribution, the model-free actor-critic Bootstrapped Dual Policy Iteration algorithm (BDPI). Thanks to its use of off-policy critics and an explicit actor, BDPI achieves high-quality exploration and sample-efficiency. Our second theoretical contribution is a formalism, based on the Options framework that allows an agent to learn partially-observable tasks (with a discrete memory) in a sample-efficient way. Finally, we focus on a real-world robot, a motorized wheelchair controlled with a joystick and propose extensions to BDPI to make it able to (safely) learn a complex navigation task, directly on the wheelchair, without a model or pre-training, in about one hour of wall-clock time.

We carefully evaluate our algorithms on numerous simulated tasks and on two robotic platforms, one of them the wheelchair mentioned above. In all our experiments, our algorithms outperform state-of-the-art approaches. Our results demonstrate the applicability of Reinforcement Learning to real-world settings, where sample-efficiency is critical.