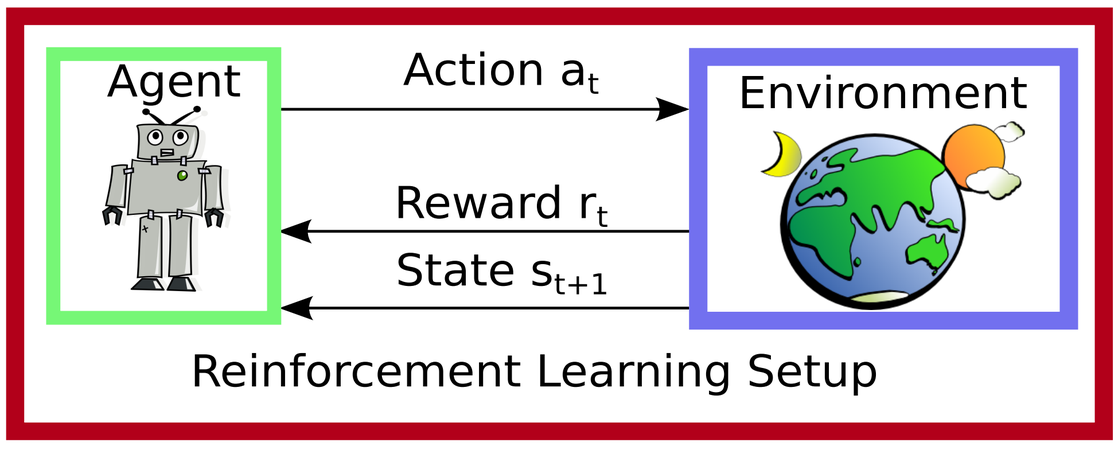
Reinforcement Learning for Movement Skills

TECHNISCHE UNIVERSITÄT DARMSTADT

Reinforcement Learning

Creating autonomous robots that can learn to assist humans in situations of daily life is a challenging task for machine learning.
Supervised learning techniques do not suffice for many motor learning problems, particularly when no expert teacher is available. Thus, the robot has to learn how to self-improve based on an external reward signal.



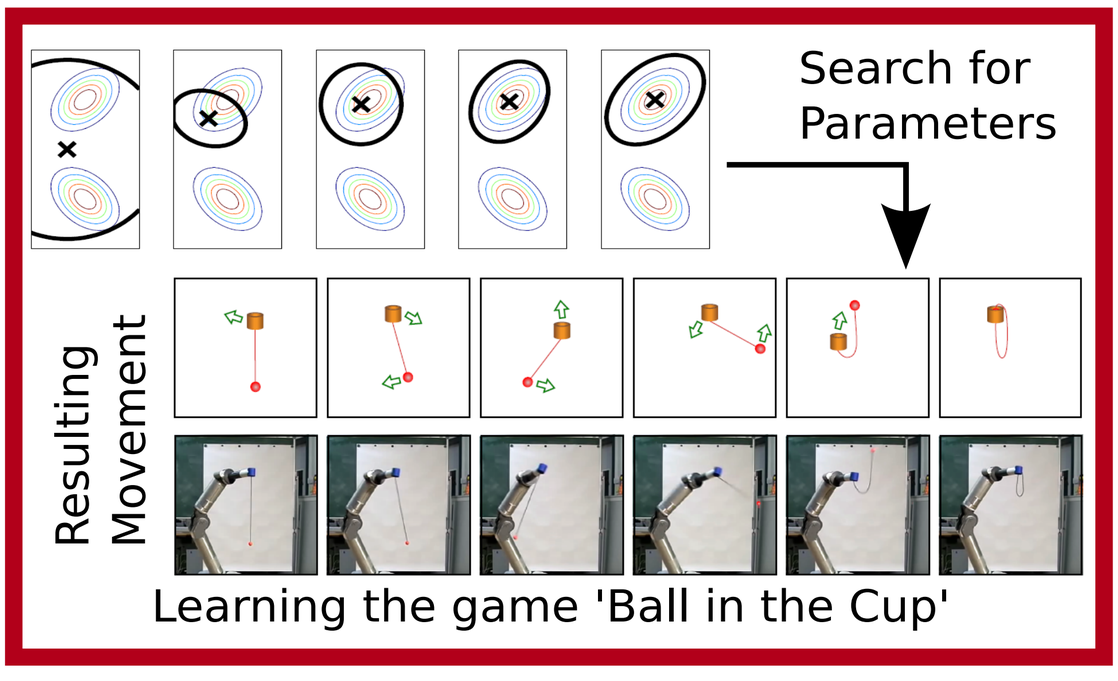
Reinforcement Learning Setup

Our general goal in reinforcement learning is the development of methods which scale into the dimensionality of humanoid robots which is a tremendous challenge for reinforcement learning as a complete exploration of the underlying state-action spaces is impossible and few existing techniques scale into this domain.
Therefore we rely upon a combination of both, watching a teacher and subsequent self-improvement. In more technical terms: first, a control policy is obtained by imitation and then improved using reinforcement learning.

Learning Movement Primitives

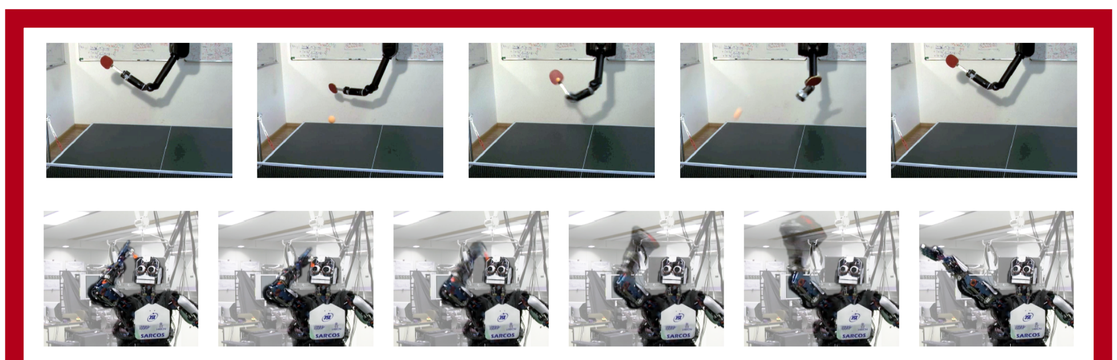
Movement primitives are parametrized movement representations which allow an efficient abstraction of the high-dimensional continuous action spaces which often occur in robotics. We can directly optimize the parameters of the primitive by the use of direct policy search methods.

Members of IAS have developed a variety of novel algorithms for this context which have been applied for learning to play table tennis, the game 'Ball in the Cup' or darts.



Learning the game 'Ball in the Cup'

- *Natural Actor-Critic(NAC)*: The NAC is currently the most efficient policy gradient method. It makes use of the fact, that a natural gradient usually beats a vanilla gradient.
- *EM-like Reinforcement Learning*: We formulated policy search as an inference problem. This has led to efficient algorithms like *reward-weighted regression* and *PoWER*.
- *Relative Entropy Policy Search (REPS)*: The optimization in policy search can rapidly change the control policy which might lead to suboptimal solutions. REPS solves this problem by bounding the Relative Entropy between two subsequent policies. This allows the derivation of a whole range of new algorithms, including learning hierarchical policies.


Robots playing Table Tennis and Darts

Christian Daniel - daniel@ias.tu-darmstadt.com
Marc Deisenroth - deisenroth@ias.tu-darmstadt.de
Jens Kober - jens.kober@tuebingen.mpg.de
Gerhard Neumann - neumann@ias.tu-darmstadt.de
Prof. Jan Peters - mail@jan-peters.net
TU Darmstadt, FB Informatik, FG IAS
Hochschulstraße 10, 64289 Darmstadt

Intelligent Autonomous Systems