Applying Reinforcement Learning in computer games and robotics is difficult because the state space is large and exploration of the state space, e.g. with $\epsilon$-greedy, to find a good policy turns out to be very difficult.

The $\epsilon$-greedy strategy explores the state space completely randomly, similar to a random-walk. Humans, on the other hand, tend to explore unknown areas more systematically by keeping track of where they have been before, and by essentially generating a map of the space in their head (see Fig.). This “simple” idea has turned out to be highly effective according to a recently published NATURE paper *First return, then explore*. In this thesis we want to explore this approach and get a better understanding how it works.

We will be able to build on the existing code, and use the existing performance as a base-line for comparison the more strategic approach.

**Requirements**

To do this work, you should have

- very good python skills

---

1 A. Ecoffet *et al.* First return, then explore. NATURE Vol. 590, Feb 2021.
• very good programming skills so that you can direct and understand the existing code
• completed a machine learning course with a good grade,
• a basic understanding for Reinforcement Learning (e.g. David Silver’s lecture)

Contact:

| Volker Krüger | Professor i Datavetenskap | volker.krueger@cs.lth.se |