# Using shaping for knowledge transfer between Reinforcement Learning and Intrinsically Motivated Reinforcement Learning

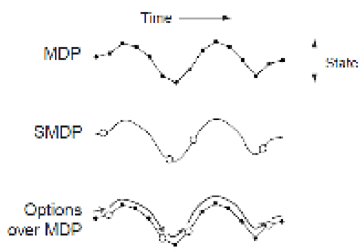
By Paul Guseman

## Introduction

Reinforcement learning is a computational approach to learning through interaction. It maps situations to actions with a goal of maximizing the numerical reward signal. The end result is the learning of a *policy*, which is a mapping from perceived states to actions to be taken from those states. Computationally, this is formalized through Markov Decision Processes (MDPs). Markov Decision Processes include a set of states S, a set of actions A that can be taken from each state s in S, a probability distribution P that state s will transition to s' upon performing action a, a real valued reward r whose expected values is a function dependent on s, s', and a, and a starting state distribution.

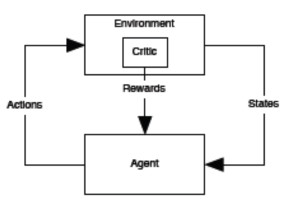
## Options

Options provide a way to add both temporal abstraction and hierarchy to the reinforcement learning model. Temporal abstraction is the extending of actions over time. Essentially you can take a sequence of actions, and represent this as only 1 action.

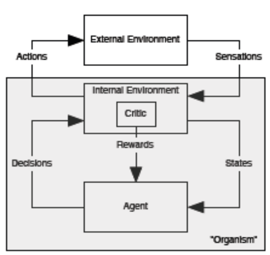



## Intrinsically Motivated Reinforcement Learning

Intrinsically Motivated Reinforcement Learning is a learning paradigm based on reinforcement learning. It internalizes the critic-reward system, transducing external reward signals so that the regular reinforcement learning framework still works. They note that their internal reward environment is favorable because an agent's motivation system "needs to be a sophisticated system that should not have to be redesigned for different problems" (Barto et Al. 2004).




External Critic-Reward System




Internalized Critic-Reward System

## Autonomous Shaping for Knowledge Transfer

Autonomous Shaping is a method for knowledge transfer in reinforcement learning. This knowledge transfer would speed up learning, reducing the time it takes for an agent to acquire a skill. Shaping can be used to speed up episodic learning over a series of similar tasks. The method proposed by Barto, Konidaris et Al. requires the agent to learn in both problem space and agent-space. While the problem space retains the Markov Property, the agent space will not necessarily be Markov. The concept is that the agent will learn to associate sensations in agent space to reward signals in problem space, and will thus learn faster in the later episodes of learning.

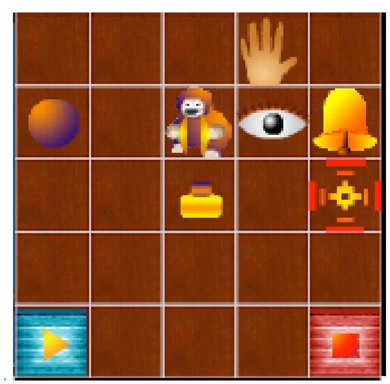
$$s_i^j = (d_i^j, c_i^j, r_i^j, v_i^j),$$

Let d be the problem space descriptor(The state, action set, and probability for transition of the MDP), while c is the agent space sensation, r is the reward at that time, and v is the value for that state.

$$L : c_i^j \mapsto v_i^j.$$

Autonomous Shaping learns the function L, a mapping from agent space sensation to problem space state values.

## Intrinsic Reinforcement Learning Example



## Drawbacks to Reinforcement Learning

Reinforcement Learning has two known detractors, the scaling of state space and the long learning time required. Because reinforcement learning is formalized with Markov Decision Processes, for a problem space sufficiently large or complex, the requirements to represent, store, and compute become daunting and even infeasible.

## Knowledge Transfer of a reinforcement learned skill.

I propose that intrinsically motivated reinforcement learning can be sped up using autonomous shaping to transfer knowledge from a reinforcement learning domain (Or even a simpler intrinsically motivated domain). I propose crafting a strong external reward signal to accomplish a specific task, one that is widely used in many options, or that is particularly slow to learn, and then training the agent strictly for that one skill in a reinforcement learning manner. This can then be used to achieve the "broad competency" desired by intrinsically motivated learning in a more expedient manner.

## Scale the domain

Additionally, I propose scaling the domain size to test the limits of the intrinsically motivated reinforcement learning paradigm, and to see in what way domain size effects the advantages of knowledge transfer proposed above.

## Requirements

To test my proposal would require the implementation of a more complex robotworld domain for the intrinsically motivated agent, and extensive testing for each option on the benefits/costs of using shaping for knowledge transfer.

## Future Work

Future work could address the scaling representation problem, perhaps using layered learning. Also, adding intrinsic motivators, because currently only saliency and transduced external rewards drive the agent, and it has been pointed out that this favors the "area of proximal development"