Behavior Sets for Adaptive Social Robotics

Chris Reinke

Inria Grenoble / Robotlearn

Socially Pertinent Robots in Gerontological Healthcare

Overall goal: Socially assistive robots

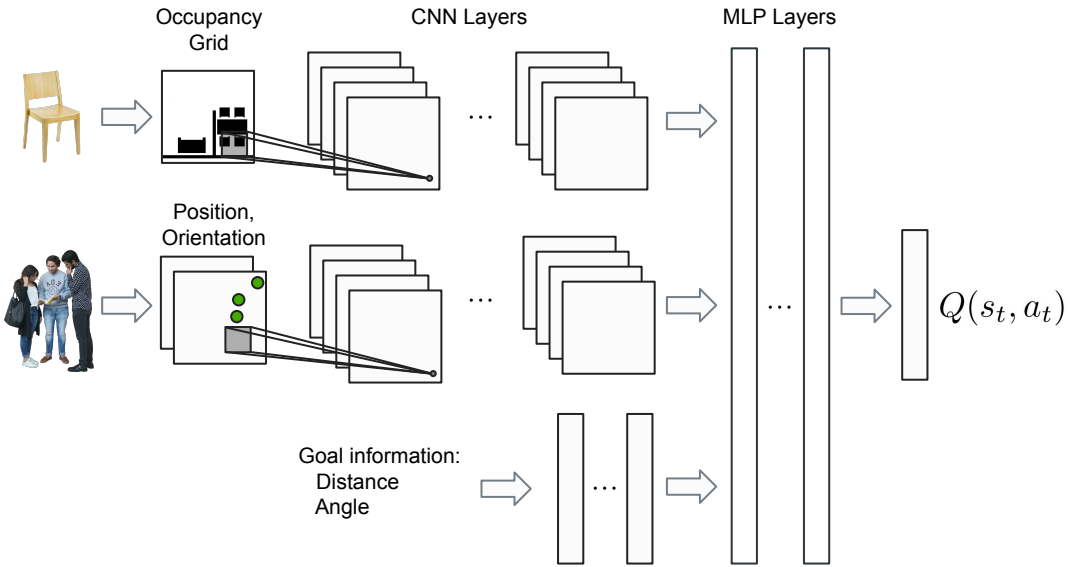
Goals of the team:

- ❏ Human-aware navigation
- ❏ Join persons and groups
- ❏ Optimal perception while navigating

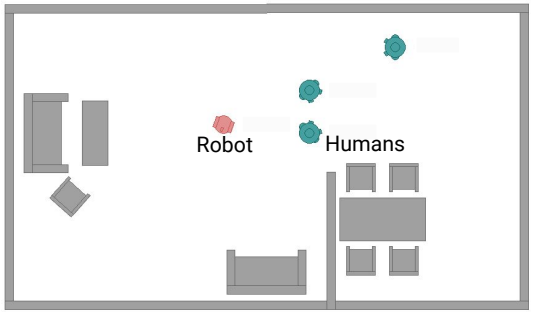
Alameda-Pineda, Addlesee, Hernandez Garcia, Reinke et al. (2024) “Socially Pertinent Robots in Gerontological Healthcare”, Journal of Social Robotics [under review]

Reinforcement Learning for Social Robotics

Deep Reinforcement Learning



Training in simulation



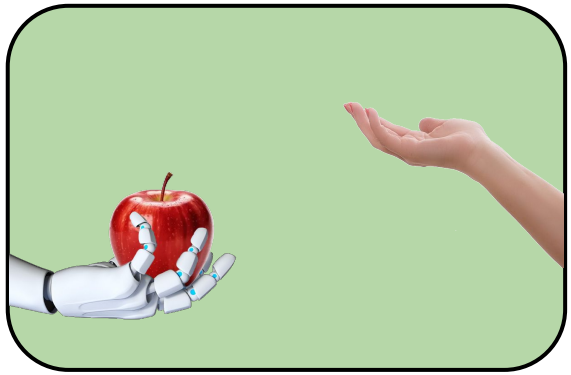
Sim2Real transfer

Sanchez; Supervison: Reinke (2023) “Deep reinforcement learning on social environment aware navigation based on maps”, KTH Masters Thesis

Challenge in Social Robotics

Social behaviors are highly user and context dependent

Task



User / Context

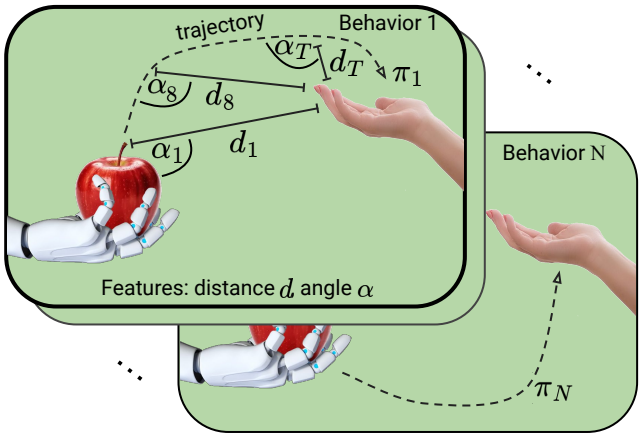
Senior

Younger adult

Research question: How to adapt behavior quickly?

Proposal: Behavior Sets

Behavior Set



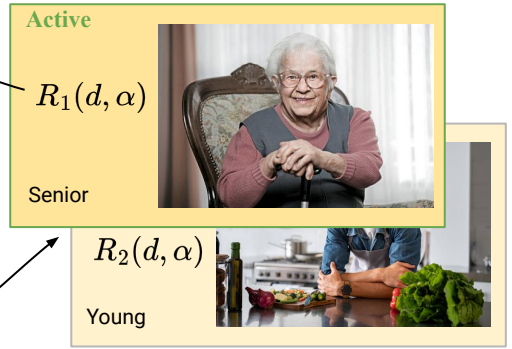
1) Evaluate each Policy

$$Q(\pi) = \sum_t R(d_t, \alpha_t)$$
$$Q(\pi_1) = 350$$
$$Q(\pi_2) = 270$$
$$\vdots$$
$$Q(\pi_N) = 290$$

2) Use most appropriate Policy

π_1

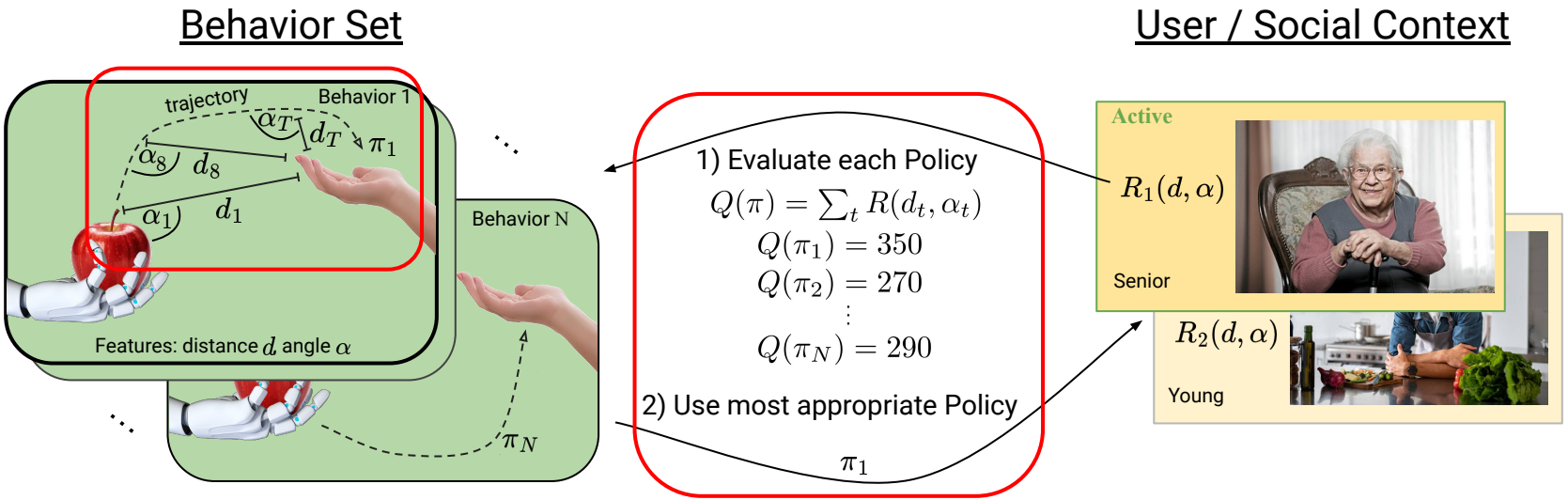
User / Social Context



Research Questions

1) How to predict and evaluate behavior outcomes?
2) How to learn behavior sets?
3) How to model user preferences?

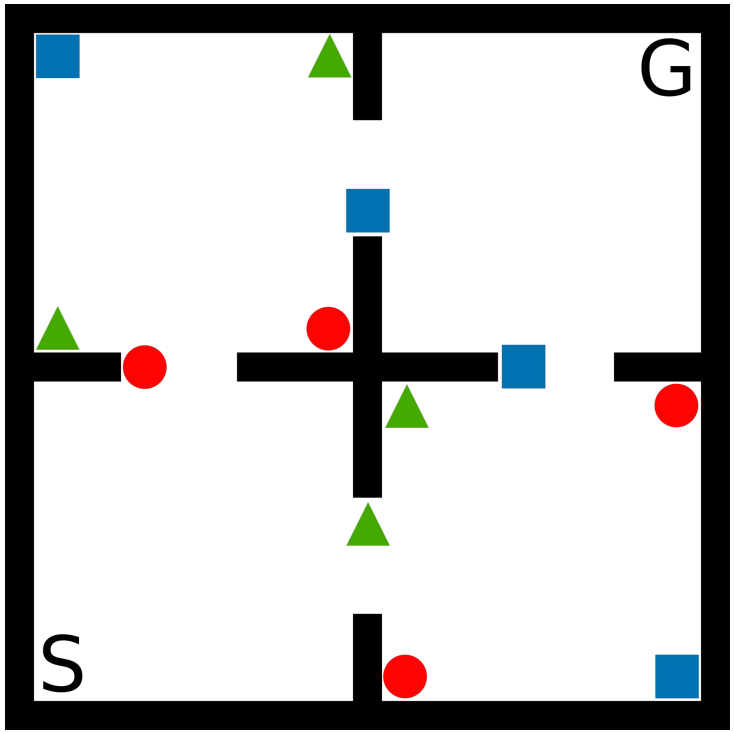
1) How to predict and evaluate behavior outcomes?



→ Successor Feature Representations

Reinke, Alameda-Pineda (2023) “Successor Feature Representations”, Transactions on Machine Learning Research

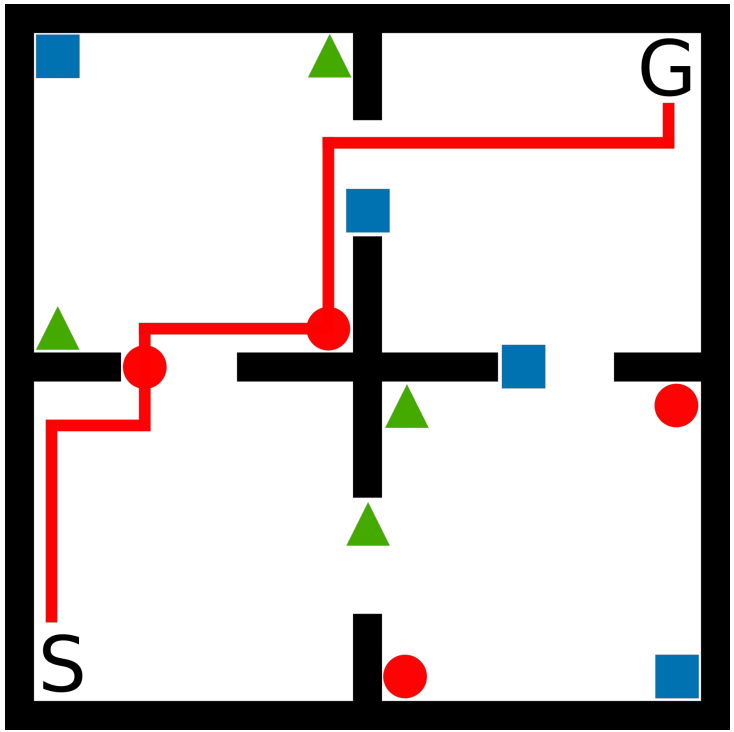
Simple Example

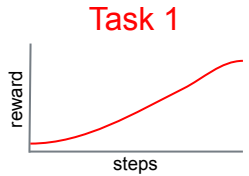


- 2D grid world
- Go from start (S) to goal (G)
- Get rewards for collecting objects
- Rewards differ between tasks

Barreto et al. (2017) “Successor features for transfer in reinforcement learning”, NeurIPS

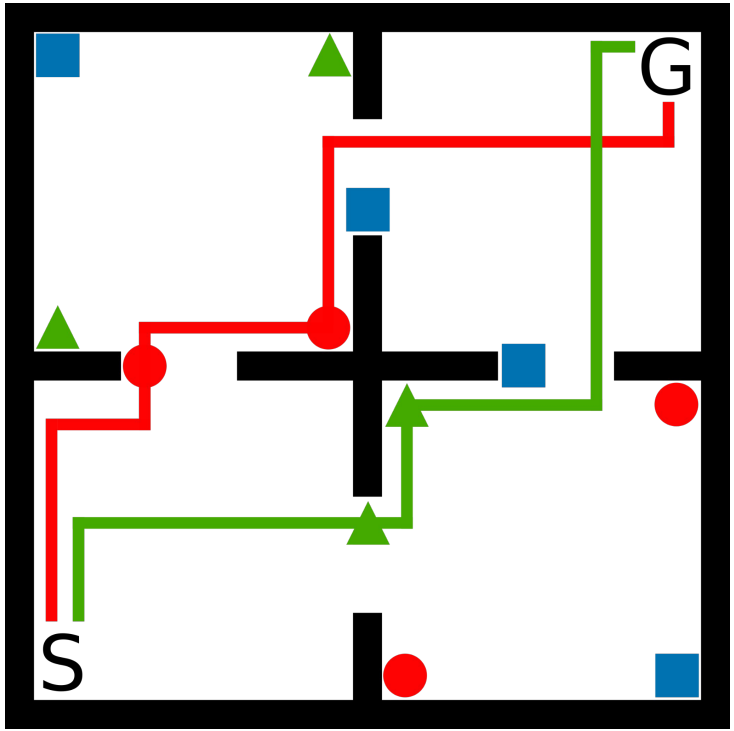
Simple Example



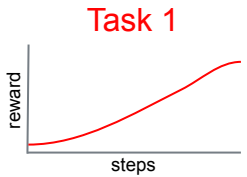


Barreto et al. (2017) “Successor features for transfer in reinforcement learning”, NeurIPS

Simple Example

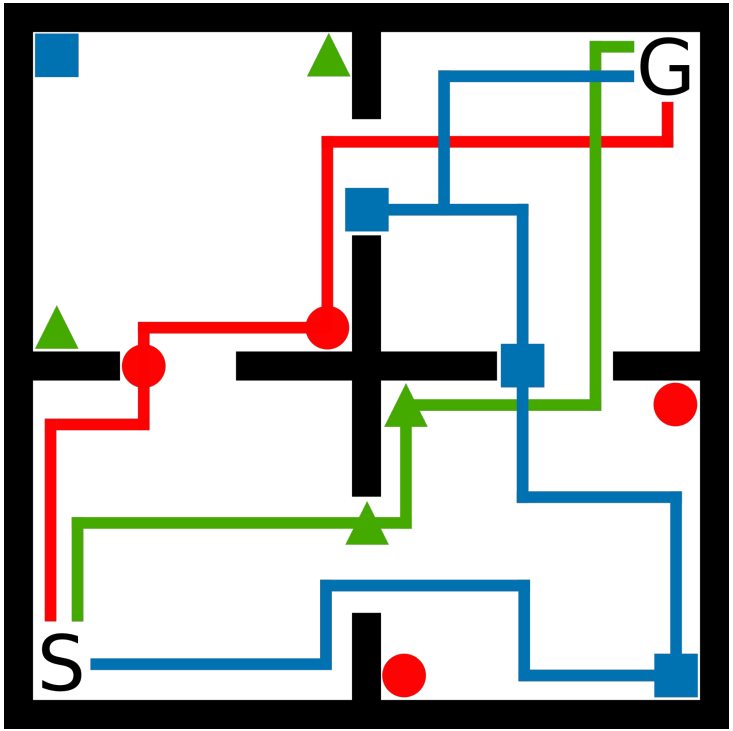


R	●	▲	■
Task 1	**1.0**	0.1	0.0
Task 2	0.1	**1.0**	0.0

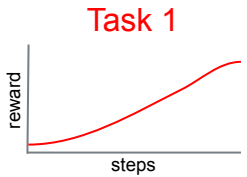


Barreto et al. (2017) “Successor features for transfer in reinforcement learning”, NeurIPS

Simple Example

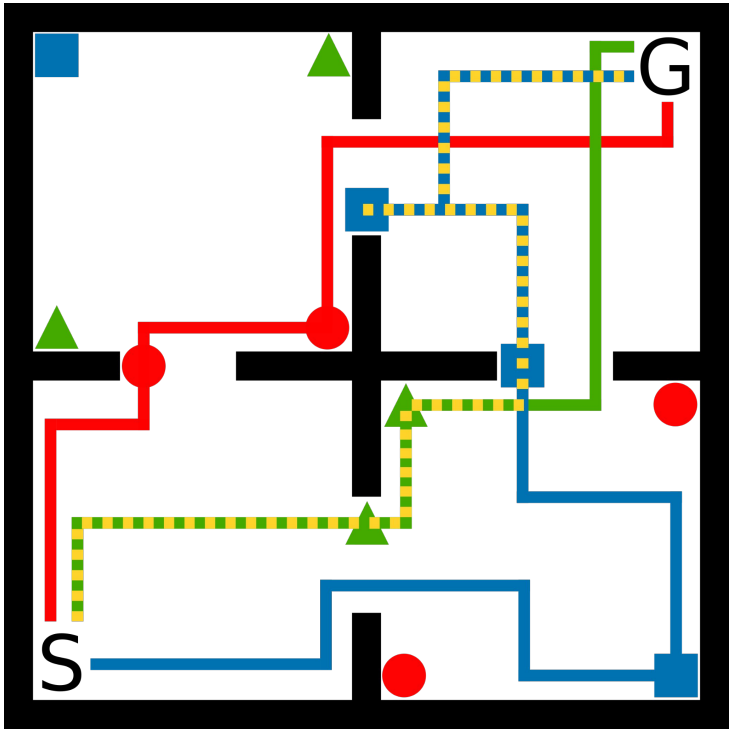


R	●	▲	■
Task 1	**1.0**	0.1	0.0
Task 2	0.1	**1.0**	0.0
Task 3	0.2	0.0	**1.0**

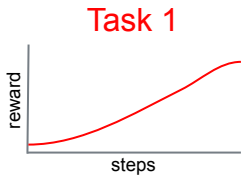


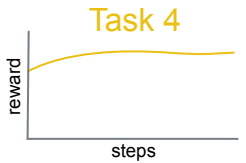
Barreto et al. (2017) “Successor features for transfer in reinforcement learning”, NeurIPS

Simple Example



R	●	▲	■
Task 1	**1.0**	0.1	0.0
Task 2	0.1	**1.0**	0.0
Task 3	0.2	0.0	**1.0**
Task 4	0.0	**0.9**	**0.6**





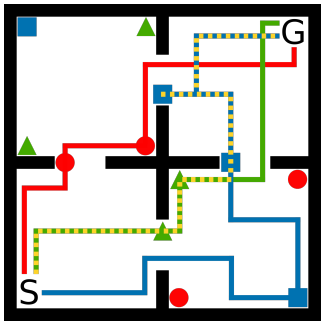
Barreto et al. (2017) “Successor features for transfer in reinforcement learning”, NeurIPS

Successor Feature Representation (SFR)

Assumption: Low-dim features to enode rewards $r_t = R(\boldsymbol{\phi}_t)$

Successor Feature Representation:

$$
\begin{aligned}
Q^{\pi}(s_t, a_t) &= \sum_{k=0}^{\infty} \gamma^k \mathbb{E}_{p(\boldsymbol{\phi}_{t+k}|s_t,a_t;\pi)} \left\{ R(\boldsymbol{\phi}_{t+k}) \right\} \\
&= \sum_{k=0}^{\infty} \gamma^k \int_{\Phi} p(\boldsymbol{\phi}_{t+k} = \boldsymbol{\phi}|s_t, a_t; \pi) R(\boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi} \\
&= \int_{\Phi} R(\boldsymbol{\phi}) \sum_{k=0}^{\infty} \gamma^k p(\boldsymbol{\phi}_{t+k} = \boldsymbol{\phi}|s_t, a_t; \pi) \mathrm{d}\boldsymbol{\phi} \\
&= \int_{\Phi} R(\boldsymbol{\phi}) \xi^{\pi}(s_t, a_t, \boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi} = \sum_{\boldsymbol{\phi} \in \Phi} R(\boldsymbol{\phi}) \xi^{\pi}(s_t, a_t, \boldsymbol{\phi})
\end{aligned}
$$

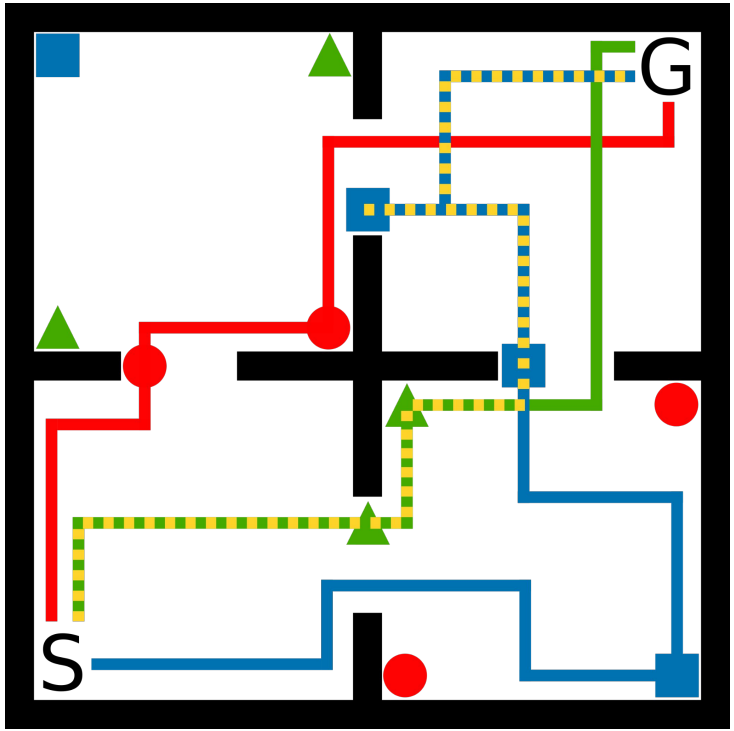


$$\boldsymbol{\phi}_t = \begin{pmatrix} \bullet ? \\ \blacktriangle ? \\ \blacksquare ? \end{pmatrix}$$

R	●	▲	■
Task 1	**1.0**	0.1	0.0
Task 2	0.1	**1.0**	0.0
Task 3	0.2	0.0	**1.0**
Task x	0.0	**0.9**	**0.6**

Reinke, Alameda-Pineda (2023) “Successor Feature Representations”, Transactions on Machine Learning Research

How to Transfer: Generalized Policy Improvement (GPI)



	R	●	▲	■
π_1	Task 1	**1.0**	0.1	0.0
π_2	Task 2	0.1	**1.0**	0.0
π_3	Task 3	0.2	0.0	**1.0**
π_4	Task 4	0.0	**0.9**	**0.6**

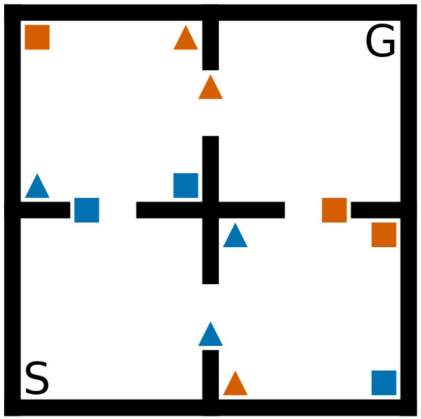
GPI

$$\tilde{\pi}_R(s) = \operatorname*{argmax}_a \max_{\pi_j} Q_R^{\pi_j}(s,a)$$

$$= \operatorname*{argmax}_a \max_{\pi_j} \sum_{\phi \in \Phi} R(\phi)\xi^{\pi_j}(s,a,\phi)$$

Reinke, Alameda-Pineda (2023) “Successor Feature Representations”, Transactions on Machine Learning Research

SFR Results - Discrete Features



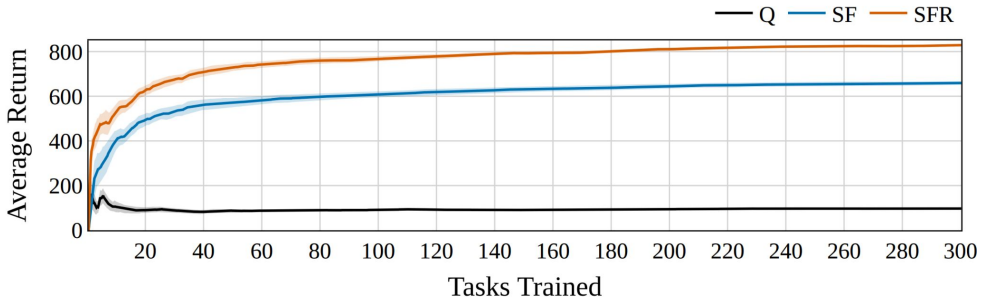
$\phi_t =$ ▲ ? ■ ? ● ? ● ? **G** ?

Form

Color

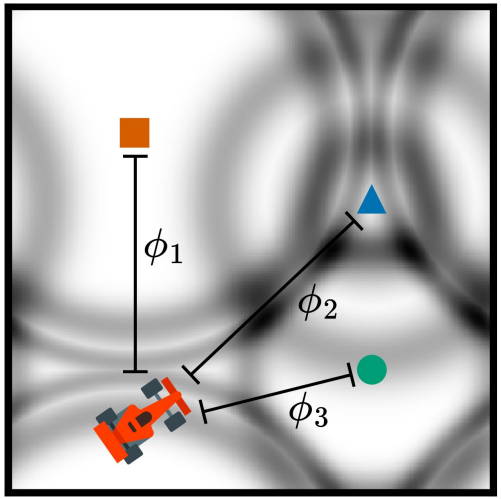
Goal

General Reward Functions: $r_t = R(\phi_t)$

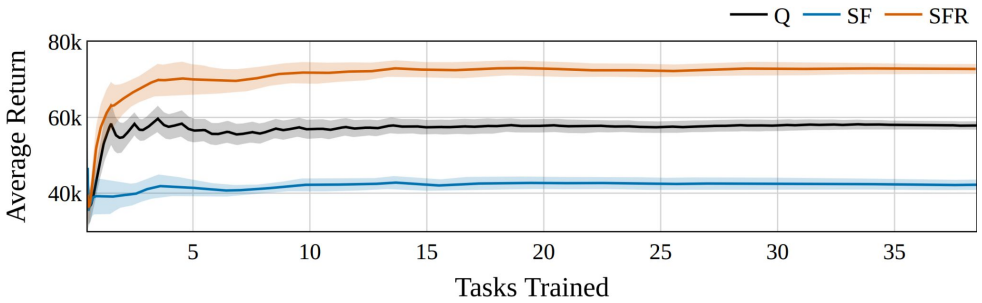


Reinke, Alameda-Pineda (2023) “Successor Feature Representations”, Transactions on Machine Learning Research

SFR Results - Continuous Features



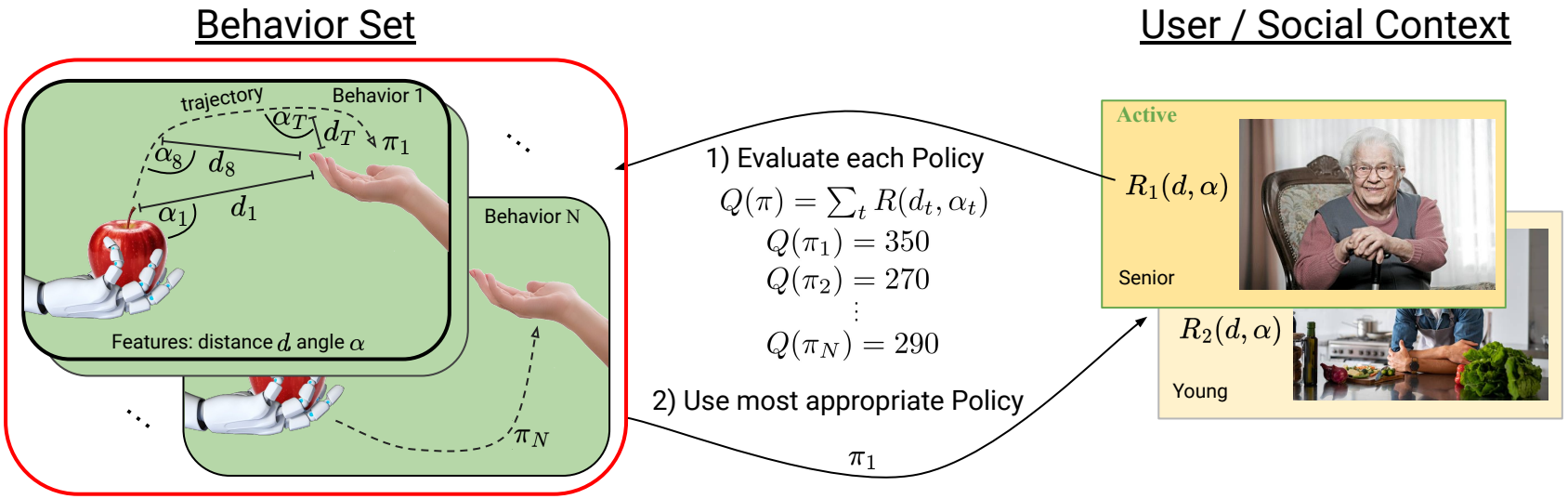
General Reward Functions: $r_t = R(\phi_t)$



$$R(\phi) = \sum_{i=1}^{3} r_i(\phi_i) \quad \text{with} \quad r_i(\phi_i) = \frac{1}{3} \max \left\{ \exp \left(-\frac{(\phi_i - \mu_i)^2}{\sigma_j} \right) \right\}_{j=1}^{m}$$

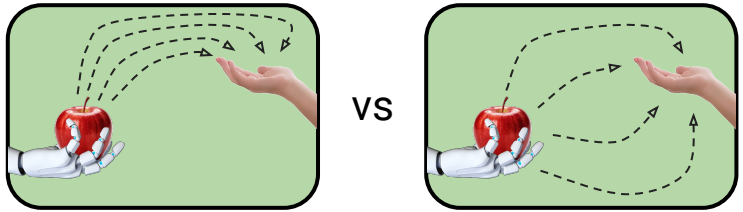
Reinke, Alameda-Pineda (2023) “Successor Feature Representations”, Transactions on Machine Learning Research

2) How to learn behavior sets?



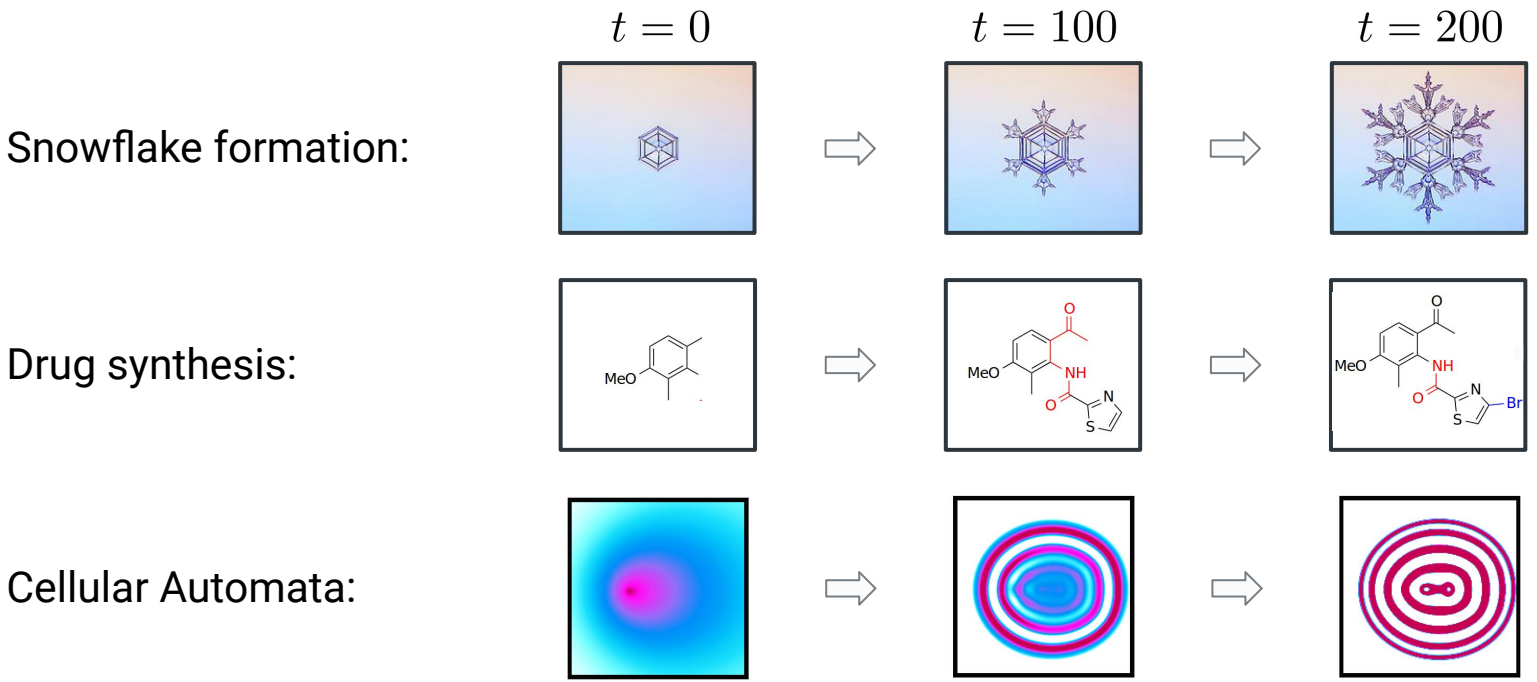
Goal: Learn a high diversity of behaviors

→ Diversity Exploration

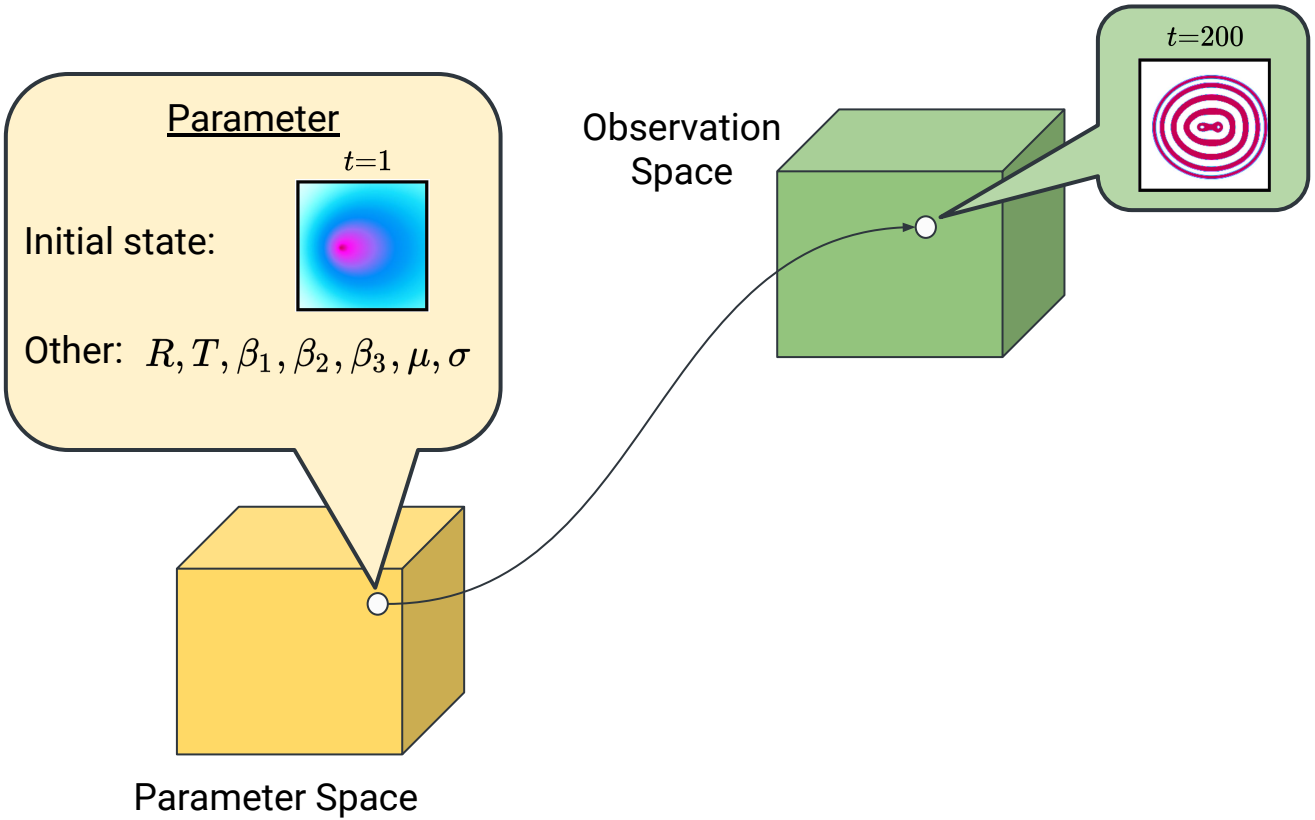


Diversity Exploration

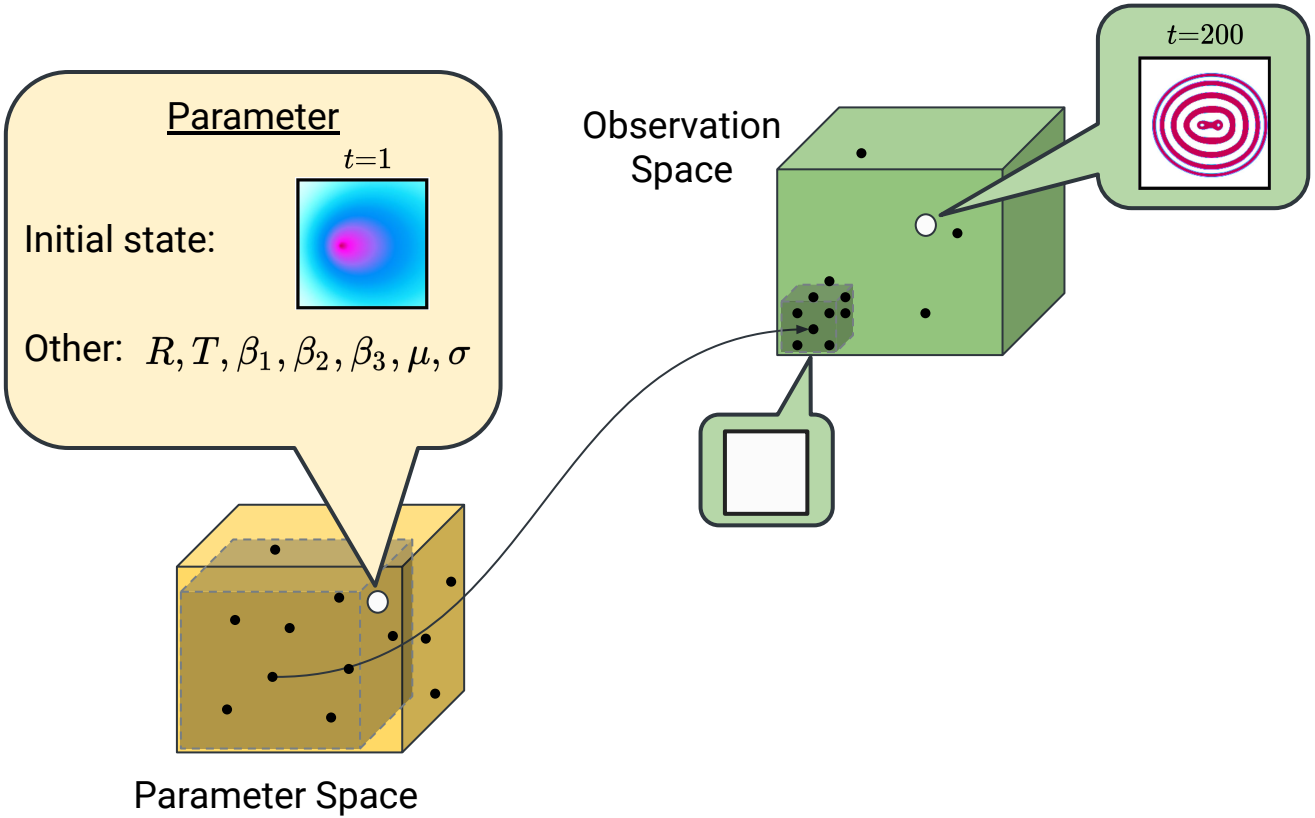
How to explore a high diversity of outcomes
for high-dimensional dynamic (black box) systems?



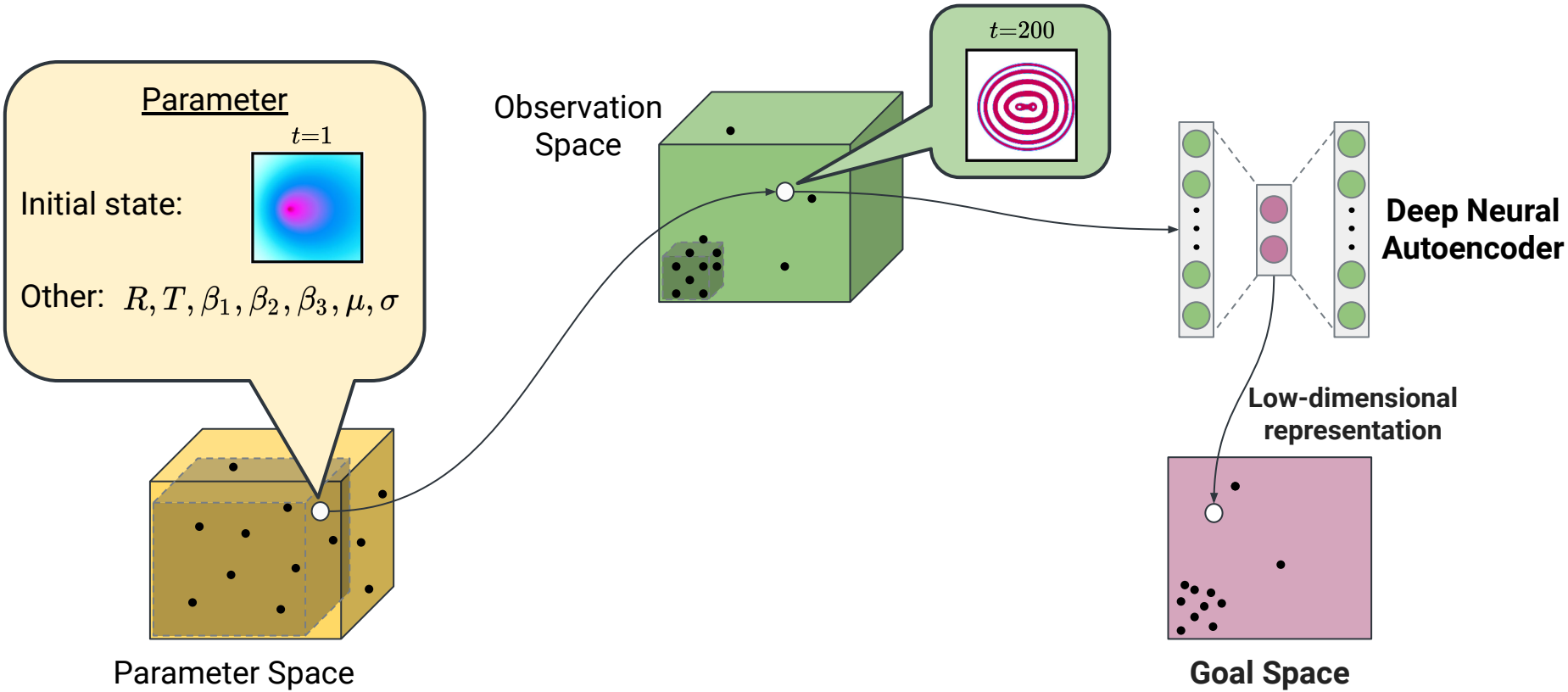
Diversity Exploration



Diversity Exploration



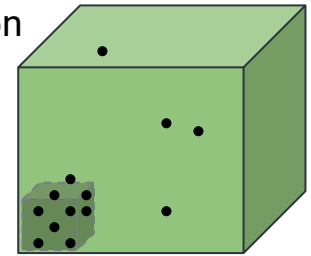
Diversity Exploration

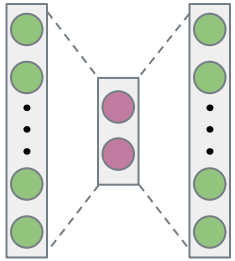


Reinke, Etcheverry, Oudeyer (2020) “Intrinsically Motivated Exploration for Automated Discovery of Patterns in Morphogenetic Systems”, ICLR

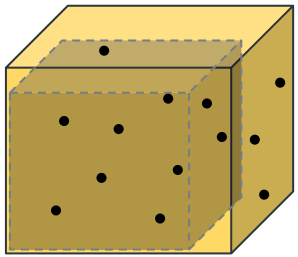
Diversity Exploration

Observation Space

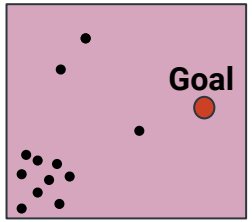




Deep Neural Autoencoder



Parameter Space

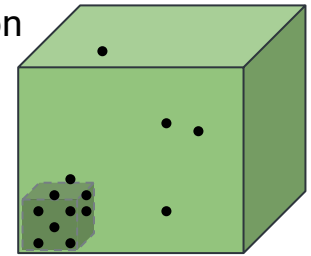


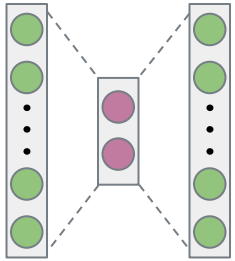
Goal Space

Reinke, Etcheverry, Oudeyer (2020) “Intrinsically Motivated Exploration for Automated Discovery of Patterns in Morphogenetic Systems”, ICLR

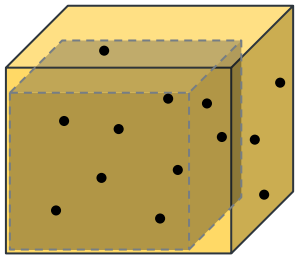
Diversity Exploration

Observation Space

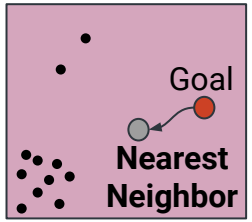




Deep Neural Autoencoder



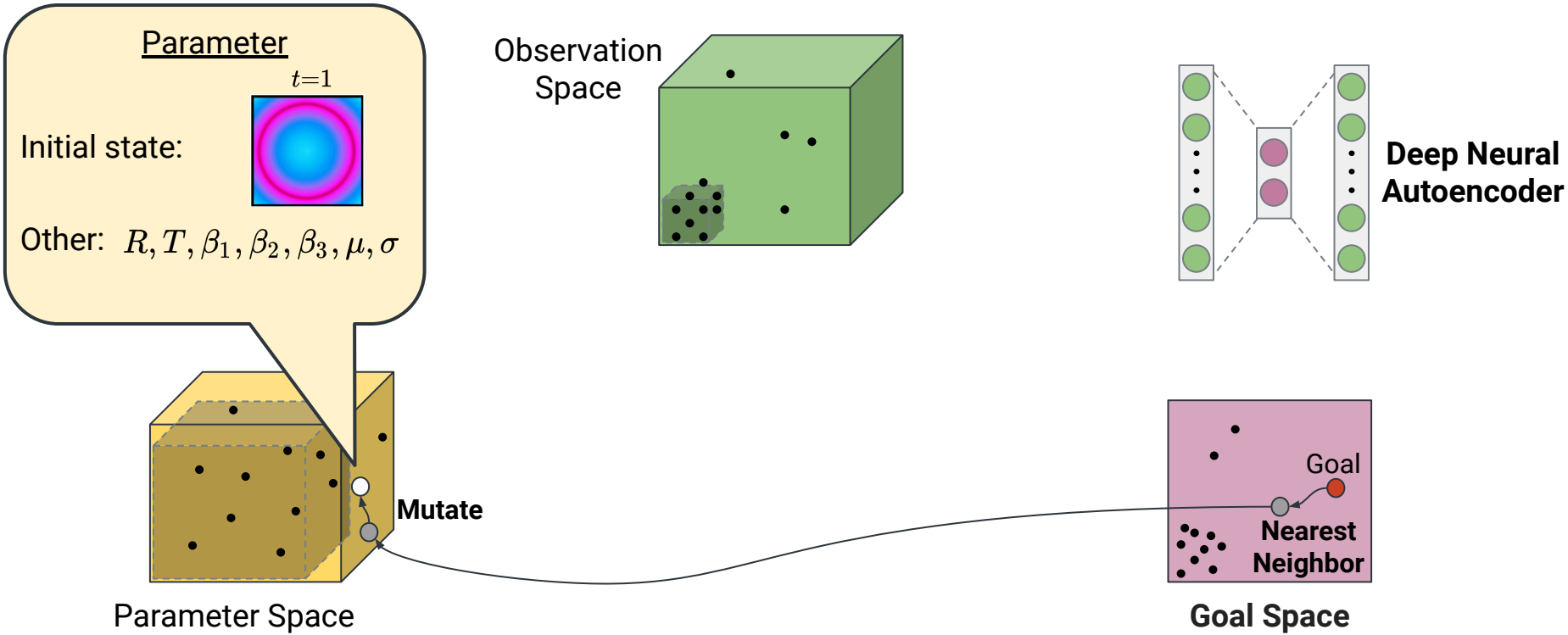
Parameter Space



Goal Space

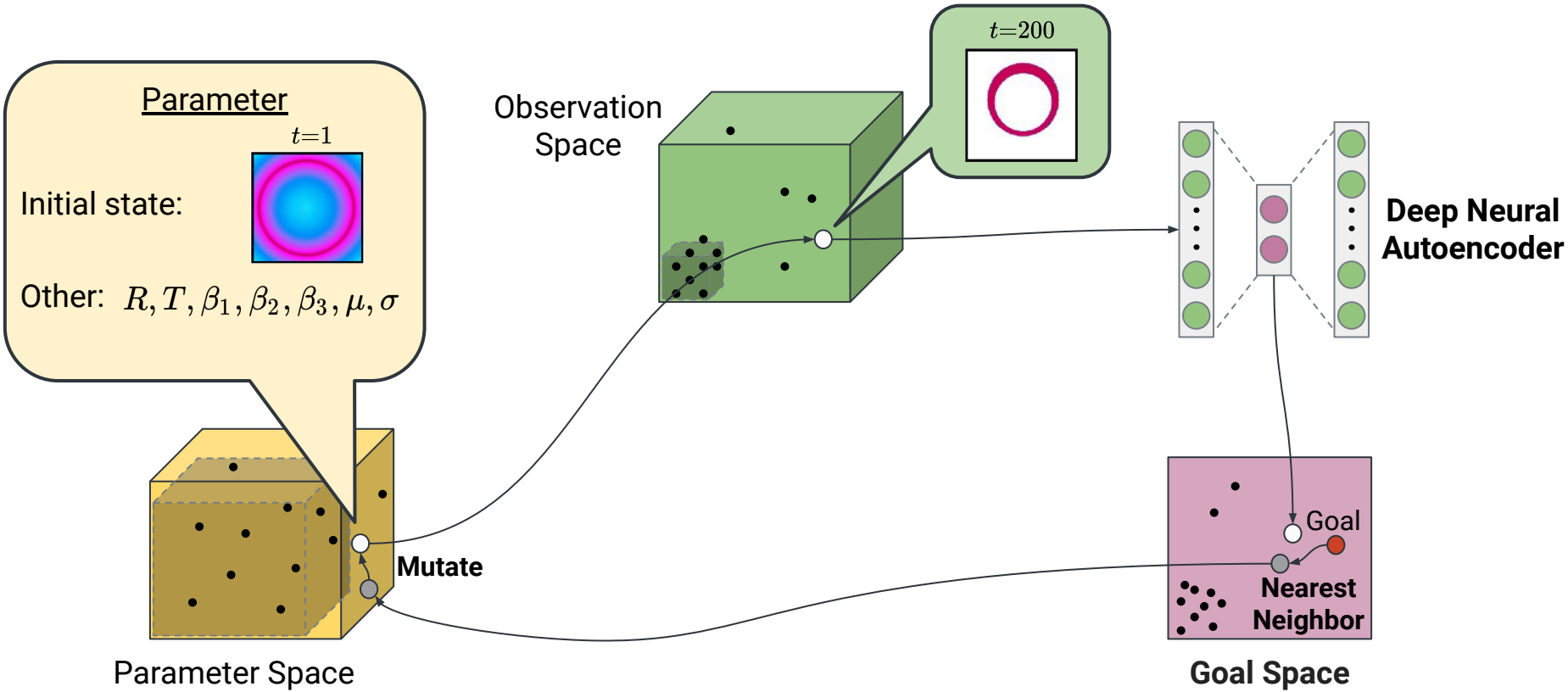
Reinke, Etcheverry, Oudeyer (2020) “Intrinsically Motivated Exploration for Automated Discovery of Patterns in Morphogenetic Systems”, ICLR

Diversity Exploration



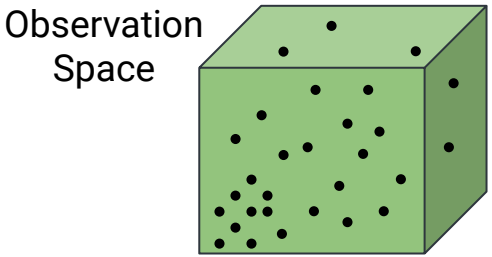
Reinke, Etcheverry, Oudeyer (2020) “Intrinsically Motivated Exploration for Automated Discovery of Patterns in Morphogenetic Systems”, ICLR

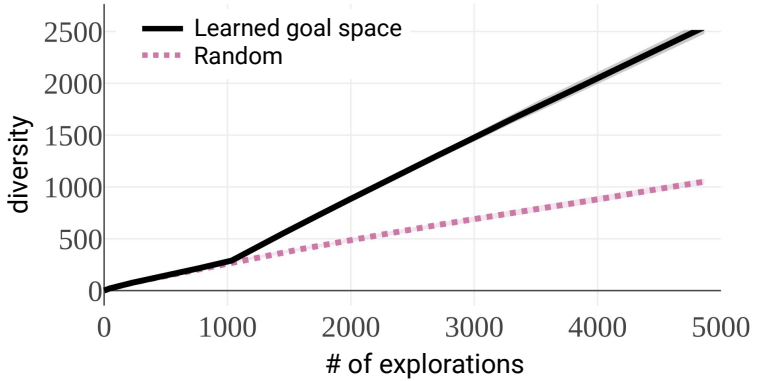
Diversity Exploration



Reinke, Etcheverry, Oudeyer (2020) “Intrinsically Motivated Exploration for Automated Discovery of Patterns in Morphogenetic Systems”, ICLR

Diversity Exploration





Reinke, Etcheverry, Oudeyer (2020) “Intrinsically Motivated Exploration for Automated Discovery of Patterns in Morphogenetic Systems”, ICLR

2) How to learn behavior sets?

Goal

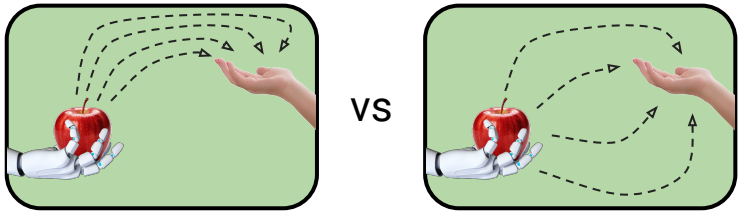
Learning a high diversity of behaviors

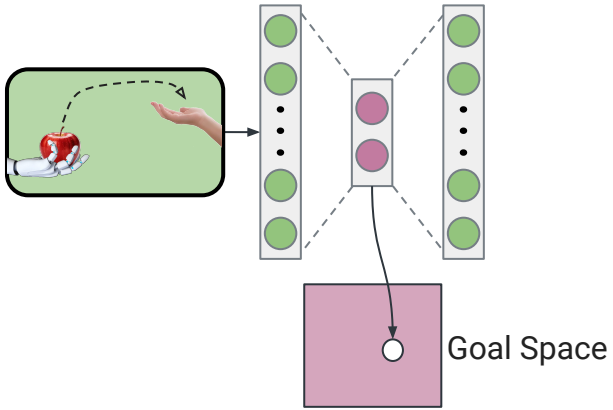
State of the art

Diversity exploration focuses on final outcomes of systems and not dynamics (trajectories)

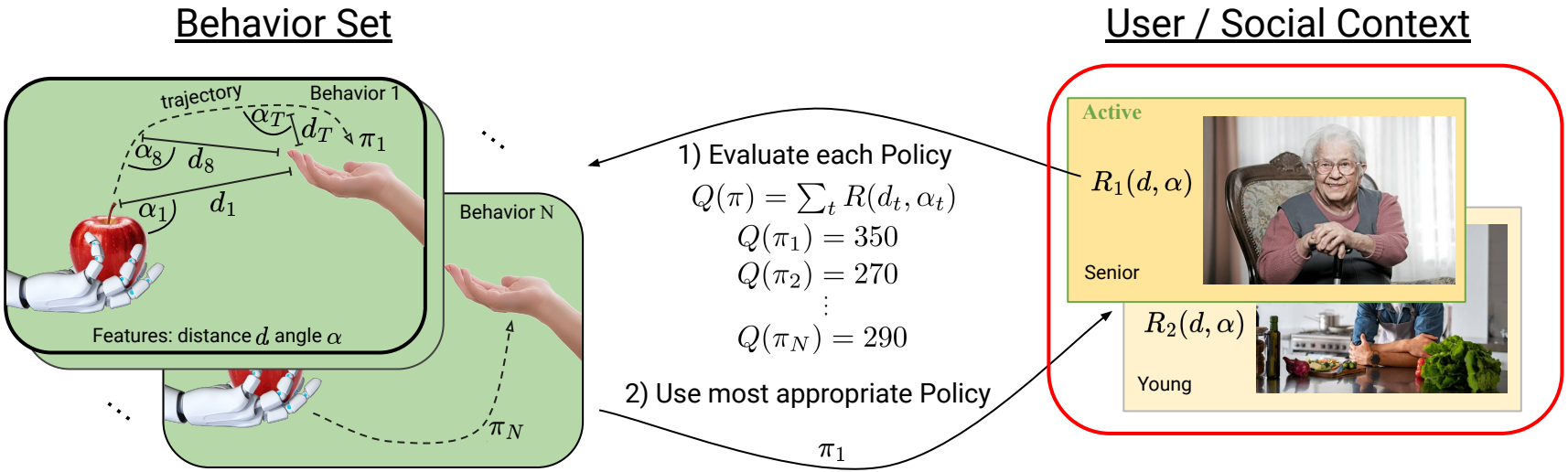
Proposed approach

Investigate unsupervised models that represent dynamics such as Dynamical VAEs





3) How to model user preferences?



3) How to model user preferences?

Goal

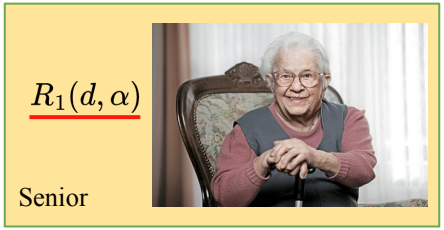
Correct and fast modeling of user preferences

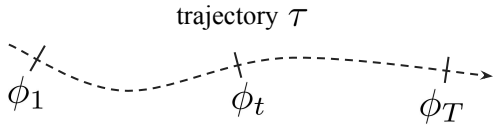
State of the art

Model preferences over full trajectories

Proposed approaches

- ❏ Model preferences over features instead of trajectories to improve data efficiency
- ❏ Directly associating time dependent indirect preference signals (e.g. attention) with features





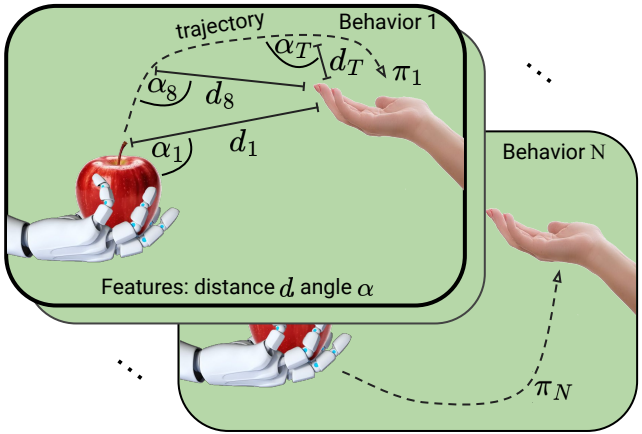
$R(\tau) \sim$ User Evaluation

vs

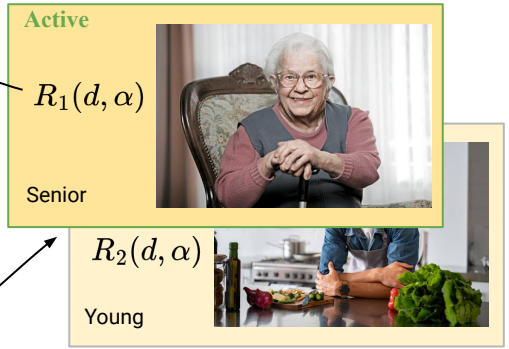
$R(\phi_1) + \ldots + R(\phi_T) \sim$ User Evaluation

Thank You

Behavior Set



User / Social Context



1) Evaluate each Policy

$$Q(\pi) = \sum_t R(d_t, \alpha_t)$$
$$Q(\pi_1) = 350$$
$$Q(\pi_2) = 270$$
$$\vdots$$
$$Q(\pi_N) = 290$$

2) Use most appropriate Policy

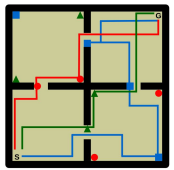
π_1

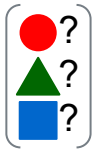
More Information

- ❏ Myself: www.scirei.net
- ❏ SFR: gitlab.inria.fr/robotlearn/sfr_learning
- ❏ Diversity Exploration: automated-discovery.github.io

Background: Successor Features

- Assumption: Rewards are composed of features and weights $r_t = \boldsymbol{\phi}_t^\top \mathbf{w}$



$\boldsymbol{\phi}_t =$ 

R				
Task 1	**1.0**	0.1	0.0	$= \mathbf{w}_1$
Task 2	0.1	**1.0**	0.0	$= \mathbf{w}_2$
Task 3	0.2	0.0	**1.0**	$= \mathbf{w}_3$
Task x	0.0	**0.9**	**0.6**	$= \mathbf{w}_x$

- Successor Features (SF) disassociate dynamics and rewards in Q-function

$$
\begin{aligned}
Q^\pi(s_t, a_t) &= \mathbb{E}\left[r_t + \gamma^1 r_{t+1} + \gamma^2 r_{t+2} + \ldots\right] \\
&= \mathbb{E}\left[\boldsymbol{\phi}_t^\top \mathbf{w} + \gamma^1 \boldsymbol{\phi}_{t+1}^\top \mathbf{w} + \gamma^2 \boldsymbol{\phi}_{t+2}^\top \mathbf{w} + \ldots\right] \\
&= \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \boldsymbol{\phi}_{t+k}\right]^\top \mathbf{w} \\
&\equiv \boldsymbol{\psi}^\pi(s_t, a_t)^\top \mathbf{w}
\end{aligned}
$$

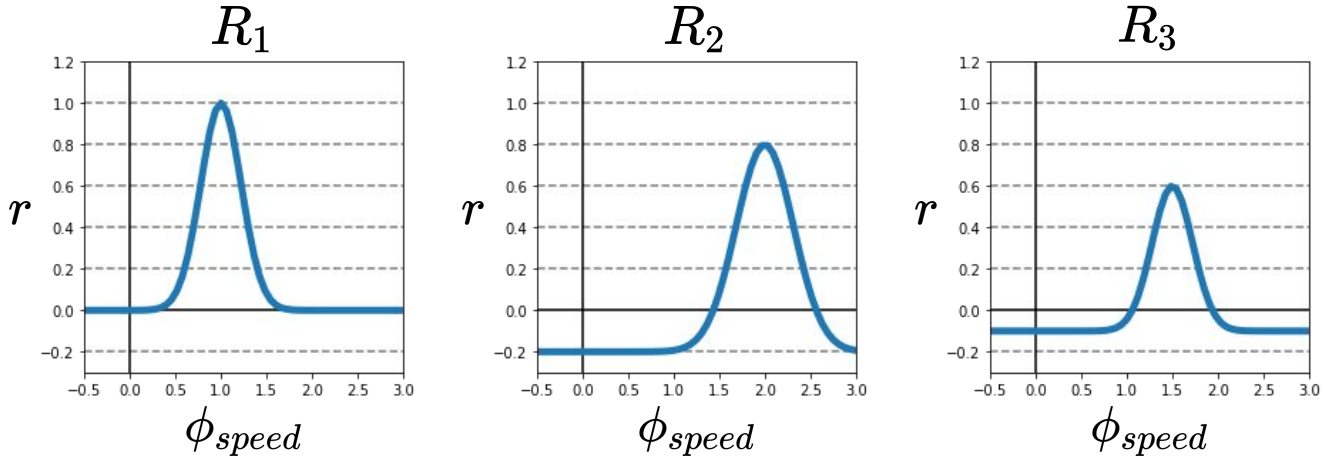
$$r_t \equiv R(s_t, a_t, s_{t+1}) \in \mathbb{R}$$
$$\boldsymbol{\phi}_t \equiv \boldsymbol{\phi}(s_t, a_t, s_{t+1}) \in \mathbb{R}^n, \quad \mathbf{w} \in \mathbb{R}^n$$

Barreto et al. (2017)

Background: Successor Features

- Problem: Assumption of linear relation of rewards $r_t = \phi_t^\top \mathbf{w}$

$$r(\phi_{speed}) = \mathrm{w} \exp\left(-\frac{(\phi_{speed} - \mu)^2}{\sigma}\right) - d$$



Proposal: Successor Feature Representations

- Goal: Allow general reward functions $r_t = \boldsymbol{\phi}_t^\top \mathbf{w} \longrightarrow r_t = R(\boldsymbol{\phi}_t)$

- Successor Feature Representation (SFR):

$$
\begin{aligned}
Q^\pi(s_t, a_t) &= \sum_{k=0}^{\infty} \gamma^k \mathbb{E}_{p(\phi_{t+k}|s_t,a_t;\pi)} \{R(\boldsymbol{\phi}_{t+k})\} \\
&= \sum_{k=0}^{\infty} \gamma^k \int_\Phi p(\boldsymbol{\phi}_{t+k} = \boldsymbol{\phi}|s_t, a_t; \pi) R(\boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi} \\
&= \int_\Phi R(\boldsymbol{\phi}) \sum_{k=0}^{\infty} \gamma^k p(\boldsymbol{\phi}_{t+k} = \boldsymbol{\phi}|s_t, a_t; \pi) \mathrm{d}\boldsymbol{\phi} \\
&= \int_\Phi R(\boldsymbol{\phi}) \xi^\pi(s_t, a_t, \boldsymbol{\phi}) \mathrm{d}\boldsymbol{\phi}
\end{aligned}
$$

Dayan (1993) “The successor representation”