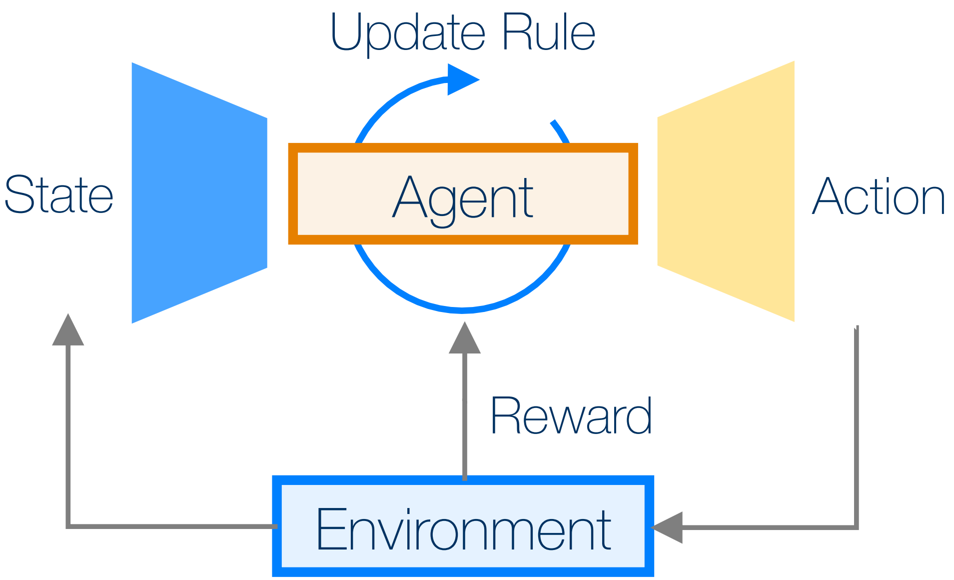
Towards Generalizable Autonomy

Structure in Reinforcement Learning for Robotics



Animesh Garg

University of Toronto
Vector Institute
NVIDIA

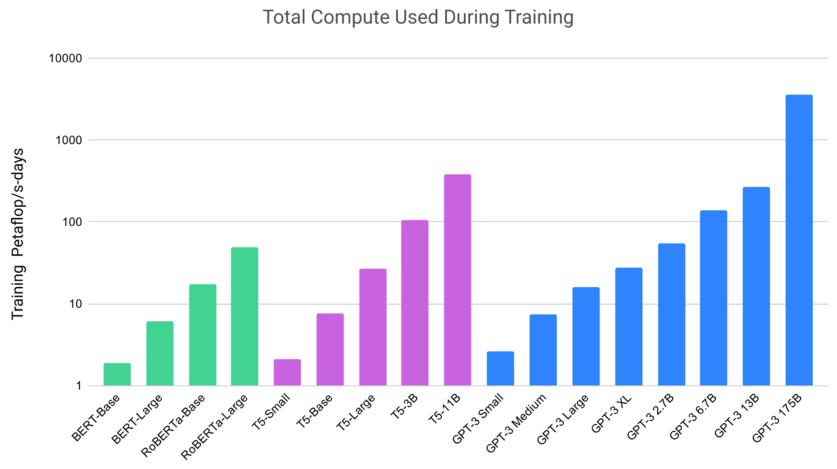
Generalizable Autonomy: Computer Vision & Language

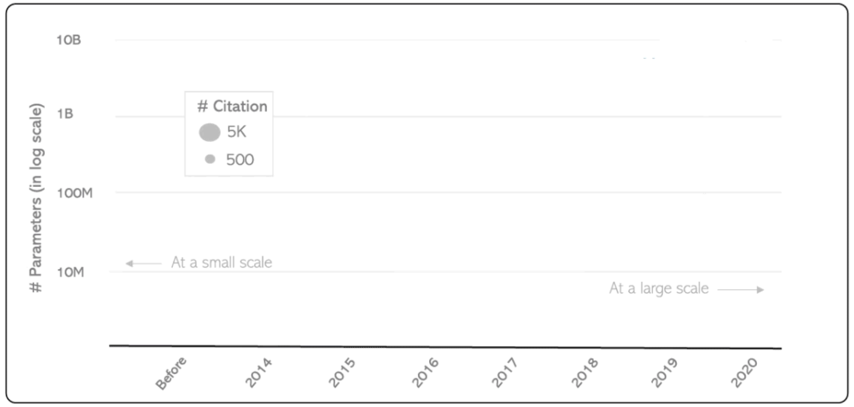
Structured Models + Data + Compute → Performance

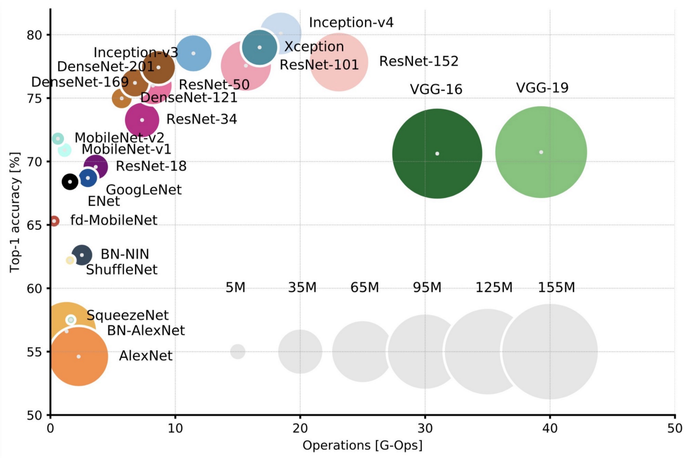
Open Images Dataset

SQuAD
The Stanford Question Answering Dataset

Common Crawl







Model	EM	F1
Human Performance *Stanford University* (Rajpurkar & Jia et al. '18)	86.831	89.452
IE-Net (ensemble) *RICOH_SRCB_DML*	**90.939**	**93.214**

Generalizable Autonomy: Computer Vision & Language

Ingredients of Modern Machine Learning & Applications

Large Structured Models

- Over-parameterized
- Structured Biases

IID Data & Datasets

- Concise problem Definition
- IID Data, easier to label

Distributed Deployment

- Large Scale Compute
- Distributed Deployment

Visual Perception

Natural Language

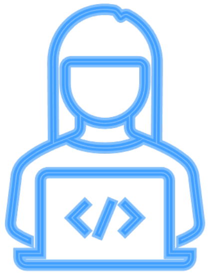
Passive Offline Decisions

Intelligent Robotics

Embodied

Generalizable Autonomy: Duality of Discovery & Bias

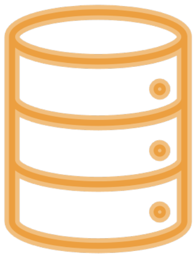
Domain Expertise

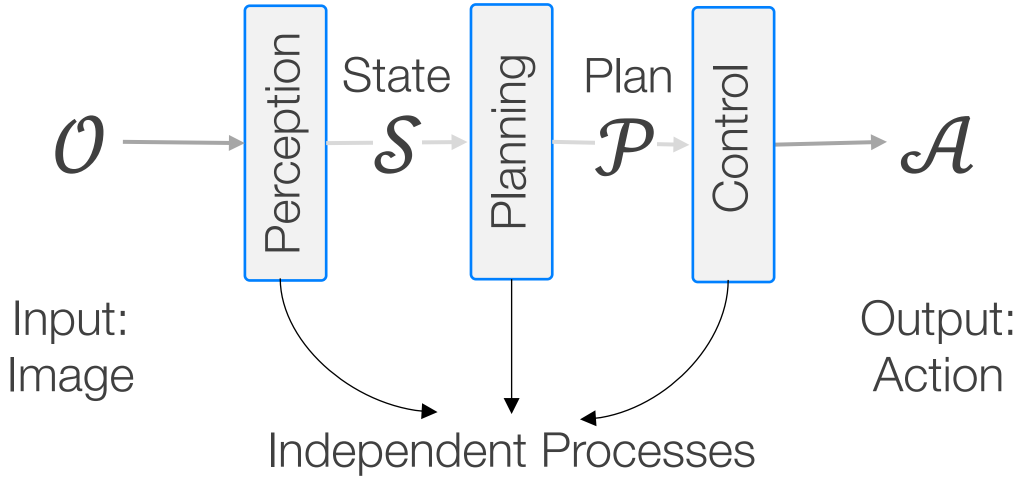


One problem,
One solution!

Structured Environments

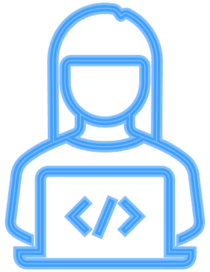
Data Driven





Generalizable Autonomy: Duality of Discovery & Bias

Domain
Expertise

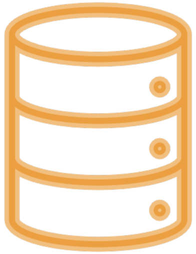


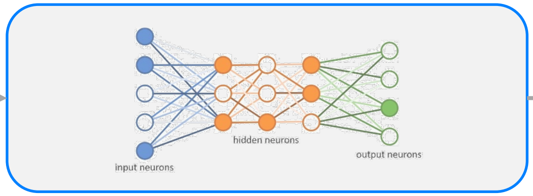
Just add
data…


The Unreasonable
Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

Data
Driven



$\mathcal{O}$ →  → $\mathcal{A}$

Input:
Image

End-to-End Policy
π

Output:
Action

Generalizable Autonomy: Duality of Discovery & Bias

Domain Expertise

One problem,
One solution!

- ✗ Need for experts
- ✗ Limited applicability
- ✗ Perf vs Flexibility

Just add data…

- ✗ Computational sustainability
- ✗ Data accessibility
- ✗ Out-of-distribution errors

Data Driven

Neither achieves
Generality at Scale

Generalizable Autonomy: Duality of Discovery & Bias

Domain Expertise

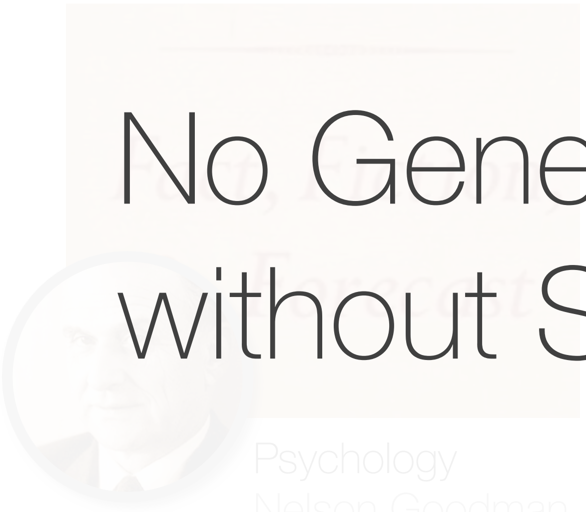
Data Driven

…make the inductive leap necessary to classify instances beyond observed…

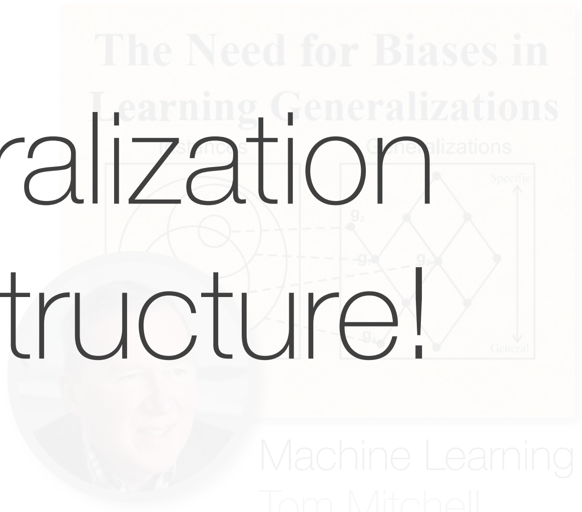
…other sources of information, or biases for choosing one generalization over the other…

No Generalization without Structure!

Philosophy
David Hume
1739



Psychology
Nelson Goodman
1955



Machine Learning
Tom Mitchell
1980

Deep Learning
Bengio, Hinton, LeCun 2020s

Generalizable Autonomy: Duality of Discovery & Bias

Domain Expertise

Data Driven

Generalizable Autonomy

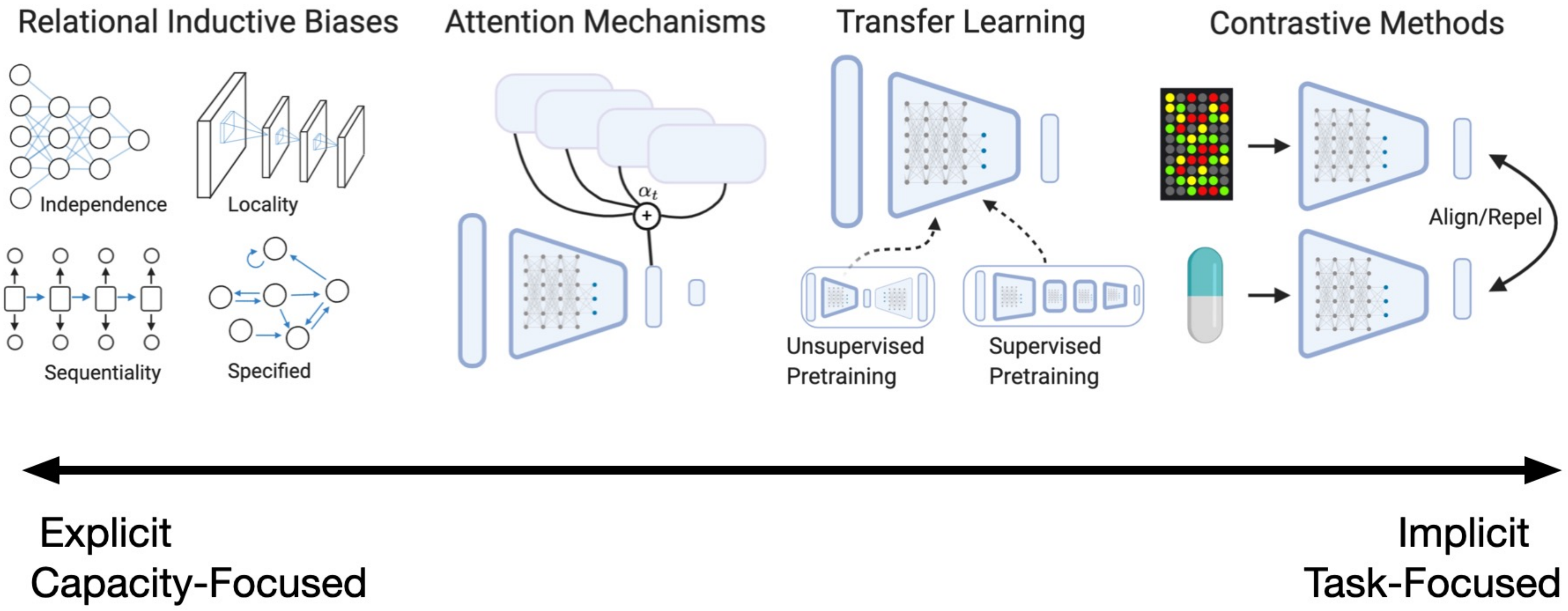
Structure + Data

- Domain knowledge,
- Inductive bias,
- Symmetries,
- Priors
- …

- Online & Offline,
- Simulation & Real,
- Labelled & self-supervised
- Human in the loop
- …

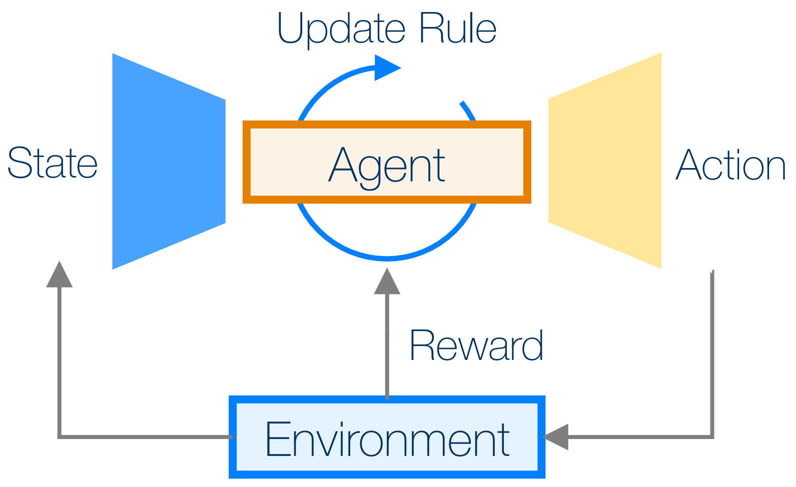
Structured Representations: Vision & Language

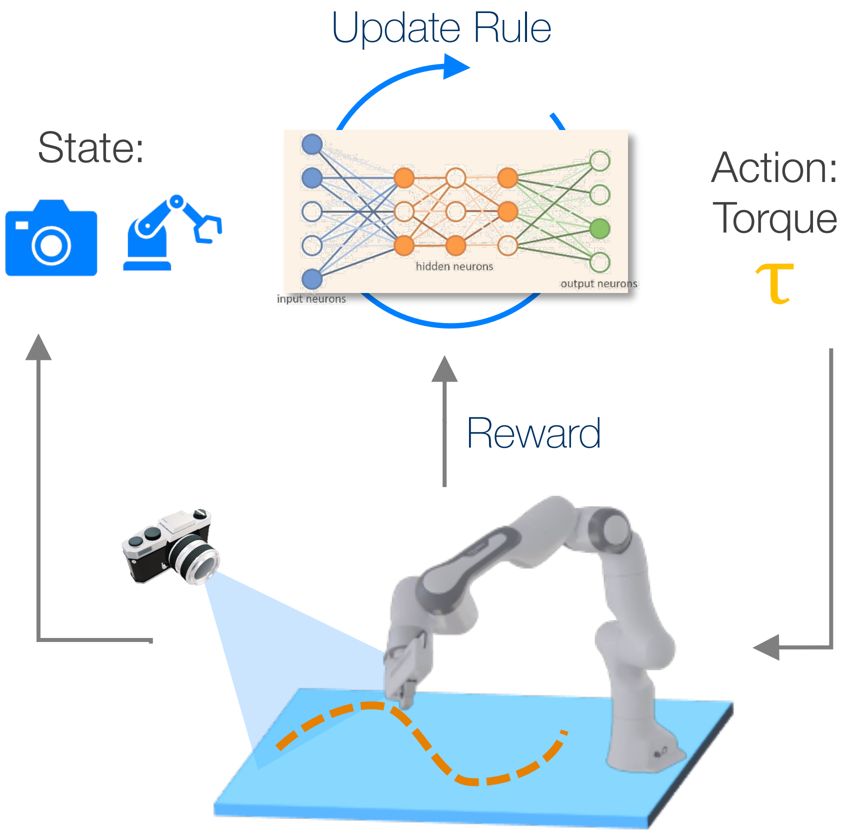
Insight: Structure makes learning possible



Sam Finlayson, 2020

Structure in Reinforcement Learning





Markov Decision Process

$$\mathcal{M} = \langle S, A, P(\cdot,\cdot), R(\cdot,\cdot), H \rangle$$

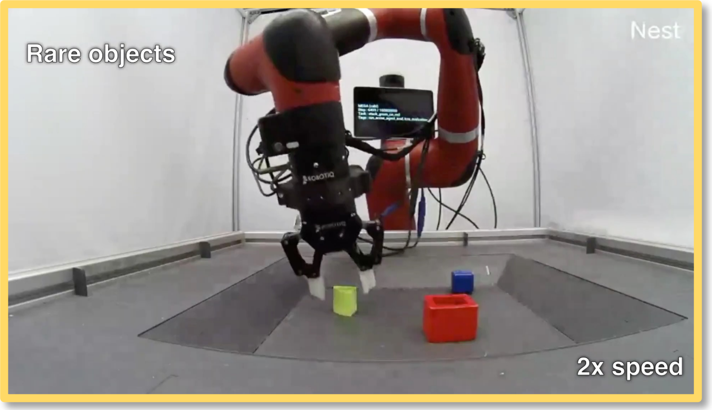
State Space · Action Space · Transition Function · Reward Function · Time Horizon

Goal: Find Optimal Policy

$$\pi^*: S \to A$$

Structure in Skill Learning: or the lack of it

Slow and Narrow:
- Specific tasks (Grasp/Stack)
- Often Supervised and rigid!

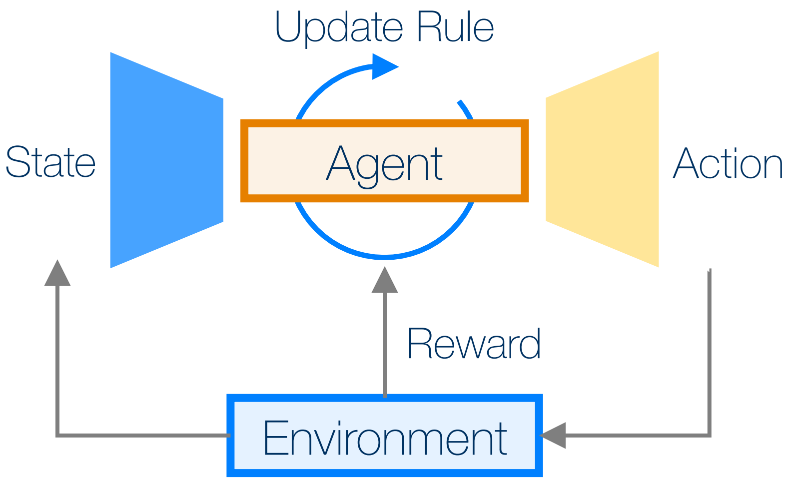


Learning Fast: Elephants Learning to use trunk

Learning Broad: Human infant learns to interact

[Kalashnikov et al. (2018), Cabi et al. (2019)]

Structure for Reinforcement Learning



Markov Decision Process

$$\mathcal{M} = \langle S, A, P(\cdot,\cdot), R(\cdot,\cdot), H \rangle$$

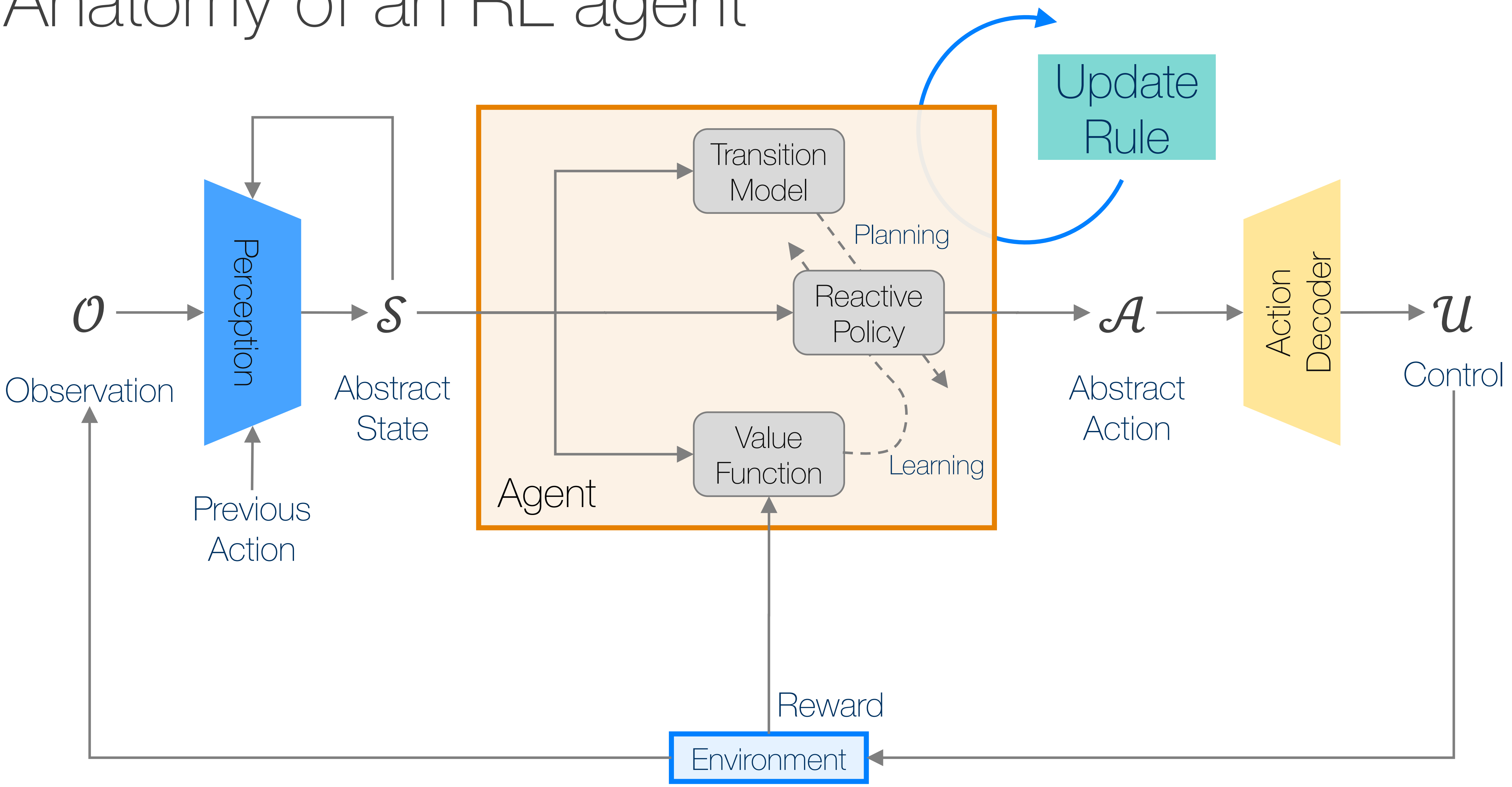
State Space | Action Space | Transition Function | Reward Function | Time Horizon

Goal: Find Optimal Policy

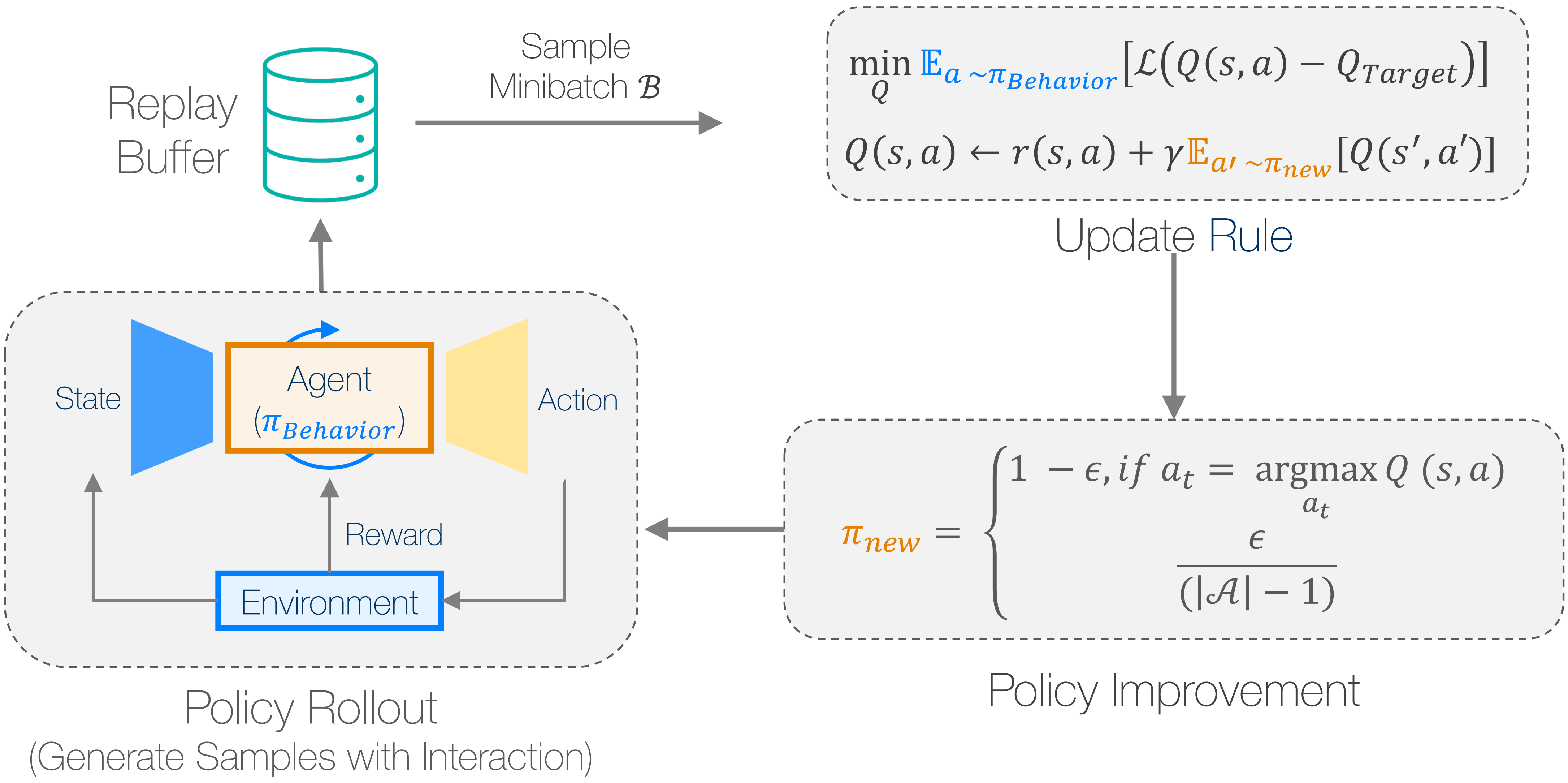
$$\pi^*: S \rightarrow A$$

Which structured biases enable generalizable autonomy in decision-making?

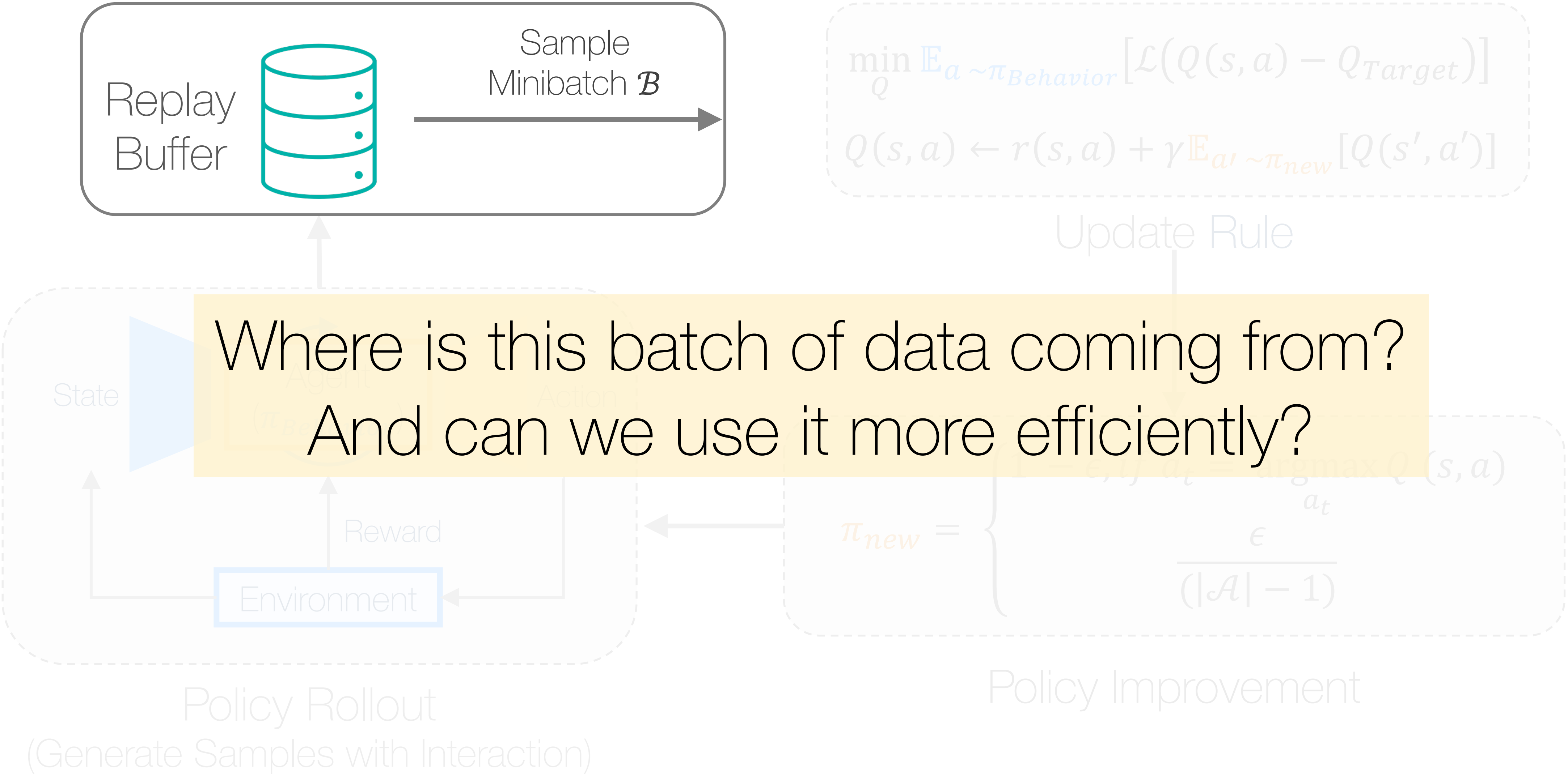
Anatomy of an RL agent



Structure for RL: Off-policy RL

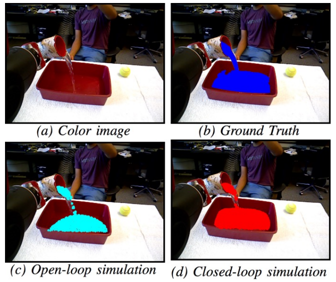


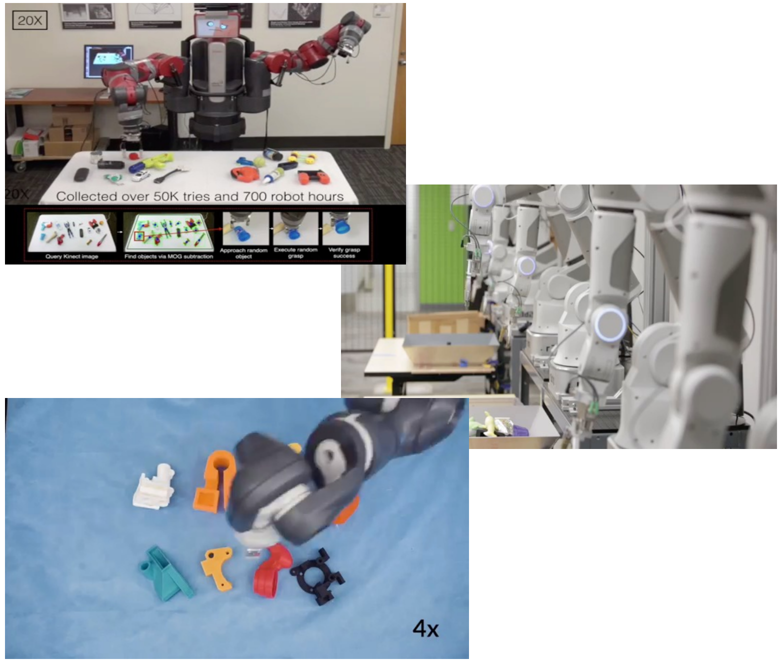
Structure for RL: Off-policy RL

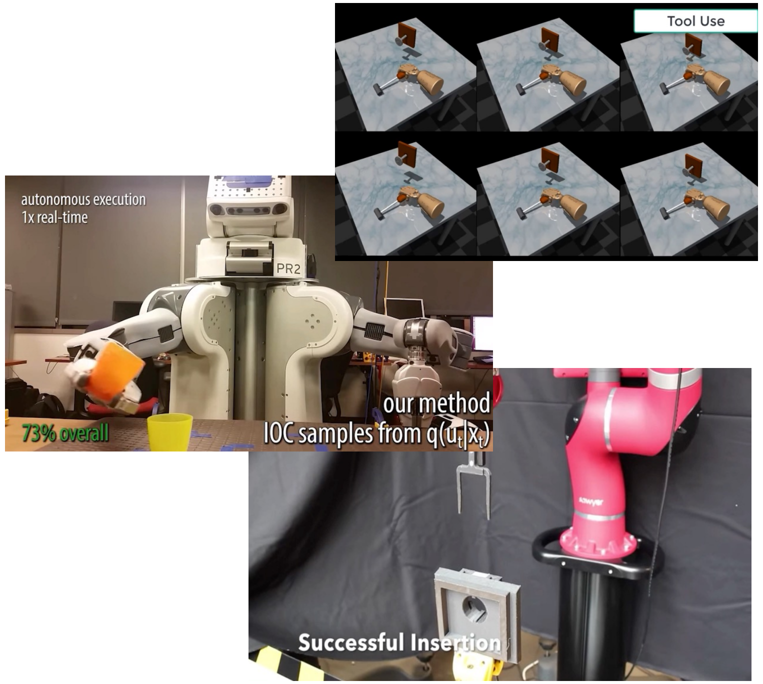


Where is this batch of data coming from?
And can we use it more efficiently?

Data in Robotics







Manipulation

Mason & Salisbury 1985
Srinivasa et al 2010
Berenson 2013
Odhner1 et al 2014
Chavan-Dafle et al 2014
Yamaguchi, et. al, 2015
…

Li , Allen et al. 2015
Yahya et al, 2016
Schenck et al. 2017
Mar et al. 2017
Laskey et al 2017
Quispe et al 2018
…

Grasping

Mishra et al 1987
Ferrari & Canny, 1992
Ciocarlie & Allen, 2009
Dogar & Srinivasa, 2011
Rodriguez et al. 2012
Bohg et al 2014

Pinto & Gupta, 2016
Levine et al 2016
Mahler et al 2017
Jang et al 2017
Viereck et al 2017
…

Imitation

Abbeel et al, 2004
Ratliff et al 2006,
Ziebart et al, 2009
Argall et al, 2009,
Boularias et al., 2011
Montfort et al 2015,
Wulfmeier et al 2015,

Krishnan et al 2017
Finn et al. 2017
Vecerik et al. 2017
Rajeswaran et al 2018
Zhu et al 2018
Ravichandar et al 2020…

Data in Robotics

Robot Data is Expensive!

✗ Short-Horizon skills
✗ Skill Specific learning

✗ Platform dependent data
✗ Scaling to other skills ?

✗ Small datasets (minutes)
✗ Low diversity ?

Manipulation

Mason & Salisbury 1985
Srinivasa et al 2010
Berenson 2013
Odhner1 et al 2014
Chavan-Dafle et al 2014
Yamaguchi, et. al, 2015
…

Li , Allen et al. 2015
Yahya et al, 2016
Schenck et al. 2017
Mar et al. 2017
Laskey et al 2017
Quispe et al 2018
…

Grasping

Mishra et al 1987
Ferrari & Canny, 1992
Ciocarlie & Allen, 2009
Dogar & Srinivasa, 2011
Rodriguez et al. 2012
Bohg et al 2014

Pinto & Gupta, 2016
Levine et al 2016
Mahler et al 2017
Jang et al 2017
Viereck et al 2017
…

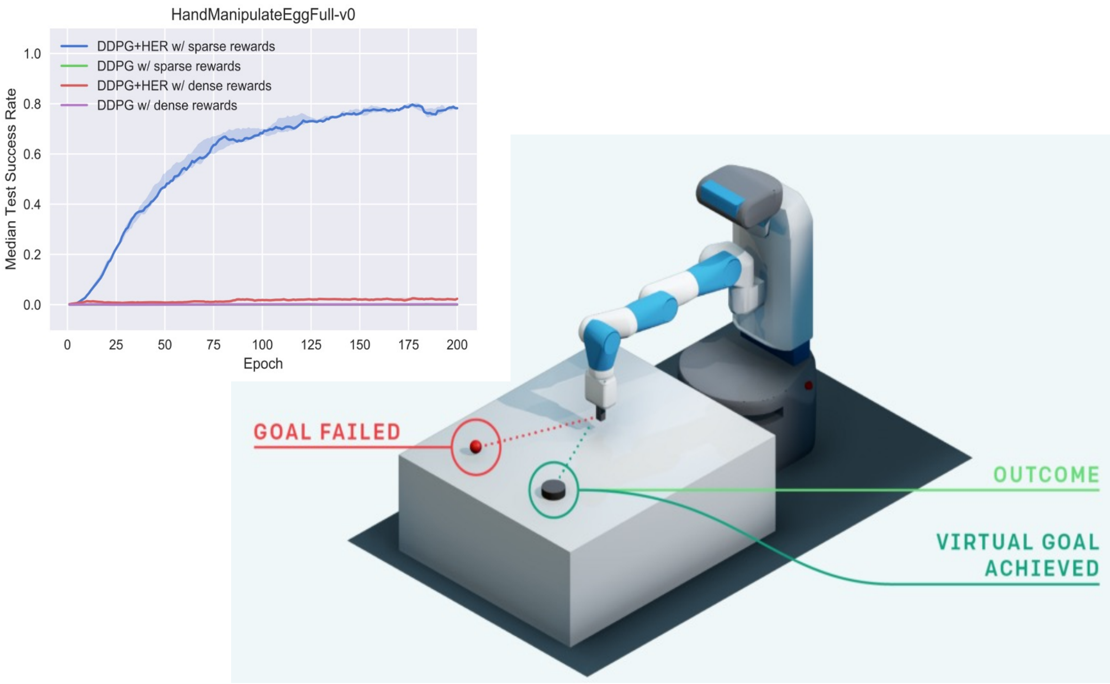
Imitation

Abbeel et al, 2004
Ratliff et al 2006,
Ziebart et al, 2009
Argall et al, 2009,
Boularias et al., 2011
Montfort et al 2015,
Wulfmeier et al 2015,

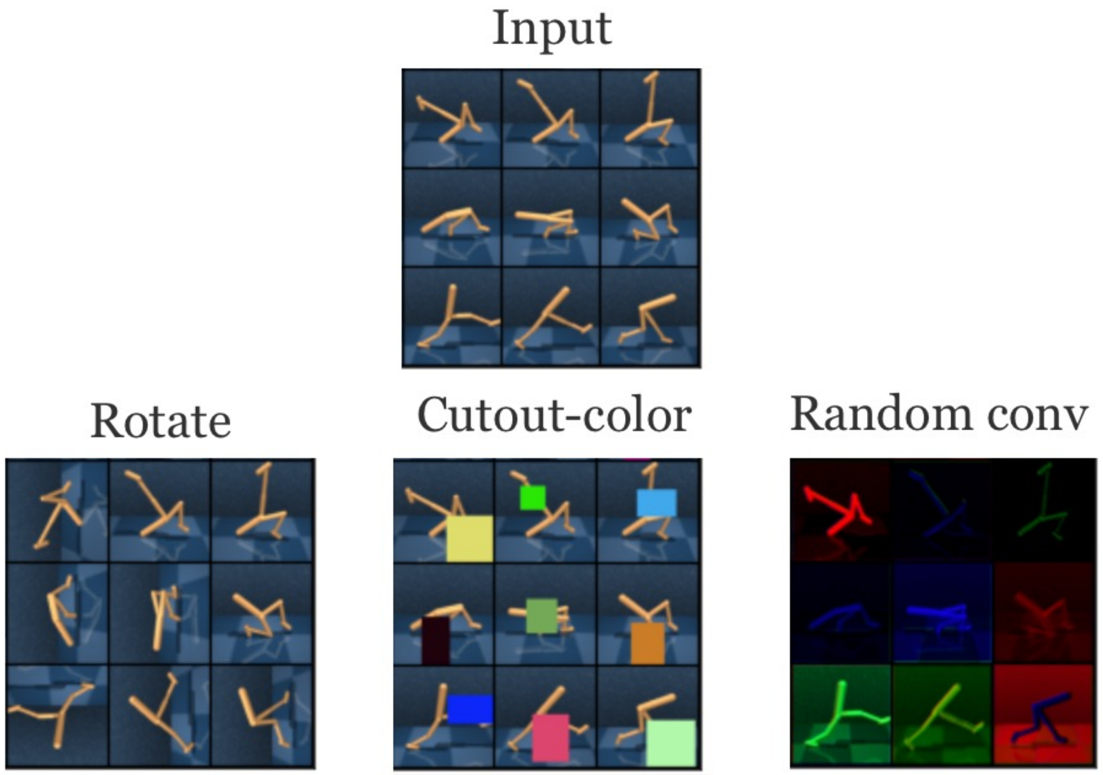
Krishnan et al 2017
Finn et al. 2017
Vecerik et al. 2017
Rajeswaran et al 2018
Zhu et al 2018
Ravichandar et al 2020…

Data Augmentation in RL
How to do this Algorithmically

- Substantial performance boosts!



Goal relabeling (e.g, HER])



Visual Input relabeling (e.g, RAD])

RAD: Laskin et al 2020, DrQ: Kostrikov et al 2020, CURL: Laskin et al. 2020, S4RL: Sinha et al. 2021, HER: Andrychowicz et al 2018

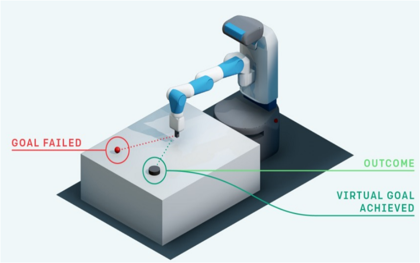
Data Augmentation in RL

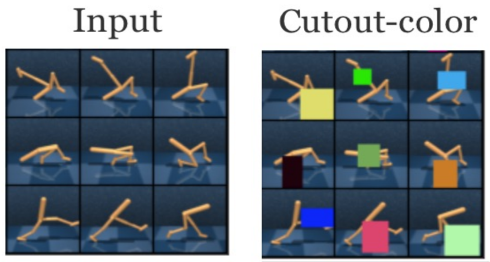
Unified View

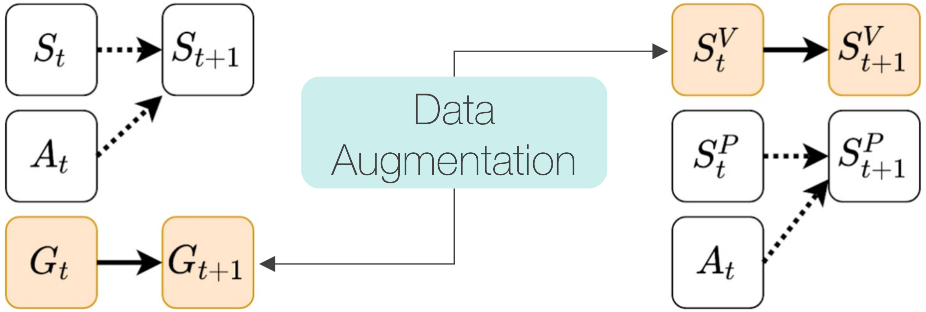
Counterfactual reasoning to generate new, causally valid (counterfactual) data!

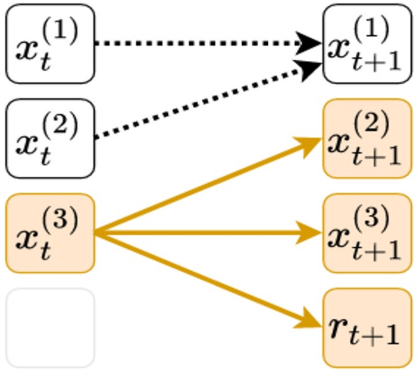
Insight

Exploit the independence of the causal mechanisms guiding transition









Goal is independent of State/Action Dynamics

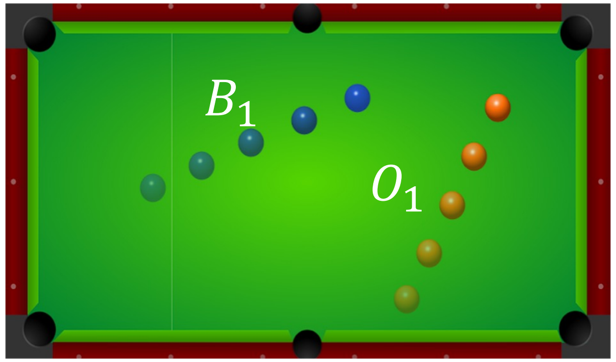
Visual characteristics (e.g., crop) are independent of physical dynamics

Given two independent mechanisms, Relabel one (conditional independence!)

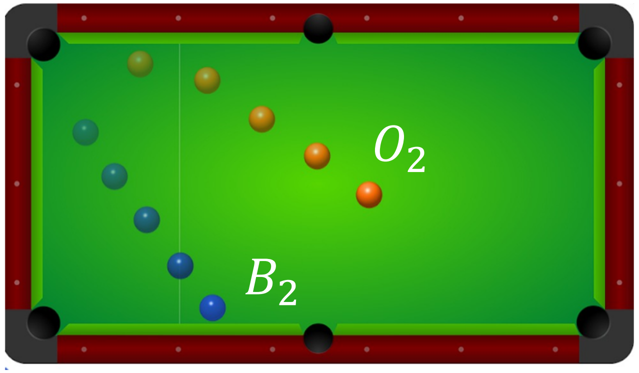
RAD: Laskin et al 2020, DrQ: Kostrikov et al 2020, CURL: Laskin et al. 2020, S4RL: Sinha et al. 2021, HER: Andrychowicz et al 2018

Data Augmentation in RL

Do more with the same data



Scenario 1



Scenario 2

Which of the following is possible (only based on observed data)

$O_1 + B_2$

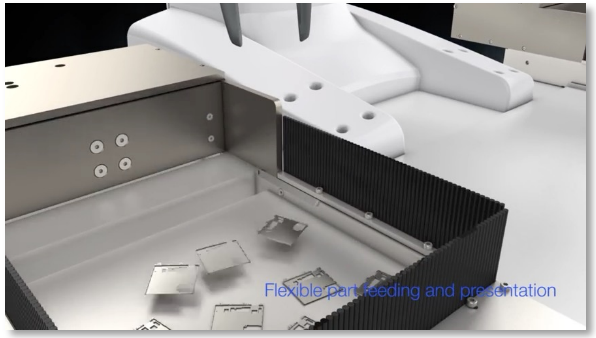
Independent
Compositional Generalization

$B_1 + O_2$

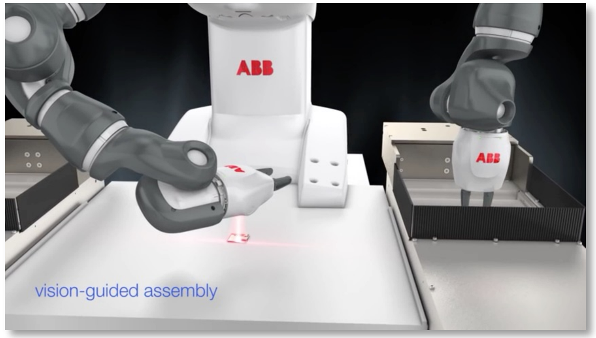
Not-Independent (!)
Hence need evidence of possibility

Data Augmentation in RL

Do more with the same data

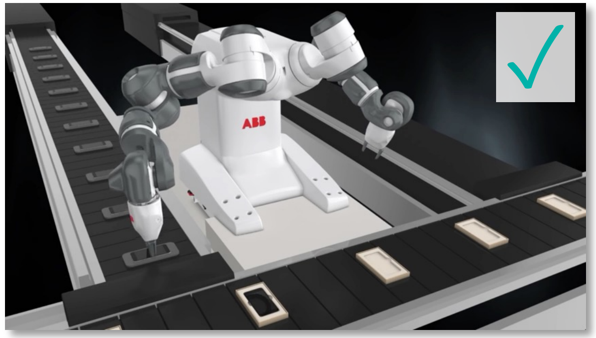


Left Arm Pick and Place

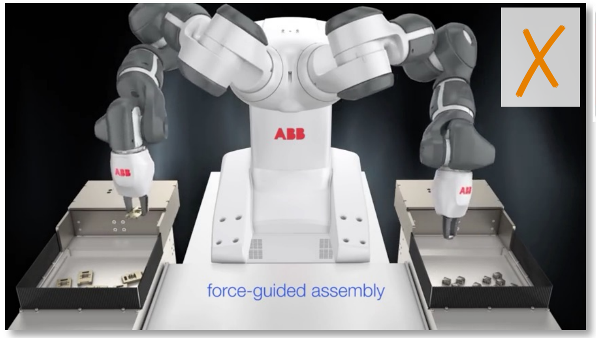


Right Arm Pick and Place

Which of the following is possible (only based on observed data)



Independent
Compositional Generalization

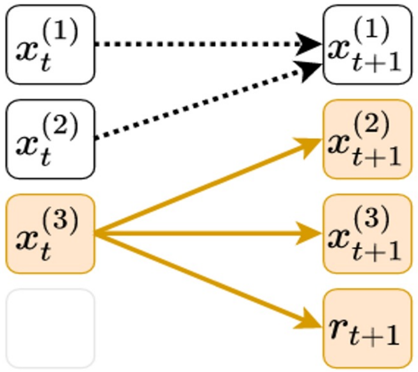


Not-Independent (!)
Hence need evidence of possibility

Counterfactual Data Augmentation

Counterfactual reasoning to generate new, causally valid (counterfactual) data!

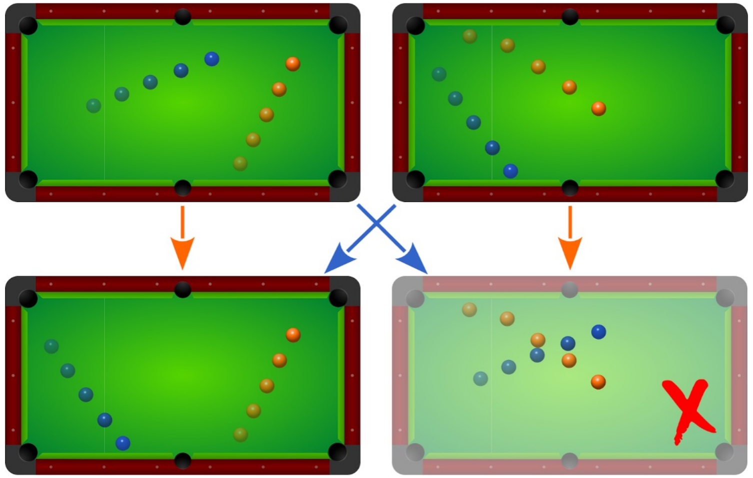
Generic CoDA



Given two independent mechanisms, Relabel one (conditional independence!)

✓ Model-Free relabelling

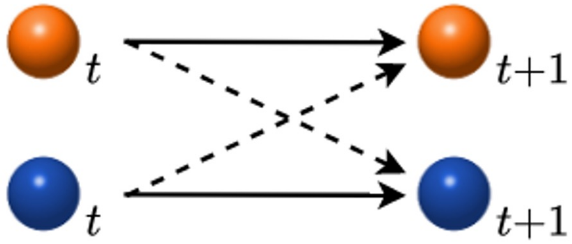
✗ But Causal Independence is not Global



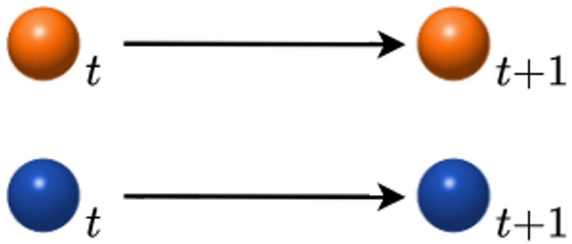
- For the most part, entitites behave independently, and we can use CoDA
- But entitites are not always independent, so this can also produce nonsense

Counterfactual Data Augmentation

Local Causal Model



Global Model



Local Model

$$\mathcal{M}_t = \langle V_t, U_t, \mathcal{F} \rangle \xrightarrow{\text{Condition on } (s_t, a_t) \in \mathcal{L}} \mathcal{M}_t^{\mathcal{L}} = \langle V_t^{\mathcal{L}}, U_t^{\mathcal{L}}, \mathcal{F}^{\mathcal{L}} \rangle$$

Structural Causal Model (SCM) that marginalizes across all possible transitions

Local Causal Model (LCM) that behaves like the global SCM in local subspace $\mathcal{L}$

Where do local models come from?

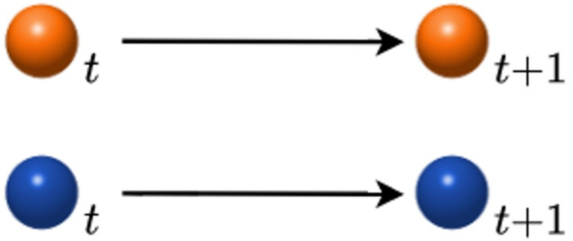
Counterfactual Data Augmentation

Learning Local Causal Model

- Input: 2 balls, each with 4 features: $[x, y, \dot{x}, \dot{y}]$

```
[[ 1.23, -0.73,  1.31,  1.07],
 [-0.6 ,  2.51, -1.51, -0.89]]
```

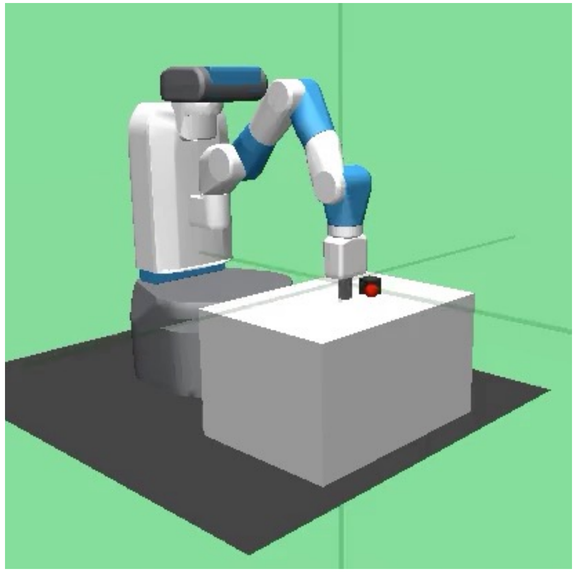
- Output: Adjacency matrix M of the causal graph (between x_t and x_{t+1})
- (intuition) M: the input-output Jacobian is non-zero

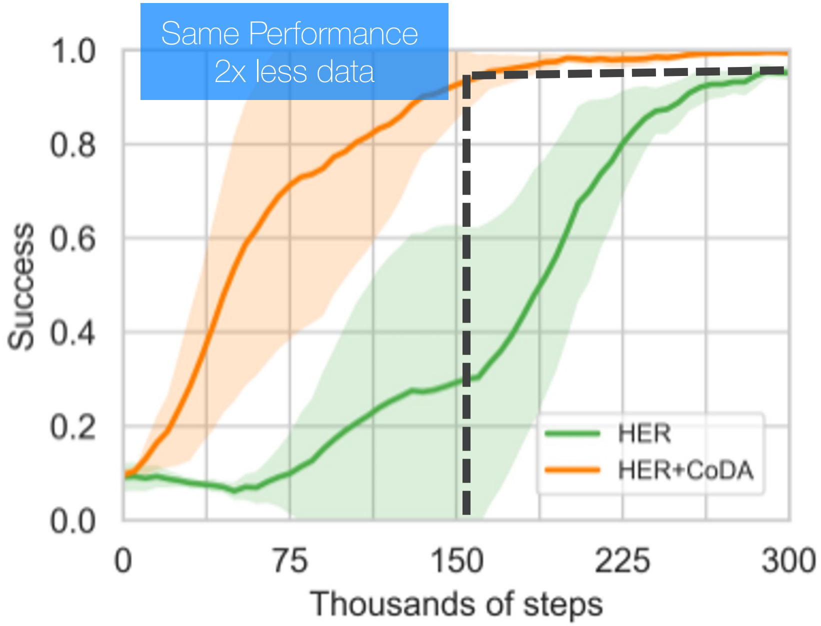


t+1 t+1

```
[[1., 1., 1., 1., 0., 0., 0., 0.],
 [1., 1., 1., 1., 0., 0., 0., 0.],
 [1., 1., 1., 1., 0., 0., 0., 0.],
 [1., 1., 1., 1., 0., 0., 0., 0.],
 [0., 0., 0., 0., 1., 1., 1., 1.],
 [0., 0., 0., 0., 1., 1., 1., 1.],
 [0., 0., 0., 0., 1., 1., 1., 1.],
 [0., 0., 0., 0., 1., 1., 1., 1.]]
```

CoDA: Goal-Conditioned (Online) RL



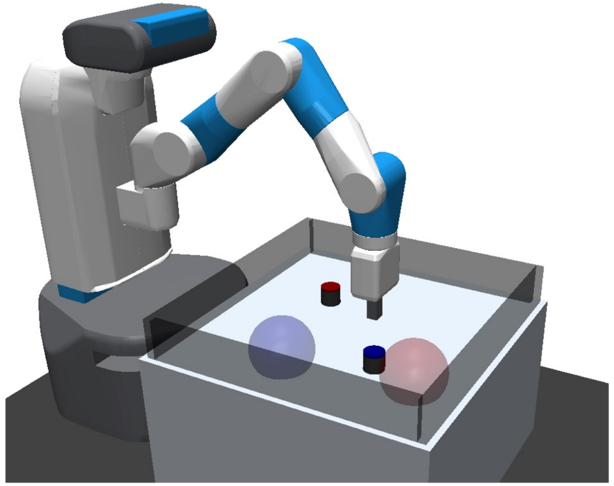


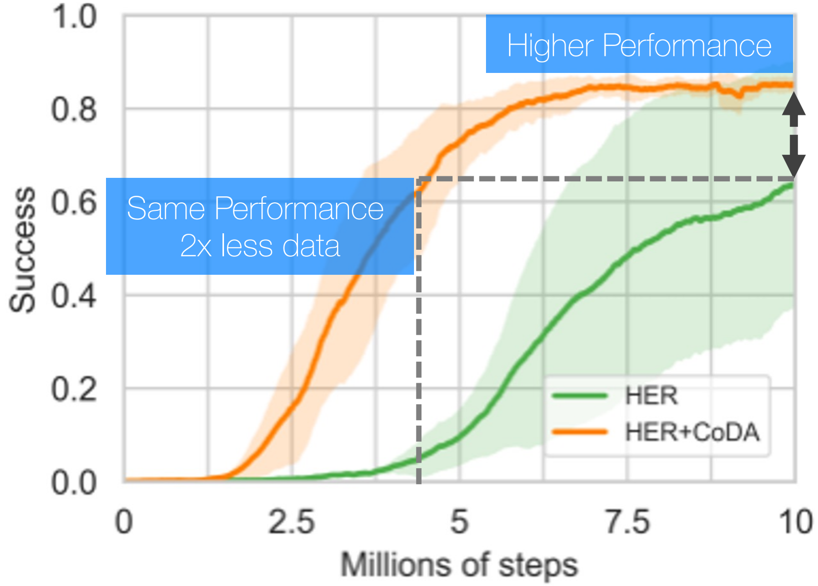
Fetch-Push-v1

state space: [Robot and **1 object**]

CoDA Heuristic for independence:

Obj ⫫ Gripper | if $\left|x_g - x_o\right| \geq 10\ cm$



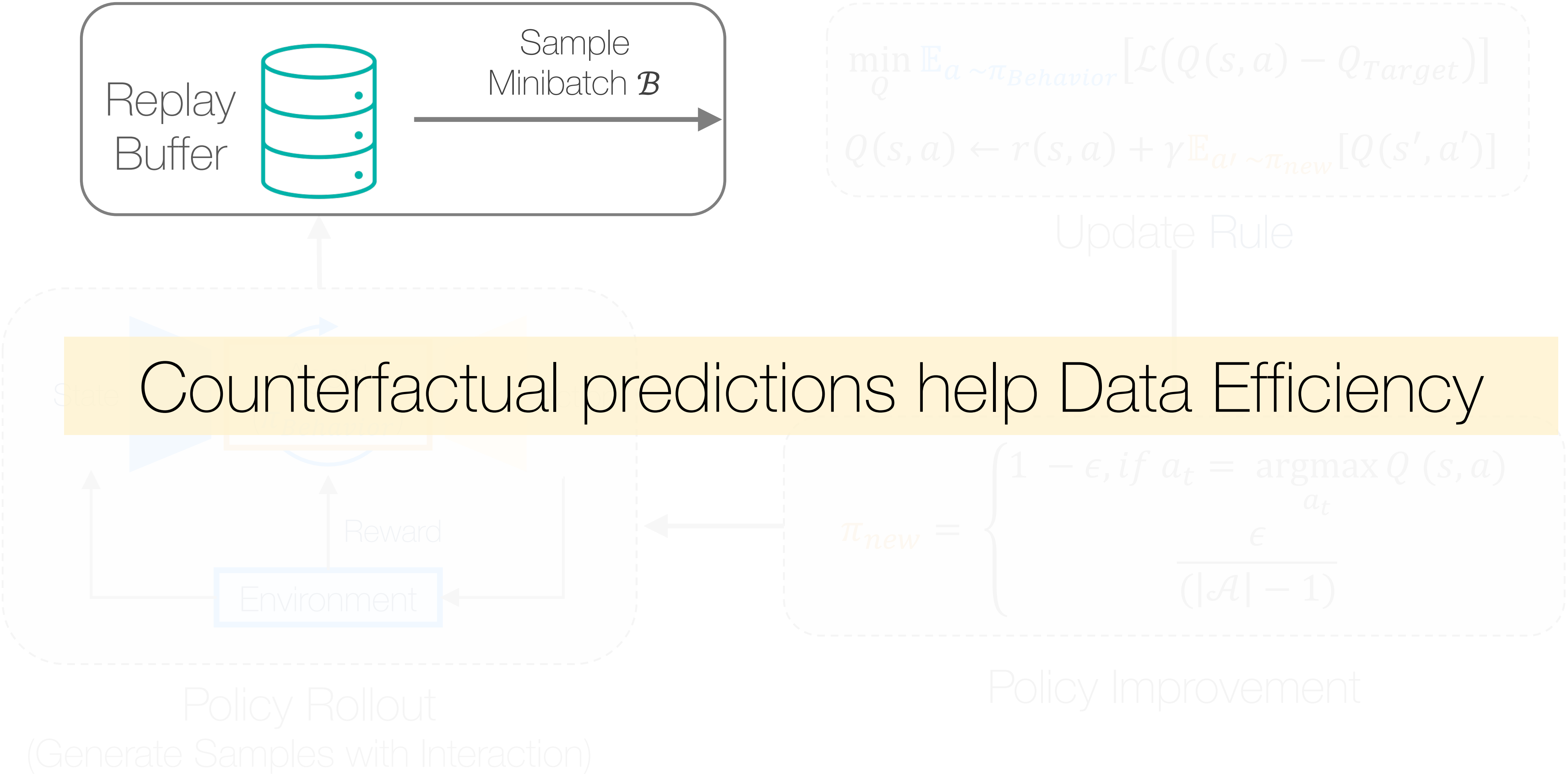


Fetch-Slide2

state space: [Robot and **2 objects**]

Harder task (30x more samples)!

Structure for RL: Off-policy RL



Counterfactual predictions help Data Efficiency

Structure for RL: Off-policy RL

Counterfactual Data Augmentation helps

How to learn this structure automatically?

Discover Causal Dynamics Structure from Visual Data

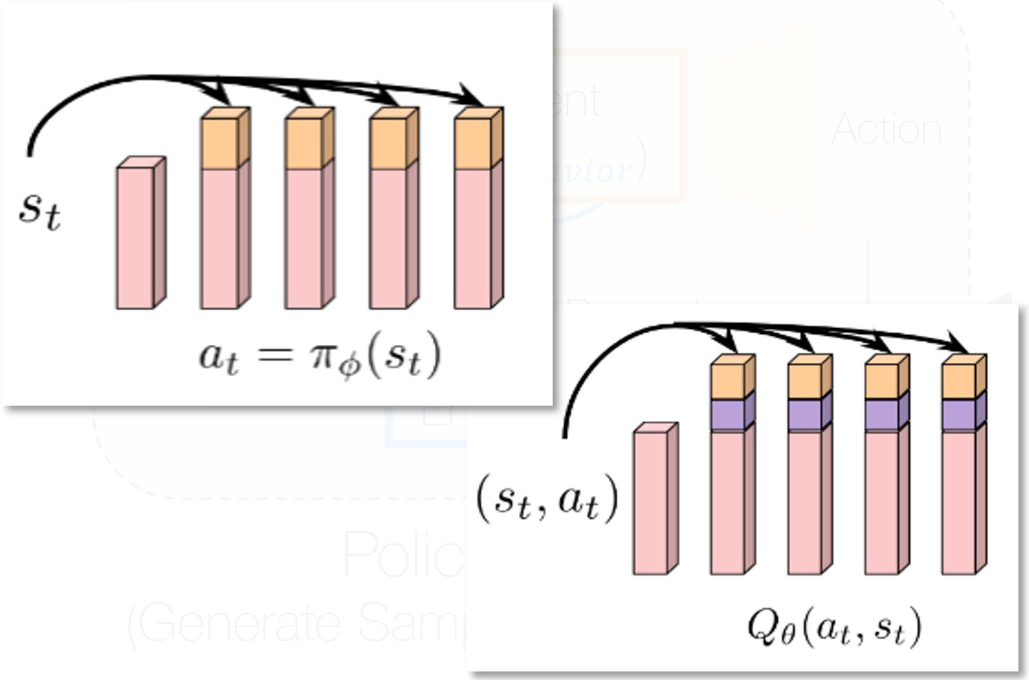
[V-CDN, Neurips 2020]

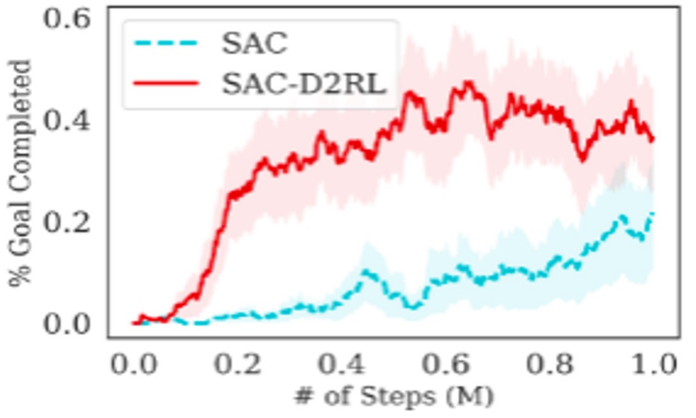
Structure for RL: Off-policy RL

Does the choice of architecture matter?

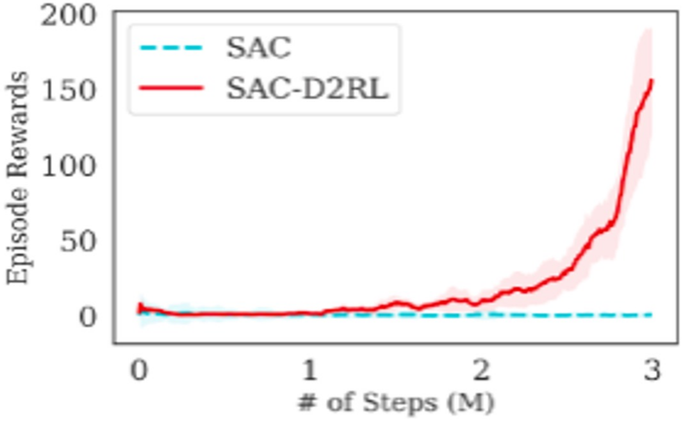
$$Q(s,a) \leftarrow r(s,a) + \gamma \mathbb{E}_{a' \sim \pi_{new}}[Q(s',a')]$$

Using Dense connections in Policy/Value improves sample efficiency





(b) Fetch Slide SAC



(d) Jaco Reach SAC

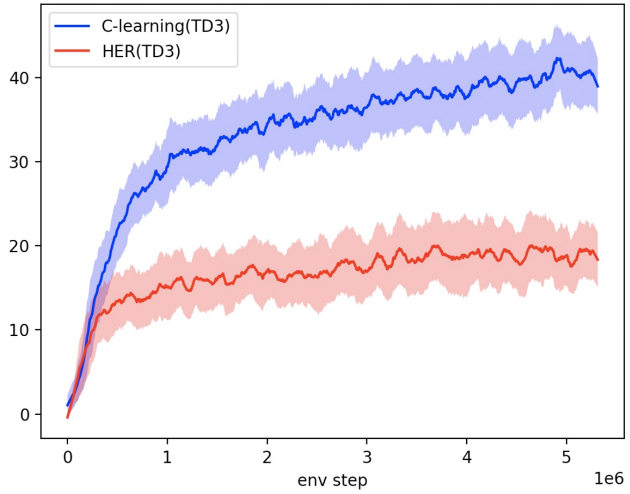
[D2RL, 2020], [C-Learning, ICLR 2021]

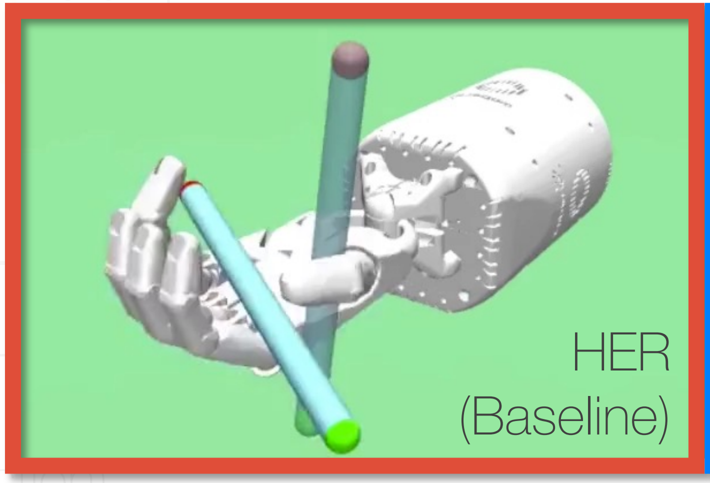
Structure for RL: Off-policy RL

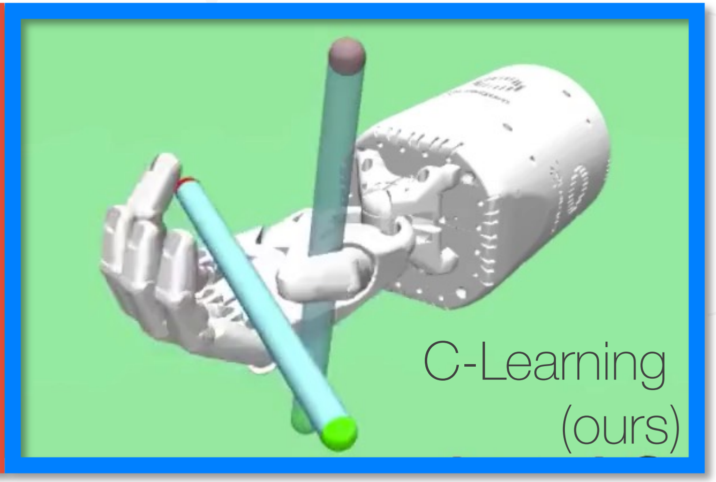
Can we use better Utility Functions?

$$Q(s,a) \leftarrow r(s,a) + \gamma \mathbb{E}_{a' \sim \pi_{new}}[Q(s',a')]$$

Learning Cumulative Accessibility $C(s,a,h)$ is better than $Q(s,a)$
Can represent multimodal, multi-goal, horizon-aware solutions as well as reachability

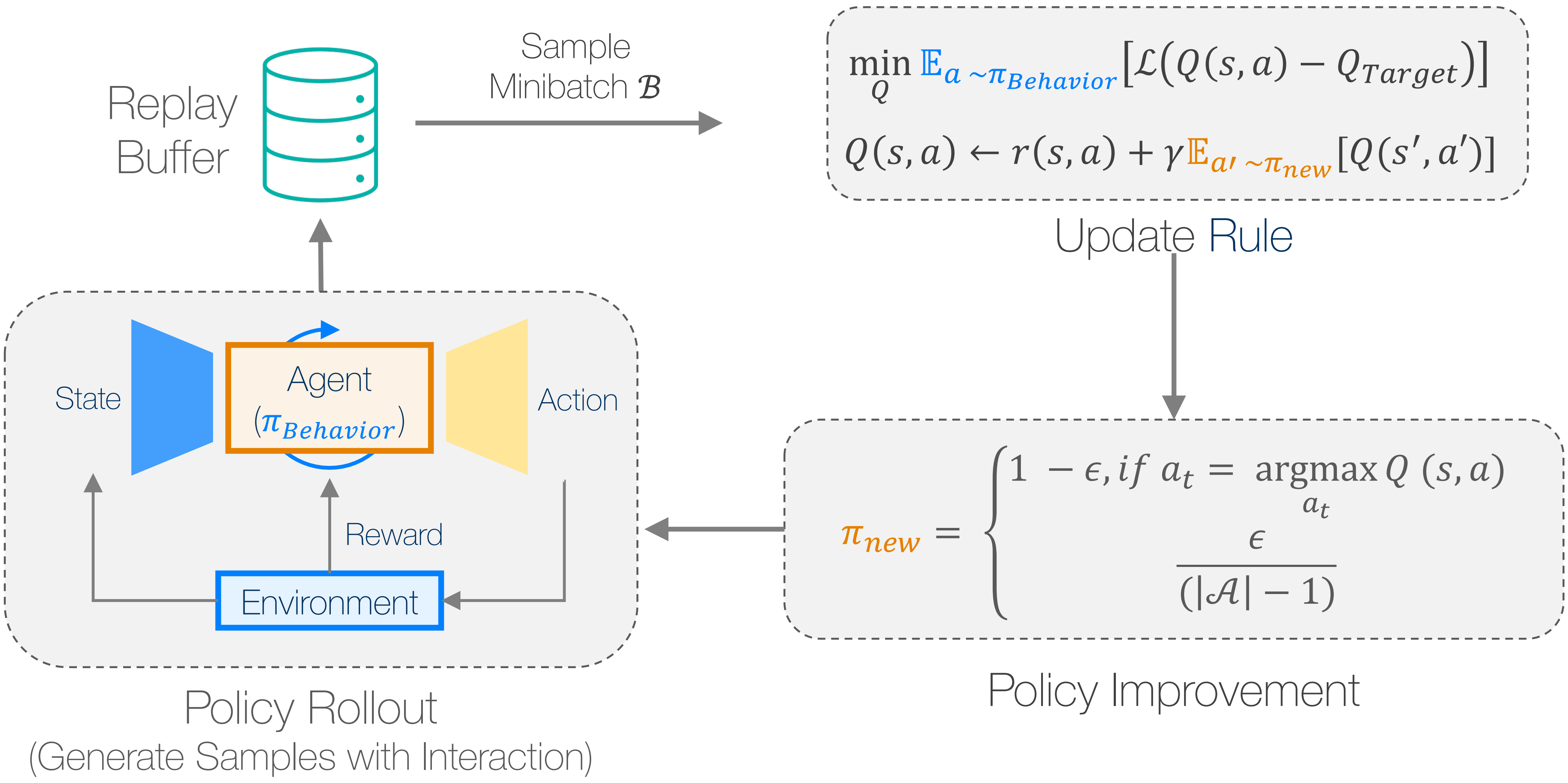




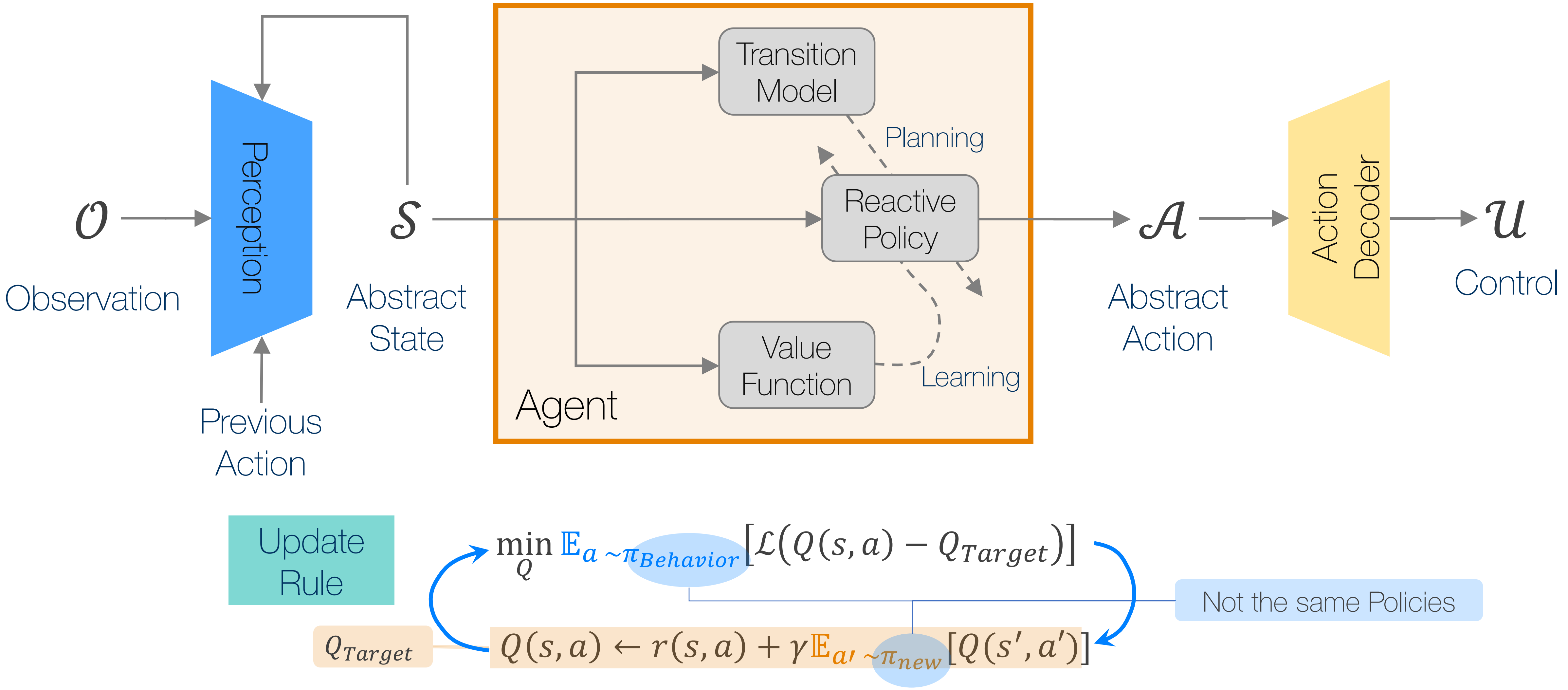


[D2RL, 2020], [C-Learning, ICLR 2021]

Structure for RL: Off-policy RL

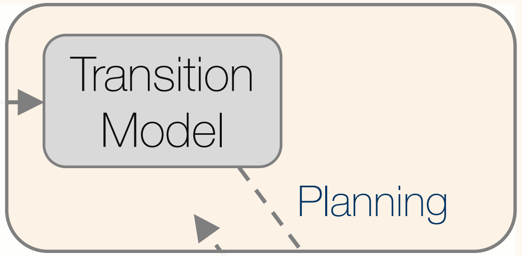


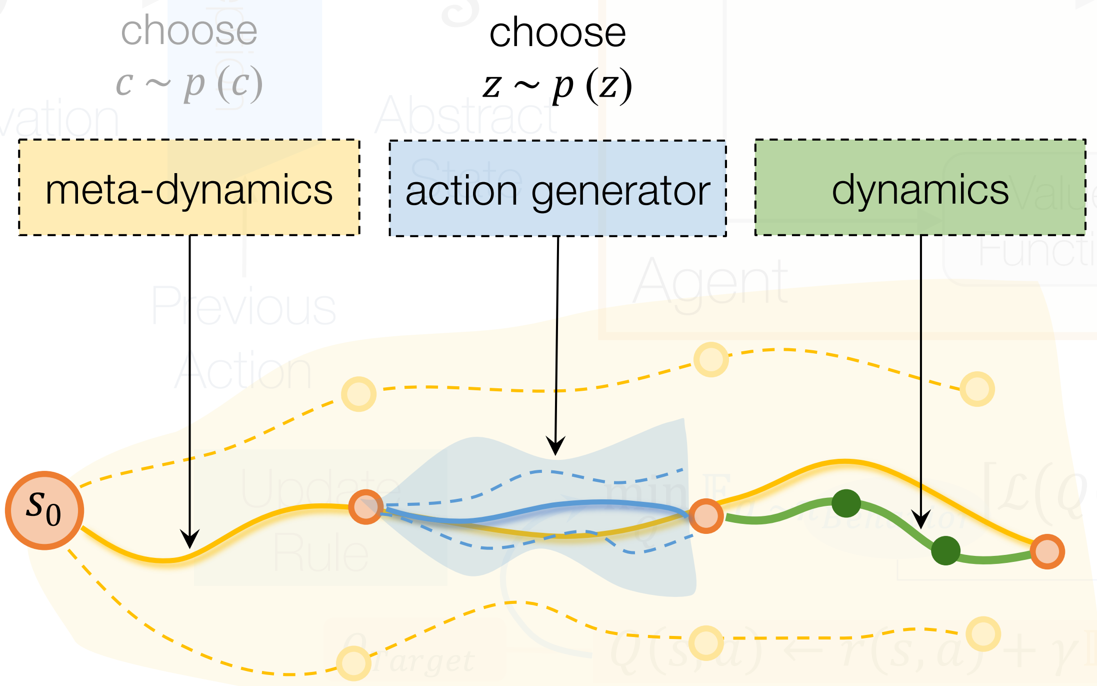
Structure for Reinforcement Learning

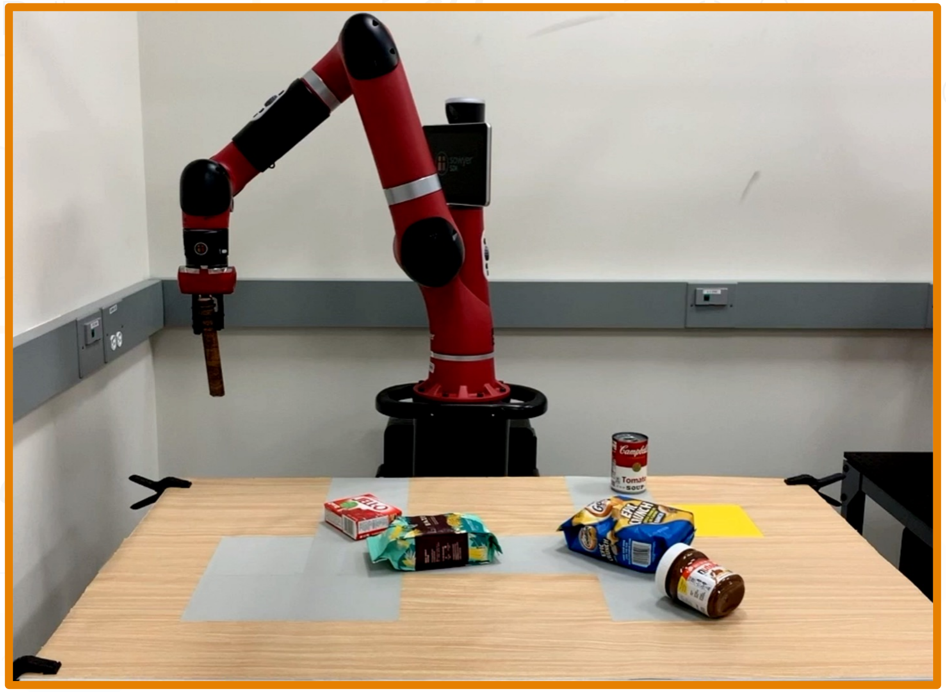


Structure for Reinforcement Learning

Structures Models for hierarchy?

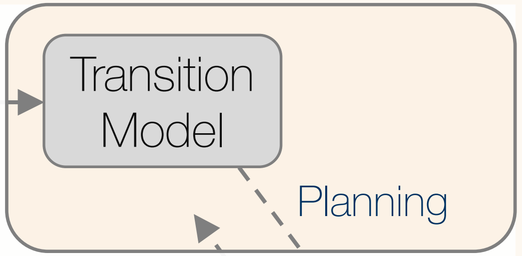




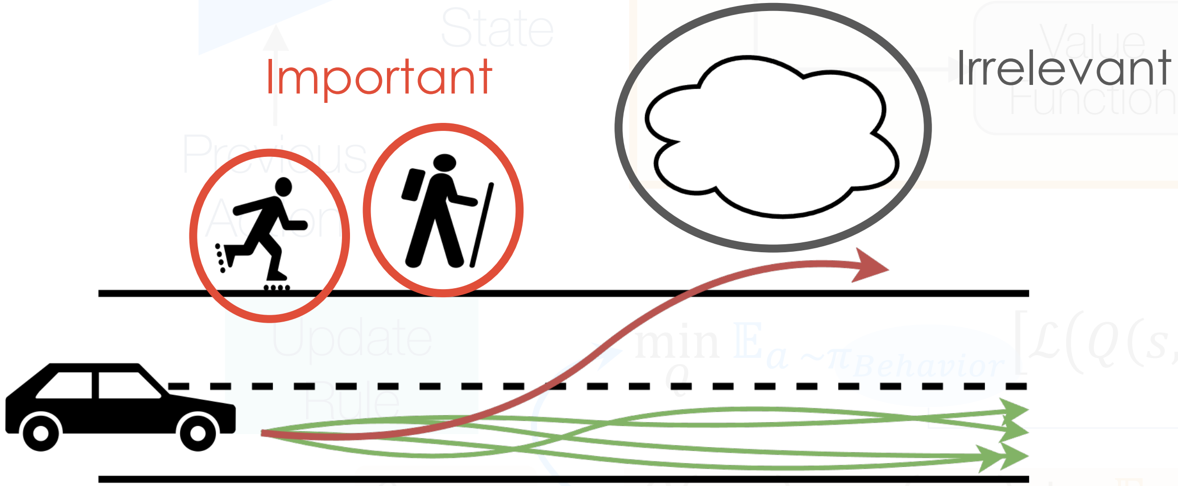


Structure for Reinforcement Learning

Dynamics Prediction,
Correct objective?

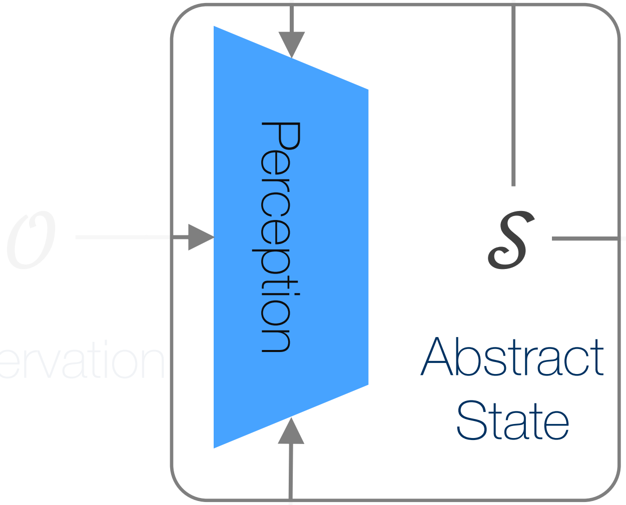


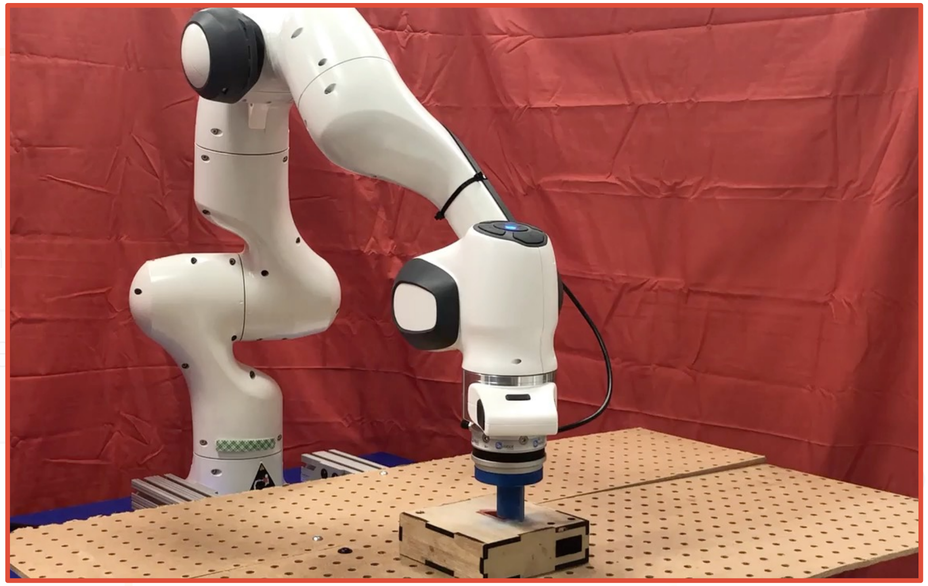
Task-Aware Objective helps efficient and targeted Model-Learning

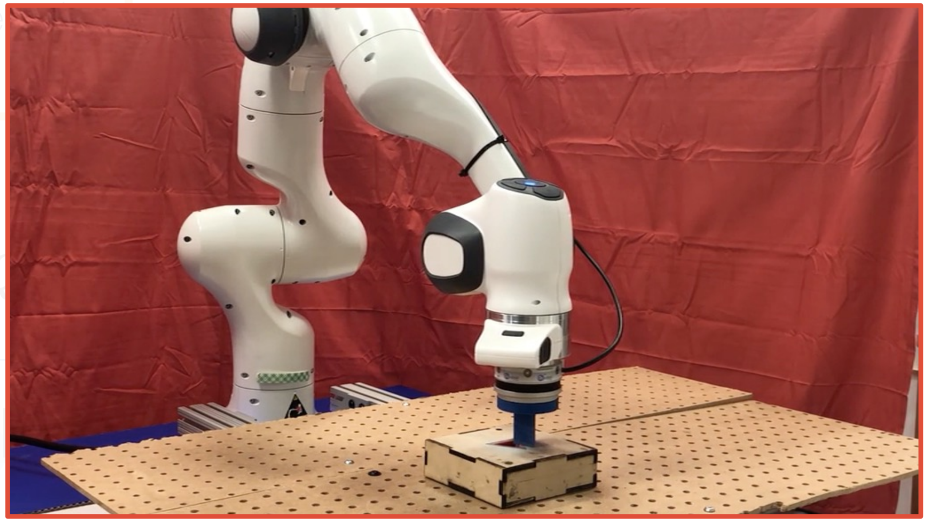


Structure for Reinforcement Learning

How can better state representations capture multimodal data?





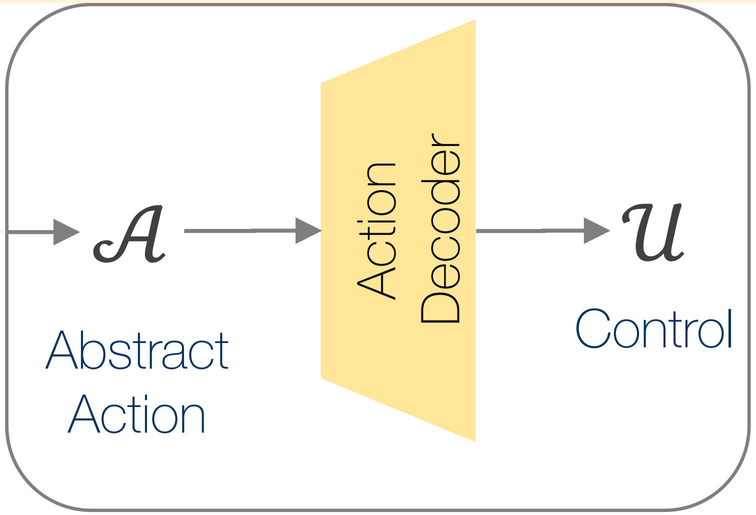


Generalizable Multi-modal State Representations

- Learn a joint Visuo-Tactile representation for Peg Transfer
- Representation transfers to new task, while Policy doesn't

Structure for Reinforcement Learning

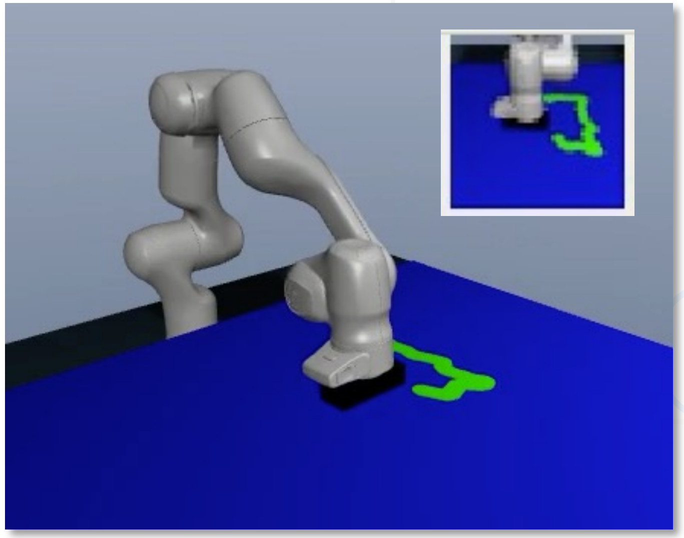
How can better action representations result in generalization?



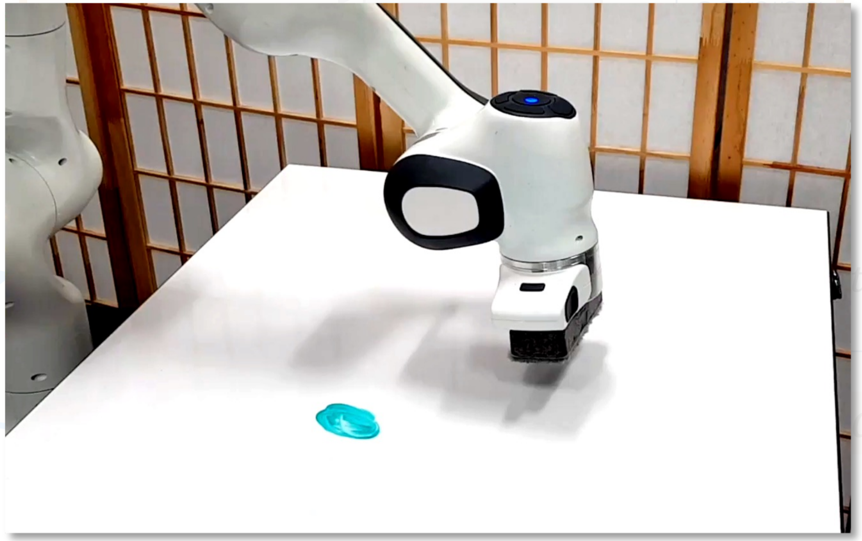
Variable Impedance Controller in End-Effector Space

- Efficiency in RL
- Ease of Sim2Real Transfer

$\tau = f_{Sim}(\pi(o_t))$



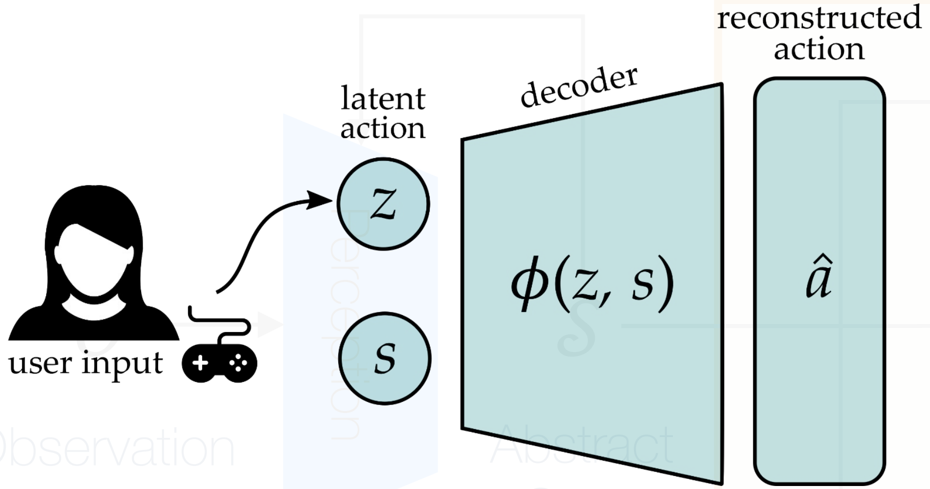
$\tau = f_{Real}(\pi(o_t))$

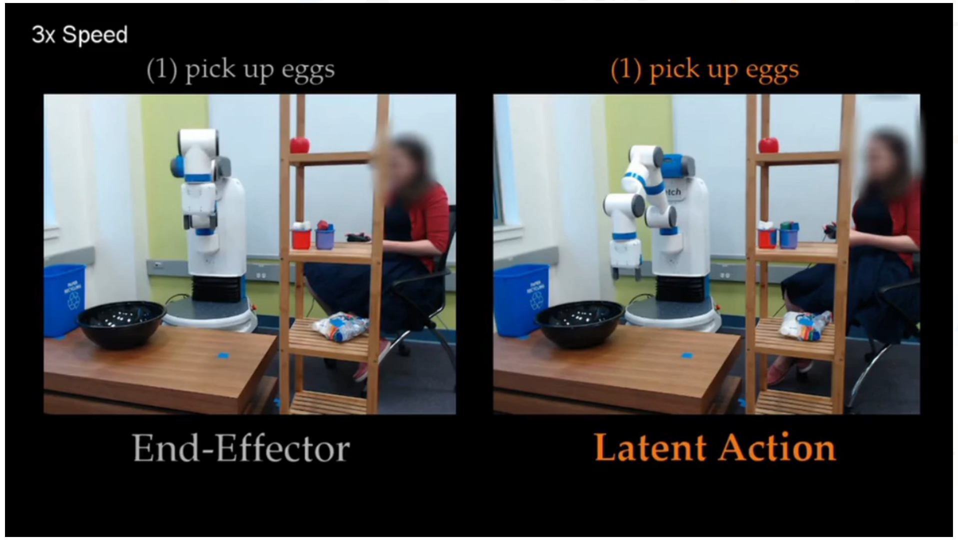


$\pi(o_t) = a$: $[\underbrace{\boxed{x_d, \dot{x}_d}}_{\text{Pose and Velocity}}, \underbrace{\boxed{K_p, K_v}}_{\text{Impedance Gains}}]$

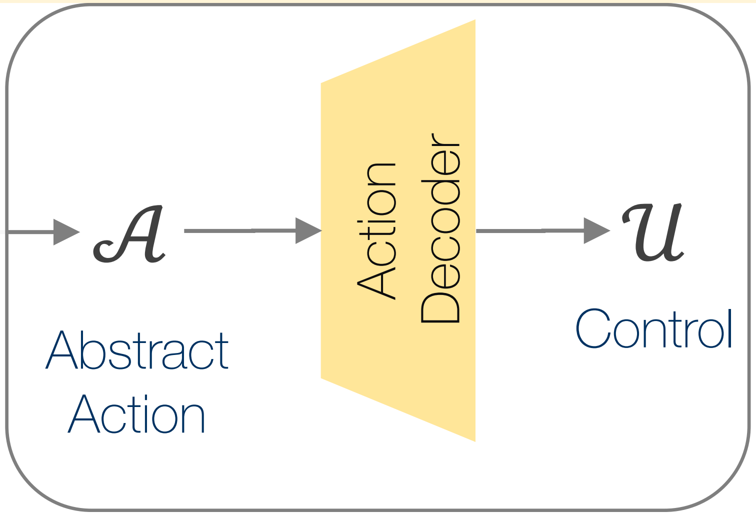
$\tau = \underbrace{\boxed{f}}_{\text{Deterministic Position-Velocity Control Jacobian } J \text{ and Inertia } M}(x_d, \dot{x}_d, K_p, K_v)$

Structure for Reinforcement Learning





Learned Action Representations for Shared Autonomy

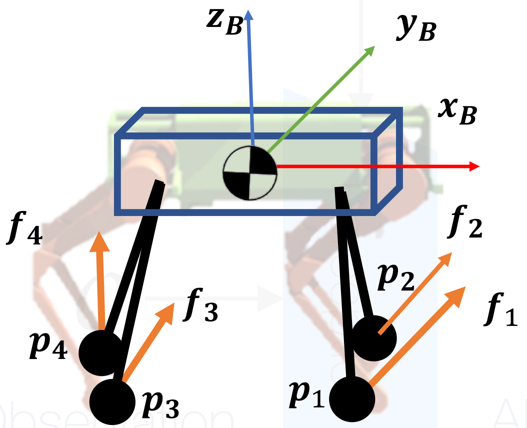


Learned Action Space

Easier to control high-dimensional robots by embedding the robot's actions into a low-dimensional latent space

Controlling Assistive Robots with Learned Latent Actions, ICRA 2020, AURO 2021

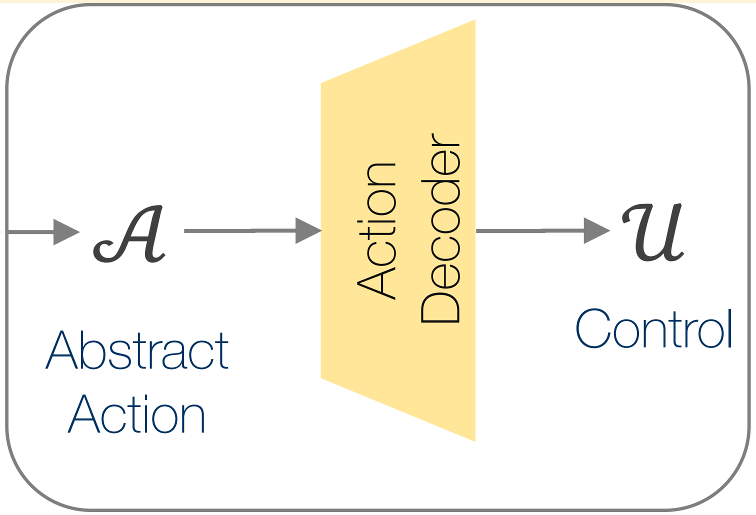
Structure for Reinforcement Learning



Centroidal Task Space

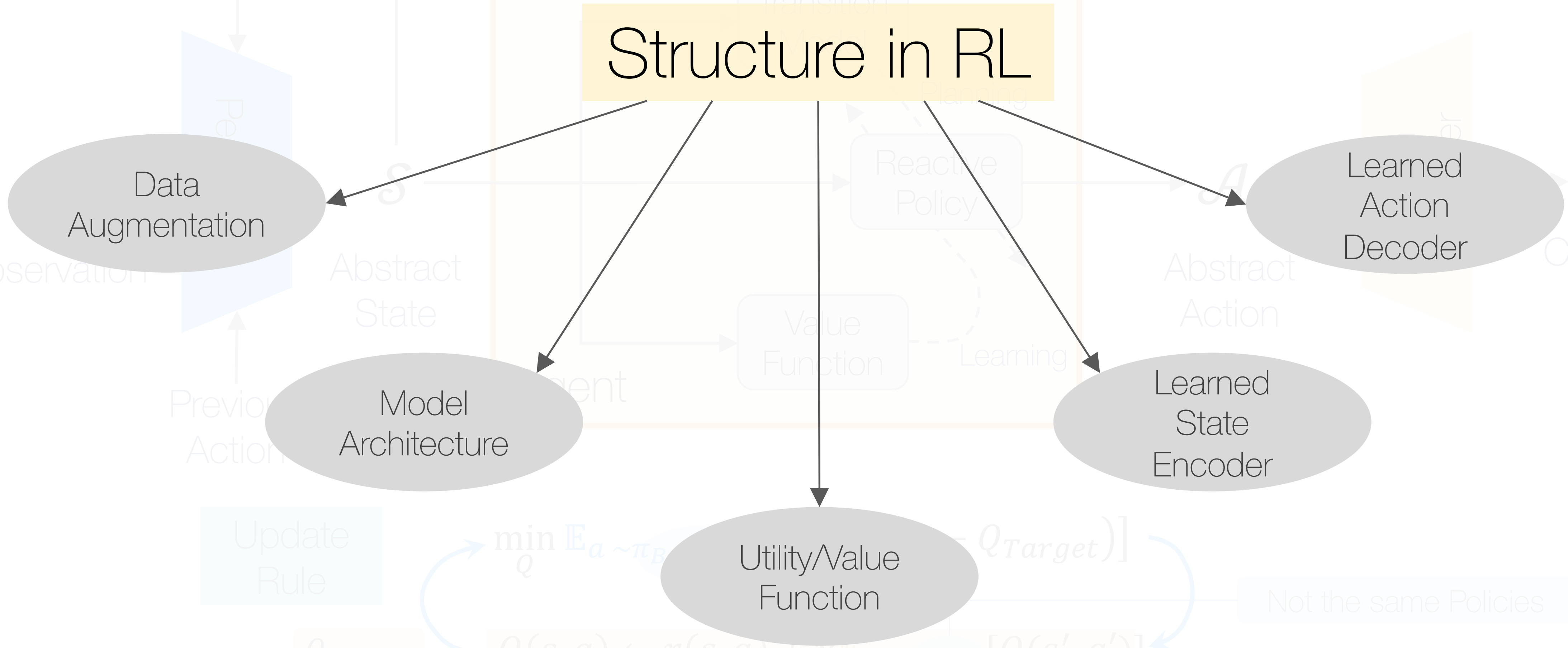
- Easier for learning
- RL + Optimal Control

Action representations for Legged Locomotion?

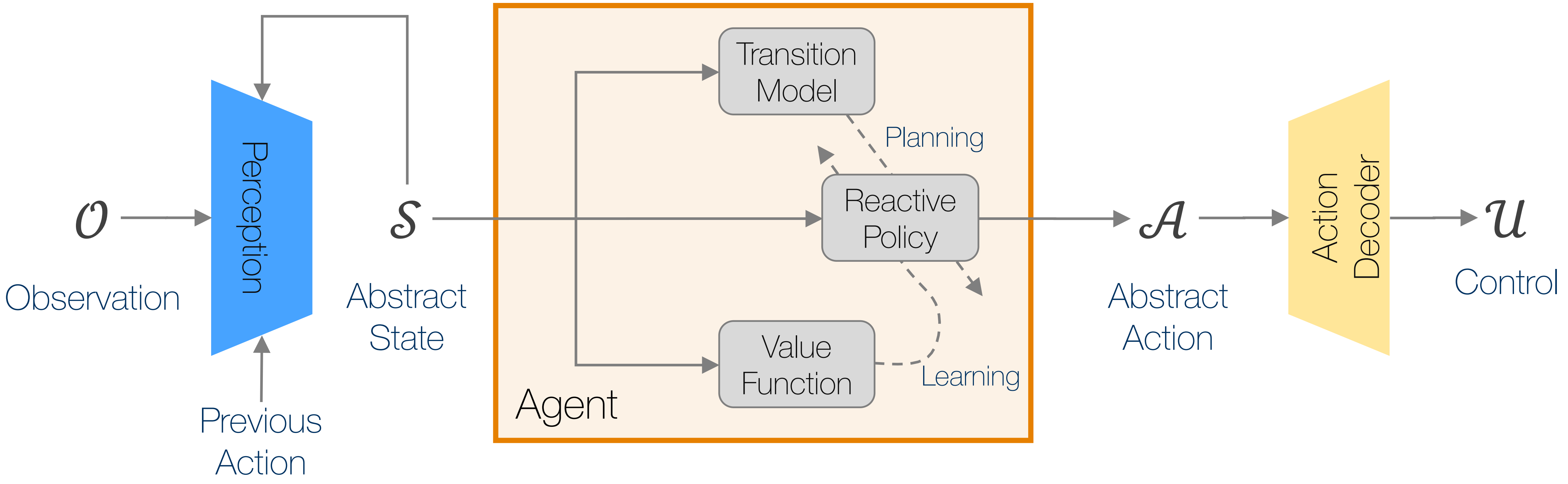


GLiDE: Generalizable Quadrupedal Locomotion, under review 2022

Structure for Reinforcement Learning



Structure for Reinforcement Learning



Update Rule

$$\min_Q \mathbb{E}_{a\sim\pi_{Behavior}}\left[\mathcal{L}\left(Q(s,a)-Q_{Target}\right)\right]$$

Q_{Target}

$$Q(s,a) \leftarrow r(s,a)+\gamma\,\mathbb{E}_{a'\sim\pi_{new}}[Q(s',a')]$$

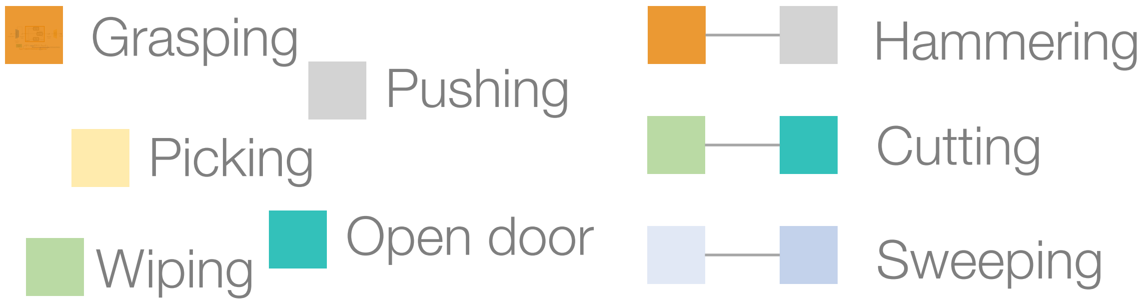
Not the same Policies

Structure in Compositional Planning

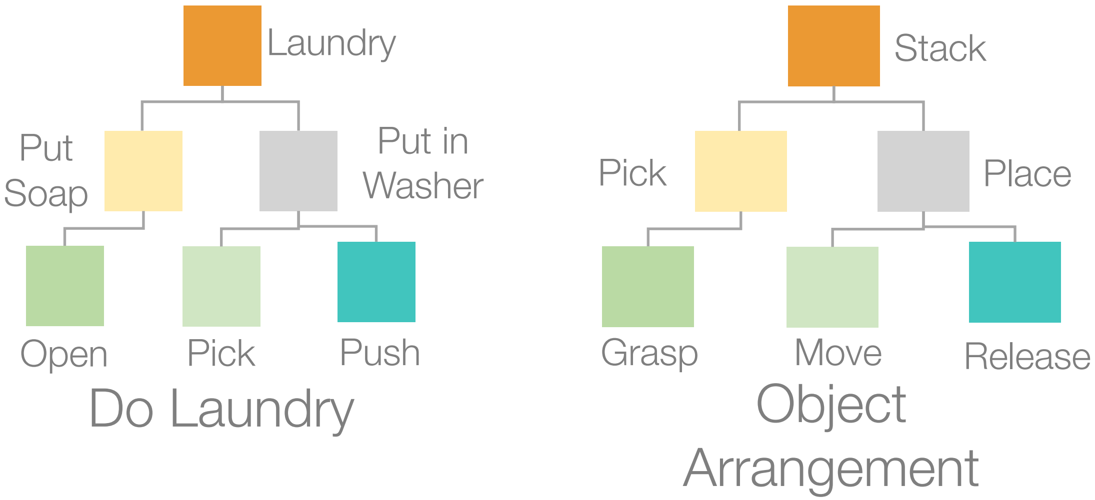
Visuo-Motor Skills

Compositional Planning

Visuo-Motor Skills



Compositional Planning



Structure in Compositional Planning

Imitation: But at which level? What should I copy?

Imitative Babbling	**Dexterity**	**Causality**	Task Specification
Movement Skills	Skill Sequencing	Semantic Purpose	{Language, Video, Kinesthetic}

Cheng et al. Sci.Rob. 2019

Structure in Compositional Planning

State s_0 → Feed-Forward Model → Latent State s or z

Object-Centric Generative Model of Action Concepts

Reference or Implicit Reward

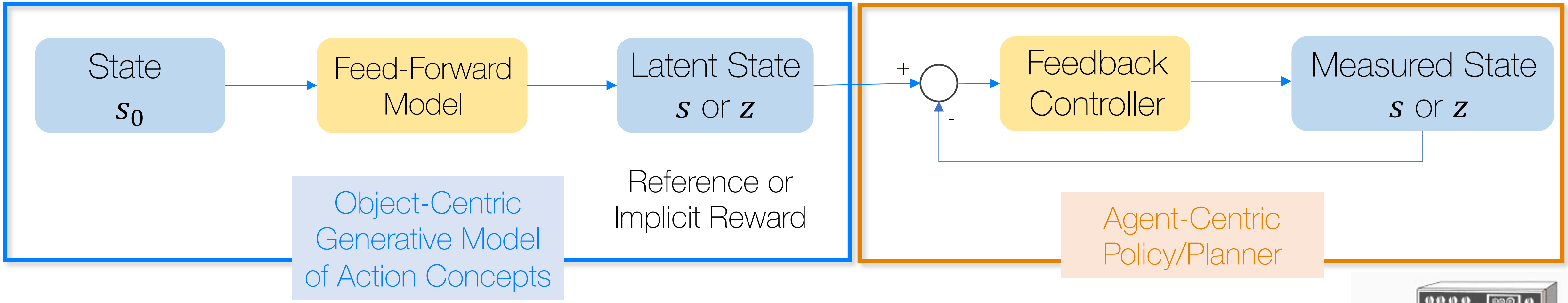
Causal Generative Model

- Learn to predict the "effect" of "action"
- Compositional & Counterfactual
- Multi-step Semantic consistency
- Pre-trainable over large problem settings

What does it mean to "open" a "door"?
"open" a "jar"?
"open" < >

→ "Take" "Jug"
→ "Open" "Fridge"
→ "Put" "Jug" in "Fridge"

Structure in Compositional Planning

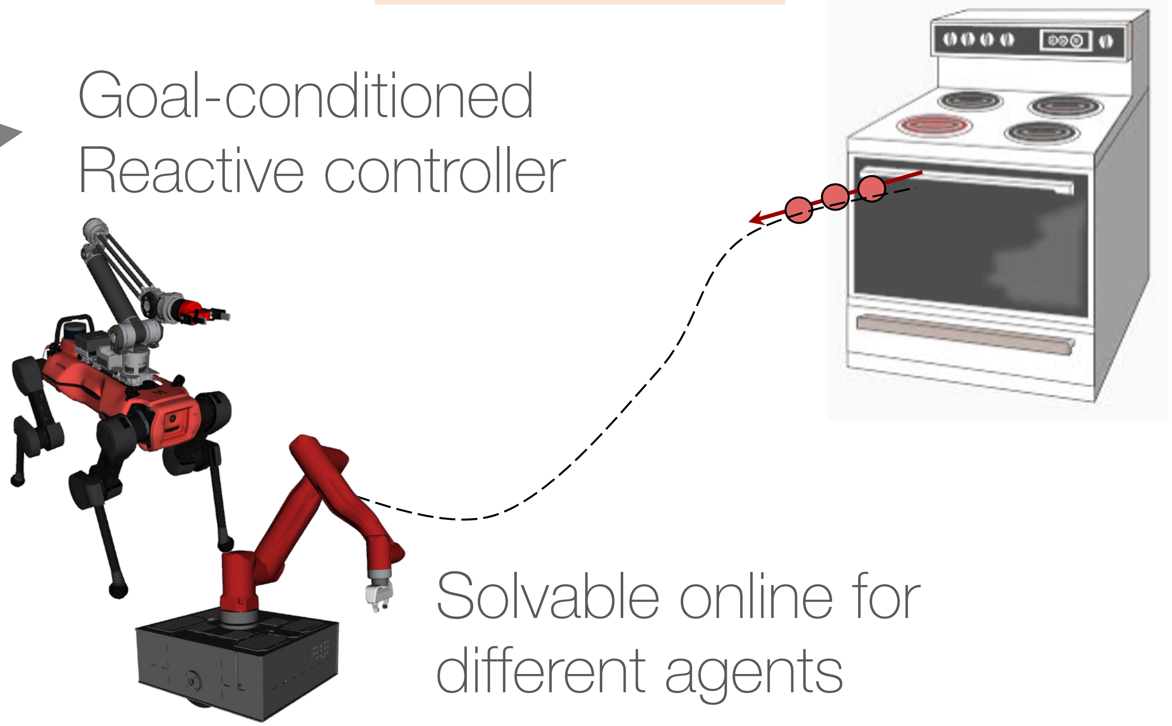


Input

→ "Take" "Jug"
→ "Open" "Fridge"
→ "Put" "Jug" in "Fridge"

Goal Generation

Goal-conditioned Reactive controller



Solvable online for different agents

Structure in Compositional Planning

State s_0 → Feed-Forward Model → Latent State s or z

Object-Centric Generative Model of Action Concepts

Reference or Implicit Reward

Feedback Controller → Measured State s or z

Agent-Centric Policy/Planner

Input

→ “Take” ”Jug”
→ “Open” ”Fridge”
→ “Put” ”Jug” in ”Fridge”

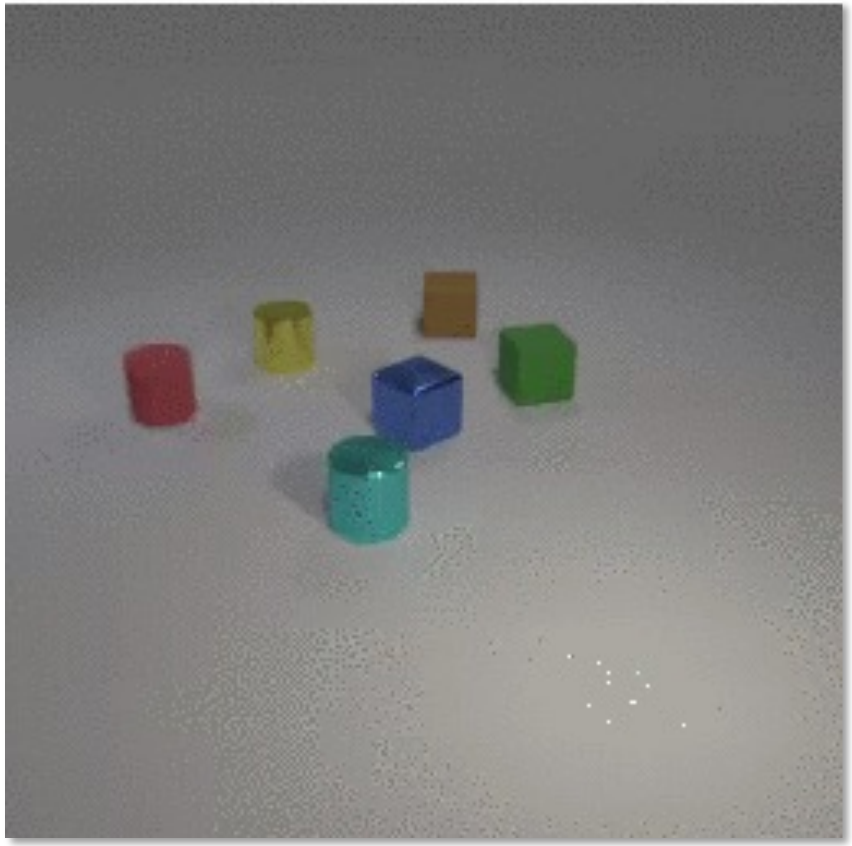
Goal Generation

Goal-conditioned Reactive controller

Solvable online for different agents

Object-Centric Causal Generative Model

Semantic + Action-Conditional



Semantic action-conditonal video prediction

Self-Supervised Modular Object Representation

Long-term Semantically Consistent Predication

No bounding box or object level supervision.

Prompt: Sequential Language Instruction

Object-Centric Causal Generative Model

Modular Action Concepts

Input: t=1

→ “Take” ”Jug”
→ “Open” ”Fridge”
→ “Put” ”Jug” in ”Fridge”

→ “Pick up” ”Green Bowl”
→ “Place in” ”large pink bowl”

Ground Truth Instruction

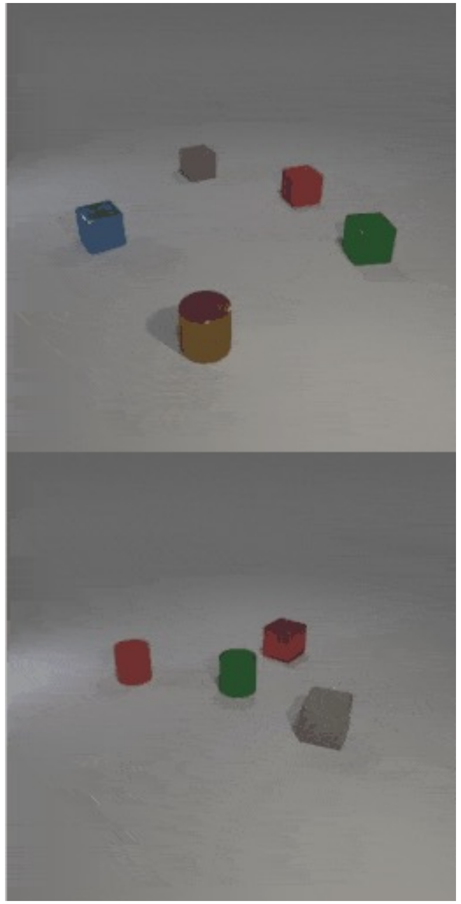
Ground Truth

MAC Prediction

Object-Centric Causal Generative Model

Systematic Generalization: Out of Distribution

Testing



All Red cubes are removed from training data

Testing:
Concurrent actions

Training:
Single action

Structure in Compositional Planning

State s_0 → Feed-Forward Model → Latent State s or z

Reference or Implicit Reward

Object-Centric Generative Model of Action Concepts

Feedback Controller → Measured State s or z

Agent-Centric Policy/Planner

Input

→ “Take” ”Jug”
→ “Open” ”Fridge”
→ “Put” ”Jug” in ”Fridge”

Goal Generation

Goal-conditioned Reactive controller

Solvable online for different agents

Structure in Compositional Planning

State s_0 → Feed-Forward Model → Latent State s or z

Object-Centric Generative Model of Action Concepts

Reference or Implicit Reward

+ → Feedback Controller → Measured State s or z → -

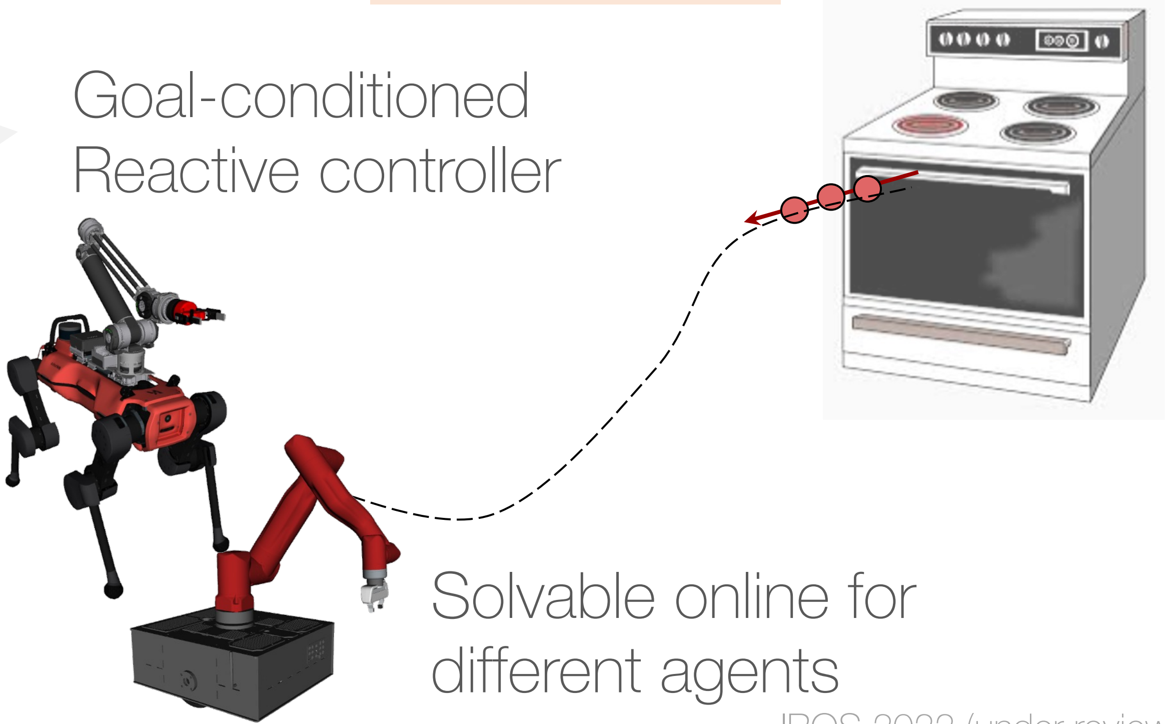
Agent-Centric Policy/Planner

Input

→ "Take" "Jug"
→ "Open" "Fridge"
→ "Put" "Jug" in "Fridge"

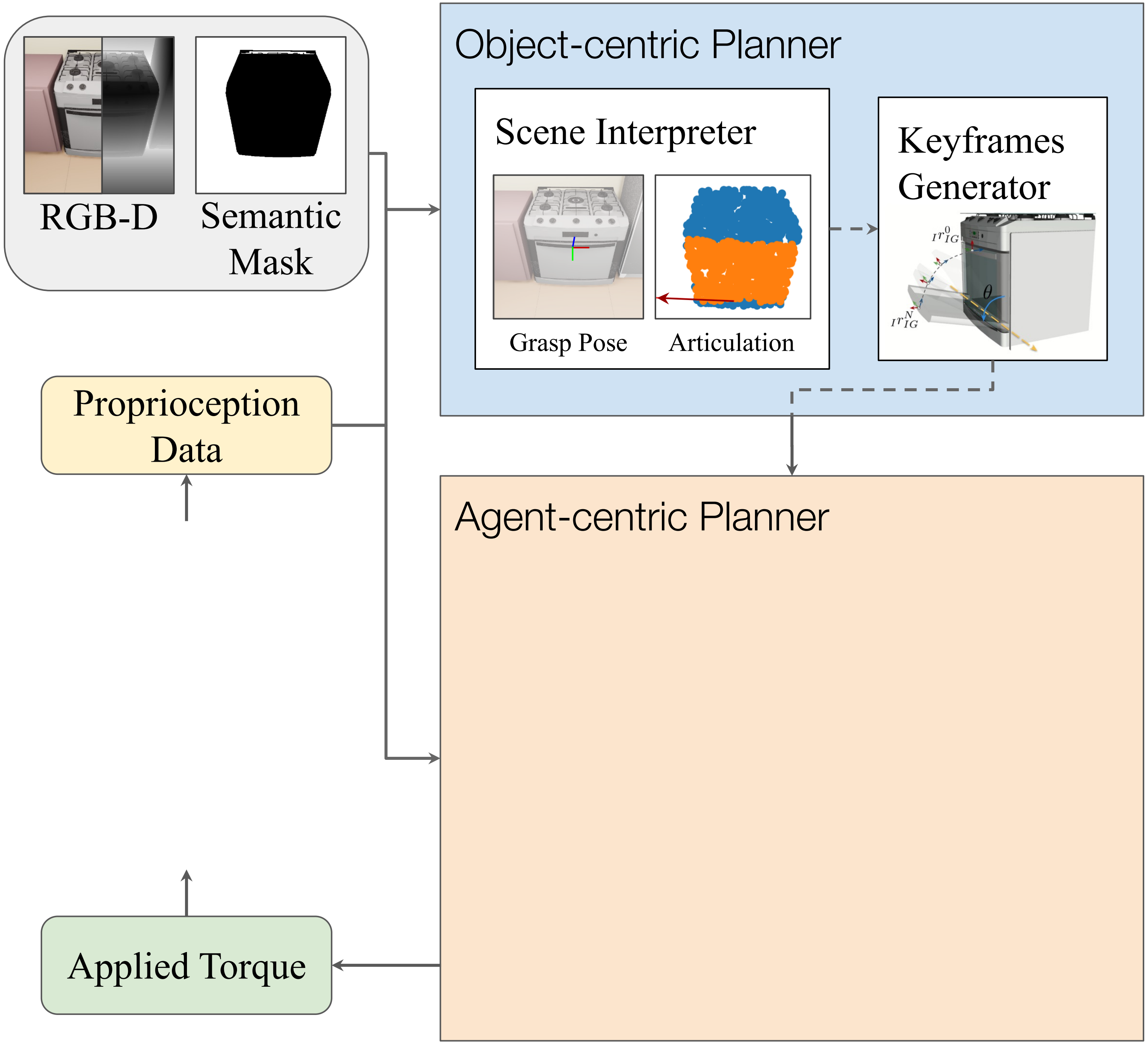
Goal Generation

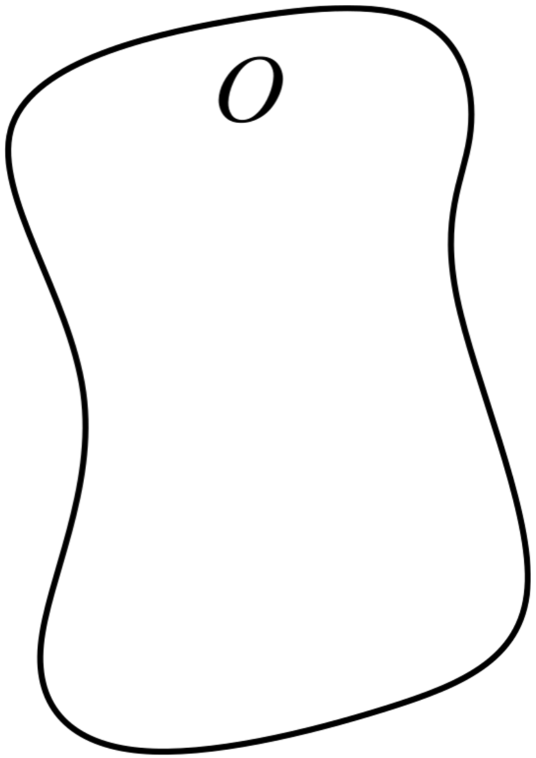
Goal-conditioned Reactive controller

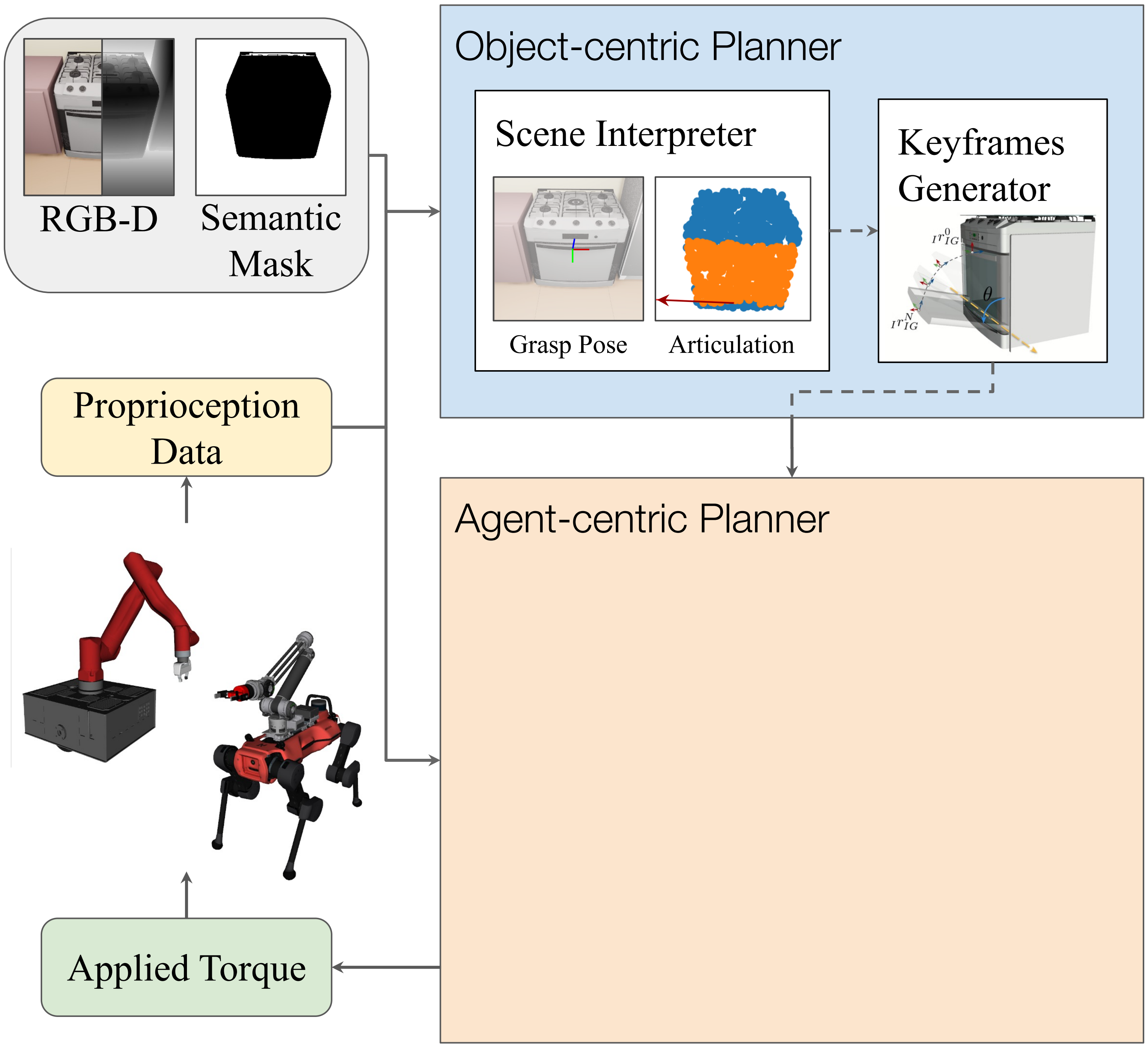


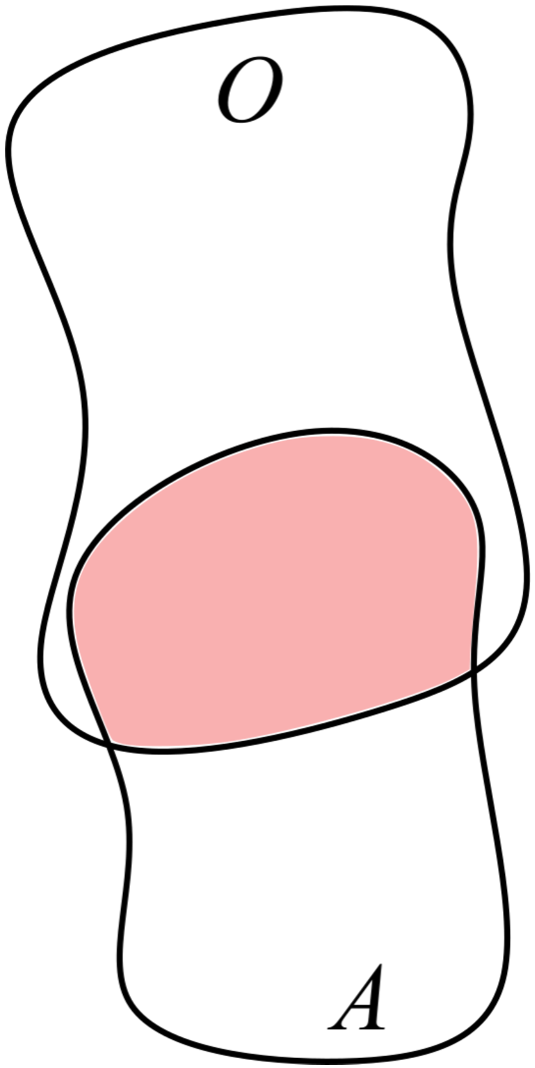
Solvable online for different agents

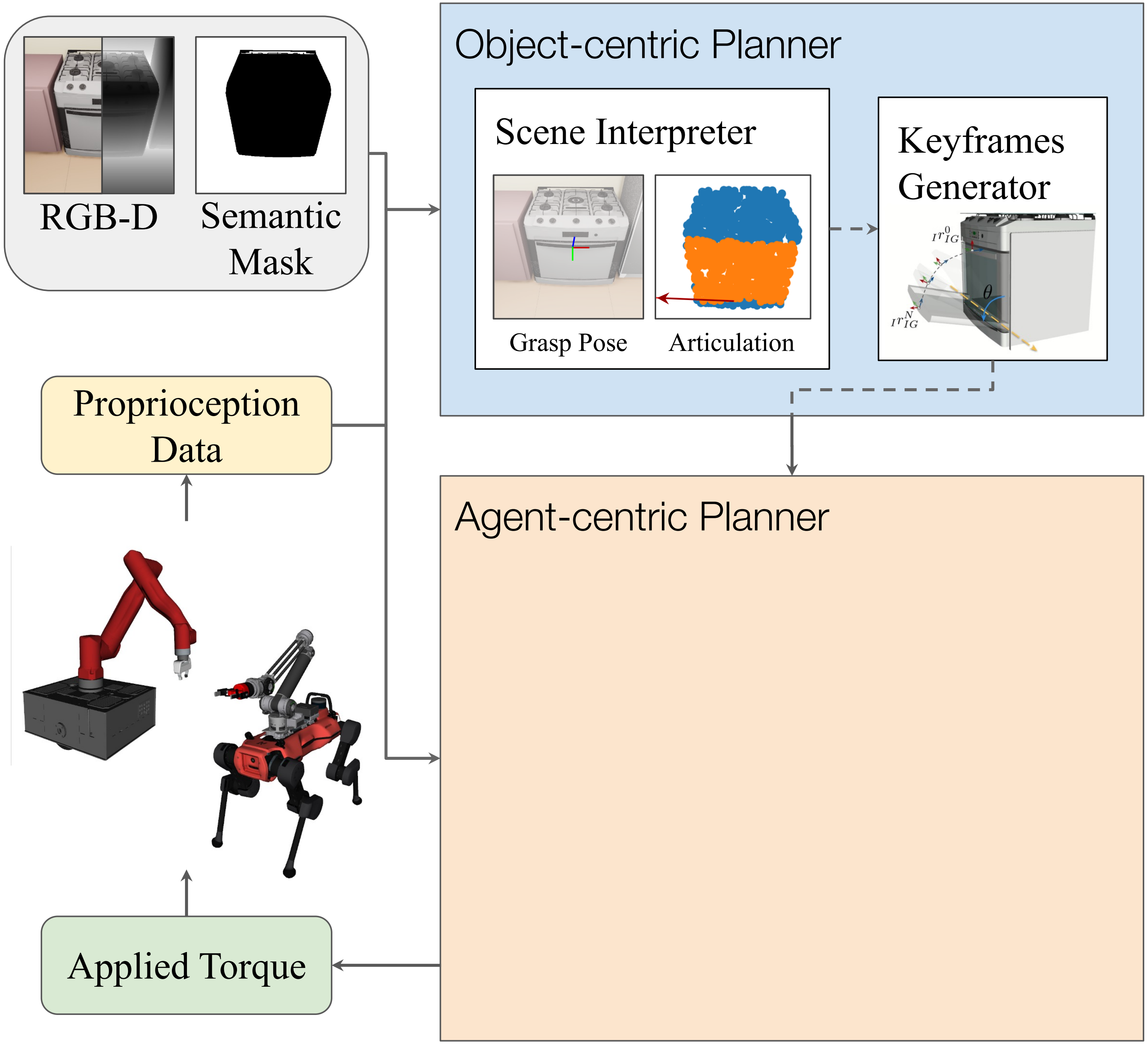
IROS 2022 (under review)


Object-centric Planner
Scene Interpreter
Keyframes Generator
RGB-D
Semantic Mask
Grasp Pose
Articulation
Proprioception Data
Agent-centric Planner
Applied Torque

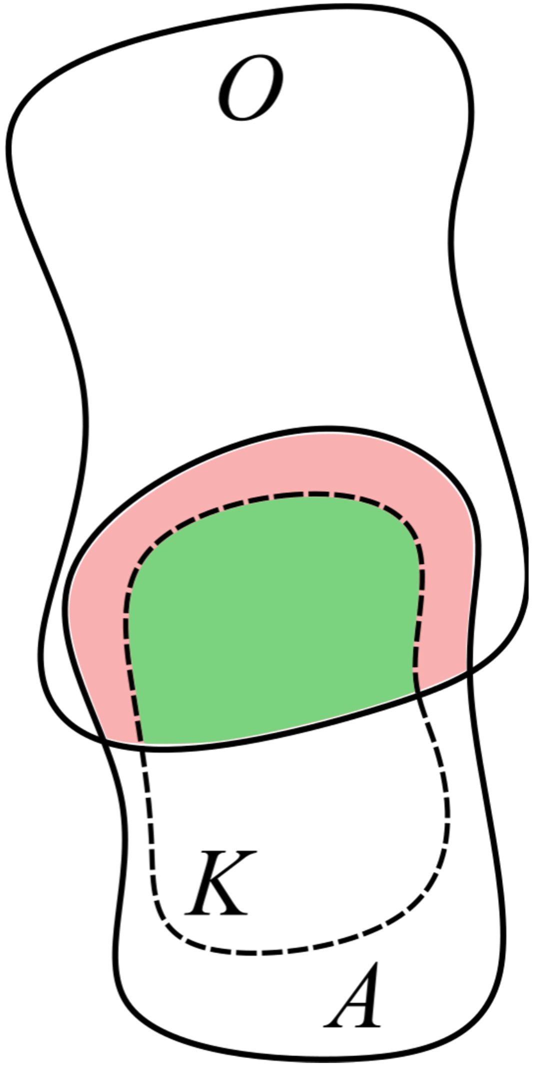

O


RGB-D
Semantic Mask
Object-centric Planner
Scene Interpreter
Grasp Pose
Articulation
Keyframes Generator
Proprioception Data
Agent-centric Planner
Applied Torque

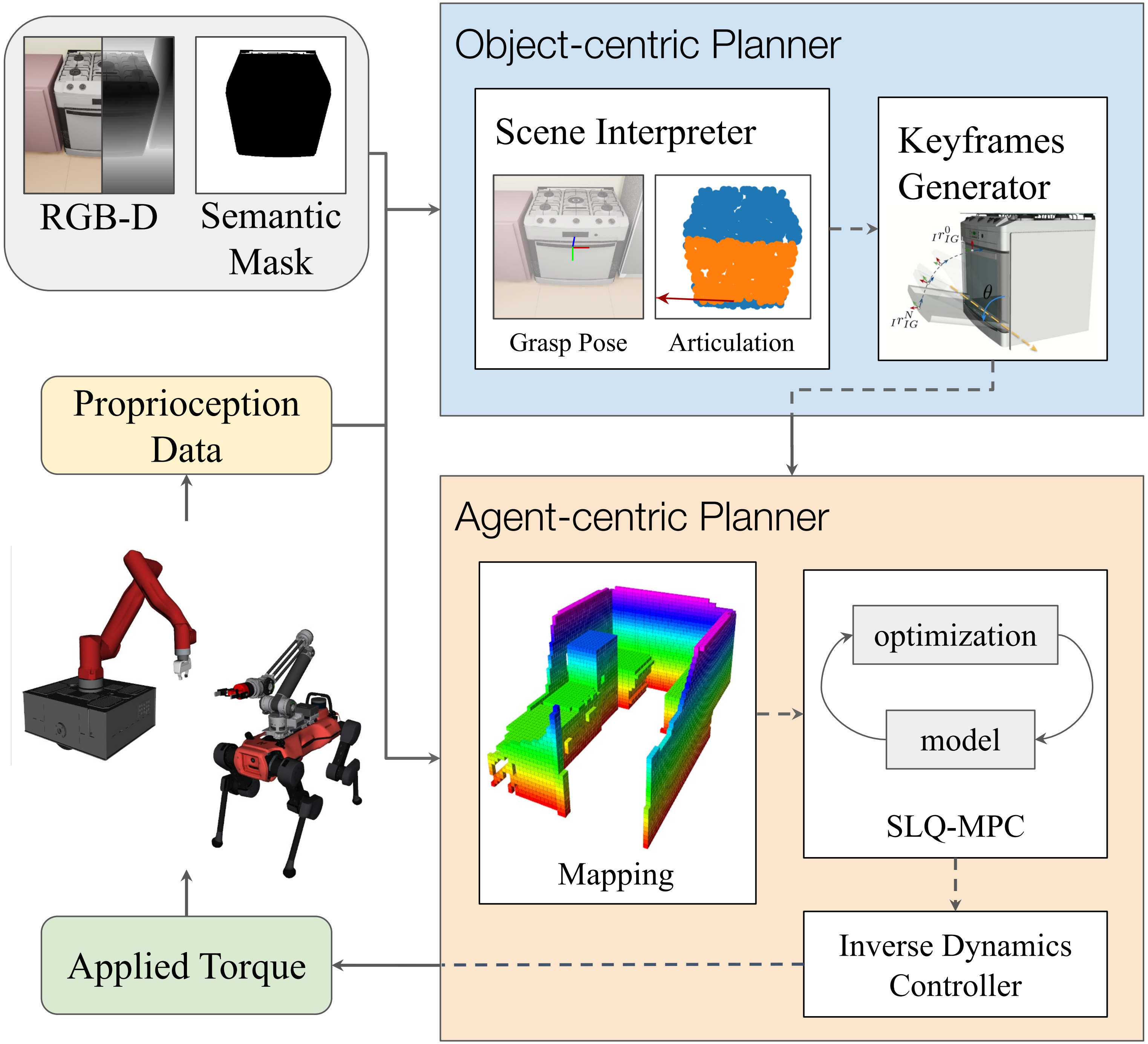

O
A

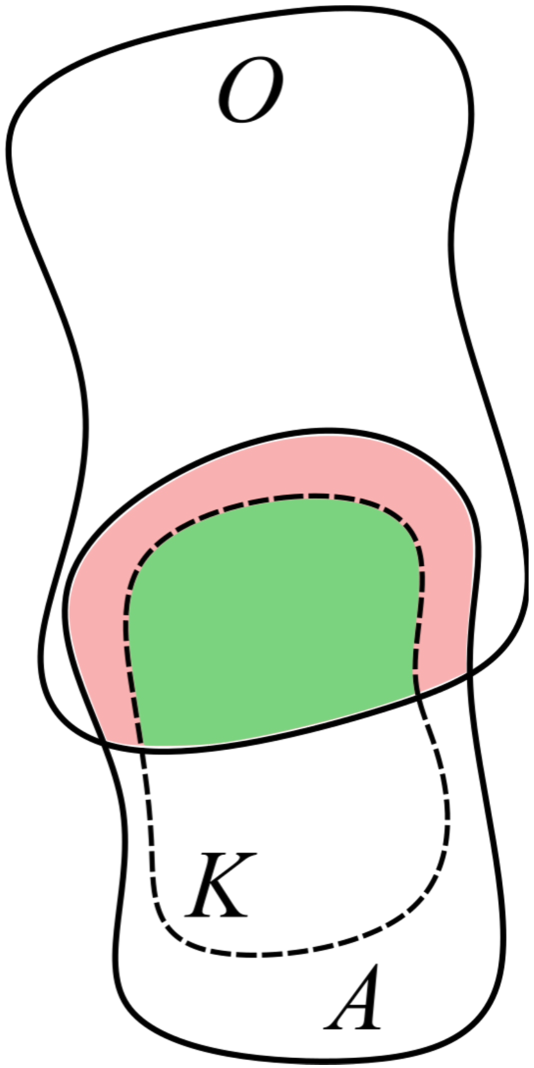


RGB-D
Semantic Mask
Object-centric Planner
Scene Interpreter
Grasp Pose
Articulation
Keyframes Generator
Proprioception Data
Agent-centric Planner
Applied Torque



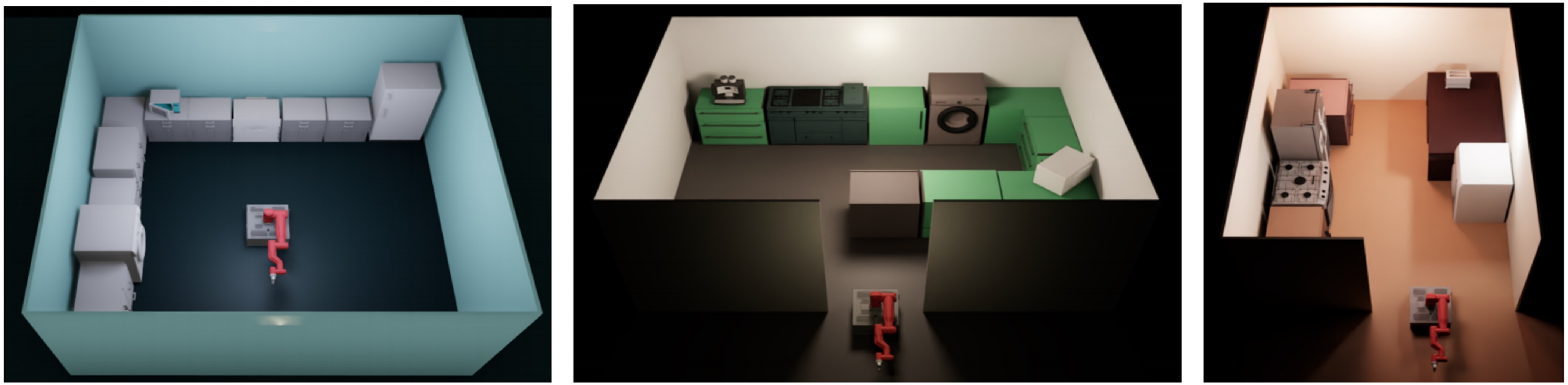
O
K
A





IROS 2022 (under review)

Structure in Compositional Planning: Setup



Different kitchen layouts designed on NVIDIA Isaac Sim using PartNet-Mobility dataset

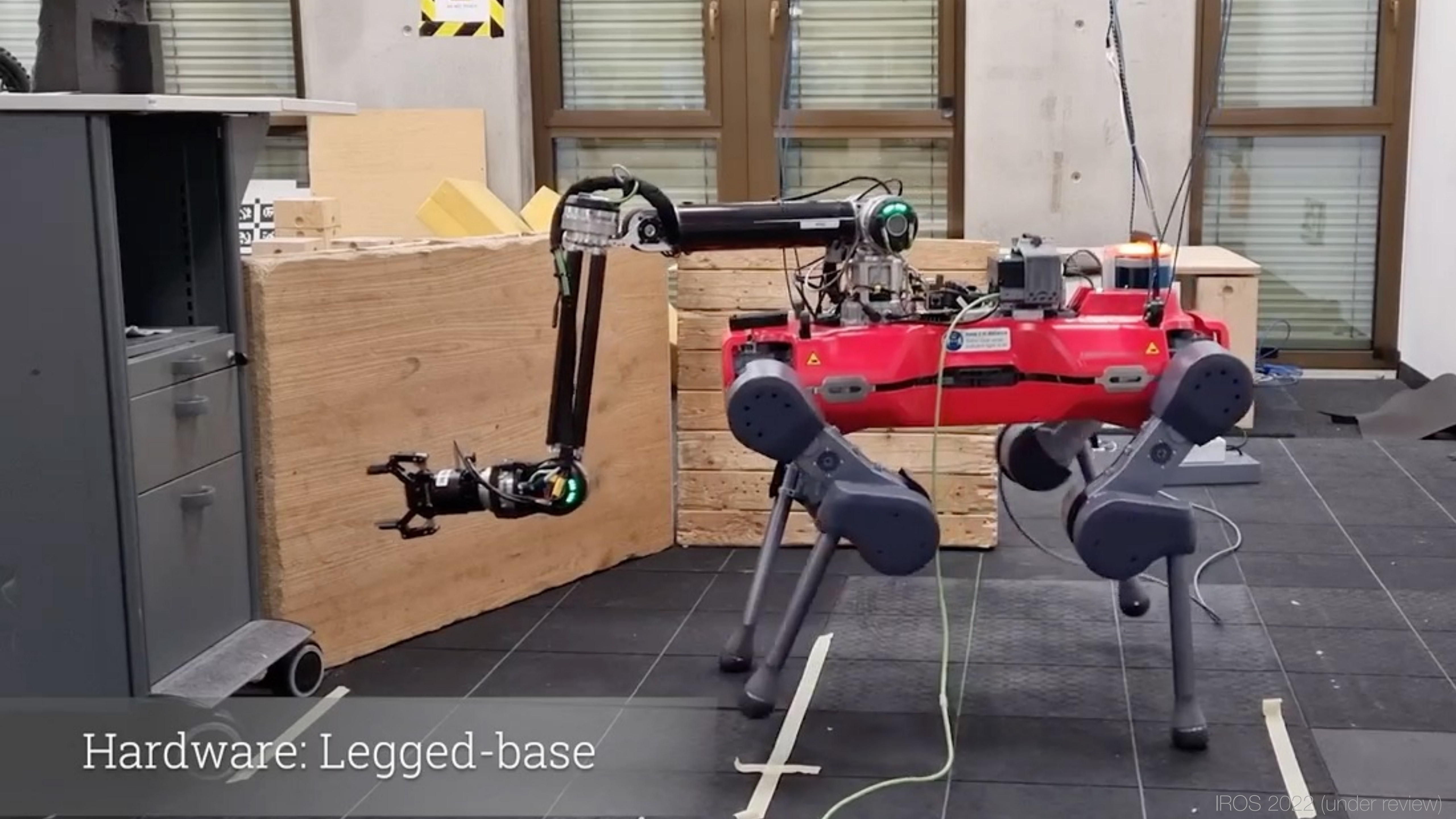
(a) Drawers

(b) Ovens

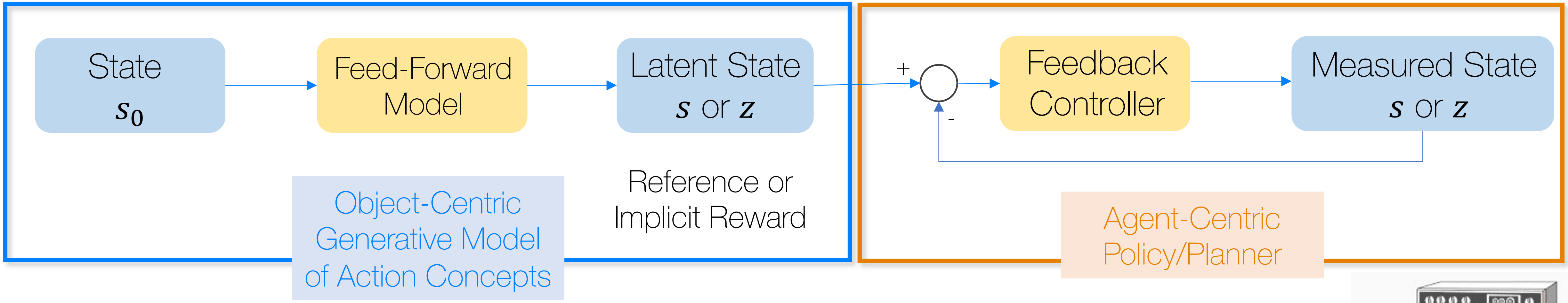
(c) Washing Machines

Static Scene: novel instances of known articulated object category


x1.5
IROS 2022 (under review)


Hardware: Legged-base
IROS 2022 (under review)

Structure in Compositional Planning

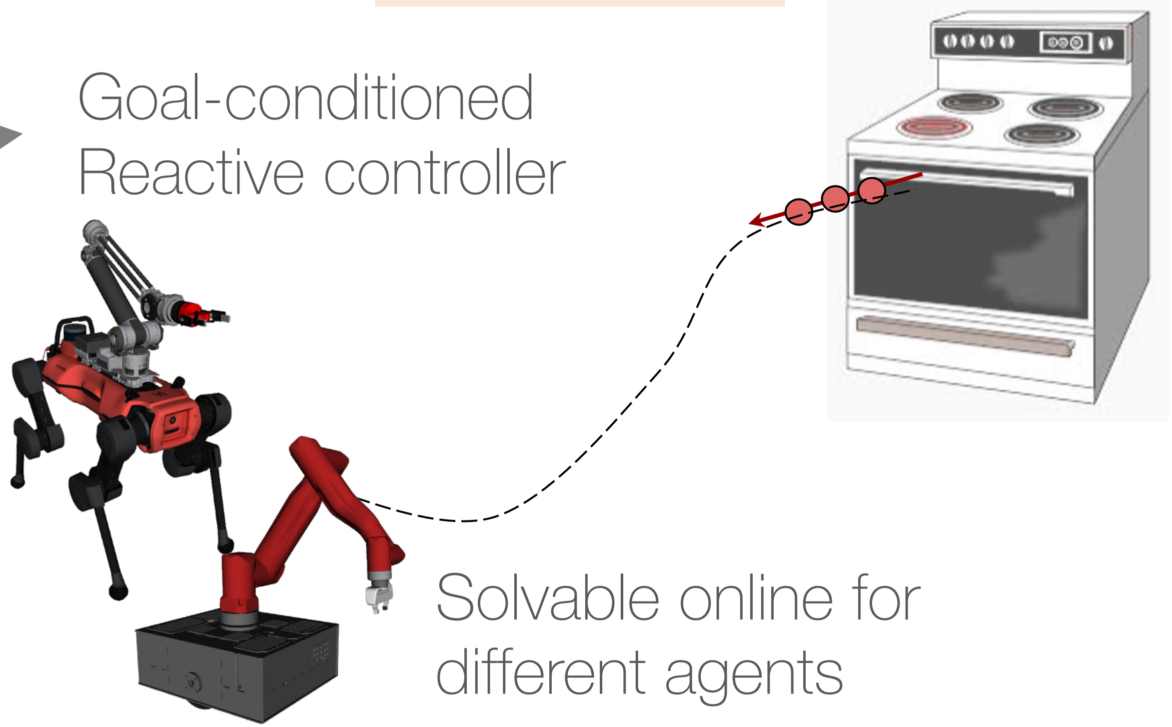


Input

→ "Take" "Jug"
→ "Open" "Fridge"
→ "Put" "Jug" in "Fridge"

Goal Generation

Goal-conditioned Reactive controller



Solvable online for different agents

Structure

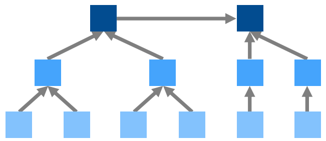
State/Action Reps.
VICES IROS19
LASER ICRA21
Making Sense ICRA19
Unsup KPs PAMI21

Inductive Biases
C-Learning ICLR21
OCEAN UAI20
D2RL arXiv20

Structure in Planning
CAVIN CORL20
Skill Hierarchy ICLR21
Finding-IT, CVPR18

Neural Programming
NTP ICRA18
NTG CVPR19
Cont.Relax IROS19

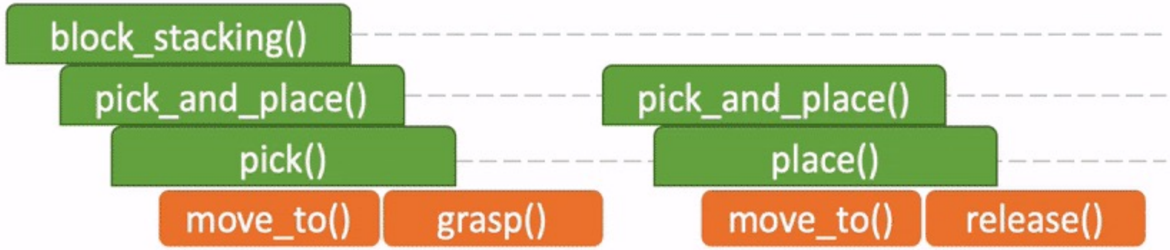
Representations for Planning

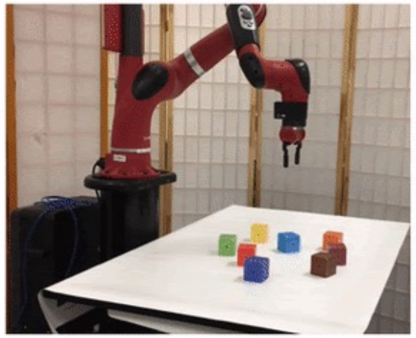


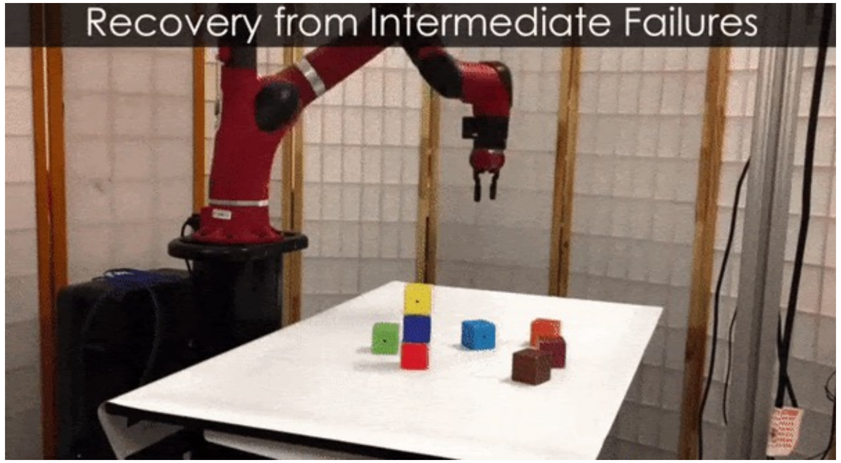
Hierarchy

What model structure enables longer term planning?

Program Induction provides a very efficient model of compositional generalization





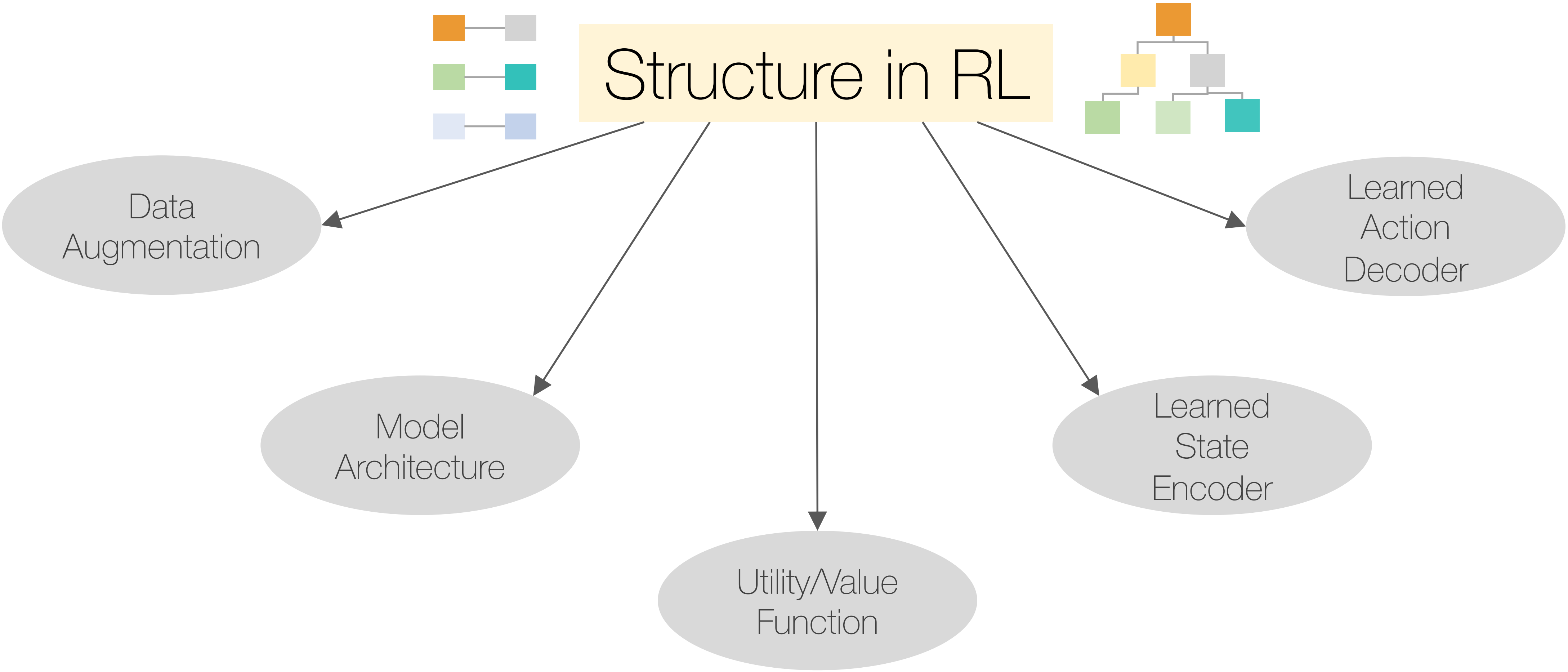


NTP, ICRA 2018, NTG, CVPR 2019

Structure for Reinforcement Learning

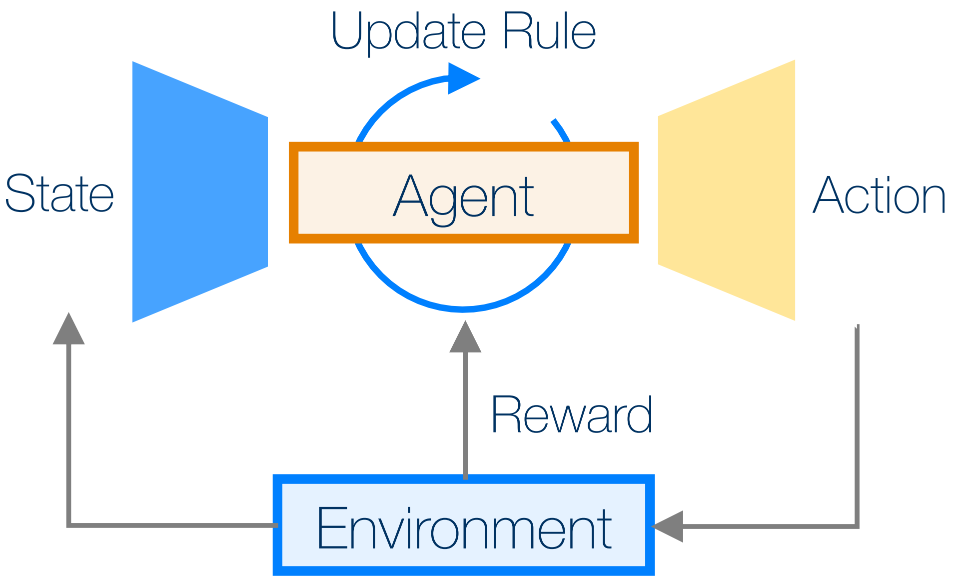
Structured Biases improve both efficiency & generalization

Robot Learning needs new ones!



Towards Generalizable Autonomy

Structure in Reinforcement Learning for Robotics



Animesh Garg

garg@cs.toronto.edu

@animesh_garg

People AI & Robots pair

UNIVERSITY OF TORONTO

VECTOR INSTITUTE

NVIDIA